

## OMNISOL: Fast Prediction of Free Energies of Solvation and Partition Coefficients

Gregory D. Hawkins,<sup>†</sup> Daniel A. Liotard,<sup>‡</sup> Christopher J. Cramer,<sup>\*,†</sup> and Donald G. Truhlar<sup>\*,†</sup>

Department of Chemistry and Supercomputer Institute, University of Minnesota, Minneapolis, Minnesota 55455-0431 and Laboratoire de Physico-Chimie Théorique, Université de Bordeaux I, 351 Cours de la Liberation, 33405 Talence Cedex, France

Received January 8, 1998

The SM5.0R model for predicting solvation energies using only geometry-dependent atomic surface tensions was developed previously for aqueous solution. Here we extend it to organic solvents. The method is based on gas-phase geometries and exposed atomic surface areas; electrostatics are treated only implicitly so a wave function or charge model is not required (which speeds up the calculations by about 2 orders of magnitude). The SM5.0R model has been parametrized for solvation free energies of solutes containing H, C, N, O, F, S, Cl, Br, and I. The training set for organic solvents consists of 227 neutral solutes in 90 organic solvents for a total of 1836 data points. The method achieves a mean unsigned error of about 0.4 kcal/mol when applied using gas-phase geometries calculated at either the Hartree–Fock level with a heteroatom-polarized valence-double- $\zeta$  basis set (HF/MIDI!) or when applied using semiempirical molecular orbital gas-phase geometries. In related work reported here, the parametrization for predicting aqueous solvation free energies is also extended to include organic solutes containing iodine. This extension is based on eight solutes and yields a mean unsigned error of 0.25 kcal/mol. The resulting SM5.0R model for solvation energies in aqueous and organic solvents can therefore be used to predict partition coefficients (including  $\log P$  for octanol/water) for any solute containing H, C, N, O, F, S, Cl, Br, and/or I.

### 1. Introduction

Free energy is the fundamental thermodynamic variable controlling equilibria, and free energy of solvation is the free energy difference between a molecule in the gas phase and in a solvent.<sup>1</sup> The free energy of solvation of a solute X in solvent Y may be used to predict its vapor pressure over a dilute solution (Henry's law), and the free energy of solvation of X in X may be used to predict the vapor pressure of a pure liquid or of the solvent of a dilute solution (Raoult's law).<sup>2</sup>

Free energy of solvation may be considered as a special case of free energy of transfer<sup>2</sup> in which one of the transfer media is an inert gas, usually called "air," and the other is a liquid. By combining the free energy of transfer of solute X from air into solvent Y with the free energy of transfer of X from air into solvent Z, we can obtain the free energy of transfer of X from Y to Z, which allows one to calculate partition coefficients, i.e., the equilibrium distribution of a solute between immiscible liquid media, which are of critical importance in pharmaceutical<sup>3</sup> and environmental<sup>4</sup> applications or for extractions. Comparing the free energies of solvation of conformational or tautomeric isomers X<sub>I</sub> and X<sub>II</sub> of molecule X in various solvents allows one to predict solvent effects on conformational<sup>5,6</sup> or tautomeric equilibria.<sup>5,7</sup>

Thus free energies of solvation are among the most broadly useful of all thermodynamic quantities referring

to molecules in solution. Their importance is further illustrated by the many qualitative concepts that have been developed to summarize trends in free energies of solvation in terms of solvophilicity of the solutes, e.g., lipophilic and lipophobic molecular surfaces, and in terms of the solvating interactions of various classes of solvents, e.g., apolar aprotic solvents, polar aprotic solvents, and protic solvents.<sup>5</sup>

Methods to estimate free energies of transfer may be classified as classical or quantum mechanical. The best developed classical method is the use of multivariate quantitative structure–property relations (multivariate QSPRs).<sup>8,9</sup> Motivated by its widespread use as a bio-availability parameter in the drug industry, the partition coefficient for 1-octanol and water has received by far the most attention,<sup>10</sup> although Abraham, for example, has developed much broader relationships applicable to almost any solvent.<sup>8</sup> Another classical model is the generalized Born/surface area (GB/SA) method employed in the MacroModel computer program;<sup>11,12</sup> the GB/SA model has been parametrized for the solvents water<sup>11</sup> and chloroform.<sup>12</sup> Quantum mechanical methods are usually based on molecular orbital theory.<sup>13,14</sup> Most such models

(5) Reichardt, C. *Solvents and Solvent Effects in Organic Chemistry*, 2nd ed.; VCH: Weinheim, 1988.

(6) Cramer, C. J.; Truhlar, D. G. In *Quantitative Treatments of Solute/Solvent Interactions*; Politzer, P., Murray, J. S., Eds.; Elsevier: Amsterdam, 1994; pp 9–54.

(7) Cramer, C. J.; Truhlar, D. G. In *Solvent Effects and Chemical Reactivity*; Tapia, O., Bertrán J., Eds.; Kluwer: Dordrecht, 1996; pp 1–80.

(8) Abraham, M. H. In ref 7, pp 83–134.

(9) Famini, G. R.; Wilson, L. Y. In ref 7, pp 213–241.

(10) Hansch, C.; Leo, A. *Exploring QSAR*; American Chemical Society: Washington, 1995.

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

(12) Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C.; MACROMODEL-version 6.0; Columbia University: New York, 1990.

<sup>†</sup> University of Minnesota.

<sup>‡</sup> Université de Bordeaux.

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

(2) Klotz, I. M.; Rosenberg, R. M. *Chemical Thermodynamics*, 4th ed.; Benjamin/Cummings: Menlo Park, CA, 1986.

(3) Silverman, R. B. *The Organic Chemistry of Drug Design and Drug Action*; Academic: San Diego, 1992.

(4) Larson, R. A.; Weber, E. J. *Reaction Mechanisms in Environmental Organic Chemistry*; Lewis Publishers: Boca Raton, 1994.

consider only electrostatics or perhaps electrostatics plus a term to estimate hydrophobic effects in water or dispersion interactions. However the Solvation-Model-5 suite, in particular SM5.4,<sup>15,16</sup> SM5.2R,<sup>17</sup> and SM5.42R,<sup>18,19</sup> is very widely applicable. These models treat long-range electrostatic effects self-consistently by molecular orbital theory (semiempirical Hartree–Fock theory,<sup>15–17</sup> ab initio Hartree–Fock theory,<sup>18</sup> or Kohn–Sham theory with gradient-corrected density functionals<sup>19</sup>) and include solvent-dependent atomic surface tensions that account for short-range interactions. These models have been parametrized against free energies of solvation for 91 solvents<sup>15–19</sup> and are applicable to water and virtually any organic solvent. A number of other self-consistent methods are also available and are reviewed with extensive references elsewhere.<sup>20</sup> We note, however, that only the SM5 models are explicitly parametrized for both water and general organic solvents.

Atomic surface tensions play an important role in the GB/SA, SM5.4, SM5.2R, and SM5.42R models and also in our earlier SM1, SM2, SM3, and SM4 models. In all these models one recognizes two contributions to the solvation free energy, one being a long-ranged electrostatic contribution associated with the volume of polarizable solvent within interaction range of the solute atomic partial charges, and the other being a short-ranged contribution associated with the first solvation shell. This partitioning cannot be accomplished rigorously, in part free energies include entropies, and entropic effects cannot be decomposed into strictly local parts. A second ambiguity in this partitioning is that the calculation of the electrostatic part requires specification of a boundary outside of which the solvent is assumed to have its bulk dielectric constant. Despite these ambiguities, the partitioning of solvation energy into bulk electrostatics and “the rest” provides a useful framework for discussion and parametrization. In our work “the rest” is usually called the CDS contribution because it is dominated by the cost

of cavity (C) deformation, the dispersion (D) interaction between solute and solvent, and the solute-induced changes in solvent structure (S), including hydrogen bonding, hydrophobic effects, and deviation of solvent properties in the first solvation shell from the bulk values. It is very natural to model first-solvation-shell effects in terms of atomic surface tensions. In such a model the atomic surface tension of atom *k* is the ratio of the CDS contribution to the free energy of solvation that may be attributed to atom *k* to the exposed or solvent-accessible surface area of atom *k*.

The present paper considers a classical model, called SM5.0R, based only on atomic surface tensions. This model has previously been parametrized for aqueous solution,<sup>21</sup> and in the present paper it is parametrized for organic solvents based on data in 90 such solvents. The SM5.0R model is simpler than GB/SA in that no electrostatic calculation is required. Furthermore, geometries are not optimized in solution; gas-phase geometries are assumed available from another source or can be calculated by any available gas-phase molecular modeling or molecular orbital procedures. Comparing SM5.0R to a model on the simpler side of the complexity scale, we note that it is more complicated than using a QSPR procedure in that it involves the full 3-D geometry of the molecule or molecular conformation under study. Nevertheless the calculations for medium-sized molecules can be completed almost instantaneously on a desktop computer using software<sup>22</sup> freely available at our World Wide Web site,<sup>23</sup> and—as we shall show—the mean deviations of calculated free energies of solvation from experimental results are remarkably small. In terms of the partitioning of solvation effects discussed in the previous paragraph, atomic surface tensions in a surface-tension-only model have a new meaning as compared to their use in models where a surface tensions term is added to an explicitly electrostatics term. In the SM5.0R model, the atomic surface tension of atom *k* is the ratio of the contribution to the entire free energy of solvation that may be attributed to atom *k* to the exposed surface area of atom *k*. Although physically we expect the CDS contributions to correlate better than the electrostatic contributions with exposed surface area, it is both intriguing and practically useful to see how well one can do with the simpler assumption that all contributions to the solvation free energy track with exposed atomic surface areas. At an intuitive level, we might expect some success for such a model because, although electrostatic forces are long ranged, the largest electrostatic contributions to the free energy of solvation are attributable to the solvent molecules closest to exposed charges, and the average amount of solvent in contact with an exposed charge does correlate with the exposed atomic surface area. A caveat about how well the model might perform in situations significantly different than those presented by our training set is given in Section 6.

Since the original SM5.0R aqueous parametrization<sup>21</sup> did not include solutes containing iodine, we present that extension as a prelude to the parametrization for organic solvents which covers solutes containing H, C, N, O, F, S, Cl, Br, and I.

(13) Tomasi, J. *ACS Symp. Ser.* **1994**, 568, 10. Storer, J. W.; Giesen, D. J.; Hawkins, G. D.; Lynch, G. C.; Cramer, C. J.; Truhlar, D. G.; Liotard, D. A. *Ibid.* p 24. Lim, C.; Chan, S. L.; Tole, P. *Ibid.* p 50. Warshel, A.; Chu, Z. T. *Ibid.* p 71. Rivail, J.-L.; Antonczak, S.; Chipot, C.; Ruiz-Lopez, M. F.; Gorb, L. G. *Ibid.* p 154. Bertrán, J.; Lluch, J. M.; González-Lafont, A.; Dillet, V.; Pérez, V. *Ibid.* p 168. Tucker, S. C.; Gibbons, E. M. *Ibid.* p 196. Gao, J.; Xia, X. *Ibid.* p 212. Venanzi, C. A.; Buono, R. A.; Luzhkov, V. B.; Zauhar, R. J.; Venanzi, T. J. *Ibid.* p 260.

(14) Tapia, O. *J. Math. Chem.* **1992**, 10, 139. Takahashi, O.; Sawahata, H.; Ogawa, Y.; Kikuchi, O. *J. Mol. Struct. (THEOCHEM)* **1997**, 393, 141. Karelson, M. M.; Katritzky, A. R.; Zerner, M. C. *J. Org. Chem.* **1991**, 56, 134. Wiberg, K. B.; Wong, M. W. *J. Am. Chem. Soc.* **1993**, 115, 1078. Orozco, M.; Bachs, M.; Luque, F. J. *J. Comput. Chem.* **1995**, 16, 563. Ford, G. P.; Wang, B. *J. Am. Chem. Soc.* **1992**, 114, 10563. Rashin, A. A.; Bukatin, M. A.; Andzelm, J.; Hagler, A. T. *Biophys. Chem.* **1994**, 51, 375. Truong, T. N.; Stefanovich, E. V. *Chem. Phys. Lett.* **1995**, 240, 253.

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

(16) Giesen, D. J.; Gu, M. Z.; Cramer, C. J.; Truhlar, D. G. *J. Org. Chem.* **1996**, 61, 8720. Giesen, D. J.; Chambers, C. C.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1997**, 101, 2061. Giesen, D. G.; Hawkins, G. D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.* **1997**, 98, 85.

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

(18) Zhu, T.; Li, J.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Phys.*, submitted.

(19) Li, J.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *Chem. Phys. Lett.*, in press.

(20) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, 94, 2027. Cramer, C. J. *Rev. Comput. Chem.* **1995**, 6, 1. Cramer, C. J.; Truhlar, D. G. In *Solvent Effects and Chemical Reactivity*; Tapia, O., Bertrán, J., Eds.; Kluwer: Dordrecht, 1996; p 1. Rivail, J.-L.; Rinaldi, D. *Comp. Chem. Rev. Curr. Trends* **1996**, 1, 139.

(21) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **1997**, 101, 7147.

(22) Hawkins, G. D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G.; OMNISOL-version 1.0, University of Minnesota: Minneapolis, 1997.

(23) <http://comp.chem.umn.edu/omnisol>.

## 2. Theory

In developing the SM5.0R model for calculating aqueous solvation free energies,<sup>20</sup> we defined the standard-state free energy of solvation to be given simply by

$$\Delta G_S^0 = \sum_k \sigma_k^A A_k \quad (1)$$

where  $\sigma_k^A$  is a geometry-dependent partial atomic surface tension for atom  $k$ , and  $A_k$  is the solvent accessible surface area for atom  $k$  using a solvent radius of zero and the atomic radii suggested by Bondi.<sup>24</sup> Equation 1 applies to the case when the standard-state concentration of the solute molecule is the same (e.g., 1 M) in both the gas phase and solution. The partial atomic surface tensions of eq 1 are defined as follows

$$\sigma_k^A|_{Z(k)=1} = \tilde{\sigma}_H + \sum_{Z(k)=6,7,8,16} \{ T(R_{HK}|\bar{R}_{HK}, W) [\tilde{\sigma}_{HK} + \tilde{\sigma}_{HO}^{(2)} \sum_{\substack{Z(k')=1 \\ k' \neq k \\ Z(k)=8}} T(R_{k'k'}|\bar{R}_{k'k'}, W) + \tilde{\sigma}_{HN}^{(2)} \sum_{\substack{Z(k')=7 \\ Z(k)=7 \\ k' \neq k'}} T(R_{k'k'}|\bar{R}_{k'k'}, W)] \} \quad (2)$$

$$\sigma_k^A|_{Z(k)=6} = \tilde{\sigma}_C + \tilde{\sigma}_{CC} \sum_{\substack{Z(k)=6 \\ k' \neq k}} T(R_{kk'}|\bar{R}_{CC}, W) + \tilde{\sigma}_{CN}^{(2)} \sum_{\substack{Z(k)=6 \\ k' \neq k}} T(R_{kk'}|\bar{R}_{CC}, W_{CC}) + \tilde{\sigma}_{CN} [ \sum_{Z(k)=7} T(R_{kk'}|\bar{R}_{CN}, W) ]^2 \quad (3)$$

$$\sigma_k^A|_{Z(k)=8} = \tilde{\sigma}_O + \tilde{\sigma}_{OC} \sum_{Z(k)=6} T(R_{kk'}|\bar{R}_{OC}^{(2)}, W_{OC}) + \tilde{\sigma}_{OO} T [- \sum_{\substack{Z(k)=8 \\ k' \neq k}} T(R_{kk'}|\bar{R}_{OO}, W) | R_{TT}, W_{TT} ] + \tilde{\sigma}_{ON} \sum_{Z(k)=7} T(R_{kk'}|\bar{R}_{ON}, W) \quad (4)$$

$$\sigma_k^A|_{Z(k)=7} = \tilde{\sigma}_N + \tilde{\sigma}_{NC} \{ \sum_{Z(k)=6} T(R_{kk'}|\bar{R}_{CN}, W) [ \sum_{\substack{k' \neq k \\ k' \neq k}} T(R_{k'k'}|\bar{R}_{CN}, W) ]^2 \}^{1.3} + \tilde{\sigma}_{NC}^{(2)} \sum_{Z(k)=6} [ T(R_{kk'}|\bar{R}_{CN}, W) \sum_{Z(k)=8} T(R_{k'k'}|\bar{R}_{CN}, W) ] \quad (5)$$

$$\sigma_k^A|_{Z(k)=9} = \tilde{\sigma}_F \quad (6)$$

$$\sigma_k^A|_{Z(k)=16} = \tilde{\sigma}_S + \tilde{\sigma}_{SS} \sum_{\substack{Z(k)=16 \\ k' \neq k}} T(R_{kk'}|\bar{R}_{SS}, W) \quad (7)$$

$$\sigma_k^A|_{Z(k)=17} = \tilde{\sigma}_{Cl} + \tilde{\sigma}_{ClC} \sum_{Z(k)=6} \left[ T(R_{kk'}|\bar{R}_{ClC}, W) \sum_{\substack{k' \neq k \\ k' \neq k}} \frac{T(R_{k'k'}|\bar{R}_{ClC}, W)}{\chi_{Z_{k'}}} \right] \quad (8)$$

$$\sigma_k^A|_{Z(k)=35} = \tilde{\sigma}_{Br} + \tilde{\sigma}_{BrC} \sum_{Z(k)=6} \left[ T(R_{kk'}|\bar{R}_{BrC}, W) \sum_{\substack{k' \neq k \\ k' \neq k}} \frac{T(R_{k'k'}|\bar{R}_{BrC}, W)}{\chi_{Z_{k'}}} \right] \quad (9)$$

and

$$\sigma_k^A|_{Z(k)=53} = \tilde{\sigma}_I \quad (10)$$

where  $Z(k)$  is the atomic number of atom  $k$ ,  $\tilde{\sigma}_Z$  and  $\tilde{\sigma}_{ZZ}$  are surface tension coefficients,  $\chi_Z$  is the Pauling electronegativity<sup>25</sup> for an atom with atomic number  $Z$ , and the geometry-dependent switching functions,  $T(R_{kk'}|\bar{R}, W)$ , and their parameters  $\bar{R}$  and  $W$  were described previously.<sup>21</sup> Note that the third and fourth term in the atomic surface tension for hydrogen and the fourth term of the atomic surface tension for nitrogen were not used in the parametrization of the SM5.0R model for aqueous solvation; they were developed in subsequent work<sup>17</sup> but are used in the present parametrization for organic solvents. The surface tension for iodine, eq 10, was also added in the current work. For aqueous solutions, the partial atomic surface tensions  $\sigma_k^A$  were actually the entire atomic surface tensions; however, that will not be the case for organic solvents. Furthermore the partial atomic surface tension coefficients will themselves become functions of other coefficients.

One aspect of eqs 2–10 is worth emphasizing here, namely the convenience of SM5 functional forms as compared to the conventional use of “typing” in molecular mechanics methods. For example, a molecular mechanics method might have different parameters for a carbonyl carbon that is bonded to  $sp^2$  carbon,  $sp^3$  carbon, or various kinds of nitrogen. The user needs to assign types to all the atoms or check that some automatic program has done this correctly. Furthermore one often encounters molecules with types of atoms for which no parameter is available. The functions in eqs 2–10 correspond to a limited amount of typing, but it is all well defined in terms of only the geometry for any solute with any bonding pattern; therefore the only input required for the solute is its geometry.

To generalize the SM5.0R model for use in calculating solvation free energies in organic solvents, we have chosen the approach that was first developed for the SM5.4 solvation model<sup>16</sup> and then modified slightly for the SM5.2R model.<sup>17</sup> In this approach, each of the partial atomic surface tension coefficients,  $\tilde{\sigma}_Z$  or  $\tilde{\sigma}_{ZZ}$ , in eqs 2–10 is expanded in terms of three solvent properties. These properties are the index of refraction<sup>26</sup> and Abraham's  $\Sigma\alpha_2^H$  and  $\Sigma\beta_2^H$  descriptors.<sup>27–29</sup> Thus, each of the surface tension coefficients can be written as

$$\tilde{\sigma}_i = \hat{\sigma}_i^{(n)} n + \hat{\sigma}_i^{(\alpha)} \alpha + \hat{\sigma}_i^{(\beta)} \beta \quad (11)$$

where  $i$  consists of the chemical symbols corresponding to  $Z$  or  $ZZ'$  (e.g., HC for  $Z, Z' = 1, 6$ ),  $n$  is the index of refraction, and  $\alpha$  and  $\beta$  are Abraham's  $\Sigma\alpha_2^H$  and  $\Sigma\beta_2^H$  descriptors, respectively.

During development of previous SM $x$  models for organic solvation,<sup>16,17</sup> four additional solvent descriptors that correlate significant components of the free energy of solvation were identified. These include the macro-

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

(25) Allred, A. L. *J. Inorg. Nucl. Chem.* **1961**, *17*, 215.

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

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

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

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

scopic surface tension (denoted  $\gamma$ ), the square of the fraction of nonhydrogenic solvent atoms which are aromatic carbon (denoted  $\phi^2$ ), the square of the fraction of the nonhydrogenic solvent atoms which are electronegative halogen (denoted  $\psi^2$ ), and the square of the solvent's  $\Sigma\beta_2^H$  descriptor (denoted  $\beta$ ). However, unlike the solvent descriptors included in eq 11, the incorporation of these descriptors was independent of any atomic number  $Z$  and was based only upon the overall exposed surface area of the solute molecule. Thus, for organic solvents, the SM5.0R expression for the solvation free energy becomes

$$\Delta G_S^{\circ} = \sum_k (\sigma_k^A + \sigma_k^M) A_k \quad (12)$$

where

$$\sigma_k^M = \hat{\sigma}_M^{(\gamma)} \gamma + \hat{\sigma}_M^{(\phi^2)} \phi^2 + \hat{\sigma}_M^{(\psi^2)} \psi^2 + \hat{\sigma}_M^{(\beta^2)} \beta^2 \quad (13)$$

and  $\sigma^M$  is a partial molecular surface tension. Note that the entire atomic surface tension of any atom  $k$  is

$$\sigma_k = \sigma_k^A + \sigma_k^M \quad (14)$$

### 3. Parametrization

**3.1. Iodine-Containing Solutes in Water.** The SM5.0R model for predicting aqueous free energies of solvation was extended to include iodine-containing solutes using the eight iodine-containing neutral organic molecules identified in earlier work<sup>15</sup> for which the experimental aqueous free energies of solvation are available. The geometries for these compounds were optimized at the Hartree–Fock level with the MIDI! basis set,<sup>30</sup> which was recently augmented to allow calculations for molecules containing iodine.<sup>31</sup> The parameters for all atoms other than iodine were fixed to the values found in the previous SM5.0R aqueous paper.<sup>21</sup> The iodine radius was set to Bondi's value of 1.98 Å,<sup>24</sup> and the single surface tension coefficient for iodine was optimized to minimize the squared error in the difference between the experimental and calculated aqueous solvation free energy for the 8 iodine-containing compounds in our training set.

**3.2. Organic Solvents.** To parametrize the SM5.0R model for predicting the free energy of solvation in organic solvents, we used the neutral-molecule training set developed for the nonaqueous parametrization of the SM5.2R model.<sup>1</sup> This training set is comprised of 90 organic solvents and 227 solutes that cover a broad range of organic functionalities for molecules composed of H, C, N, O, F, S, Cl, Br, and I. Altogether there are 1836 experimental data points used in the parametrization of the SM5.0R model for organic solvents. The sources for this experimental data have been given elsewhere;<sup>16,17,32</sup> we simply note that the majority of the data is obtained by combining aqueous free energies of solvation with water/organic partition coefficients from the MedChem database.<sup>33</sup>

As in the SM5.0R model for aqueous solvation,<sup>21</sup> the SM5.0R model for organic solvation is designed to be used with any realistic gas-phase geometry, and the parameters of the model are designed to absorb the effects of any structural changes that occur upon placing a solute molecule in solution. The gas-phase geometries used in the parametrization were optimized by performing ab initio Hartree–Fock (HF) calculations with the MIDI! basis set. The MIDI! basis is a heteroatom-polarized split valence basis, whose polarization functions were selected to provide accurate gas-phase geometries for a reasonable cost.<sup>30,31</sup> We believe that the resulting SM5.0R model for organic solvents is relatively stable to small perturbations in the gas-phase geometry, and to demonstrate this we have also included results for the SM5.0R model using geometries calculated by semiempirical molecular orbital theory, in particular by Austin Model 1<sup>34</sup> (AM1), and by molecular mechanics, in particular by the MM3\* model, which is the MacroModel<sup>12</sup> version of MM3.<sup>35</sup>

Within the SM5.0R model for organic solvents, there are three types of parameters including:

(a) 17 nonlinear parameters that control the dependence of the atomic surface tensions on the local molecular geometry;

(b) 9 van der Waals atomic radii used in calculating the exposed van der Waals surface area for a particular atom within the solute molecule;

(c) the linear surface tension coefficients.

Type a and type b parameters were taken to be identical to those used in the SM5.2R model.<sup>17</sup> The linear parameters of type c were fit as part of the current work. The 48 surface tension coefficients defined in eqs 2–13 were fit to minimize the squared difference between the predicted solvation free energies and the experimental solvation free energy for the 1836 data points in our training set.

The parametrization was accomplished using the procedures explained in detail in a preceding paper<sup>17</sup> in the SM5 series, and it was based on gas-phase geometries.

## 4. Results and Discussion

**4.1. Parameters.** The surface tension coefficient  $\tilde{\sigma}_I$  that we obtained for iodine solutes in water is  $-41.64$  cal mol<sup>-1</sup> Å<sup>-2</sup>. The surface tension coefficients for organic solvents are given in Tables 1 and 2.

**4.2. Performance: I Solutes in Water.** Table 3 contains the results for the extension of the SM5.0R-aqueous model for solutes containing iodine. The mean unsigned error over the eight compounds is 0.25 kcal/mol when HF/MIDI! geometries are employed and 0.26 kcal/mol using AM1 geometries.

The mean unsigned errors are about 28% of the mean unsigned solvation energy (0.91 kcal/mol) of the training set molecules. In 25% of the cases, the unsigned error exceeds 49% of the unsigned solvation energy; but in these cases with large relative errors, the mean unsigned absolute error is still only 0.33 kcal/mol.

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

(31) Li, J.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.*, in press.

(32) Giesen, D. J.; Storer, J. W.; Cramer, C. J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1995**, *117*, 1057. Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1995**, *99*, 7137.

(33) Leo, A. J. *Masterfile* 1994 from MedChem Software, BioByte Corp., P.O. 517, Claremont, CA 91711-0157.

(34) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902. Dewar, M. J. S.; Zoebisch, E. G. *J. Mol. Struct. (THEOCHEM)* **1988**, *180*, 1. Dewar, M. J. S.; Jie, C. *J. Mol. Struct. (THEOCHEM)* **1989**, *187*, 1. Dewar, M. J. S.; Yate-Ching, Y. *Inorg. Chem.* **1990**, *29*, 3881.

(35) Allinger, N. L.; Yuh, Y. H.; Lii, J.-H. *J. Am. Chem. Soc.* **1989**, *111*, 8551. Allinger, N. L.; Rahman, M.; Lii, J. H. *J. Am. Chem. Soc.* **1990**, *112*, 8293.

**Table 1. Partial Atomic Surface Tension Coefficients (cal mol<sup>-1</sup> Å<sup>-2</sup>) Optimized for SM5.0R for Organic Solvents**

<i>i</i>	$\hat{\sigma}_i^{(n)}$	$\hat{\sigma}_i^{(a)}$	$\hat{\sigma}_i^{(p)}$
H	43.22		
C	45.12	31.18	-41.37
N	-33.08	-169.60	20.75
O	-116.87	-126.31	
F	-2.82		
S	-83.63	-76.37	42.15
Cl	-18.06		
Br	-41.56		
I	-51.36		
H, C	-92.91		
H, N	-94.47		-164.33
H, N (2)	-142.16		
H, O	-19.33	-192.63	-545.28
H, O (2)	129.24		
H, S	53.43		
C, C	-63.92		
C, C (2)	-6.13		
O, C	60.47		-105.57
O, O	17.54	136.78	-13.92
C, N	-82.08	161.37	
N, C	-11.80	-53.66	3.79
O, N	109.28	38.20	-5.26
S, S	10.10		
Cl, C	-38.76		
Br, C	0.85		-62.93

**Table 2. Molecular Surface Tension Coefficients (cal mol<sup>-1</sup> Å<sup>-2</sup>) Optimized for SM5.0R for Organic Solvents**

coefficient	value
$\hat{\sigma}_{CS}^{(v)}$	0.1410
$\hat{\sigma}_{CS}^{(p)}$	16.68
$\hat{\sigma}_{CS}^{(a)}$	-1.45
$\hat{\sigma}_{CS}^{(v)}$	-12.30

**Table 3. Calculated and Experimental Free Energies of Solvation,  $\Delta G_s^\circ$  (kcal/mol), for the Parametrization of the SM5.0R-Aqueous Model for Iodine<sup>a</sup>**

molecule	SM5.0 //HF/MIDI!	SM5.0 //AM1	expt
diiodomethane	-2.52	-2.51	-2.49
iodomethane	-0.38	-0.40	-0.89
iodoethane	-0.54	-0.50	-0.72
1-iodopropane	-0.44	-0.37	-0.59
2-iodopropane	-0.30	-0.21	-0.46
1-iodobutane	-0.30	-0.21	-0.25
1-iodopentane	-0.17	-0.05	-0.12
iodobenzene	-2.58	-2.46	-1.73
mean signed error	0.00	-0.07	
mean unsigned error	0.25	0.26	

<sup>a</sup> In all tables “//” denotes “at the geometry computed by”.

For the full set of 248 neutral solutes (including iodine-containing solutes) that was used as the training set for the SM5.2R aqueous model,<sup>17</sup> the SM5.0R model achieves a mean unsigned error in the aqueous solvation free energies of 0.53 kcal/mol using HF/MIDI! geometries and 0.57 kcal/mol when AM1 geometries are employed.

**4.3 Performance: Organic Solvents.** Table 4 shows the performance of the SM5.0R model for various solvent classes over the 1836 nonaqueous-solvent data points in our training set. Using geometries optimized at the Hartree–Fock level with the MIDI! basis set, the mean unsigned error over our training set is 0.38 kcal/mol, and when using AM1 geometries this error increases to only 0.39 kcal/mol. Neither the HF/MIDI! or AM1 geometries produce a systematic error larger than 0.46 kcal/mol for any solvent class in the training set. The

largest systematic error is for the 36 data points in ester solvents where the mean signed error is 0.43–0.46 kcal/mol using either the HF/MIDI! or AM1 geometries. The mean unsigned errors in each of the 18 solvent classes are less than 0.6 kcal/mol when using either set of gas-phase geometries. The last two columns of Table 4 are discussed in Section 4.4.

The errors in the free energies of solvation for the SM5.0R-organic model are broken down by solute class in Table 5. Of the 31 identified solute classes, only the amides, ureas, and bifunctional compounds containing H, C, O, and/or N have mean unsigned errors larger than 0.8 kcal/mol. Of particular interest in Table 5 is the performance of the SM5.0R model using molecular mechanics geometries from the MM3\* force field as coded into MacroModel-version 6.0.<sup>12</sup> MM3\* in MacroModel-version 6.0 does not have parameters for all types of atoms; in particular it does not have the parameters required to optimize the geometry for H<sub>2</sub> or molecules containing a nitro group, but for the 1802 data points in our training set for which MM3\* was able to obtain a gas-phase geometry, SM5.0R achieves a mean unsigned error of only 0.39 kcal/mol and a mean signed error of 0.05 kcal/mol. For this same subset of 1802 compounds, the mean unsigned and mean signed errors for HF/MIDI! geometries are 0.38 and 0.00 kcal/mol, respectively, and the mean unsigned and mean signed errors for AM1 geometries are 0.39 and 0.01 kcal/mol, respectively. The general agreement of the values calculated with the various geometries indicates that the SM5.0R model is relatively stable to the choice of method for calculating the gas-phase geometries, and thus it can be used to predict solvation free energies from any good gas-phase geometry including those produced by molecular mechanics methods.

The results in Tables 4 and 5 confirm that a successful solvation model for nonionic species can be built up based on accessible surface areas computed from 3-D geometries without explicit consideration of electrostatics.

**4.4. Comparison to Other Models.** The only other 3-D solvation models parametrized for as wide a range of solvents as the SM5.0R model are the SM5.4<sup>16</sup> and the SM5.2R<sup>17</sup> models also produced by our group, although other research groups have parametrized models for organic solvents such as chloroform<sup>12,36</sup> and carbon tetrachloride<sup>37</sup> where there is an appreciable amount of experimental data available. Luque et al. in Barcelona have parametrized molecular orbital solvation models based on the formulation of Miertius, Scrocco, and Tomasi<sup>38</sup> for chloroform<sup>36</sup> and carbon tetrachloride;<sup>37</sup> we will label their model B-MST. A second model to which we can compare solvation free energies in chloroform is the GB/SA model.<sup>11,12</sup> The SM5.4, SM5.2R, B-MST, and GB/SA models all use explicit representations of the electrostatic interactions, and they all include surface tensions as well.

Table 6 contains free energies of solvation in chloroform for these models and compares them to the present results. Note that the B-MST and SM5.2R calculations are carried out at gas-phase geometries, as in the present

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

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

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

**Table 4. Performance of the SM5.0R Model for Nonaqueous Solvents by Solvent Class**

solvent class	number of			SM5.0R//HF/MIDI!		SM5.0R//AM1		SM5.4/AM1	
	solvents <sup>a</sup>	solute classes <sup>b</sup>	data <sup>c</sup>	MSE <sup>d</sup>	MUE <sup>e</sup>	MSE <sup>d</sup>	MUE <sup>e</sup>	MSE <sup>d</sup>	MUE <sup>e</sup>
alkanes	11	30	475	-0.05	0.29	-0.06	0.32	-0.10	0.33
cycloalkanes	2	24	106	-0.07	0.35	-0.08	0.36	-0.02	0.41
arenes	12	16	256	-0.05	0.28	-0.07	0.29	0.20	0.42
aliphatic alcohols	12	31	299	-0.03	0.44	0.00	0.45	-0.12	0.57
aromatic alcohols	2	7	12	0.14	0.48	0.16	0.48	0.51	0.71
ketones	4	10	35	0.25	0.42	0.27	0.41	-0.35	0.48
esters	2	8	36	0.43	0.51	0.46	0.53	0.27	0.44
aliphatic ethers	4	19	99	0.09	0.44	0.11	0.43	0.02	0.57
aromatic ethers	3	5	15	-0.33	0.37	-0.29	0.34	-0.02	0.32
amines	2	6	12	-0.21	0.55	-0.18	0.55	0.56	0.66
pyridines	3	5	15	0.03	0.32	0.10	0.35	0.10	0.33
nitriles	2	5	10	0.06	0.50	0.11	0.47	-0.54	0.54
nitro compounds	4	8	27	0.25	0.47	0.24	0.49	-0.65	0.68
tertiary amides	2	5	10	0.28	0.43	0.37	0.47	-0.12	0.33
haloaliphatics	12	27	269	0.04	0.53	0.02	0.53	0.03	0.56
haloaromatics	6	11	106	0.14	0.29	0.15	0.28	-0.30	0.50
miscellaneous acidic solvents	3	5	15	0.10	0.31	0.15	0.36	0.02	0.41
miscellaneous nonacidic solvents	4	12	39	-0.21	0.58	-0.15	0.59	-0.07	0.37
total:	90	31	1836	0.00	0.38	0.01	0.39	-0.05	0.46

<sup>a</sup> Number of solvents in this solvent class. <sup>b</sup> Number of solute classes (identified in Table 5) for which data exists in this solvent class. <sup>c</sup> Total number of solute/solvent data involving this solvent class. <sup>d</sup> Mean signed error over data in this solvent class. <sup>e</sup> Mean unsigned error over data in this solvent class.

**Table 5. Performance of the SM5.0R Model for Nonaqueous Solvents by Solute Functional Group Class**

solute class	number of			SM5.0R//HF/MIDI!		SM5.0R//AM1		SM5.0R//MM3*	
	solutes <sup>a</sup>	solvent classes <sup>b</sup>	data <sup>c</sup>	MSE <sup>d</sup>	MUE <sup>e</sup>	MSE <sup>d</sup>	MUE <sup>e</sup>	MSE <sup>d</sup>	MUE <sup>e</sup>
unbranched alkanes	9	18	76	0.12	0.40	0.25	0.44	0.22	0.43
branched alkanes	5	2	7	0.49	0.52	0.57	0.58	0.56	0.57
cycloalkanes	4	5	13	-0.06	0.35	0.02	0.35	0.01	0.35
alkenes	8	3	18	-0.04	0.19	0.01	0.21	0.02	0.21
alkynes	5	2	9	0.02	0.12	0.05	0.14	0.08	0.14
arenes	9	18	126	-0.17	0.38	-0.12	0.36	-0.11	0.36
alcohols	17	18	369	-0.02	0.35	0.01	0.36	-0.04	0.37
ethers	12	18	81	0.12	0.41	0.16	0.43	0.13	0.41
aldehydes	7	7	32	-0.12	0.54	-0.19	0.57	-0.13	0.54
ketones	12	17	191	-0.10	0.36	-0.16	0.38	-0.09	0.35
carboxylic acids	5	13	119	0.03	0.39	-0.07	0.40	0.06	0.39
esters	13	7	236	-0.02	0.28	-0.14	0.30	-0.03	0.29
bifunctional compounds containing H, C, O	4	7	23	0.41	0.92	0.36	0.88	0.44	0.92
inorganic compounds containing H and O	2	8	20	0.00	0.62	0.00	0.62	0.01 <sup>f</sup>	0.68 <sup>f</sup>
aliphatic amines	11	9	153	0.00	0.24	0.10	0.27	0.09	0.26
aromatic amines	11	11	71	0.16	0.43	0.23	0.44	0.23	0.44
nitriles	4	5	18	0.14	0.47	0.16	0.46	0.17	0.47
nitrohydrocarbons	6	7	32	-0.03	0.22	-0.06	0.23	<i>g</i>	<i>g</i>
amides and ureas	2	5	7	1.32	1.66	1.29	1.69	1.38	1.73
bifunctional compounds containing N	4	2	6	-0.88	0.97	-0.81	0.93	-0.88	0.96
inorganic compounds containing N	2	7	13	-0.21	0.66	-0.26	0.66	-0.24	0.66
thiols	3	4	10	0.37	0.40	0.46	0.46	0.40	0.42
sulfides	6	5	17	-0.21	0.71	-0.09	0.70	-0.18	0.70
disulfides	2	2	3	0.00	0.27	0.10	0.25	0.10	0.28
fluorinated hydrocarbons	5	4	13	0.02	0.51	0.07	0.52	0.07	0.51
chloroalkanes	7	4	22	0.05	0.35	0.04	0.36	0.06	0.37
chloroalkenes	4	3	11	0.70	0.70	0.74	0.74	0.74	0.74
chloroarenes	6	5	29	-0.15	0.30	-0.09	0.29	-0.10	0.29
brominated hydrocarbons	14	5	36	-0.02	0.36	0.00	0.36	0.01	0.36
oiodinated hydrocarbons	9	5	20	-0.02	0.50	0.06	0.47	0.03	0.49
multifunctional halogenated solutes	19	8	55	0.12	0.63	0.12	0.62	0.14	0.65
subtotal applicable to MM3* geometries <sup>h</sup>	220	18	1802	0.00	0.38	0.00	0.39	0.05	0.39
total:	227	18	1836	0.00	0.38	0.01	0.39		

<sup>a</sup> Number of solutes in this solute class. <sup>b</sup> Number of solvent classes (identified in Table 4) for which there are data for this solute class. <sup>c</sup> Total number of solute/solvent data involving solutes in this solute class. <sup>d</sup> Mean signed error over data in this solute class. <sup>e</sup> Mean unsigned error over data in this solute class. <sup>f</sup> The MM3\* model implemented in MACROMODEL-version 6.0 does not have atom types sufficient for optimizing the geometry for H<sub>2</sub>, so the 2 hydrogen molecule data points were omitted when calculating this number. <sup>g</sup> The MM3\* model implemented in MACROMODEL-version 6.0 does not have atom types sufficient to optimize geometries for nitrohydrocarbons. <sup>h</sup> These number were calculated for the 1802 data points for which there are adequate types in the MM3\* model to calculate the gas-phase geometries.

model, but the GB/SA and SM5.4 calculations involve geometries relaxed in solution. The GB/SA model adds solvation to the MM3\* Hamiltonian, and the SM5.4

models in Table 6 add it to either the AM1 or PM3 Hamiltonian. For the 25 molecules for which results are available for all of the models, the SM5.4 models perform

**Table 6.** Experimental and Calculated Free Energies of Solvation in Chloroform,  $\Delta G_S^\circ$  (kcal/mol), for Several Models<sup>a</sup>

solute	SM5.2R/ <sup>b</sup>												
	SM5.0R//		MNDO/d//				...//HF/MIDI!		SM5.4/ <sup>c</sup>		B-MST// <sup>d</sup>		expt $\Delta G_S^\circ$
	HF/MIDI!	AM1	HF/MIDI!	AM1	AM1	PM3	AM1	PM3	6-31G*	AM1	GB/SA <sup>e</sup>		
$\Delta G_S^\circ$	$\Delta G_S^\circ$	$\Delta G_S^\circ$	$\Delta G_S^\circ$	$\Delta G_S^\circ$	$\Delta G_S^\circ$	$\Delta G_S^\circ$	$\Delta G_S^\circ$	$\Delta G_S^\circ$	$\Delta G_S^\circ$	$\Delta G_S^\circ$			
methanol	-2.5	-2.5	-2.7	-2.7	-2.8	-2.7	-2.6	-2.7	-3.3	-2.5	-3.3	-3.3	
ethanol	-3.5	-3.5	-3.6	-3.6	-3.8	-3.7	-3.7	-3.7	-3.7	-3.5	-3.9	-3.9	
1-propanol	-4.3	-4.3	-4.3	-4.3	-4.4	-4.4	-4.3	-4.2	-4.5	-4.2	-4.4	-4.4	
1-butanol	-5.0	-5.0	-5.1	-5.1	-5.1	-5.1	-4.9	-5.1	-5.1	-5.0	-4.9	-5.3	
1-pentanol	-5.8	-5.7	-5.8	-5.8	-5.7	-5.8	-5.7	-5.8	-5.9	-5.8	-5.4	-5.9	
phenol	-6.7	-6.7	-6.8	-6.9	-7.2	-6.9	-7.2	-7.1	-7.3	-7.2	-6.4	-7.1	
1-hexanol	-6.5	-6.4	-6.5	-6.5	-6.4	-6.5	-6.7	-6.9	-6.6	-6.6	-5.9	-6.7	
<i>o</i> -cresol	-7.1	-7.1	-7.2	-7.3	-7.7	-7.4	-7.5	-7.2	-7.5	-7.6	-6.6	-7.6	
<i>p</i> -cresol	-7.2	-7.2	-7.2	-7.3	-7.6	-7.4	-7.8	-7.6	-7.9	-7.8	-6.6	-7.6	
1-heptanol	-7.2	-7.2	-7.2	-7.2	-7.1	-7.2	-7.4	-7.6	-7.4	-7.5	-6.4	-7.5	
propanone	-4.0	-4.1	-4.4	-4.8	-4.5	-4.6	-4.4	-4.4	-5.1	-4.9	-4.6	-4.4	
methyl phenyl ketone	-7.9	-8.0	-8.2	-8.7	-8.4	-8.4	-7.9	-8.0	-8.5	-8.8	-7.4	-7.8	
ethanoic acid	-4.2	-4.4	-5.0	-5.8	-5.2	-5.1	-4.7	-4.8	-5.0	-5.0	-3.8	-4.7	
propanoic acid	-5.0	-5.1	-5.4	-6.2	-5.6	-5.6	-5.2	-5.3	-5.5	-5.5	-4.2	-5.4	
butanoic acid	-5.7	-5.8	-6.1	-6.8	-6.1	-6.1	-5.7	-5.8	-6.2	-6.2	-4.7	-6.0	
methyl ethanoate	-3.8	-4.0	-4.2	-4.9	-4.4	-4.3	-4.8	-4.8	-5.3	-5.4	-4.5	-4.9	
ethyl ethanoate	-4.7	-4.9	-5.1	-5.8	-5.2	-5.2	-5.8	-5.7	-5.9	-5.0	-5.1	-5.6	
ethylamine	-3.5	-3.4	-3.4	-3.4	-3.3	-3.4	-4.2	-4.0	-4.0	-4.1	-3.7	-4.0	
propylamine	-4.2	-4.2	-4.1	-4.1	-4.1	-4.7	-4.8	-4.7	-4.6	-4.9	-4.2	-4.7	
trimethylamine	-3.7	-3.3	-3.6	-3.3	-3.6	-3.5	-4.2	-4.1	-3.6	-3.9	-4.0	-3.9	
butylamine	-5.0	-4.9	-4.9	-4.8	-4.8	-4.8	-5.5	-5.4	-5.4	-5.7	-4.7	-5.4	
diethylamine	-4.7	-4.5	-4.6	-4.5	-4.3	-4.6	-4.7	-5.0	-4.4	-4.7	-4.6	-5.2	
pyridine	-5.5	-5.4	-5.6	-5.7	-6.0	-5.6	-6.6	-6.2	-6.3	-6.4	-6.7	-6.5	
aniline	-6.4	-6.3	-6.5	-6.6	-7.1	-6.9	-7.2	-7.5	-7.4	-6.4	-6.9	-7.3	
<i>p</i> -hydroxybenzaldehyde	-8.9	-9.1	-9.1	-9.6	-9.1	-9.2	-9.5	-9.4	-10.3	-9.6	-6.3	-10.3	
subtotal MSE <sup>f</sup>	0.50	0.50	0.35	0.15	0.24	0.28	0.10	0.10	-0.05	0.05	0.65		
subtotal MUE <sup>f</sup>	0.50	0.51	0.42	0.48	0.37	0.40	0.21	0.18	0.22	0.33	0.69		
<i>m</i> -nitrophenol <sup>g</sup>	-9.1	-9.3	-9.5	-10.1	-9.0	-9.4	-10.4	-10.1	-10.0	-10.5	<i>h</i>	-10.5	
<i>p</i> -nitrophenol <sup>g</sup>	-9.1	-9.2	-9.7	-10.4	-9.2	-9.7	-10.6	-10.3	-10.1	-10.6	<i>h</i>	-11.0	
mean Signed Error <sup>i</sup>	0.58	0.57	0.41	0.17	0.34	0.34	0.11	0.13	0.00	0.06	<i>h</i>		
mean Unsigned Error <sup>i</sup>	0.59	0.59	0.47	0.49	0.47	0.46	0.21	0.21	0.26	0.32	<i>h</i>		
MSEfor SM5.0R set <sup>j</sup>	0.32	0.36	0.17	-0.04	0.04	0.11	0.06	0.08	<i>k</i>	<i>k</i>	0.28 <sup>l</sup>		
MUEfor SM5.0R set <sup>j</sup>	0.64	0.61	0.58	0.59	0.57	0.58	0.45	0.42	<i>k</i>	<i>k</i>	1.00 <sup>l</sup>		

<sup>a</sup> // denotes "calculated at a geometry obtained by". <sup>b</sup> from Hawkins et al.<sup>17</sup> <sup>c</sup> from Giesen et al.<sup>16</sup> <sup>d</sup> from Luque et al.<sup>35</sup> <sup>e</sup> The partial charges and geometries used in these calculations were obtained from MM3\*, the geometries were optimized in solution by adding solvation to the MM3\* Hamiltonian, and the calculations were performed with MACROMODEL-version 6.0. <sup>f</sup> mean signed error (MSE) or mean unsigned error (MUE) in 25 molecules listed above. <sup>g</sup> Not used in the parametrization of the SM5.0R model. <sup>h</sup> The MM3\* model implemented in MACROMODEL-version 6.0 does not have atom types sufficient to optimize geometries for nitrohydrocarbons. <sup>i</sup> For the 27 molecules listed above. <sup>j</sup> Mean signed error (MSE) or mean unsigned error (MUE) for the 92 molecules in the SM5.0R chloroform training set. <sup>k</sup> Not available. <sup>l</sup> The two nitrohydrocarbon compounds were excluded when calculating this number.

best with mean unsigned errors of 0.21 and 0.18 kcal/mol for the AM1 and PM3 parametrizations, respectively, although we caution the reader that small differences in mean errors are not necessarily meaningful if one considers the statistical uncertainties inherent in any selection of a test set. The B-MST model has only slightly larger mean unsigned errors of 0.22 and 0.33 kcal/mol when using HF/6-31G\* and AM1 geometries, respectively. The SM5.2R model has mean unsigned errors of between 0.37 and 0.48 kcal/mol for this set, while the SM5.0R model has a mean unsigned error of about 0.50 kcal/mol for both HF/MIDI! and AM1 geometries. The GB/SA model performs least well over this set with a mean unsigned error of 0.69 kcal/mol. In assessing these results, one should note that the SM5.4, B-MST, and GB/SA models were specially parametrized for chloroform, whereas in the SM5.0R and SM5.2R models, the parametrization was carried out for all organic solvents simultaneously. Thus, in any single solvent it is possible for SM5.0R to have a significant systematic error even though the solvent descriptors used in the model attempt to minimize such errors across the full range of solvents. In this respect, we have found that chloroform is a particularly difficult solvent in that it does not follow systematic expectations as well as most solvents. In particular, the SM5.0R model predicts chloroform solva-

tion free energies which are typically about 0.3 kcal/mol higher (less negative) than the experimental value over the applicable portion of the data set. Nevertheless, we believe that simultaneously parametrizing over a large group of solvents provides a model which is more robust over the universe of all organic solvents and can to a certain extent compensate for the limited experimental free energy data available for most solutes in most solvents.

The difference between the performance of the SM5.0R model and the quantum mechanical SM5 models lessens somewhat when we consider all 92 compounds which were used in the chloroform portion of our training set, where the mean unsigned errors for all the SM5 models listed range from 0.42 to 0.64 kcal/mol. Over the 90 molecules from this expanded set that do not contain a nitro group, the GB/SA model achieves a mean unsigned error of 1.00 kcal/mol, compared to a mean unsigned error of 0.62 kcal/mol for the SM5.0R//HF/MIDI! model for this same set. Most of the mean signed errors are positive corresponding to systematic undersolvation. This is less surprising in a model like SM5.0R that is parametrized for a wide range of solvents than for a model like GB/SA that is specifically parametrized for chloroform.

Overall, the SM5.0R model for predicting solvation free energies in organic solvents performs similarly to the

**Table 7. Free Energies of Solvation (kcal/mol) of Eight Solutes in Ten Solvents**

solute	solvent									
	hexane	decane	benzene	butanol	octanol	4M2P <sup>a</sup>	butyl acetate	ethyl ether	CHCl <sub>3</sub>	CCl <sub>4</sub>
toluene	-4.64 <sup>b</sup>	-4.63	-5.37	-4.58	-4.63	-5.40	-5.32	-5.34	-5.32	-5.76
	-4.84 <sup>c</sup>	-4.65	-5.32	-4.50	-4.55	na <sup>d</sup>	na	-5.23	-5.76	-5.12
ethanol	-2.50	-2.49	-3.27	-5.01	-5.04	-4.24	-4.03	-3.96	-3.51	-3.18
	-2.61	-2.44	-3.42	-5.02	-4.36	na	-3.97	-4.41	-3.94	-2.96
phenol	-5.50	-5.51	-6.75	-8.22	-8.30	-7.96	-7.67	-7.52	-6.70	-6.59
	-5.49	-5.50	-7.12	na	-8.69	-9.38	-8.96	-8.75	-7.14	-6.14
ethanoic acid	-3.20	-3.20	-4.28	-6.49	-6.53	-5.86	-5.54	-5.38	-4.25	-3.93
	-2.83	na	-4.02	-6.81	-6.35	-6.33	-6.11	-4.26	-4.74	-3.64
propylamine	-3.11	-3.08	-3.71	-4.73	-4.76	-3.96	-3.86	-3.89	-4.24	-3.99
	-3.13	-2.96	-3.68	-5.04	-4.77	na	na	-3.65	-4.73	-3.59
pyridine	-4.62	-4.64	-5.23	-5.18	-5.25	-5.06	-5.01	-5.02	-5.51	-5.57
	-3.81	na	-5.28	na	-5.34	-5.33	-5.31	-4.81	-6.45	-5.01
chlorobenzene	-5.04	-5.04	-5.75	-4.93	-5.00	-5.63	-5.56	-5.58	-5.74	-6.15
	-5.14	-4.93	na	na	-5.00	na	na	-5.42	-5.45	-5.21
bromobenzene	-5.69	-5.70	-6.56	-5.92	-6.00	-6.64	-6.53	-6.51	-6.47	-6.87
	-5.66	-5.43	na	na	-5.46	na	na	-5.99	-6.07	-5.85

<sup>a</sup> 4-Methyl-2-pentanone. <sup>b</sup> Upper value: SM5.0R/HF/MIDI! <sup>c</sup> Lower value: experiment. <sup>d</sup> na: not available.

previous SM5.4 and SM5.2R models over the full training set molecules in all the solvents. Thus the main advantage of the SM5.0R model is *not* higher accuracy but rather that the model is very fast. By virtue of its being reasonably stable to the choice of method used for obtaining the gas-phase geometry, the SM5.0R model can be useful for quick estimations of solvation effects on many systems including those where quantum mechanical methods for optimizing geometries are prohibitively expensive. In fact the method does not even require a quantum mechanical wave function at a single geometry.

Considerable previous work has been focused on the octanol/water partition coefficient (usually called  $\log P$  or  $\log P_{o/w}$ ), especially using quantitative structure-activity relationships.<sup>39</sup> This subject is especially important because octanol is considered a simple mimic of cell membranes. Although the present study is more general, the calculation of  $\log P_{o/w}$  values is one possible application. However, we can also adjust the solvent descriptors to reflect partitioning into solvents that mimic biomembranes even better than octanol does.<sup>40</sup>

**4.5. Examples of Solvent Dependence.** Table 7 illustrates the kind of results we obtain for the solvent dependence of solvation free energies by comparing theory and experiment for a few selected molecules in a few selected solvents. (Data for *all* cases are in Supporting Information.) Examination of the table shows that the trends with solvent are predicted, on average, about as accurately as the absolute values of the free energies of solvation.

There is no particular tendency for the trend with solvent to be more accurate because errors do not tend to cancel out in any systematic fashion. However, the

(39) Abraham, M. H.; Chadha, H. S.; Lee, A. J. *J. Chromatogr. A* **1994**, *685*, 203. Basak, S. C.; Gute, B. D.; Grundwald, G. D. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1054. Mannhold, R.; Rekker, R. F.; Sonntag, C.; Ter Laak, A. M.; Dross, K.; Polymeropoulos, E. E. *J. Pharm. Sci.* **1995**, *84*, 1410. Leo, A. J. *Methods Princ. Med. Chem.* **1996**, *4*, 157. Mannhold, R.; Dross, K. *Quant. Struct.-Act. Relat.* **1996**, *15*, 403. van de Waterbeemd, H.; Mannhold, R. *ibid.* **1996**, *15*, 410. Gombar, V. K.; Enslin, K. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1127. Wang, R.; Fu, Y.; Lai, L. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 615. Bodor, N.; Buchwald, P. *J. Phys. Chem. B* **1997**, *101*, 3404. Tackos-Novak, K. *Hung. Acta Pharm. Hung.* **1997**, *67*, 179. Devillers, J.; Domine, D. *SAR QSAR Environ. Res.* **1997**, *7*, 195.

(40) Chalmers, C. C.; Giesen, D. J.; Hawkins, D. G.; Vaes, W. H. J.; Cramer, C. J.; Truhlar, D. G. In *Rational Drug Design*; Truhlar, D. G., Howe, W. J., Hopfinger, A. J., Blaney, J. M., Demmkoehler, R. A., Eds.; Springer: New York, in press.

**Table 8. Average Computer Time<sup>a</sup>**

method	type of gradient <sup>b</sup>	computer program	average time (s)
SM5.4/AM1	numerical	AMSOL 6.0 <sup>c</sup>	18.4
SM5.4/AM1	analytical	AMSOL 6.5 <sup>d</sup>	2.47
SM5.2R/AM1	not required <sup>e</sup>	AMSOL 6.5	1.32
SM5.0R	not required <sup>e</sup>	OMNISOL 1.0	0.020

<sup>a</sup> All computer times were obtained on the same machine, an SGI IRIS Indigo with an R10000 chip; the time is an average over all 1836 data points in the present parametrization suite. <sup>b</sup> The gradient of the energy with respect to the nuclear coordinates is required for calculations involving geometry optimization. <sup>c</sup> Timings for runs based on numerical gradients would be essentially the same with later versions (6.1.1, 6.5) of AMSOL. <sup>d</sup> Analytical gradients were introduced in AMSOL in version 6.1 for aqueous solvents and in version 6.5 for nonaqueous solvents. The timings would be essentially the same with either version. <sup>e</sup> For the R models we assume a gas-phase geometry is available as input, and the timings refer to the calculation of only the free energy of solvation.

sign of the free energy of transfer tends to be correct in most cases. For example, the model correctly predicts that transfers of ethanoic acid, propylamine, and pyridine from octanol to ethyl ether are endergonic, whereas transfers of toluene, chlorobenzene, and bromobenzene from octanol to ethyl ether are exergonic.

## 5. Computer Program and Timings

The method proposed here and all the parameters required for calculations on aqueous and nonaqueous solutions have been incorporated in a computer code called OMNISOL.<sup>22</sup> The only computationally intensive step is the calculation of accessible surface areas of the atoms; this step is carried out by the Analytic Surface Area (ASA) algorithm<sup>41</sup> presented elsewhere. The OMNISOL computer code is freely available over the Internet.<sup>23</sup>

Table 8 compares computer timings for four types of SM5 calculations. SM5.4 calculations include quantum mechanical electrostatics and geometry optimization in solution, SM5.2R calculations include the former but not the latter, and SM5.0R calculations include neither. In all cases we assume that an optimized gas-phase geom-

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



etry is available as a starting point. The first two rows refer to SM5.4 calculations. Even with an inexpensive semiempirical Hamiltonian (AM1) and analytic gradients, the computer time for the SM5.4 calculations on the 1836 data in the present parametrization suite is 4435 s (2.47 s per free energy), as compared to 37 s (0.020 s per free energy) for SM5.0R, a difference of a factor of 120. The timings for SM5.2R show that a factor of 1.9 comes from the elimination of geometry optimization, and a factor of 66 comes from the elimination of quantum mechanics and electrostatics. Only in light of these timings may the success of the present model be fully appreciated. We note that modern desktop computers are almost as fast as the computer used for the timings. Furthermore the memory requirements for the SM5.0R models are much less than those for models employing quantum mechanics so the calculations can be run even on low-memory desktop computers with slow bus speed. Furthermore, SM5.0R does not involve atom typing, which is inconvenient and often leads to the "missing parameter" problem (atom type not supported) in conventional molecular mechanics. Thus one can obtain useful accuracy for essentially any organic molecule containing the heteroatoms N, O, F, S, Cl, Br, and/or I in virtually any organic solvent faster than one can remove one's finger from the return key on a desktop computer.

## 6. Concluding Remarks

As discussed in a preceding paper in this series,<sup>17</sup> it is useful to have a variety of levels available for modeling molecular structures and energetics in the condensed phase, ranging from the most complete models (which are usually also the most complicated, most expensive, and—one hopes, although it is not always true, the most accurate) to the simplest models (which are also usually the easiest to use, least expensive, and most widely applicable). For aqueous solutions, we now have models based on class IV and class II charges, SM5.4<sup>15</sup> and SM5.2R,<sup>17</sup> respectively, and also a model based on only using atomic surface tensions with the electrostatics implicit, SM5.0R.<sup>21</sup> We have previously extended the SM5.4<sup>16</sup> and SM5.2R<sup>17</sup> models to all organic solvents, and the present paper extends the SM5.0R model to all

organic solvents. We have summarized some of the considerations involved in choosing levels and methodologies for solvation calculations in the final section of a preceding paper<sup>17</sup> in the SM5 series. For example, where explicit consideration of partial charges and geometry changes in solution are important, one would prefer the SM5.4 model. The main advantages of the method in the present paper are simplicity and speed of computation. Considering the simplicity of the method, the results are amazingly accurate for the data used for parametrization. The reverse side of the coin is that one is not sure how much to trust the method for compounds that are significantly different from any in our parametrization set. But there are many cases one does not expect conformation changes or other significant geometrical changes upon dissolution, and our parametrization set is very broad, containing all common organic functional groups for which any free energy of solvation data are available, and so we anticipate a wide range of usefulness of the model even with this caveat. As discussed previously,<sup>21</sup> the SM5.0R model is capable of very *rapid* evaluations of solvation free energies. Therefore, it is especially well suited for calculations on large systems, such as proteins, or for situations where a large number of calculations must be performed, such as scoring hypothetical or virtual combinatorial libraries.

**Acknowledgment.** This work was supported in part by the National Science Foundation through grant no. CHE94-23927, by the Army Research Office through grant no. DAAH-04-93-G-0036, and by the National Institute of Standards and Technology through the general computational methodology provisions of an Advanced Technology Project subcontract with Phillips Petroleum Company.

**Supporting Information Available:** The Supporting Information contains a table that gives the calculated and experimental free energies of solvation at the HF/MIDI!, AM1, and MM3\* geometries for the 1836 cases used in the parametrization of the SM5.0R model (65 pages). This material is contained in libraries on microfiche, immediately follows this article in the microfilm version of the journal, and can be ordered from the ACS; see any current masthead page for ordering information.

JO980046Z