Accurate Transferable Model for Water, *n*-Octanol, and *n*-Hexadecane Solvation Free Energies

A. J. Bordner, C. N. Cavasotto,† and R. A. Abagyan*

The Scripps Research Institute, 10550 North Torrey Pines Road, Mail TPC-28, San Diego, California 92037

Received: July 3, 2002; In Final Form: August 13, 2002

We present a fast continuum method for the calculation of solvation free energies. It is based on a continuum electrostatics model with MMFF94 atomic charges combined with a nonelectrostatic term, which is a linear function of the solvent-accessible surface area. The model's parameters have been optimized using sets of 410, 382, and 2116 molecules for gas−water, gas−hexadecane, and water−octanol transfer, respectively. These are the largest, most diverse sets of molecules used to date for a similar solvation model. The model's predictive power was verified by using 90% of the molecule set for training and the remainder as a test set. The average test set errors differed by only about 1% from the average training set error, thus demonstrating the transferability of the parameters. The root-mean-square error for gas−water, gas−hexadecane, and water−octanol transfer are 0.53, 0.38, and 0.58 log *P* units, respectively. Because the solvation calculation takes on average only about 0.34 s per molecule on a 700 MHz Pentium CPU and contains atom types for essentially all drug molecules, it is suitable for real-time calculations of the ADME properties of molecules in virtual ligand screening libraries.

1. Introduction

The value of the free energy to transfer a solute from vapor phase into a solvent is important for understanding a wide range of biochemical processes. For example, the logarithm of the partition coefficient between water and *n*-octanol (log P_{ow}), which is proportional to the corresponding transfer free energy, is used extensively as a predictor of the degree of adsorption of drugs and environmental toxins by the body.[1−4] Another example is the use of gas−water transfer free energies of small molecules to parametrize the solvation energy term in the potential energy function used in protein molecular dynamics simulations.[5]

Most of the methods currently used for evaluating water−octanol partition coefficients are based on the assumed additivity of chemical fragments. The most widely used approach involves a "fragment constant" methodology in which log P_{ow} is calculated as the sum of individual contributions by atom types, bond types, or chemical fragments.[6] For large and flexible molecules, in which a conformationally dependent estimate of solvation energy is required, the contribution of each group may be weighted according to its solvent-accessible surface area.[5,7] Unfortunately, all of these methods share serious conceptual and practical limitations. Because the contributions of individual groups are generally not additive, a large number of empirical correction factors are required to attain good agreement with experimental data. In addition, such methods cannot be applied to compounds containing unusual chemical fragments that are not present in the training set.[8] Finally, it is difficult to evaluate or improve many recent fragment-based methods because their parameters are not published in the literature.

A second approach, which avoids some of the problems listed above, is to construct a quantitative structure−activity relationship (QSAR) by identifying empirical correlations between log P_{ow} and a variety of whole-molecule descriptors.[8−12] Attempts to compare the accuracy of these diverse methods can be somewhat misleading. Published performance parameters (such as standard deviation and regression coefficient) depend heavily on the validation sets used, which vary considerable in their size, degree of overlap with the training set, and inherent complexity. For complex methods that employ hundreds of parameters and extensive training sets, such as CLOGP, extensive out-of-sample results are often difficult to obtain. Also, the fact that the dispersion in different experimental log P_{ow} values for some compounds is much larger than the accuracy of the method suggests overfitting of the model.[13]

Macroscopic solvation models provide an alternative method for describing solvent−solute interactions that is both physically reasonable and computationally tractable. Several investigators have used these models with atomic point charges to estimate gas−water transfer free energies.[14−17] More recently, this approach has been extended to include gas−octanol[18] and water−octanol[19] systems. Although the results do not attain the accuracy that has been claimed for some empirical methods, the level of agreement is close enough to be of practical use. In addition to offering physical insight into the energetic contributions to the solvation free energy, these models have fewer empirical parameters and are applicable to a wider range of chemical structures.

There are also methods to include solvent polarization in ab initio quantum mechanical calculations of the electrostatic component of the solvation free energy. These methods include the polarizable continuum (PCM) model,[20−22] the polarizable conductor (COSMO) model[23−25] and the isodensity PCM model.[26] Although these methods calculate the solute charge density more accurately, they require considerably more computational time than atomic point charge models. This is one crucial issue in applications in which the solvent free energy

† Present address: Molsoft LLC, 3366 North Torrey Pines Court, San Diego, CA 92037.

10.1021/jp0264477 CCC: $22.00
Published on Web 10/02/2002

calculation needs to be done for a large number of molecules, such as a combinatorial library for screening candidate drug compounds. Another issue is that a nonelectrostatic contribution must be added to the ab initio electrostatic contribution to calculate the experimentally observable solvation free energy. Although there has been an attempt to calculate part of this contribution (the dispersion and repulsion part) using quantum mechanical methods,[27] the cavitation part still needs to be estimated by nonquantum mechanical methods from the Pierotti−Claverie formula.[28,29] These arguments render the ab initio approach as not the best choice when a fast determination of partition coefficients needs to be performed on a very large compound database.

Because we are interested in a fast yet reasonably accurate method to calculate the solvation free energy, we will use a model with atomic point charges. We have applied this continuum solvation model to predicting gas−water, gas−hexadecane, and water−octanol transfer free energies. It is based on an electrostatics contribution using MMFF94 atomic charges and a fast boundary element method solution of the Poisson equation[30] combined with a nonelectrostatic surface tension term. The parameters of the model have been optimized using experimental transfer free energy data. The data sets are larger than those used in previous studies of continuum solvation models and contain compounds with a diverse set of functional groups. Because the MMFF94 atom types include all of those found in drug molecules, the method should be useful in the real-time screening of large virtual ligand screening libraries. The accuracy of the model in predicting the solvation free energies of molecules outside the training set has also been studied by optimizing the parameters using 90% of the molecules and then calculating the solvation free energy for the remainder. We find that the average test set error differs by only about 1% from the average training set error, thus demonstrating the transferability of the fit parameters to molecules outside of the training set.

First, we review the theory of the solvation model and the MMFF94 force field parameters that are used. Next, the selection of the molecule sets, parameter optimization method, and cross-validation scheme are described. Finally, we present the results, discuss the physical basis for the parameter values, and give suggestions for future improvements.

2. Solvation Free Energies

The solvation free energy of tranferring a compound between two phases, A and B, may be decomposed into a sum of contributions along a thermodynamic path in which the molecule is first discharged in phase A, then transferred into the phase B, and finally charged in the phase B.[31] Thus the total solvation energy is decomposed as $\Delta G_{solv} = \Delta G_{el} + \Delta G_{nel}$ where the electrostatic contribution, ΔG_{el}, is from the first and last steps and the remaining nonelectrostatic term, ΔG_{nel}, from the second step. The phases may be either the gaseous state of the compound or a solution of the compound in a liquid. Because the limit of low concentration of the compound is assumed, interactions between solute molecules are not included.

For calculating ΔG_{el}, the molecule is represented by atomic point charges inside a cavity with a uniform dielectric constant inside and an external region with the corresponding dielectric constant of the solvent. The molecular cavity is defined by solvent-excluded surface, the region in which a spherical probe of the approximate size of a solvent molecule is excluded when the minimum separation of the probe and other atoms is the sum of their respective radii.[32] The electrostatic contribution to the solvation free energy is

$$\Delta G_{el} = \frac{1}{2}\sum_i q_i[\phi(\vec{x}_i;\{r_i\},\epsilon_{in},\epsilon_B) - \phi(\vec{x}_i;\{r_i\},\epsilon_{in},\epsilon_A)] \quad (2.1)$$

where $\{q_i\}$ are the atomic charges, $\{r_i\}$ are the atomic radii, ϵ_{in} is the internal dielectric constant, $\epsilon_{A,B}$ are the dielectric constants of the two phases, and $\phi(\vec{x}_i)$ is the electric potential at the charge center, $\vec{x}_i$. The model parameters are shown explicitly because they will be optimized. $\epsilon_A = 1$ for the case of transfer from the gaseous state.

A simple constant surface tension model is used for the nonelectrostatic contribution from the second step of the thermodynamic path

$$\Delta G_{nel} = \sigma A + b \quad (2.2)$$

where A is the solvent-accessible surface area, which is the surface defined by the center of a probe atom as it moves along the surface. This model is based on the assumption that the nonelectrostatic contribution to the free energy is proportional to the size of the first solvation shell surrounding the solute molecule, which is approximated by the solvent-accessible surface area.[33,34] This model for ΔG_{nel}, while simple, avoids the problems inherent in the approximate cancellation of a large negative dispersion term and large positive cavity term, as used in some approaches. The calculation of the solvation free energy described in the following sections used the algorithm described in ref 35 to calculate the solvent accessible surface.

The partition coefficient for solvents A and B is defined as the ratio of solute concentrations in each solvent at equilibrium, $P_{AB} = X_B/X_A$, and it is related to the transfer free energy by

$$\Delta G_{AB} = -RT \ln P_{AB} \quad (2.3)$$

At an ambient temperature of $T = 298$ K, $\Delta G_{AB} = -1.3635 \log P_{AB}$ (kcal/mol).

2.1. MMFF94 Force Field. Atomic charges and van der Waals (VdW) parameters from the MMFF94 force field were used.[36] This force field was chosen because it contains parameters for a diverse set of molecular functional groups, which encompass all molecules in our test sets, and is publically available. The atomic charges in the MMFF94 force field are calculated using bond charge increments ω_{IJ}, which describe the polarity of the bond between atom types I and J. The charge q_i on a particular atom of type I is

$$q_i = q_I^0 + \sum_{j=\text{bonded atoms}} \omega_{JI} \quad (2.4)$$

where q_I^0 is the formal charge, which is zero for all molecules considered here. The bond charge increments, ω_{IJ}, for MMFF94 were derived by fitting ab initio scaled vacuum dipole moments, calculated using HF/6-31G(d), to experimental values. The ab initio dipole moments were scaled by a factor of 1.1 before fitting to reproduce the larger dipole moments in water. One may argue that this scaling results in MMFF94 charges that are inappropriate for molecules that are not in aqueous solution. We attempted to compensate for this in the calculation of the gas−water solvation free energy by multiplying all charges in the calculation of the gaseous phase electrostatic energy by a factor of 0.91, however this did not improve the agreement with experimental data. We discuss below the cases in which overestimation of charges can be a problem.

The MMFF94 force field was also used to determine the solvent-accessible surface used for the nonelectrostatic contribution. The distance from a surface atom to the nearest point on the solvent-accessible surface was defined to be the separation between the surface atom and a water oxygen atom (atom type 70) at which the van der Waal's potential is $kT = 0.6$ kcal/mol above the minimum. These distances for all MMFF atom types included in the molecule sets are given in Table 1.

2.2. Molecule Sets. A set of 410 molecules for which the experimental gas−water log *P* values were given in ref 37 and references therein was used to determine the gas−water solvation parameters. The molecular conformations for all molecule sets were optimized using the MMFF94 force field in a vacuum by the ICM program.[38] The experimental gas−hexadecane log *P* values for 382 compounds in this set are given in the same reference and were used for calculating the gas−hexadecane solvation parameters. Both of these sets contained molecules with 25 MMFF atom types. The molecule set for determining the water−octanol parameters consisted of compounds with recommended (starred) log P_{ow} values in ref 13 and for which we could obtain the structures from public sources such as the NCI Open Database (http://cactvs.cit.nih.gov/ncidb2/download.html). Also because, according to ref 39, log P_{ow} may be reliably measured by the shake-flask method for $-2 <$ log $P_{ow} < 4$ and by high performance liquid chromatography for $0 <$ log $P_{ow} < 6$, all compounds with log P_{ow} values outside the intersection of these ranges, namely, $0 <$ log $P_{ow} < 4$, were removed. The remaining set had 2116 molecules comprising 42 MMFF atom types. The compounds include many biologically active compounds such as drugs that are predominantly complex polycyclic molecules. The experimental free energy values and optimized molecular structures for all compounds are provided as Supporting Information.

2.3. Parameter Optimization and Cross-Validation. The following parameters in the solvation free energy model of eqs 1 and 2 were optimized for each of the gas−water, gas−hexadecane, and water−octanol molecule sets: ϵ_{in}, σ, b, and the electrostatic radii, $\{r_i\}$. The optimum parameters were calculated by performing a local minimization[40] of the root-mean-square (rms) difference between the calculated and experimental ΔG_{solv} starting from 20 random starting values. The electrostatic potential in eq 1 was calculated using a fast boundary element method solution of the Poisson equation as implemented in ICM.[30] The resulting optimal electrostatic radii, $\{r_i\}$, are given in Table 1 and the optimal values of ϵ_{in}, σ, and b are given in Table 2. The solvent dielectric constants 78.5, 2.04, and 9.87 were used for water, hexadecane, and octanol, respectively.

To cross-validate the parameters, each of the molecule sets were divided into 10 randomly selected, approximately equal parts and 10 pairs of training and test sets were created by including each of the subsets in a test set and the remaining molecules in the corresponding training set. The parameters were then optimized using each training set, and the rms error in ΔG_{solv} was evaluated for the corresponding test set.

3. Results and Discussion

3.1. Reproducibility of Model Accuracy. Plots of the correlations and residual errors between the calculated and experimental free energies for gas−water, gas−hexadecane, and water−octanol transfer are shown in Figures 2, 3, and 4, respectively. The statistics for the deviation between the calculated and experimental transfer free energies are given in Table 3. The averages of the test and training set rms errors for the 10 different sets are given. The standard deviation in the rms error values for the test set is also given as a measure of the reproducibility of the stated rms error. As expected, this value is smaller for the larger sets used in the water−octanol calculations. As mentioned above, the test set solvation free energy values are calculated using the parameters optimized with the corresponding training set. Because the average test set error differs by only about 1% from the average training set error, this suggests that these solvation parameters can be used to predict solvation free energies for molecules of comparable size outside the training set with similar accuracy.

3.2. Outliers and Anomalous Electrostatic Radii. The number of outliers the calculated ΔG_{solv} of which differed from the experimental value by more than 1 log *P* unit (1.36 kcal/mol) were 23, 7, and 158 for gas−water, gas−hexadecane, and water−octanol transfer, respectively. There was a preponderance

TABLE 1: Minimum Solvent-Accessible Surface (SAS) Distance and Optimal Electrostatic Radii, $\{r_i\}$, for Each of the Solvents

element	MMFF atom type	SAS distance (Å)	water−octanol $\{r_i\}$ (Å)	gas−water $\{r_i\}$ (Å)	gas−hexadecane $\{r_i\}$ (Å)
H	5	2.58	0.81	1.18	1.14
	21	2.47	1.42	0.96	1.56
	23	2.47	1.20	1.34	1.18
	24	2.47	1.89	1.40	1.46
	27	2.47	0.99		
	28	2.47	0.92	0.94	0.87
	29	2.47	1.49	1.01	0.80
	71	2.47	0.99		
C	1	3.04	1.21	1.46	1.89
	2	3.15	1.72	1.92	1.23
	3	3.06	1.18	2.18	1.08
	4	3.13	1.85	1.84	1.67
	22	3.06	1.66	1.91	1.90
	37	3.15	1.82	1.71	0.76
	63	3.15	1.52		
	64	3.15	1.46		
N	8	3.08	1.01	1.17	1.43
	9	2.99	1.69		
	10	3.03	1.09		
	38	2.97	1.22	1.70	1.75
	39	3.07	1.35		
	40	3.03	1.50	1.50	1.45
	42	3.03	1.28	1.90	2.16
	43	3.03	1.13		
	45	3.09	2.15	1.87	1.47
	46	3.14	1.27		
	65	3.03	1.46		
	66	2.92	1.27		
	67	3.01	1.39		
O	6	2.90	0.99	1.38	1.48
	7	2.88	1.23	1.35	2.20
	32	2.93	1.34	1.76	1.66
	59	2.88	1.28		
F	11	2.71	1.66	2.14	2.31
Cl	12	3.25	2.21	1.93	1.34
Br	13	3.40	2.28	1.55	1.00
I	14	3.63	2.16	1.17	1.38
P	25	3.09	1.86		
S	15	3.38	1.91	1.62	1.46
	16	3.52	1.92		
	18	3.21	1.78		
	44	3.38	2.26		

TABLE 2: Optimal Internal Dielectric Constant, ϵ_{in}, and Surface Tension Parameters for Each of the Solvents[a]

parameter	gas−water	gas−hexadecane	water−octanol
ϵ_{in}	2.21	2.95	3.52
σ (kcal/(mol Å^2))	0.00387	−0.0242	−0.0158
b (kcal/mol)	0.698	3.55	0.919

[a] Determined along with the electrostatic radii of Table 1.

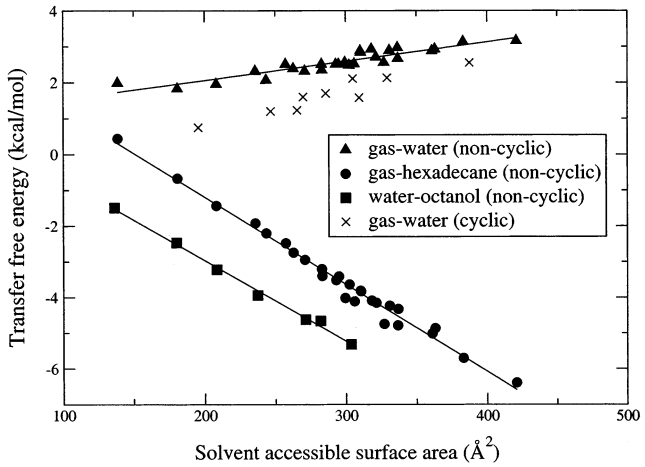

Figure 1. Alkane transfer free energy. Experimental gas−water, gas−hexadecane, and water−octanol transfer free energies for noncyclic alkanes and gas−water transfer free energies for cyclic alkanes are plotted versus solvent-accessible surface area. Linear regression fits to the noncyclic alkane data are shown by lines, and fit parameters are given in the text.

TABLE 3: Statistics for the Deviation of the Calculated Transfer Free Energy, ΔG_{solv}^{calcd}, from the Experimental Value, ΔG_{solv}^{exptl}[a]

$\Delta G_{solv}^{calcd} - \Delta G_{solv}^{exptl}$ statistic	gas−water	gas−hexadecane	water−octanol
complete set rms value	0.716	0.514	0.785
average training set rms value	0.714	0.512	0.785
average test set rms value	0.723	0.507	0.788
standard deviation in test set rms value	0.134	0.084	0.028

[a] All values are in kcal/mol. Their method of calculation is described in the text.

of molecules containing a nitro group and halogenated compounds in the outliers, the corresponding atom types of which have anomalous electrostatic radii.

Most of the electrostatic radii in Table 1 do not drastically differ from van der Waals radii. One unusual trend is the decrease in the radii of halogens with increasing atomic number for gas−water and gas−hexadecane transfer. Starting the minimization from random values for these radii did not significantly change these values. This is due to the general trend of decreasing (more favorable) solvation free energy for higher atomic mass halogen compounds. For example, ΔG_{solv} decreases from −0.55 to −0.82 to −0.89 kcal/mol for chloro-, bromo-, and iodomethane gas−water transfer, respectively. Likewise, the gas−hexadecane transfer free energies for the same compounds are −1.59, −2.22, and −2.87, respectively. This trend is repeated for all monohalogenated alkanes and, except for the low ΔG_{solv} of fluorobenzene, for monohalogenated benzene as well. This same effect may be seen in the noble gas gases, the solvation free energies of which monotonically decrease from 2.76 kcal/mol for helium to 0.881 kcal/mol for radon and the gas−hexadecane solvation free energies of which also monotonically decrease from 2.37 kcal/mol for helium to −1.20 kcal/mol for radon.[37] Because the enthalpy of transfer also decreases with increasing atomic mass, this implies that the solvent−solute interaction becomes more favorable as the atomic weight increases. A possible explanation for this is that the increased polarizability of the larger halogen atoms leads to a larger favorable van der Waals interaction with the solvent. Because this effect is not included in the solvation model, it

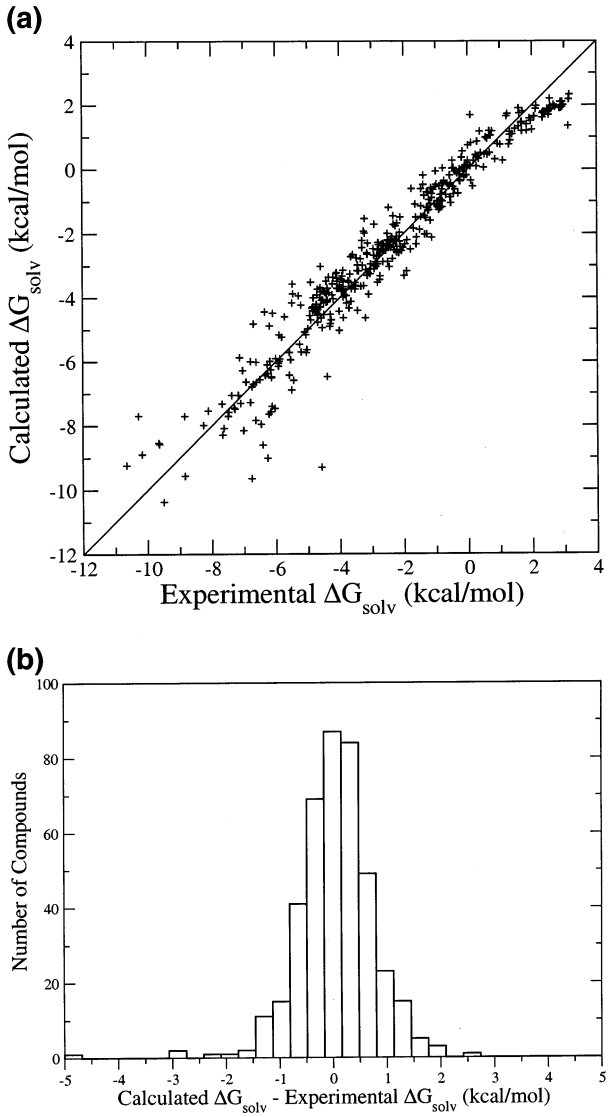

Figure 2. Gas−water transfer free energy: (a) calculated versus experimental gas−water transfer free energy values for the complete 410 molecule set; (b) a histogram of the differences between these values for the same set.

may be compensated by a decrease in the electrostatic radii of the higher atomic mass halogens and consequent decrease in ΔG_{solv}.

Another anomalous electrostatic radius for water−octanol transfer is the large value, 2.15 Å, for the nitro-group nitrogen (type 45). An ab initio calculation of the charges for the 243 compounds containing this atom type was performed for comparison. The geometry was first optimized using B3LYP/6-31G(d), and then the atomic charges were calculated using the Merz−Singh−Kollman scheme[41,42] with HF/6-31G(d,p) as implemented in Gaussian 98.[43] The average charge for type 45 atoms was 0.71 for the Merz−Singh−Kollman charges and 0.90 for the MMFF94 charges. Because the MMFF94 charges are, on average, about 20% larger, which is more than the 10% increase over HF-derived charges used in MMFF94 charge derivation, the large electrostatic radius may be compensating for the larger charge. Although the type 45 radius for gas−water transfer is not particularly large, the high charge for this atom type may explain why the largest outlier, 2-nitrophenol, has a calculated ΔG_{solv} about 4.7 kcal/mol lower than the experimental value. This error is significantly larger than that of the second largest outlier, with an error of 2.9 kcal/mol. The

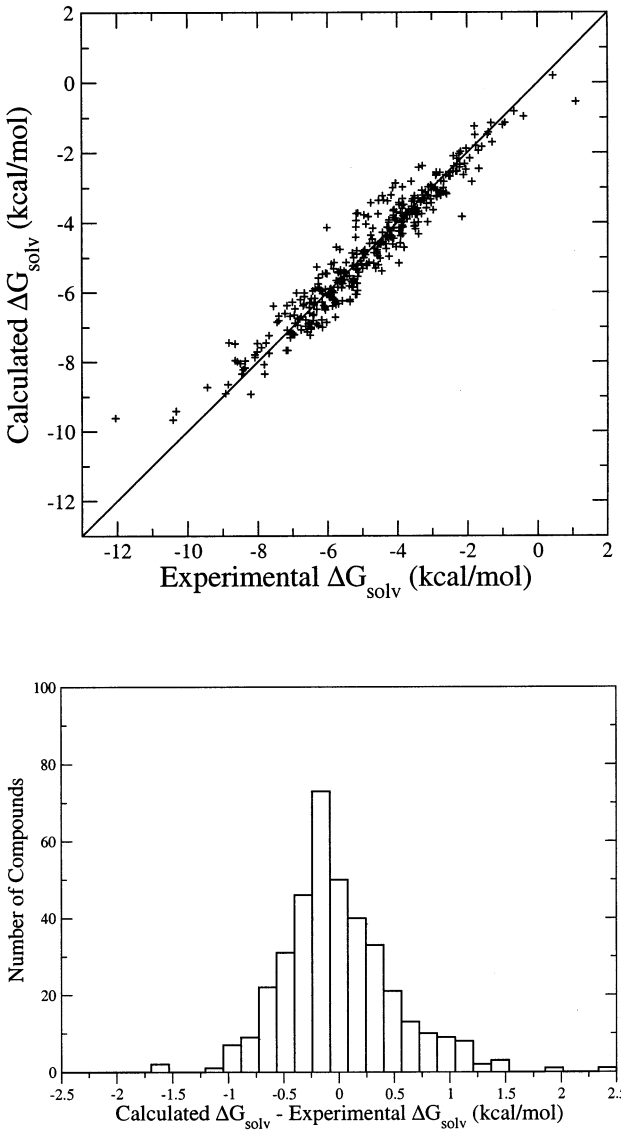

Figure 3. Gas−hexadecane transfer free energy: (a) calculated versus experimental gas−hexadecane transfer free energy values for the complete 382 molecule set; (b) a histogram of the differences between these values for the same set.

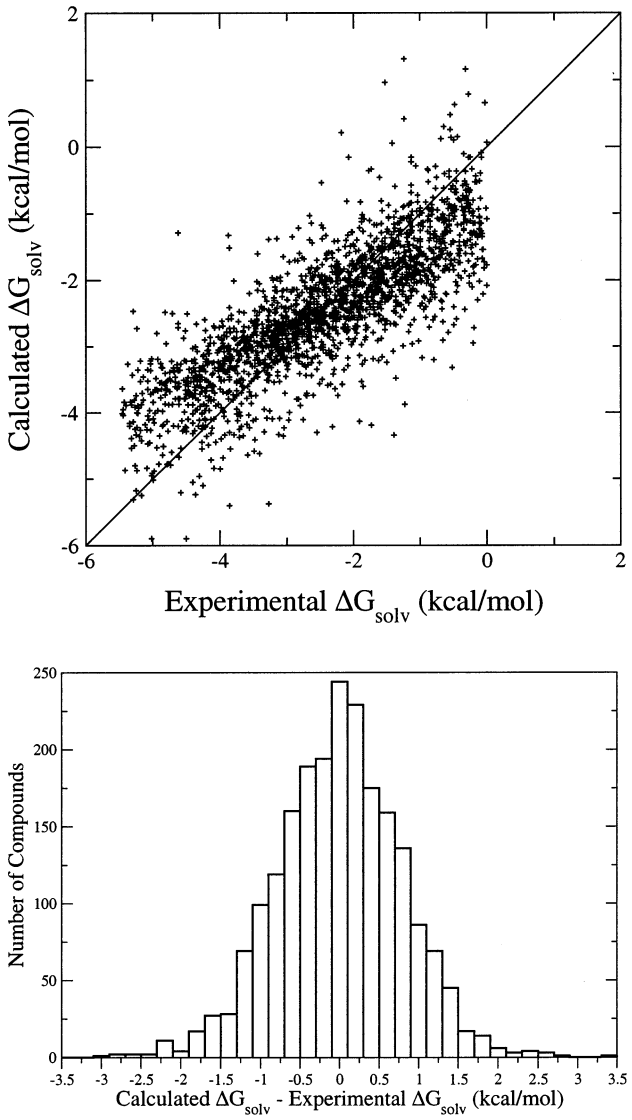

Figure 4. Water−octanol transfer free energy: (a) calculated versus experimental water−octanol transfer free energy values for the complete 2116 molecule set; (b) a histogram of the differences between these values for the same set.

type 45 radius for gas−hexadecane transfer is likewise unexceptional; however, the electrostatic contribution to the free energy is smallest for this solvent.

Finally, the small radius of 0.76 Å for the aromatic carbon, type 37, for gas−hexadecane transfer may be explained by its low atomic charge, an average of −0.084 with a standard deviation of 0.12 for the molecules in this set. The low charge combined with the small difference in the external dielectric constants makes the electrostatic contribution from this atom type small, and hence, the corresponding radius is relatively unconstrained in the parameter optimization.

3.3. Comparison with Alkane Surface Tension Parameters. It is useful to compare the optimal surface tension parameters, σ and b, to those derived from a linear regression fit to the experimental ΔG_{solv} values versus solvent-accessible surface area dependence for noncyclic, that is, linear and branched, alkanes. Because the solvation free energy of alkanes is dominated by the nonelectrostatic surface tension term, this is sometimes used to determine the surface tension parameters for eq 2. In the case of gas−water transfer, this plot of ΔG_{solv} versus solvent-accessible surface area for all alkanes also reveals that the simple linear dependence of the solvation free energy on area fails for cyclic alkanes, the solvation free energies of which are lower than noncyclic alkanes with the same solvent-accessible surface area. One possible explanation for the lower solvation free energy for cyclic alkanes compared with their linear counterparts is the additional contribution of attractive pairwise atomic hydrophobic interactions between the carbon atoms, which are closer to one another in the cyclic compound. The plot is shown for the three different choices of solvent in Figure 1 along with linear regression fits to the linear alkane data. The linear regression fits to the noncyclic data give $\sigma = 0.005\,35$, $b = 0.989$ ($r^2 = 0.842$) for gas−water, $\sigma = -0.0244$, $b = 3.69$ ($r^2 = 0.985$) for gas−hexadecane, and $\sigma = -0.0226$, $b = 1.55$ ($r^2 = 0.996$) for water−octanol transfer.

Comparing these surface tension parameters derived from only the noncyclic alkane data with those obtained from the complete molecule set data shows that, while the parameters for gas−hexadecane are essentially the same, those for the other two solvents differ in magnitude. In particular, σ determined by a fit to the complete data set is lower than the value obtained from a fit to the noncyclic alkane data alone for gas−water transfer and higher in the case of water−octanol transfer.

Because, unlike water and octanol, hexadecane is a nonpolar solvent that does not form hydrogen bonds, this effect may be due to the neglect of solvent−solute hydrogen-bond interactions in the simple solvation model used here. Because the solute−solvent hydrogen-bond interactions have a nonelectrostatic component and should be, on the average, stronger for larger molecules, the lack of these interactions in the solvation model may be compensated by a change in the surface tension parameters. The fact that solute−solvent hydrogen-bond interactions are more favorable for water then explains the observed changes in the surface tension parameters when nonalkane compounds, many of which are expected to form hydrogen bonds with water, are included in the data set.

3.4. Conformational Dependence of Solvation Free Energy. The energy terms of both eqs 1 and 2 depend on the conformation of the molecule used in the calculation. A more rigorous, though computationally expensive, approach is to sum the energies of the lowest-energy configurations weighted by the Boltzmann factor, $\exp(-E/(kT))$. Actually, because the space of conformational degrees of freedom is continuous, the average energy should be an integral over these variables, but we assume that there are a small number of low-energy configurations that lie within narrow potential energy wells. To study the approximation of using only the lowest-energy conformation, Monte Carlo sampling of the molecular conformations was performed using ICM. A total of 10^5 conformations of the water−octanol molecule set in each of water and octanol were sampled and those of which the all-atom rms deviations differ by more than 1.0 Å were retained. The solvation free energies were calculated using the water−octanol radii in Table 1 and the dielectric constant in Table 2. The surface tension parameters for gas−water and the sum of the gas−water and water−octanol parameters in Table 2 were used to calculate the nonelectrostatic contributions in water and octanol, respectively. The water−octanol solvation free energy was then calculated as the difference between the Boltzmann-weighted solvation energies in octanol and water. The rms deviation between the water−octanol solvation free energy calculated using the Boltzmann-weighted energies and that calculated using only the lowest-energy conformation was 0.035 kcal/mol. However, only one conformation was found for about 45% of the molecules in the set indicating that many of them are relatively rigid with perhaps a few rotatable bonds. The small error implies that including multiple low-energy conformations does not significantly change the calculated free energy value, at least for molecules with a comparable number of free torsion angles. Boltzmann sampling is expected to become important for larger, more flexible molecules.

3.5. Calculation Time. One feature of this method is its speed. The calculation of the solvation free energy for the water−octanol molecule set takes approximately 0.34 s per molecule on a 700 MHz Pentium III workstation. Most of this time is for the boundary element calculation, which takes longer for larger molecules. However, because this molecule set contains a significant portion of large drug molecules, with an average of 22 atoms per molecule in the set, this time should be typical for molecules of comparable size such as those in a virtual screening library.

3.6. Conclusion. In conclusion, the continuum solvation model of the type considered here, with atomic point charges and a surface tension term, is reasonably accurate and fast. In addition, cross-validation has shown that it is not overfit and therefore its accuracy is retained even for molecules outside of the training set. Possible improvements to the model include derivation of new atom types and charges, particularly for the problematic atom types discussed above, inclusion of solute−solvent hydrogen-bonding interactions, and a more accurate nonelectrostatic free energy term. In addition, it would be interesting to extend the model to include larger molecules, such as peptides, by including Boltzmann averaging over low-energy molecular conformations.

Acknowledgment. We thank M. Totrov for many useful discussions and A. Morrill for the initial compilation of the gas−water and water−hexadecane solvation data sets. This work was supported by grants from the Department of Energy (Grant DOE/DE-FG03-00ER6304) and the Department of Defense (Grant DOD/DAMD17-99-1-9318).

Supporting Information Available: Experimental log *P* values and optimized molecular structures (MOL2 format) for all compounds in the data set. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Hansch, C.; Leo, A. *Exploring QSAR Fundamentals and Application in Chemistry and Biology*; American Chemical Society: Washington, DC, 1995.

(2) Pliska, V.; Testa, B.; van de Waterbeemd, H. *Lipophilicity in Drug Action and Toxicology*; VCH: Weinheim, Germany, 1996.

(3) Carrupt, P.; Testa, B.; Gaillard, P. In *Reviews in Computational Chemistry*; Lipkowitz, K., Boyd, D., Eds.; Wiley-VCH: New York, 1997; Vol. 11.

(4) Sangster, J. *Octanol−Water Partition Coefficients: Fundamentals and Physical Chemistry*; Wiley: Chichester, U.K., 1997.

(5) Wesson, L.; Eisenberg, D. *Protein Sci.* **1992**, *1*, 227−235.

(6) Howard, P.; Meylan, W. *J. Pharm. Sci.* **1995**, *84*, 83−92.

(7) Abagyan, R. In *Computer simulation of biomolecular systems: Theoretical and experimental applications*; van Gunsteren, W. F., Weiner, P. K., Wilkinson, A. J., Eds.; Kluwer Acadeic Publishers: Dordrecht, Boston, London, 1997, pp 363−394.

(8) Huang, M.; Bodor, N. *J. Pharm. Sci.* **1990**, *81*, 272−281.

(9) Chadra, H.; Abraham, M.; Whiting, G.; Mitchell, R. *J. Pharm. Sci.* **1994**, *83*, 1085−1100.

(10) Kubodera, H.; Sasaki, Y.; Matsuzaki, T.; Umeyama, H. *J. Pharmacobio-Dyn.* **1991**, *14*, 207−214.

(11) Koehler, M.; Dunn, W., III; Grigoras, S. *J. Med. Chem.* **1987**, *30*, 1121−1126.

(12) Gabanyi, N. B. Z.; Wong, C. *J. Am. Chem. Soc.* **1989**, *111*, 3783−3786.

(13) Hansch, C.; Leo, A.; Hoekman, D. *Exploring QSAR − Hydrophobic, Electronic, and Steric Constants*; American Chemical Society: Washington, DC, 1995.

(14) Hall, N.; Smith, B. *J. Comput. Chem.* **1998**, *19*, 1482−1493.

(15) Honig, B.; Nicholls, A. *J. Comput. Chem.* **1990**, *12*, 435−445.

(16) Svensson, B.; Fushiki, M.; Jonsson, B.; Woodward, C. *Biopolymers* **1991**, *31*, 1149−1158.

(17) Honig, B.; Nicholls, A. *Science* **1995**, *268*, 1144−1149.

(18) Best, S.; Merz, K.; Reynolds, C. *J. Phys. Chem. B* **1997**, *101*, 10479−10487.

(19) Fine, R.; Schmidt, A. *Biopolymers* **1995**, *36*, 599−605.

(20) Miertus, S.; Scrocco, E.; Tomasi, J. *Chem. Phys.* **1981**, *55*, 117−129.

(21) Amovilli, C.; Barone, V.; Cammi, R.; Cances, E.; Cossi, M.; Mennucci, B.; Pomelli, C.; Tomasi, J. In *Recent advances in the description of solvent effects with the polarizable continuum model*; Wilson, S., Marvani, J., Grout, P. J., Mc Weeny, R., Smeyers, Y. G., Eds.; Advances in Quantum Chemistry, Vol. 32; Academic Press: San Diego, CA, 1999.

(22) Cances, E.; Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3032−3041.

(23) Klamt, A.; Schuurmann, G. *J. Chem. Soc., Perkin Trans. 2* **1993**, *5*, 799−805.

(24) Truong, T.; Stefanovich, E. *J. Chem. Phys.* **1995**, *103*, 3709−3717.

(25) Barone, V.; Cossi, M. *J. Phys. Chem. A* **1998**, *102*, 1995−2001.

(26) Foresman, J. B.; Keith, T. A.; Wiberg, K.; Snoonian, J.; Frisch, M. J. *J. Phys. Chem.* **1996**, *100*, 16098−16104.

(27) Amovilli, C.; Mennucci, B. *J. Phys. Chem. B* **1997**, *101*, 1051−1057.

(28) Pierotti, R. *Chem. Rev.* **1976**, *76*, 717−726.

(29) Langlet, J.; Claverie, P.; Caillet, J.; Pullman, A. *J. Phys. Chem.* **1988**, *92*, 1617−1631.

(30) Totrov, M.; Abagyan, R. *Biopolymers* **2001**, *60*, 124−133.
(31) Sitkoff, D.; Sharp, K.; Honig, B. *J. Phys. Chem.* **1994**, *98*, 1978−1988.
(32) Totrov, M.; Abagyan, R. *J. Struct. Biol.* **1996**, *116*, 138−143.
(33) Hermann, R. *J. Phys. Chem.* **1972**, *76*, 2754−2759.
(34) Honig, B.; Sharp, K.; Yang, A. *J. Phys. Chem.* **1993**, *97*, 1101−1109.
(35) Abagyan, R.; Totrov, M.; Kuznetsov, D. *J. Comput. Chem.* **1994**, *15*, 488−506.
(36) Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 490−641.
(37) Abraham, M. H.; Whiting, G. S.; Fuchs, R.; Chambers, E. J. *J. Chem. Soc., Perkin Trans. 2* **1990**, *2*, 291−300.
(38) *ICM software manual*, version 2.8; Molsoft, LLC: San Diego, CA, 2002.
(39) *EPA Product Properties Test Guidelines OPPTS 830.7570, Partition coefficient (n-Octanol/Water) estimation by liquid chromatography*; United States Government Printing Office: Washington, DC, 1996.
(40) Nelder, J. A.; Mead, R. *Comput. J.* **1965**, *7*, 308−313.
(41) Singh, U.; Kollman, P. *J. Comput. Chem.* **1984**, *5*, 129−145.
(42) Besler, B.; Merz, K.; Kollman, P. *J. Comput. Chem.* **1990**, *11*, 431−439.
(43) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, revision A.9; Gaussian, Inc.: Pittsburgh, PA, 1998.