Article

Development of Surface-SFED Models for Polar Solvents

Sehan Lee,[†] Kwang-Hwi Cho,[‡] William E. Acree, Jr.,[§] and Kyoung Tai No*[,†,∥]

[†]Department of Biotechnology, Yonsei University, Seoul, 120-749, Korea

[‡]Department of Bioinformatics, Soongsil University, Seoul, 156-743, Korea

[§]Department of Chemistry, University of North Texas, 1155 Union Circle Drive #305070, Denton, Texas 76230-5017, United States

[∥]Bioinformatics and Molecular Design Research Center, Seoul, 120-749, Korea

S *Supporting Information*

ABSTRACT: We developed surface grid-based solvation free energy density (Surface-SFED) models for 36 commonly used polar solvents. The parametrization was performed with a large and diverse set of experimental solvation free energies mainly consisting of combinations of polar solvent and multipolar solute. Therefore, the contribution of hydrogen bonds was dominant in the model. In order to increase the accuracy of the model, an elaborate version of a previous hydrogen bond acidity and basicity prediction model was introduced. We present two parametrizations for use with experimentally determined (Surface-SFED/HB^{exp}) and empirical (Surface-SFED/HB^{cal}) hydrogen bond acidity and basicity values. Our computational results agreed well with experimental results, and inaccuracy of empirical hydrogen bond acidity and basicity values was the main source of error in Surface-SFED/HB^{cal}. The mean absolute errors of Surface-SFED/HB^{exp} and Surface-SFED/HB^{cal} were 0.49 and 0.54 kcal/mol, respectively.
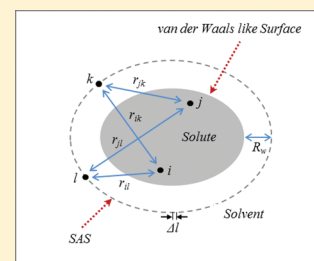
INTRODUCTION

Solvation effect plays an important role in a wide variety of phenomena in chemical and biological processes. The free energy of solvation of a solute is a fundamental quantity that represents the effect of solvation. Differences in the solvation free energies of a solute in immiscible solvents can be used to predict the partitioning of the solute between these solvents. For this reason, the development of models to calculate solvation free energies for diverse solvents may contribute much to the fields of computational biology and physical and biophysical chemistry.

Because of the complexity of solute−solvent interactions, prediction of the solvation free energy is considered as one of the most challenging energy terms to calculate. Though many computational models have been proposed, from detailed ab initio models to fast empirical models,[1−8] most have so far been applied to a limited number of solvents, such as 1-octanol, chloroform, and particularly water.[9−13]

The solvation free energy density (SFED) model, an empirical continuum solvation model proposed by No et al., was developed to predict solvation free energies.[14] The model has been applied to calculate solvation free energies in 10 solvents including water and 1-octanol,[14−16] to compute 1-octanol/water partition coefficients,[15] and to generate the hydration free energy density tensor.[17] In a recent paper,[16] we extensively modified this model, as follows: (1) A basis function for electrostatics was modified to include ionic molecules. (2) Two basis functions were introduced to express hydrogen bond contributions. (3) A surface grid was introduced, instead of the shell grid, to avoid the bumping of the grid between molecules

and improve computational efficiency. The parametrization was performed for over 10 solvents that have more than 50 solvation free energies, and the mean absolute errors (MAEs) for 1200 solvation free energies of 379 neutral molecules in these 10 solvents and 90 hydration free energies of ionics were 0.40 and 1.70 kcal/mol, respectively. Although the data set of neutral molecules mainly consists of nonpolar or unipolar molecules, the accuracy of empirical hydrogen bond acidity and basicity was a critical factor in the model.

Here, we present the extended application of the surface grid-based SFED (Surface-SFED) model to 36 polar solvents including alcohol, haloaliphatic, haloaromatic, ketone, aliphatic ether, ester, and tertiary amide solvent classes. The data set includes a variety of multipolar molecules ("multi" in this study refers to more than two), which form complex hydrogen bonds. Therefore, the accurate prediction of hydrogen bond contributions is very important in this work. The hydrogen bond acidity and basicity computation model was refined to improve the predictability of the model. We optimized linear expansion coefficients suitable for use with experimental and empirical hydrogen bond acidity and basicity, and these parametrizations of the model are differentiated by the notations/HB^{exp} and/HB^{cal}, respectively.

METHODOLOGY

A. Description of the Surface-SFED Model. The details of the Surface-SFED model are well described in our previous

paper.[16] However, some modifications are needed to describe the solvation free energies of polar solutes in polar solvents in which the hydrogen bonding contributes significantly. On the basis of the assumption that the solvation free energy can be partitioned into the different contributions, the solvation free energy of a molecule is described as a sum of three terms.

$$\Delta G_{solv} = \Delta G_{inter} + \Delta G_{HB} + \Delta G_{cav} \tag{1}$$

The free energy of interactions, ΔG_{inter}, is described as a linear combination of four basis functions, $h_j(r_k)$, representing the contribution of the interaction between interacting compartments of the solute and the kth grid point on the cavity surface (Figure 1).
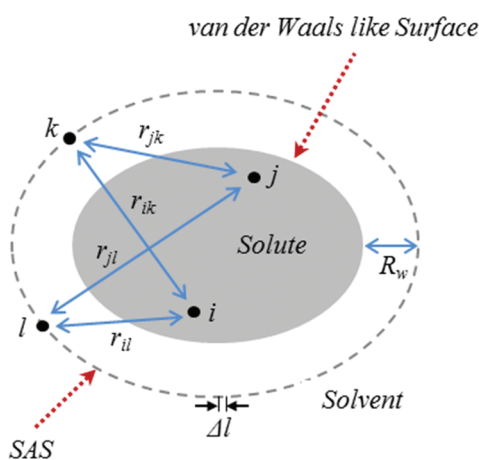

Figure 1. Solute and solvent of a solution are described as an assemblage of interacting compartments. Furthermore, the solvent accessible surface (SAS), R_w, R_{shell}, Δl, and van der Waals-like surface of the model are described.

$$\Delta G_{inter} = \sum_{k}^{S} \sum_{j}^{4} C_j h_j(r_{ik})$$

$$= \sum_{k}^{S} \left(C_1 \left| \sum_{i=1}^{N_A} \frac{q_i}{r_{ik}^2} \right| + C_2 \sum_{i=1}^{N_A} \frac{q_i^2}{r_{ik}^3} \right.$$

$$\left. + C_3 \sum_{i=1}^{N_A} \frac{\alpha_i}{r_{ik}^3} + C_4 \sum_{i=1}^{N_A} \frac{\alpha_i}{r_{ik}^6} \right) \tag{2}$$

where S and N_A are the number of grid points on the cavity surface and the number of atoms in the solute, respectively. Atom-centered net atomic charges, q_i, were calculated with an empirical net atomic charge calculation method, the modified partial equalization of orbital electronegativity (MPEOE),[18,19] and effective atomic polarizabilities, α_i, were calculated using an empirical method, the charge dependent effective atomic polarizability (CDEAP).[20]

The hydrogen bond stabilization term, ΔG_{HB}, was divided into two terms on the basis of the role of the solvent in hydrogen bonding, with the acceptor and donor represented with the subscripts "a" and "d", respectively. These two terms were simple functions of the hydrogen bond acidity and basicity of the solute, A_{solute} and B_{solute}, respectively.

$$\Delta G_{HB} = \Delta G_a + \Delta G_d = C_a A_{solute} + C_d B_{solute} \tag{3}$$

The values of A_{solute} and B_{solute} were obtained from Abraham's et al.,[21-23] A^{exp} and B^{exp}, and empirical model, A^{cal} and B^{cal}. Application of A^{exp} and B^{exp} to the Surface-SFED model produced results in good agreement with the experimental solvation free energy.[16] However, because there are a limited number of Abraham's hydrogen bond parameters, application of the Surface-SFED model can be limited to a molecule for which A^{exp} and B^{exp} were determined. For the general purpose of the Surface-SFED model, A^{cal} and B^{cal} are given by the sum of the acidities, α_i, and basicities, β_i, of the acidic and basic atoms of the solute, respectively.

$$A^{cal} = \sum_i \alpha_i \text{ and } B^{cal} = \sum_i \beta_i \tag{4}$$

where α_i and β_i were meant to reproduce the values published by Abraham, et al.

The free energy of cavity formation by the solute, ΔG_{cav}, was expressed by a term proportional to the SAS area, S_S, of the solute as follows

$$\Delta G_{cav} = C_S S_S + C \tag{5}$$

The final formula for the SFED computation of a solute is

$$\Delta G_{solv}^o = \Delta G_{inter} + \Delta G_{HB} + \Delta G_{cav}$$

$$= \sum_{k}^{S} \sum_{j}^{4} C_j h_j(r_{ik}) + C_d A_{solute} + C_a B_{solute} + C_S S_S + C \tag{6}$$

B. Solvation Free Energy Database Construction. Most of the experimental solvation free energies were obtained from the experimental gas/solvent partition coefficients database of Acree et al.[24-35] The partition coefficient is the ratio of concentrations of a solute between solvent A and solvent B at equilibrium.

$$P_{A/B} = \frac{[solute]_A}{[solute]_B} \tag{7}$$

The gas/solvent partition coefficient, K, is a specific case of eq 7 where solvent B is replaced with gas. The base-10 logarithm of the gas/solvent partition coefficient is related to the standard state free energy of solvation as follows

$$\log K = \log \frac{[solute]_A}{[solute]_G} \tag{8}$$

and

$$\Delta G_{solv}^o = -2.303 RT \log K \tag{9}$$

The gas/solvent partition coefficient was calculated from either the Raoult's law infinite dilution activity coefficient, γ_{solute}^∞, from the Henry's law constant, K_{Henry}, for solutes dissolved in anhydrous organic solvents, or from partition coefficients of a solute between water and organic solvent saturated with water, and the gas/water partition coefficient of the solute, K_w, as shown in eqs 10−12.

$$\log K = \log \left(\frac{RT}{\gamma_{solute}^\infty P_{solute}^o V_{solvent}} \right) \tag{10}$$

$$\log K = \log\left(\frac{RT}{K_{\mathrm{Henry}} V_{\mathrm{solvent}}}\right) \tag{11}$$

$$\log K = \log P + \log K_w \tag{12}$$

Here R is the universal gas constant, T is the system temperature, $P^{\mathrm{o}}_{\mathrm{solute}}$ is the vapor pressure of the solute at T, and V_{solvent} is the molar volume of the solvent. In the case of crystalline solutes, log K values have also been calculated from the ratio of the molar solubilities of the solute in the organic solvent and in water, in accordance with eq 13.

$$\log K = \log(C_{\mathrm{solute, organic\ solvent}}/C_{\mathrm{solute, water}}) + \log K_w \tag{13}$$

The log K calculations and the individual data sets are discussed in greater detail elsewhere.[24−36]

In compiling the data, the gas/solvent partition coefficients were classified as "wet" or "dry", according to the experimental method, to ensure that the solubilizing properties of anhydrous solvent, denoted as "dry", and water-saturated solvent, denoted as "wet", were the same before the measured values were combined into a single data set.

We excluded the following types of data: (1) chemical elements such as helium and oxygen, (2) molecules containing metal or phosphorus, (3) "wet" data pertaining to solvents for which the "wet" and "dry" experimental values were not the same. After these exclusions, 3927 free energies of solvation for 766 solutes in 36 solvents were obtained.

Large experimental solvation free energies were directly taken from our previously published data set.[7] This data set contained 2193 free energies of solvation for 387 solutes in 91 solvents.

After averaging the solvation free energies of the solute/solvent combination that were common to both data sets and combining into a single data set, we were able to extract 4190 free energies of solvation for 787 solutes in 36 solvents that have more than 50 solvation free energies. The collected experimental data set encompassed a broad range of molecules including diverse drugs and natural molecules. A total of 32 solvents have newly been introduced in this study, and four solvents, 1-octanol, diethyl ether, chloroform, and carbon tetrachloride, are part of the expanded set from our previous SFED study. The geometries of the molecules were obtained by energy minimization in the gas-phase, using HF/MIDIx.

The Abraham's hydrogen bond acidity, A^{exp}, and basicity, B^{exp}, were taken from the Acree database referenced above. In the literature, A^{exp} and B^{exp} values of some solutes have been altered slightly due to periodic update/revision or rounding/truncating to the third decimal place. The most recent A^{exp} and B^{exp} values are given in Table S1 of the Supporting Information.

C. Refinement of Hydrogen Bond Acidity and Basicity Prediction Model. Because the collected set of solvation free energies includes a wide variety of multipolar molecules, the parameter set of α_i and β_j determined in our previous work[16] was expanded and reoptimiezed. Primary, secondary, and tertiary sulfonamide, amide in nucleobases, alkylated aromatic hydrocarbon, and 3-valent aromatic nitrogen are newly defined.

The strength of the hydrogen bond can be influenced by substituents near the hydrogen bond donor or acceptor. The presence of substituent induces charge rearrangements through the covalent framework of the molecule and obstructs the

formation of other hydrogen bond. This suggests that α_i and β_j in eq 4 have to be corrected in order to prevent the unphysical condition that all the acidic and basic atoms can be participated in hydrogen bond independently. The effect due to charge rearrangements was not considered because A^{cal} and B^{cal} are calculated on the basis of the summation of empirical parameters, α_i and β_j. Thus, the effective hydrogen bond acidity and basicity, α_i^* and β_j^*, of acidic and basic atoms in the molecule were computed by including a sort of attenuation factor

$$\alpha_i^* = \alpha_i - \sum_j k \times \alpha_j \times \exp(-d_{ij} \times t) \text{ and}$$

$$\beta_i^* = \beta_i - \sum_j k \times \beta_j \times \exp(-d_{ij} \times t) \tag{14}$$

where d_{ij} is the distance between atom i and atom j, and conventionally d_{ii} is set to infinity. Adjustable parameters, k and t, are defined 0.5 and 0.77, respectively. Tables 1 and 2 list the

Table 1. Hydrogen Bond Acidity Parameters and Their Optimized Value

type	$\alpha_i \times 10$	type	$\alpha_i \times 10$
C_{sp}−H	1.20	C_{ar}N−H	1.37
RO-H	3.30	RCONH−H	2.71
c−O−H[a]	2.57	RCONR−H	3.63
HO−H	4.86	HCONH−H	3.10
C_{ar}O−H	4.02	HCONR−H	4.00
RCOO−H	6.10	CONR−H[b]	2.20
HCOO−H	7.60	N_{ar}−H	5.67
RHN−H	0.94	SO_2NH−H	3.19
R_2N−H	0.80	SO_2NR−H	3.30
H_2N−H	0.75		

[a]H in cycloalcohol. [b]H bonded to amide in nucleobases.

hydrogen bond parameters and their determined values, respectively. The effective hydrogen bond acidity, A^*, and basicity, B^*, of a solute is

$$A^* = \sum_i \alpha_i^* \text{ and } B^* = \sum_i \beta_i^* \tag{15}$$

D. Model Development. Two parametrizations are present here, one for use with the Abraham's hydrogen bond acidity and basicity (Surface-SFED/HB$^{\mathrm{exp}}$) and one for use with empirical effective hydrogen bond acidity and basicity calculated using eq 15 (Surface-SFED/HB$^{\mathrm{cal}}$). The linear expansion coefficients, C_j, C_d, C_a, C_S, and C, were determined for each solvent by minimizing the difference between the calculated and experimental solvation free energies. The optimum values for atomic radii, R_{vdw}, effective solvent shell thickness, R_w, and grid interval, Δl, have been defined in our previous publications.

RESULTS AND DISCUSSION

The experimental and calculated solvation free energies in each solvent are shown in Table S2 of the Supporting Information. All Surface-SFED results quoted in this work are calculated with the same functional forms for electrostatic, dispersion−repulsion, hydrogen bond, and cavitation effects. Two Surface-SFED models were parametrized against the same training set including diverse multipolar molecules.

Table 2. Hydrogen Bond Basicity Parameters and Their Optimized Value

type	$\beta_i \times 10$	type	$\beta_i \times 10$
$-C_{sp2}$	0.39	$RC(=O)-OH$	3.73
$c-C_{sp2}{}^a$	0.61	$HC(=O)-OH$	3.30
$-C_{ar}$	0.54	$-NH_2$	6.10
$C_{sp3}-C_{ar}$	1.59	$-NRH$	6.90
$-C_{sp}$	0.62	$-NR_2$	7.67
$C_{sp3}-F$	0.01	H_3N	5.60
$C_{sp3}-Cl$	0.56	$C_{ar}-NR_2$	4.85
$C_{sp3}-Br$	0.43	$RCO-NH_2$	6.87
$C_{sp3}-I$	1.40	$RCO-NHR$	6.82
$ClC_{sp3}-Cl$	0.50	$RCO-NR_2$	8.47
$Cl_2Cl_{sp3}-Cl$	0.20	$HCO-NH_2$	6.00
$C_{ar}-F$	0.04	$HCO-NHR$	5.50
$C_{ar}-Cl$	0.13	$HCO-NR_2$	7.40
$C_{ar}-Br$	0.01	$RC\equiv N$	3.43
$C_{ar}-I$	0.01	$C_{sp3}NO-O$	1.72
$-OH$	4.80	$C_{ar}NO-O$	0.98
$c-OH^b$	5.60	N_{ar}	6.76
H_2O	3.50	$CONR_2{}^d$	6.76
$C_{ar}-OH$	3.10	$RN_{ar}{}^e$	5.45
R_2O	4.06	$-SH$	2.00
$R_2C=O$	4.90	R_2S	2.89
$RCHO$	4.50	$RSSR^f$	1.56
$c-C=O^c$	5.24	RSO_2-NH_2	4.50
$RC(OR)=O$	4.50	RSO_2-NHR	4.13
$HC(OR)=O$	3.80	RSO_2-NR_2	4.48

^{a}C$_{sp2}$ in cycloalkene. ^{b}O$_{sp3}$ in cycloalcohol. ^{c}C$_{sp2}$ in cycloketone. ^{d}N$_{sp2}$ in amide in nucleobases. ^{e}3-valent N$_{ar}$. ^{f}S$_{sp3}$ in disulfide.

Table 3 presents the mean absolute errors (MAEs) of the free energies of solvation for each solvent. The MAEs of Surface-SFED/HB^{exp} and Surface-SFED/HB^{cal} parametrizations are 0.49 and 0.54 kcal/mol, respectively, over the entire data set of 4190 free energies of 787 solutes in 36 solvents. This shows that the Surface-SFED model with A^{exp} and B^{exp} does well for all organic solvents, and the largest MAE is 0.66 kcal/mol in dichloromethane. Most of the alcohol, ketone, and tertiary amide solvents have MAEs less than 0.5 kcal/mol, and ester, and haloaliphatic solvents have relatively larger MAEs. However, because the latter two solvent classes mainly consist of polar solutes, the errors might not be significant. A plot of the computational results versus the observed solvation free energies is shown in Figure 2, where we have distinguished the points for each parametrization.

The only difference between the various solvent models is in the linear expansion coefficients. The optimized linear expansion coefficients for each solvent model are presented in Tables 4 and 5. The basis functions represent the solute effect on various solute–solvent interactions, and the linear expansion coefficients correspond to the complementary effect of the solvent on these interactions. The first two terms, $|\sum_{i=1}^{N_A} q_i/r_{ik}^2|$ and $\sum_{i=1}^{N_A} q_i^2/r_{ik}^3$, are introduced to describe the dipole and induced dipole interactions between the solute and the solvent. The latter two terms, $\sum_{i=1}^{N_A} \alpha_i/r_{ik}^3$ and $\sum_{i=1}^{N_A} \alpha_i/r_{ik}^6$, describe the interactions originating from first- and second-order perturbation and dispersion–repulsion effects. Because all basis functions have positive values, the sign and magnitude of the coefficients indicate the contribution of the characteristic features of each solvent to

Table 3. Errors in Predicted Solvation Free Energies from the Surface-SFED Model Parameterizations by Solvent

solvent	data^a	Surface-SFED/ HB^{exp}	Surface-SFED/ HB^{cal,b}
alcohols			
methanol	138	0.47	0.55
ethanol	124	0.51	0.64
1-propanol	137	0.51	0.64
isopropanol	133	0.41	0.58
1-butanol	156	0.49	0.62
isobutanol	92	0.39	0.50
sec-butanol	92	0.44	0.54
tert-butanol	99	0.33	0.44
1-pentanol	125	0.49	0.56
2-pentanol	55	0.43	0.49
3-methyl-1-butanol	66	0.41	0.44
1-hexanol	89	0.53	0.63
1-heptanol	69	0.49	0.57
1-octanol	297	0.50	0.51
1-decanol	78	0.41	0.48
ethylene glycol	71	0.45	0.56
aliphatic ethers			
diethyl ether	116	0.50	0.53
di-n-butyl ether	66	0.53	0.59
methyl tert-butyl ether	50	0.49	0.51
tetrahydrofuran	101	0.49	0.58
1,4-dioxane	117	0.53	0.59
esters			
methyl acetate	75	0.59	0.63
ethyl acetate	118	0.54	0.57
n-butyl acetate	90	0.55	0.60
propylene carbonate	62	0.40	0.35
haloaliphatics			
chloroform	368	0.63	0.61
dichloromethane	73	0.66	0.64
carbon tetrachloride	254	0.37	0.37
1-chlorobutane	59	0.39	0.38
haloaromatics			
chlorobenzene	168	0.52	0.53
bromobenzene	102	0.55	0.53
ketones			
acetone	92	0.43	0.47
2-butanone	86	0.47	0.44
tertiary amides			
N,N-dimethylformamide	162	0.55	0.59
N,N-dimethylacetamide	97	0.45	0.50
N-methyl-2-pyrrolidone	113	0.42	0.40
total	4190	0.49	0.54

^{a}Number of solutes in a solvent. ^{b}Hydrogen bond acidity and basicity were computed using eq 15.

solvation. All Surface-SFED models were developed for polar solvents; thus, the differences of coefficients between solvents were not significant. C_1 and C_2 show tendency to roughly depend on the polarity of solvents because the major solute–solvent interaction is due to electrostatic interaction. C_4 of all solvents was positivity; however, the sum of the latter two terms has favorable contributions to solvation free energies. C_d and C_a became larger as the polarity of the solvent grew, implying formation of stronger hydrogen bonds. Some examples of contribution of each basis function are shown in Table 6.

In this work, more than 400 multipolar molecules were used for parametrizations; thus, the accurate prediction of A
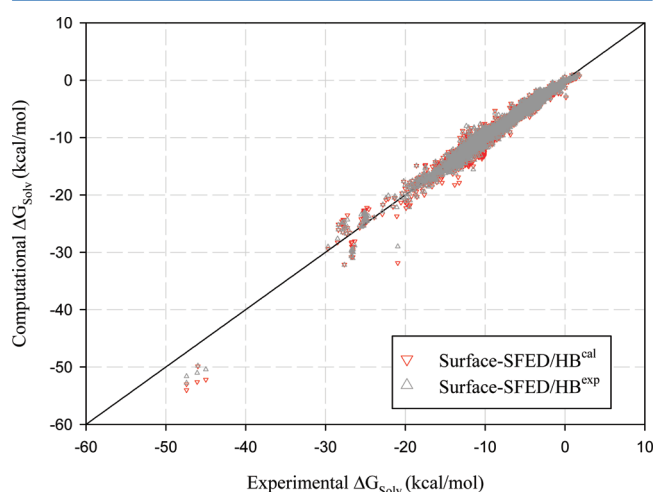

Figure 2. Plot of calculated solvation free energies against experimental values.

and B is crucial factor determining the performance of the Surface-SFED/HB^{cal} model. For the purpose of accurate description of hydrogen bond contributions in the solvation free energy, the influence of substituents on hydrogen bond acidity and basicity was considered, eq 14. Plots of calculated hydrogen bond acidities and basicities against experimental values are shown in Figure 3. There are clear improvements in the molecules that have large A and/or B. For instance, in rutin, a citrus flavonoid glycoside with 10 donors and 16 acceptors, the values of A^{cal} and B^{cal} predicted using eq 4 are about double those of A^{exp} and B^{exp}, the experimental and predicted values being 2.50 and 4.23 for A and 3.73 and 7.29 for B, respectively. $A*$ and $B*$ predicted using eq 15 are 2.92 and 4.09, respectively. Some examples of the comparison between experimental and empirical hydrogen bond acidity and basicity are shown in Table 7. The effect of the inaccuracy of empirical hydrogen bond acidity and basicity on the accuracy of the predicted solvation free energy depends on the solute/solvent combination. For example, these errors represent about 10 kcal/mol in alcohol solvents and about 3 kcal/mol in chloroaliphatic solvents. Table 8 gives a summary of the errors in prediction across solvent classes. Overall, the Surface-SFED model

Table 4. Linear Expansion Coefficients for Each Solvent Model Optimized with Experimental Hydrogen Bond Acidity and Basicity

solvent	$C_1 \times 10$	$C_2 \times 10$	$C_3 \times 10^2$	$C_4 \times 10$	$C_S \times 10^3$	$C \times 10$	C_d	C_a
methanol	−2.218	−2.304	−3.456	5.012	0.922	2.664	−4.142	−0.981
ethanol	−2.118	−2.024	−3.551	5.165	0.811	3.023	−4.159	−0.957
1-propanol	−2.100	−1.894	−3.562	4.414	1.047	2.834	−4.496	−0.712
isopropanol	−2.093	−1.889	−3.484	4.734	0.675	3.889	−4.784	−0.191
1-butanol	−2.083	−1.778	−3.659	4.279	1.239	2.373	−4.466	−0.668
isobutanol	−2.086	−1.815	−3.413	4.434	0.579	5.232	−4.156	−0.839
sec-butanol	−2.075	−1.747	−3.841	5.563	1.190	3.639	−4.265	−1.548
tert-butanol	−2.035	−1.802	−3.425	4.765	0.577	4.204	−4.911	−0.356
1-pentanol	−2.064	−1.728	−3.682	4.493	1.290	2.583	−4.077	−1.696
2-pentanol	−2.062	−1.677	−3.687	4.737	1.146	3.511	−4.084	−1.598
3-methyl-1-butanol	−2.079	−1.718	−3.634	4.512	1.134	3.417	−3.691	−1.687
1-hexanol	−2.061	−1.643	−3.789	4.593	1.439	2.559	−4.272	−1.534
1-heptanol	−2.036	−1.862	−3.672	3.345	1.538	1.408	−4.171	−0.587
1-octanol	−1.998	−1.750	−3.757	3.783	1.683	2.489	−4.233	−1.142
1-decanol	−1.990	−1.504	−4.134	4.815	1.840	2.776	−4.216	−1.093
ethylene glycol	−2.340	−2.627	−4.399	4.361	5.144	−7.723	−3.801	−3.195
diethyl ether	−2.055	−1.552	−3.828	5.467	0.192	7.210	−3.724	0.000
di-n-butyl ether	−1.778	−1.367	−4.022	4.497	0.959	4.243	−3.537	0.000
methyl tert-butyl ether	−2.025	−1.462	−3.743	4.939	0.295	4.174	−2.398	0.000
tetrahydrofuran	−2.049	−1.790	−4.410	4.419	1.613	1.529	−4.149	0.000
1,4-dioxane	−2.070	−1.808	−4.726	4.909	2.410	−1.330	−4.243	0.000
methyl acetate	−2.052	−2.049	−4.552	6.948	1.258	2.378	−3.384	0.000
ethyl acetate	−2.074	−1.959	−4.093	5.265	1.049	2.632	−3.535	0.000
n-butyl acetate	−2.062	−1.927	−4.050	5.046	1.195	2.587	−3.193	0.000
propylene carbonate	−2.565	−3.437	−4.111	4.101	3.372	−5.239	−1.065	0.000
chloroform	−2.102	−1.604	−3.904	3.831	0.975	1.130	−0.112	−1.175
dichloromethane	−2.075	−1.836	−4.130	4.340	1.451	0.808	−1.071	−1.112
carbon tetrachloride	−2.005	−1.391	−4.023	3.735	1.368	1.566	0.000	0.000
1-chlorobutane	−1.971	−1.484	−4.128	4.455	1.009	3.297	−0.372	0.000
chlorobenzene	−1.996	−1.486	−4.196	2.759	2.144	−1.200	−0.717	0.000
bromobenzene	−1.991	−1.899	−4.803	2.824	3.248	−1.960	−0.633	0.000
acetone	−2.167	−2.073	−4.140	5.551	1.214	2.632	−3.355	−0.741
2-butanone	−2.084	−2.602	−3.704	4.681	0.885	2.132	−2.745	0.000
N,N-dimethylformamide	−2.404	−2.899	−5.010	5.513	3.160	−2.061	−4.069	0.000
N,N-dimethylacetamide	−2.335	−3.009	−4.547	4.558	2.560	−1.195	−4.100	0.000
N-methyl-2-pyrrolidone	−2.235	−3.295	−5.068	5.898	3.564	−5.867	−4.657	0.000

Table 5. Linear Expansion Coefficients for Each Solvent Model Optimized with Empirical Hydrogen Bond Acidity and Basicity Computed Using Eq 15

solvent	$C_1 \times 10$	$C_2 \times 10$	$C_3 \times 10^2$	$C_4 \times 10$	$C_S \times 10^3$	$C \times 10$	C_d	C_a
methanol	−2.371	−2.155	−3.448	5.218	0.989	0.886	−4.094	−0.981
ethanol	−2.415	−1.841	−3.678	4.773	1.306	0.485	−3.964	−0.895
1-propanol	−2.165	−1.833	−3.594	4.441	1.056	2.671	−4.232	−0.747
isopropanol	−2.284	−1.704	−3.392	4.444	0.489	5.369	−4.484	−0.021
1-butanol	−2.054	−1.699	−3.720	4.257	1.302	1.396	−4.407	−0.668
isobutanol	−2.090	−1.996	−3.366	4.493	0.318	6.851	−4.053	−0.839
sec-butanol	−2.105	−2.046	−3.992	4.928	1.575	3.093	−4.209	−0.987
tert-butanol	−2.057	−2.156	−3.436	4.726	0.604	4.657	−4.756	−0.554
1-pentanol	−1.989	−1.909	−4.024	4.678	1.888	0.925	−4.099	−1.715
2-pentanol	−2.055	−1.524	−3.777	4.697	1.078	3.415	−4.589	−0.785
3-methyl-1-butanol	−2.047	−1.540	−3.654	4.501	0.897	3.783	−4.249	−1.010
1-hexanol	−2.030	−1.532	−3.706	0.280	2.851	0.554	−4.168	−1.823
1-heptanol	−2.122	−1.618	−3.346	1.380	1.511	1.206	−4.544	−0.300
1-octanol	−2.018	−1.600	−3.604	3.675	1.413	1.332	−4.785	−0.638
1-decanol	−1.958	−1.345	−4.332	5.324	1.999	2.098	−4.741	−1.093
ethylene glycol	−2.484	−2.365	−4.384	2.996	5.123	−4.544	−3.337	−3.020
diethyl ether	−2.023	−1.753	−4.162	6.000	1.065	1.646	−3.752	0.000
di-n-butyl ether	−1.778	−1.250	−3.945	4.372	0.972	1.469	−1.885	0.000
methyl tert-butyl ether	−1.983	−1.605	−3.627	5.113	0.226	0.963	−2.046	0.000
tetrahydrofuran	−2.064	−2.089	−4.315	3.782	1.755	0.598	−3.735	0.000
1,4-dioxane	−2.105	−2.161	−4.653	4.849	2.397	−1.659	−3.853	0.000
methyl acetate	−2.110	−2.468	−4.895	8.890	1.391	0.633	−2.981	0.000
ethyl acetate	−2.101	−2.142	−4.176	6.445	0.940	0.655	−3.272	0.000
n-butyl acetate	−2.102	−2.248	−4.058	5.319	1.207	0.763	−2.642	0.000
propylene carbonate	−2.294	−2.901	−5.278	1.815	6.641	−15.442	−1.596	0.000
chloroform	−2.165	−1.290	−4.107	6.940	0.307	1.021	−0.128	−1.511
dichloromethane	−2.124	−1.610	−4.223	4.769	1.475	0.489	−0.383	−1.424
carbon tetrachloride	−1.997	−1.309	−4.175	4.583	1.386	0.475	0.000	0.000
1-chlorobutane	−1.956	−1.346	−4.414	5.659	1.118	1.696	−0.051	0.000
chlorobenzene	−1.947	−1.461	−4.211	2.797	2.100	−1.333	−0.631	0.000
bromobenzene	−1.899	−2.204	−4.825	2.664	3.252	−1.960	−0.012	0.000
acetone	−2.167	−2.500	−4.410	7.062	1.381	0.598	−2.615	−1.098
2-butanone	−1.988	−3.450	−4.238	9.601	0.224	0.229	−1.913	0.000
N,N-dimethylformamide	−2.614	−3.137	−4.909	5.775	2.828	−0.509	−2.612	0.000
N,N-dimethylacetamide	−2.496	−2.883	−4.559	4.616	2.551	−0.598	−3.417	0.000
N-methyl-2-pyrrolidone	−2.419	−3.556	−4.965	6.455	3.547	−8.472	−4.043	0.000

Table 6. Contributions of Each Basis Function

	methanol solvation free energy							
mol name	$\sum\|C_1 h_1\|$	$\sum C_2 h_2$	$\sum C_3 h_3$	$\sum C_4 h_4$	$C_S S_S$	C	$C_d A^{exp}$	$C_a B^{exp}$	ΔG^{cal}
n-pentane	−0.057	−0.010	−5.026	1.745	1.020	0.266	0.000	0.000	−2.062
1-pentanol	−1.202	−1.208	−5.397	1.869	1.076	0.266	−1.533	−0.471	−6.600
dimethylamine	−0.811	−0.744	−3.083	1.202	0.771	0.266	−0.331	−0.647	−3.378
rutin	−7.292	−13.672	−27.066	6.162	2.828	0.266	−10.355	−3.659	−52.787
	chloroform solvation free energy							
mol name	$\sum\|C_1 h_1\|$	$\sum C_2 h_2$	$\sum C_3 h_3$	$\sum C_4 h_4$	$C_S S_S$	C	$C_d A^{exp}$	$C_a B^{exp}$	ΔG^{cal}
n-pentane	−0.054	−0.007	−5.679	1.334	1.078	0.113	0.000	0.000	−3.215
1-pentanol	−1.140	−0.841	−6.098	1.428	1.138	0.113	−0.042	−0.564	−6.005
dimethylamine	−0.769	−0.518	−3.483	0.919	0.815	0.113	−0.009	−0.775	−3.707
rutin	−6.911	−9.518	−30.580	4.709	2.990	0.113	−0.280	−4.382	−40.640

parametrized with A^{exp} and B^{exp} has greater accuracy, and this tendency becomes clearer with increasing solvent polarity.

The absence of the influence of solvent accessibility on hydrogen bonding is another shortcoming of eq 4. Atoms must be exposed to solvent to interact with the solvent, and one way of measuring the degree of solvent accessibility of a solute atom is to compute the exposed surface area for the atom. This solvent-accessible area of an atom depends on the atomic radii, R_w, and optimized solute geometry. Because the parametrizations were performed with small molecules where most atoms
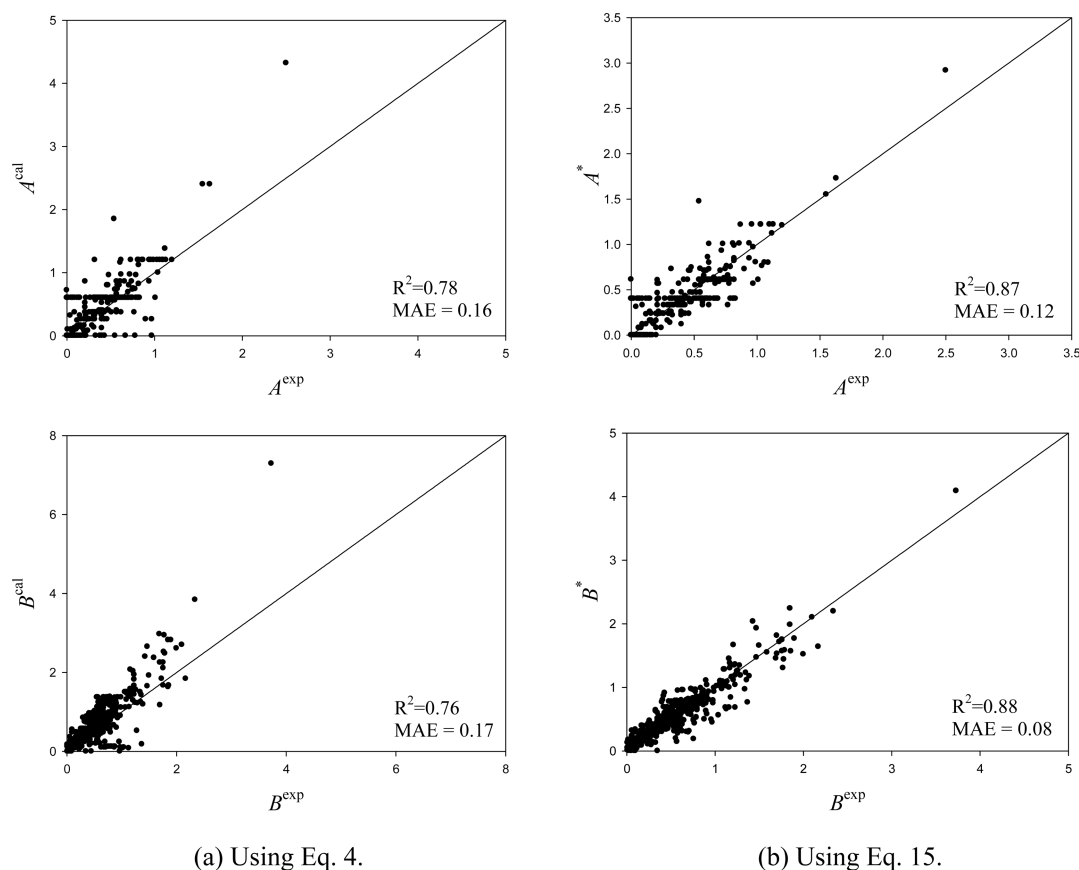
(a) Using Eq. 4. (b) Using Eq. 15.

Figure 3. Plots of empirical A and B against experimental A and B. R^2 and MAE are denoted on the figures.

Table 7. Solvation Free Energies (kcal/mol) and Hydrogen Bond Acidity and Basicity for Selected solutes

solute	$\Delta G^{exp\,a}$	Surface-SFED/HB^{exp}			Surface-SFED/HB^{cal}					
		A^{exp}	B^{exp}	ΔG^{cal}	A^{cal}	B^{cal}	ΔG^{cal}	A^*	B^*	ΔG^{cal}
rutin	−44.99	2.50	3.73	−50.42	4.32	7.29	−56.62	2.92	4.09	−52.19
xylitol	−13.25	0.54	1.43	−13.47	1.85	2.40	−18.51	1.47	2.03	−17.88
luteolin	−25.19	1.55	1.16	−25.37	2.40	2.07	−28.65	1.55	1.00	−25.19
diapocynin	−21.06^b	0.32	1.77	−22.15	1.20	2.52	−24.98	0.73	1.30	−23.71
salicylamide	−12.09	0.61	0.51	−13.16	1.16	0.98	−15.49	0.86	0.78	−14.39
trimethoprim	−19.47	0.21	1.69	−19.81	0.52	2.97	−21.20	0.46	1.45	−20.82
gallic acid	−19.73^c	1.63	0.86	−18.38	2.40	1.11	−19.72	1.73	0.80	−18.49
barbital	−11.58^d	0.58	1.12	−12.06	0.80	1.89	−12.95	0.71	0.68	−11.27
o-vanillin	−9.58^e	0.07	0.61	−9.42	0.60	1.20	−9.55	0.40	0.79	−10.15

^{a}Experimental solvation free energy in 1-butanol (kcal/mol), unless otherwise noted. ^{b}Experimental solvation free energy in 1-propanol. ^{c}Experimental solvation free energy in ethanol. ^{d}Experimental solvation free energy in chloroform. ^{e}Experimental solvation free energy in bromobenzene.

are exposed to solvent, we did not consider solvent accessibility. However, if we apply the SFED model to large molecules, such as proteins, the effect of solvent accessibility on hydrogen bonding must be considered.

We have expanded the application of the SFED model to diverse solvents in a series of studies. Among the 36 solvents dealt with in the present study, four solvents are 2−5 times enlarged set of our current work.[7] For the validation of the robustness of the model, the performance of parametrization of our prior study[7] (old parametrization) was tested with the data set from this work. The results are presented in Table 9. The MAEs of the old parametrization are between 0.5 and 0.7 and

are overall 14% higher than those seen in parametrization of this work (new parametrization). This increase in accuracy may be considered small in that (1) the test set is same as the training set used in this study; therefore, the MAE of the new parametrization is almost minimal for Surface-SFED/HB^{exp}. (2) The old parametrization was performed with a data set that was 1/2−1/5 the size of the current test set, consisting mainly of small nonpolar or unipolar small molecules. Therefore, the Surface-SFED model is a robust model for prediction of solvation free energies encompassing a broad range of common functional groups present in biological and drug-like molecules.

Table 8. Performance of the Surface-SFED Model by Solvent Functional class

	number of			MAE[a]	
solvent class	solvents[b]	solutes[c]	data[d]	Surface-SFED/HB[exp]	Surface-SFED/HB[cal,e]
alcohols	16	410	1824	0.46	0.55
aliphatic ethers	5	209	450	0.51	0.56
esters	4	177	345	0.53	0.55
haloaliphatics	4	473	755	0.52	0.51
haloaromatics	2	178	270	0.53	0.53
ketones	2	135	178	0.45	0.46
tertiary amides	3	215	372	0.48	0.51
total	36	787	4190	0.49	0.54

[a]Mean absolute error over data in this solvent class. [b]Number of solvents in this solvent class. [c]Number of solute molecules in this solvent class. [d]Total number of solvation free energies in this solvent class. [e]Hydrogen bond acidity and basicity were computed using eq 15.

Table 9. Comparison of Performance of the Surface-SFED/HB[exp] Model Parameterizations

		old parametrization[a]		new parametrization[b]
solvent	test set	training set	MAE	MAE
1-octanol	300	147	0.53	0.48
chloroform	369	79	0.69	0.63
carbon tetrachloride	254	64	0.44	0.38
diethyl ether	116	56	0.65	0.51
total	1039	346	0.58	0.51

[a]From ref 16. [b]This work. Test set is same with training set.

CONCLUSIONS

We have extended the application of the Surface-SFED model to 36 polar organic solvents. Each solvent consists of more than 50 neutral molecules, including diverse multipolar drug and natural molecules. The free energy of solvation was obtained by a linear combination of simple empirical functions for the electrostatic, dispersion−repulsion, hydrogen bonding, and cavitation terms. The linear expansion coefficients were optimized for each solvent model.

In the model, the contribution of hydrogen bond to the solvation free energy was dominant. The inaccuracy of the model with empirical hydrogen bond acidity and basicity calculated using eq 4 is due to the nonadditive character of A and B. The performance of the model is improved by reflecting the influence of substituents near the hydrogen bond donor and acceptor.

We have presented two parametrizations for solvation free energies in 36 solvents on the basis of the experimental and empirical hydrogen bond acidities and basicities, respectively. The MAEs for 4190 solvation free energies of 787 solutes in 36 solvents were 0.49 and 0.54 kcal/mol for Surface-SFED/HB[exp] and Surface-SFED/HB[cal] parametrizations, respectively. Errors arising from inaccuracies in $A*$ and $B*$ depend on the solute/solvent combination and are the main source of errors in the Surface-SFED/HB[cal] model

The Surface-SFED model has yielded accurate results for a very large set of solvation free energies of neutral molecules in diverse solvents and ionics in aqueous solvents. Each solvent model can be combined with other solvent models to calculate partition coefficients of interest. Because of the simplicity of the empirical functions and of the surface model, the Surface-SFED model conserves computational resources and is flexible in its application.

ASSOCIATED CONTENT

Ⓢ Supporting Information

Most recent Abraham's hydrogen bond acidity, A^{exp}, and basicity, B^{exp}, used in Surface-SFED/HB[exp] (Table S1) and experimental and calculated solvation free energies in each solvent (Table S2). This information is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Phone: 82-2-393-9551. Fax: 82-2-393-9554. E-mail: ktno@yonsei.ac.kr.

ACKNOWLEDGMENTS

REFERENCES

(1) Williams, R. L.; Vila, J.; Perrot, G.; Scheraga, H. A. Empirical solvation models in the context of conformation energy searches: Application to bovine pancreatic trypsin inhibitor. *Proteins: Struct., Funct., Genet.* **1992**, *14*, 110−119.

(2) Tomasi, J.; Persico, M. Molecular interactions in solution: An overview of methods based on continuous distributions of the solvent. *Chem. Rev.* **1994**, *94*, 2027−2094.

(3) Cramer, C. J.; Truhlar, D. G. Implicit solvation models: Equilibria, structure, spectra, and dynamics. *Chem. Rev.* **1999**, *99*, 2161−2200.

(4) Silvestrelli, P. L.; Parrinello, M. Structural, electronic, and bonding properties of liquid water from first principles. *J. Chem. Phys.* **1999**, *111*, 3572−3580.

(5) Tomasi, J.; Mennucci, B.; Cammi, R. Quantum mechanical continuum solvation models. *Chem. Rev.* **2005**, *105*, 2999−3093.

(6) Sharma, M.; Resta, R.; Car, R. Dipolar correlations and the dielectric permittivity of water. *Phys. Rev. Lett.* **2007**, *98*, 247401.

(7) Haberthür, U.; Caflisch, A. FACTS: Fast analytical continuum treatment of solvation. *J. Comput. Chem.* **2008**, *29*, 701−715.

(8) Lange, A. W.; Herbert, J. M. A simple polarizable continuum solvation model for electrolyte solutions. *J. Chem. Phys.* **2011**, *134*, 204110.

(9) Rashin, A; Young, L.; Atopol, I. Qunatitative evaluation of hydration thermodynamics with a continuum model. *Biophys. Chem.* **1994**, *51*, 359−374.

(10) Luque, F. J.; Zhang, Y.; Alemán, C.; Bachs, M.; Gao, J.; Orozco, M. Solvation effects in chloroform solution: parameterization of the MST/SCRF continuum model. *J. Phys. Chem.* **1996**, *100*, 4269−4276.

(11) Giesen, D. J.; Chambers, C. C.; Cramer, C. J.; Truhlar, D. G. Solvation model for chloroform based on class IV atomic charges. *J. Phys. Chem. B* **1997**, *101*, 2061−2069.

(12) Curutchet, C.; Orozco, M; Luque, F. J. Solvation in ocatnol: Parameterization of the continuum MST model. *J. Comput. Chem.* **2001**, *22*, 1180−1193.

(13) Bidon-Chanal, A.; Huertas, O.; Orozco, M.; Luque, F. J. Solvation enthalpies of neutral solutes in water and octanol. *Theor. Chem. Acc.* **2009**, *123*, 11−20.

(14) No, K. T.; Kim, S. G.; Cho, K. -H.; Scheraga, H. A. Description of hydration free energy density as a function of molecular physical properties. *Biophys. Chem.* **1999**, *78*, 127−145.

(15) In, Y.; Chai, H. H.; No, K. T. A partition coefficient calculation method with the SFED model. *J. Chem. Inf. Model.* **2005**, *45*, 254−263.

(16) Lee, S.; Cho, K. -H.; Lee., C. J.; Kim., G. E.; Na, C. H.; In, Y.; No, K. T. Calculation of the solvation free energy of neutral and ionic molecules in diverse solvents. *J. Chem. Inf. Model.* **2011**, *51*, 105−114.

(17) Son, S. H.; Han, C. K.; Ahn., S. K.; Yoon, J. H.; No, K. T. Development of three-dimensional descriptors represented by tensors: Free energy of hydration density tensor. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 601−609.

(18) No, K. T.; Grant, J. A.; Scheraga, H. A. Determination of net atomic charges using a modified partial equalization of orbital electronegativity method. 1. Application to neutral molecules as models for polypeptides. *J. Phys. Chem.* **1990**, *94*, 4732−4739.

(19) No, K. T.; Grant, J. A.; Jhon, M. S.; Scheraga, H. A. Determination of net atomic charges using a modified partial equalization of orbital electronegativity method. 1. Application to ionic and aromatic molecules as models for polypeptides. *J. Phys. Chem.* **1990**, *94*, 4140−4746.

(20) No, K. T.; Cho, K. H.; Jhon, M. S.; Scheraga, H. A. An empirical method to calculate average molecular polarizabilities from the dependence of effective atomic polarizabilities on net atomic charge. *J. Am. Chem. Soc.* **1993**, *115*, 2005−2014.

(21) Abraham, M. H. Scales of solute hydrogen-bonding: Their construction and application to physicochemical processes. *Chem. Soc. Rev.* **1993**, *22*, 73−83.

(22) Abraham, M. H. Hydrogen bonding. 31. Construction of a scale of solute effective or summation hydrogen-bond basicity. *J. Phys. Org. Chem.* **1993**, *6*, 660−684.

(23) Abraham, M. H. Hydrogen bonding. 32. An analysis of water−octanol and water−alkane partitioning and the log *p* parameter of seiler. *J. Pharm. Sci.* **1994**, *83*, 1085−1100.

(24) Grubbs, L. M.; Saifullah, M.; De La Rosa, N. E.; Ye, S.; Achi, S. S.; Acree, W. E.; Abraham, M. H. Mathematical correlations for describing solute transfer into functionalized alkane solvents containing hydroxyl, ether, ester or ketone solvents. *Fluid Phase Equilib.* **2010**, *298*, 48−53.

(25) Sprunger, L. M.; Achi, S. S.; Acree., W. E. Jr.; Abraham, M. H. Development of correlations for describing solute transfer into acyclic alcohol solvents based on the Abraham model and fragment-specific equation coefficients. *Fluid Phase Equilib.* **2010**, *288*, 139−144.

(26) Sprunger, L. M.; Achi, S. S.; Pointer, R.; Blake-Taylor, B. H.; Acree, W. E. Jr.; Abraham, M. H. Development of Abraham model correlations for solvation characteristics of linear alcohols. *Fluid Phase Equilib.* **2009**, *286*, 170−174.

(27) Sprunger, L. M.; Achi, S. S.; Pointer, R.; Acree, W. E. Jr.; Abraham, M. H. Development of Abraham model correlations for solvation characteristics of secondary and branched alcohols. *Fluid Phase Equilib.* **2010**, *288*, 121−127.

(28) Abraham, M. H.; Acree, W. E. Jr. The transfer of neutral molecules, ions and ionic species from water to ethylene glycol and to propylene carbonate; descriptors for pyridinium cations. *New J. Chem.* **2010**, *34*, 2298−2305.

(29) Abraham, M. H.; Acree, W. E. Jr.; Cometto-Muniz, J. E. Partition of compounds from water and from air into amides. *New J. Chem.* **2009**, *33*, 2034−2043.

(30) Abraham, M. H.; Acree, W. E. Jr.; Leo, A. J.; Hoekman, D. Partition of compounds from water and from air into wet and dry monohalobenzenes. *New J. Chem.* **2009**, *33*, 1685−1692.

(31) Sprunger, L. M.; Gibbs, J.; Acree, W. E. Jr.; Abraham, M. H. Correlation and prediction of partition coefficients for solute transfer to 1,2-dichloroethane from both water and from the gas phase. *Fluid Phase Equilib.* **2008**, *273*, 78−86.

(32) Sprunger, L. M.; Proctor, A.; Acree, W. E. Jr.; Abraham, M. H.; Benjelloun-Dakhama, N. Correlation and prediction of partition coefficients between the gas phase and water, and the solvents dry methyl acetate, dry and wet ethyl acetate, and dry and wet butyl acetate. *Fluid Phase Equilib.* **2008**, *270*, 30−44.

(33) Sprunger, L. M.; Achi, S. S.; Acree, W. E. Jr.; Abraham, M. H.; Leo, A. J.; Hoekman, D. Correlation and prediction of solute transfer to chloroalkanes from both water and the gas phase. *Fluid Phase Equilib.* **2009**, *281*, 144−162.

(34) Abraham, M. H.; Acree, W. E. Jr.; Leo, A. J.; Hoekman, D. The partition of compounds from water and from air into wet and dry ketones. *New J. Chem.* **2009**, *33*, 568−573.

(35) Saifullah, M.; Ye, S.; Grubbs, L. M.; De La Rosa, N. E.; Acree, W. E. Jr.; Abraham, M. H. Abraham model correlation for transfer of neutral molecules to tetrahydrofuran and to 1,4-dioxane, and for transfer of ions to tetrahydrofuran. *J. Solution Chem.* Accepted for publication.

(36) Abraham, M. H.; Smith, R. E.; Luchtefeld, R.; Boorem, R. E.; Luo, R.; Acree, W. E. Jr. Solubility of drugs and other compounds in organic solvents. *J. Pharm. Sci.* **2010**, *99*, 1500−1515.