———LETTER———

Comments on "Solvation Parameters. 2. A Simplified Molecular Topology To Generate Easily Optimized Values"

Christina Mintz,[†] William E. Acree, Jr.,*,[†] and Michael H. Abraham[‡]

Department of Chemistry, P.O. Box 305070, University of North Texas, Denton, Texas 76203-5070, and Department of Chemistry, University College London, 20 Gordon Street, London WC1H 0AJ, U.K.

Abstract: A polemic is presented in regards to a recent paper by Laffort and Héricourt [*J. Chem. Inf. Model.* **2006**, *46*, 1723−1734].

In a recent paper appearing in this Journal Laffort and Héricourt[1] presented a generalized method to establish the numerical values of the solvation parameters of solutes. The solvation parameters, when combined with the five solvent solvation parameters, describe the intermolecular forces present in fluid solutions. As part of their discussion the authors suggested two sets of optimized values of solute solvation parameters. The first set of parameters was based on the published Abraham solute descriptors (**E**, **S**, **A**, **B**, and **L**), modified by the authors

$$\delta_2 = \mathbf{L} - 0.532\mathbf{S} - 0.894\mathbf{R} - 0.115 \tag{1}$$

$$\omega_2 = 1.523\mathbf{S} - 0.538\mathbf{B} - 0.837\mathbf{E} \tag{2}$$

$$\epsilon_2 = \mathbf{E} \tag{3}$$

$$\alpha_2 = 2.825\mathbf{A} \tag{4}$$

$$\beta_2 = 0.728\mathbf{B} \tag{5}$$

to include scaling factors and greater independence (orthogonality). In eqs 1−5, δ_2 denotes the solute's Laffort et al. dispersion parameter, ω_2 is the solute's orientation factor, ϵ_2 refers to the polarizability-induction parameter of the solute, and α_2 and β_2 represent the solute's acidity and basicity parameters, respectively. Laffort et al.[1,2] deduced their second set of solvation parameters from experimental gas−liquid chromatographic retention indices of the solute on five selected stationary phases. The selected stationary phases were prepared in limited quantity and are not commercially available. Numerical solvation parameters were reported for 133 substances for the second of the two

proposed methods. The learning experimental data set used in the Laffort and Héricourt paper had solvation parameters for 241 of the 369 total compounds based on the recommended tranformation equations (eqs 1−5).

One of the main purposes of the transformation of the Abraham parameters by Laffort and Héricourt[1] was to obtain a set of descriptors that were as orthogonal as possible. Indeed, one of the inherent difficulties of using descriptors as the independent variables in multiple linear regression analysis is cross-correlation of the descriptors. Of course, such cross-correlation can be avoided by the use of principal component analysis, but then the coefficients of the resulting correlation equation are difficult to interpret. Our intent is not to criticize the work of Laffort and Héricourt[1] as there are definite advantages associated with having an orthogonal set of solute descriptors/parameters. Rather, we wish to point out some of the undesired consequences of using the published transformation expressions in regards to the Abraham solvation parameter model. The basic Abraham model employs two linear free energy relationships (LFERs) to describe solute transfer from the gas phase[3,4]

$$\mathrm{SP} = c + e\mathbf{E} + s\,\mathbf{S} + a\mathbf{A} + b\mathbf{B} + l\mathbf{L} \tag{6}$$

and solute transfer between two condensed phases

$$\mathrm{SP} = c + e\mathbf{E} + s\,\mathbf{S} + a\mathbf{A} + b\mathbf{B} + v\mathbf{V} \tag{7}$$

where SP denotes a solute property such as the logarithm of a gas−liquid partition coefficient or gas chromatographic retention indices. Eqs 6 and 7 use the updated symbolism for the solute descriptors and equation process coefficients. The published transformations involve the Abraham **E**, **S**, **A**, **B**, and **L** solute descriptors, hence our discussion will focus on solute transfer from the gas phase (eq 6).

Advantages that the Abraham model offers over other published LFERs is that the same set of descriptors are used to describe a wide range of chemical and biological properties, and both the equation coefficients and solute descriptors encode meaningful chemical information. For example, the excess molar refraction, **E**, is derived from the solute refractive index, and hence the *e*-coefficient gives a measure of general solute−solvent dispersion interactions. The **V** and **L** descriptors were set up as measures of the endoergic effect of disrupting solvent−solvent bonds. However, solute volume is always well correlated with polarizability, and so the *v*- and *l*-coefficients will include not only an endoergic cavity effect but also exoergic solute−solvent effects that arise through solute polarizability. The **S** descriptor is a measure of dipolarity and polarizability, and hence the *s*-coefficient will reflect the ability of a solvent to undergo dipole−dipole and dipole-induced dipole interactions with a solute. The **A** descriptor is a measure of solute hydrogen bond acidity, and hence the *a*-coefficient will reflect the complimentary solvent hydrogen bond basicity. Similarly the *b*-coefficient will be a measure of the solvent hydrogen bond acidity. All this is

* Corresponding author fax: (940)565-4318; e-mail: acree@unt.edu.
† University of North Texas.
‡ University College London.

1880 *J. Chem. Inf. Model., Vol. 46, No. 5, 2006*

LETTER

straightforward for gas-to-solvent partitions, because there are no interactions to consider in the gas phase. For partitions between two condensed phases, the coefficients in eq 7 then refer to differences between the properties of the two phases.

Such information is lost in transforming the Abraham solute descriptors to the solvation parameters suggested in the Laffort and Héricourt paper.[1] To illustrate this point, we write the transformation equations as follows:

$$\mathbf{A} = 0.3540\alpha_2 \tag{8}$$

$$\mathbf{B} = 1.3736\beta_2 \tag{9}$$

$$\mathbf{E} = \epsilon_2 \tag{10}$$

$$\mathbf{S} = 0.6566\omega_2 + 0.4852\beta_2 + 0.5496\epsilon_2 \tag{11}$$

$$\mathbf{L} = \delta_2 + 0.3493\omega_2 + 0.2581\beta_2 + 1.1864\epsilon_2 + 0.115 \tag{12}$$

Substitution of eqs 8–12 into eq 6, followed by suitable algebraic manipulation and grouping of like terms gives

$$
\begin{aligned}
\mathrm{SP} = {}& (c_{\text{abraham}} + 0.115 l_{\text{abraham}}) + \\
& (e_{\text{abraham}} + 0.5496 s_{\text{abraham}} + 1.1864 l_{\text{abraham}})\epsilon_2 + \\
& (0.6566 s_{\text{abraham}} + 0.3493 l_{\text{abraham}})\omega_2 + \\
& (0.3540 a_{\text{abraham}})\alpha_2 + (1.3736 b_{\text{abraham}} + 0.4852 s_{\text{abraham}} + \\
& 0.2581 l_{\text{abraham}})\beta_2 + l_{\text{abraham}}\delta_2 \tag{13}
\end{aligned}
$$

for the solute's gas-to-condensed phase transfer property. Similarly, an expression could be developed for converting the equation coefficients for correlations deduced using the Laffort et al. descriptors to equation coefficients based on the Abraham descriptors.

In Table 1 we have converted our existing Abraham equation coefficients[5–7] for 43 organic solvents to values that would be obtained if one were to use the transfer equations suggested by Laffort and Héricourt.[1] Careful examination of the tabulated numerical entries reveals that all of the alkane solvents have a nonzero β_2-coefficient. The β_2 solvation parameter coefficient should provide a meaningful measure of the solvent's hydrogen-bond acidity. Alkane solvents are incapable of hydrogen-bond formation, and the β_2-coefficient should be zero. As can be seen, the transformed β_2-coefficient provides incorrect information regarding the ability of saturated alkane solvents to function as hydrogen-bond acids.

A more dramatic example of how the suggested tranformation leads to incorrect chemical information is found by examining the correlation equations for dry diethyl ether, dry tetrahydrofuran, and dry 1,4-dioxane. The equations obtained by application of eq 6 are as follows, where K is the gas to solvent partition coefficient, equivalent to the Ostwald solubility coefficient.

$$
\begin{aligned}
\mathrm{Log}\,K_{\text{diethyl ether}} = {}& 0.288 - 0.347\mathbf{E} + 0.775\mathbf{S} + \\
& 2.985\mathbf{A} + 0.000\mathbf{B} + 0.973\mathbf{L} \tag{14}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Log}\,K_{\text{THF}} = {}& 0.189 - 0.347\mathbf{E} + 1.238\mathbf{S} + 3.289\mathbf{A} + \\
& 0.000\mathbf{B} + 0.982\mathbf{L} \tag{15}
\end{aligned}
$$

Table 1. Equation Coefficients Based on the Transformation Expressions and Scaling Factors Suggested by Laffort and Héricourt[1]

process/solvent	con	eps	omeg	alph	bet	del
1,4-dioxane (dry)	0.072	1.663	1.420	1.069	1.049	0.919
1-butanol (dry)	0.069	1.128	0.680	1.338	1.869	0.934
1-decanol (dry)	−0.027	1.264	0.544	1.301	1.456	0.947
1-heptanol (dry)	0.045	1.168	0.605	1.254	2.070	0.927
1-hexanol (dry)	0.073	1.157	0.738	1.319	1.547	0.936
1-octanol (dry)	−0.011	1.220	0.696	1.266	1.479	0.940
1-octanol (wet)	−0.099	1.410	0.765	1.246	2.528	0.858
1-pentanol (dry)	0.065	1.118	0.671	1.338	1.846	0.932
1-propanol (dry)	0.072	1.202	0.729	1.424	1.971	0.869
2-butanol (dry)	0.091	1.048	0.829	1.329	2.074	0.906
2-methyl-1-propanol (dry)	0.115	1.023	0.753	1.290	2.318	0.895
2-methyl-2-propanol (dry)	0.175	0.985	0.853	1.399	1.761	0.905
2-propanol (dry)	0.043	1.110	0.773	1.422	2.000	0.893
3-methyl-1-butanol (dry)	0.092	1.045	0.668	1.298	1.999	0.925
2-pentanol (dry)	0.077	1.056	0.652	1.342	1.888	0.934
diethyl ether (dry)	0.400	1.233	0.849	1.057	0.627	0.973
ethanol (dry)	0.110	1.240	0.816	1.287	2.404	0.853
ethyl acetate (dry)	0.309	1.440	1.142	1.044	0.844	0.917
tetrahydrofuran (dry)	0.302	1.498	1.156	1.164	0.854	0.982
1-decanol (dry)	−0.027	1.264	0.544	1.301	1.456	0.947
acetonitrile	0.078	1.633	1.874	0.738	1.959	0.738
benzene	0.224	1.476	1.048	0.162	1.006	1.020
cyclohexane	0.280	1.092	0.354	0.000	0.261	1.013
decane	0.270	1.030	0.345	0.000	0.255	0.989
ethylene glycol (dry)	−0.812	1.726	1.135	1.623	4.307	0.558
heptane	0.388	1.004	0.343	0.000	0.254	0.983
hexadecane	0.115	1.186	0.349	0.000	0.258	1.000
hexane	0.405	0.992	0.342	0.000	0.253	0.979
isooctane	0.387	0.909	0.340	0.000	0.251	0.972
methanol (dry)	0.085	1.342	1.039	1.310	2.735	0.769
methyl *tert*-butyl ether (dry)	0.392	1.129	0.873	0.883	0.645	0.993
methylcyclohexane	0.434	0.986	0.353	0.000	0.261	1.012
nonane	0.313	1.018	0.342	0.000	0.253	0.980
octane	0.326	1.098	0.338	0.000	0.250	0.967
toluene	0.237	1.494	0.969	0.165	0.852	1.012
undecane	0.225	1.152	0.339	0.000	0.251	0.971
1,2-dichloroethane	0.119	1.750	1.270	0.230	1.949	0.936
2,2,4-trimethylpentane	0.387	0.909	0.340	0.000	0.251	0.972
chlorobenzene	0.173	1.371	1.187	0.129	0.877	1.041
chloroform	0.230	1.373	1.137	0.049	2.807	0.994
dichloromethane	0.229	1.587	1.429	0.143	2.136	0.940
tetrachloromethane	0.403	1.192	0.668	0.000	0.493	1.047
dodecane	0.166	1.170	0.344	0.000	0.254	0.986
(gas to water)	−1.296	2.077	1.727	1.382	7.888	−0.213

$$
\begin{aligned}
\mathrm{Log}\,K_{\text{dioxane}} = {}& -0.034 - 0.347\mathbf{E} + 1.674\mathbf{S} + 3.021\mathbf{A} + \\
& 0.000\mathbf{B} + 0.919\mathbf{L} \tag{16}
\end{aligned}
$$

Eqs 14–16 all make chemical sense. In particular, the *a*-coefficients are all quite large, thus reflecting the hydrogen bond basicity of the ethers, and the *b*-coefficients are all zero, as expected for solvents that all have zero hydrogen bond acidity.

If we now use the Laffort and Héricourt [1] transformations, the equations are restated as

$$
\begin{aligned}
\mathrm{Log}\,K_{\text{diethyl ether}} = {}& 0.400 + 1.233\epsilon_2 + 0.8408\omega_2 + \\
& 1.057\alpha_2 + 0.627\beta_2 + 0.973\delta_2 \tag{17}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Log}\,K_{\text{THF}} = {}& 0.302 + 1.498\epsilon_2 + 1.156\omega_2 + 1.164\alpha_2 + \\
& 0.854\beta_2 + 0.982\delta_2 \tag{18}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Log}\,K_{\text{dioxane}} = {}& 0.072 + 1.663\epsilon_2 + 1.420\omega_2 + 1.069\alpha_2 + \\
& 1.049\beta_2 + 0.919\delta_2 \tag{19}
\end{aligned}
$$

The calculated β_2-coefficients are now quite large and positive, so that the transformed equations indicate, quite incorrectly, that ethers are hydrogen bond acids.

There are advantages associated with having an orthogonal set of solute descriptors. However, we note that the transformed descriptors of Laffort and Héricourt[1] will be orthogonal only for their original data set of 369 compounds. For any other data set, there will be loss of orthogonality. For example, in the data set used in deriving eq 19, the largest value of R^2 between descriptors is 0.682 for **E** and **L** in our descriptors, and R^2 is 0.421 between ϵ and δ in the Laffort descriptors. The dioxane database contains 105 compounds. For compounds not listed in Table C1 of the Laffort et al.[2] paper, we calculated the Laffort descriptors using the suggested transformation equations (eqs 1−5). PCA on any set of solute descriptors will always yield orthogonal PCs that can be used in linear regressions but has the disadvantage that the equation coefficients cannot be interpreted easily. The procedure of Abraham has the disadvantage that the solute descriptors may be too cross-correlated to yield a stable regression equation but has the advantage that the equation coefficients are easily interpretable. The transformations of Laffort and Héricourt[1] when applied to other data sets than the original 369 compound set will yield descriptors that are not exactly orthogonal and with the disadvantage that the equation coefficients cannot easily be interpreted. We do not believe that the advantage of a possible increase in orthogonality outweighs the chemical information that is lost when the coefficients take on unrealistic values. In several past studies[8−11] we have used a number of methods to compare derived correlation equations for biological processes with various correlation equations for gas-to-organic solvent and water-to-organic solvent partitions in order to determine the molecular interactions responsible for the observed biological property. Indeed, such analyses become meaningless if the equation coefficients were to lose their encoded chemical information.

So useful is the comparison of equation coefficients that Ishihama and Asakawa[12] developed a rigorous mathematical procedure for such comparison. A rather different mathematical procedure has been given by Abraham and Martins.[13,14] Given the procedures that have been developed for the comparison of equation coefficients, based on the coefficients encoding specific chemical information, and in view of our own observations, we prefer the existing solute descriptors over the calculated values based on the transformations suggested in the Laffort and Héricourt paper. We are unable at the present time to evaluate the second set of optimized solvation parameters recommended by Laffort and co-workers. There are too few compounds with known solvation parameters for us to determine what the calculated equation coefficients might be. The databases used in deriving the existing Abraham gas-to-organic solvent partitions contain a large number of compounds not found in the 133 compound method 1 solvation parameter database.

REFERENCES AND NOTES

(1) Laffort, P.; Héricourt, P. Solvation parameters. 2. A simplified molecular topology to generate easily optimized values. *J. Chem. Inf. Model.* **2006**, *46*, 1723−1734.

(2) Laffort, P.; Chauvin, F.; Dallos, A.; Callegari, P.; Valentin, D. Solvation parameters. Part 1: Mutual improvements of several approaches and determination of two first sets of optimized values. *J. Chromatogr., A* **2005**, *1100*, 90−107.

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

(4) Abraham, M. H.; Ibrahim, A.; Zissimos, A. M. The determination of sets of solute descriptors from chromatographic measurements. *J. Chromatogr., A* **2004**, *1037*, 29−47.

(5) Acree, W. E., Jr.; Abraham, M. H. Solubility predictions for crystalline nonelectrolyte solutes dissolved in organic solvents based upon the Abraham general solvation model. *Can. J. Chem.* **2001**, *79*, 1466−1476.

(6) Stovall, D. M.; Acree, W. E., Jr.; Abraham, M. H. Solubility of 9-fluorenone, thianthrene and xanthene in organic solvents. *Fluid Phase Equilib.* **2005**, *232*, 113−121.

(7) Stovall, D. M.; Hoover, K. R.; Acree, W. E., Jr.; Abraham, M. H. Solubility behavior of crystalline polycyclic aromatic hydrocarbons (PAHs): Prediction of fluorene in organic solvents with the Abraham solvation parameter model. *Polycyclic Aromat. Compd.* **2005**, *25*, 313−326.

(8) Abraham, M. H.; Ibrahim, A.; Acree, W. E., Jr. Air to blood distribution of volatile organic compounds: A linear free energy analysis. *Chem. Res. Toxicol.* **2005**, *18*, 904−911.

(9) Hoover, K. R.; Acree, W. E., Jr.; Abraham, M. H. Chemical toxicity correlations for several fish species based on the Abraham solvation parameter model. *Chem. Res. Toxicol.* **2005**, *18*, 1497−1505.

(10) Bowen, K. R.; Flanagan, K. B.; Acree, W. E., Jr.; Abraham, M. H. Correlating toxicities of organic compounds to select protozoa using the Abraham model. *Sci. Total Environ.* **2006**, available online as an article in press.

(11) Mintz, C.; Acree, W. E., Jr.; Abraham, M. H. Correlation of minimum inhibitory concentrations toward oral bacterial growth based on the Abraham model. *QSAR Comb. Sci.* **2006**, in press.

(12) Ishihama, Y.; Asakawa, N. Characterization of lipophilicity scales using vectors from solvation energy descriptors. *J. Pharm. Sci.* **1999**, *88*, 1305−1312.

(13) Abraham, M. H. Can we identify models for intestinal absorption, blood-brain barrier distribution and intestinal absorption? In *EuroQSAR 2002. Designing drugs and crop protectants: processes, problems and solutions*; Ford, M., Livingstone, D., Dearden, J., van de Waterbeemd, H., Eds.; Blackwell: Oxford, 2003; pp 5−7.

(14) Abraham, M. H.; Martins, F. Human skin permeation and partition; general linear free energy relationship analyses. *J. Pharm. Sci.* **2004**, *93*, 1508−1523.