The Correlation and Prediction of the Solubility of Compounds in Water Using an Amended Solvation Energy Relationship

MICHAEL H. ABRAHAM* AND JOELLE LE

Contribution from *Department of Chemistry, University College London, 20 Gordon Street, London, United Kingdom WC1H OAJ.*

Abstract □ The aqueous solubility of liquids and solids, as log S_W, has been correlated with an amended solvation equation that incorporates a term in $\Sigma\alpha_2^H \times \Sigma\beta_2^H$, where the latter are the hydrogen bond acidity and basicity of the solutes, respectively. Application to a training set of 594 compounds led to a correlation equation with a standard deviation, SD, of 0.56 log units. For a test set of 65 compounds, the SD was 0.50 log units, and for a combined correlation equation for 659 compounds, the SD was 0.56 log units. The correlation equations enable the factors that influence aqueous solubility to be revealed. The hydrogen-bond propensity of a compound always leads to an increase in solubility, even though the $\Sigma\alpha_2^H \times \Sigma\beta_2^H$ term opposes solubility due to interactions in the liquid or solid. Increase in solute dipolarity/polarizability increases solubility, whereas an increase in solute excess molar refraction, and especially, volume decrease solubility. The solubility of Bronsted acids and bases is discussed, and corrections for the fraction of neutral species in the saturated solution are graphically presented.

Introduction

The solubility of liquids and solids in water is a very important molecular property that influences the release, transport, and extent of absorption of drugs in the body and that is a key determinant of the environmental fate of agrochemicals and pollutants in the environment. Not surprisingly, numerous methods for the prediction of aqueous solubilities have been suggested. We restrict our discussion primarily to methods that include solid solutes because methods that predict only liquid solubilities are of limited use.

One of the first predictive methods for aqueous solubilities was that of Irmann,[1] who set up a group contribution scheme for liquid hydrocarbons and halocompounds. For solids, Irmann used an additional term, $\Delta S_m(T_m - T)/1364$, where ΔS_m is the entropy of fusion (melting) at the melting point T_m. A value of 13 cal deg^{-1} mol^{-1} was taken for ΔS_m, leading to the following simplified correction term

$$-0.0095(\text{mp} - 25) \tag{1}$$

In eq 1, mp is the melting point in °C; for liquids, the term (mp − 25) is taken as zero. Irmann[1] gave no statistical analysis, but we have used Irmann's original data, excluding compounds for which the observed solubility was given as approximate, and give details in Table 1. Several other group contribution schemes have been constructed,[2−5] some of which[2] do not require any mp correction term.

The UNIFAC and UNIQUAT methods are also group contribution schemes and have been used to estimate aqueous solubilities.[6,7] Because the reference state for solutes in these methods is the pure liquid, they require a knowledge of the solute enthalpy of fusion or an ap-proximate mp correction term for solids. Another type of group contribution scheme is used in the AQUAFAC program,[8−13] which was applied to 970 compounds.[10] Again, either the entropy of fusion or the mp is needed for solid solutes.

A number of correlations are based on theoretically calculated descriptors.[14−18] None of these require any mp correction term for solids and therefore are capable of predicting aqueous solubilities from structure. Interestingly, there is no discussion[14−18] on why it is not necessary to include a correction for solids.

Quite different types of calculation were initiated by Hansch and co-workers,[19] who showed that there was a relationship (eq 2) between log S_W (the solubility in mol dm^{-3}) and the water−octanol partition coefficient (log P_{oct}) for a training set of 156 liquids

$$\log S_W = -1.339 \log P_{oct} + 0.978$$

$$n = 156, \text{SD} = 0.472, r^2 = 0.874 \tag{2}$$

Yalkowsky and Valvani[20] extended the applicability of this relationship by incorporation of similar terms to those used by Irmann[1] for solids. They showed that the entropy of fusion could be estimated and that the entropy of fusion term could be replaced by a mp correction term as in eq 3 (compare eq 1). Several related equations were put forward:[10,21,22]

$$\log S_W = -1.05 \log P_{oct} - 0.012(\text{mp} - 25) + 0.87$$

$$n = 155, \text{SD} = 0.308, r^2 = 0.979 \tag{3}$$

$$\log S_W = -1.00 \log P_{oct} - \Delta S_m(\text{mp} - 25)/1364 + 0.87$$

$$n = 873 \tag{4}$$

$$\Delta S_m = 13.5 - 4.6(\log \sigma) \tag{5}$$

In eq 3, and elsewhere, n is the number of data points, SD is the standard deviation, r is the correlation coefficient and F is the F-statistic. Values of log P_{oct} in eq 4 were not experimental ones but were calculated by the CLOGP program. The entropies of fusion were a combination of experimental and calculated values, using eq 5 where σ is the rotational symmetry number. However, the compound mp is still needed to apply eq 4, so log S_W values cannot be calculated from structure.

Mobile Order Theory[23−25] has recently been applied to the estimation of aqueous solubility with impressive results.[24] However, the method requires not only the entropy of fusion of solid solutes (or a mp correction term), but also a modified nonspecific solute cohesion parameter. The latter is obtained either from experimental solubilities in hydrocarbon solvents or is "...deduced by analogy to similar compounds."[24]

Table 1—Models for the Correlation and Prediction of Aqueous Solubility (log S_W) that Require Additional Data (ΔS_f, mp, δ')

	training set			test set		
N	SD	outliers	N	SD	outliers	ref
168	0.31[a] AAE (0.17)	none				1
694	AAE (0.38)	none				4
68	0.61[b] AAE (0.45)	none				7
167	0.24	none				20
205	0.40[c]	none	none			24
873	0.56[d]	none	97	0.56[d] AAE (0.41)	none	10
873	0.80[e]	none	97	0.80[e] AAE(0.61)	none	10

[a] Calculated in this work. The AAE is given by AAE = Σ(|log $S_W$obs − log $S_W$calc|)/n. [b] Calculated in this work from data in Tables 2 and 3 in ref 7. [c] Calculated in this work. Note that the value in ref 24 at the foot of Table 4 is incorrect. [d] Aquasolve method. The SD value for the test set has been calculated in this work, and we have taken the SD value for the training set to be the same. [e] Equation 4. The SD value for the test set has been calculated in this work, and we have taken the SD value for the training set to be the same.

Table 2—Models for the Correlation and Prediction of Aqueous Solubility (log S_W) that Do Not Require Additional Data

	training set			test set		
N	SD	outliers	N	SD	outliers	ref
469[a]	0.46	none	25	0.50	none	3
			12	0.37	1	
483[b]	0.53	none	25	0.55	none	3
			19	0.86	2	
123[c]	0.22	4	13	0.23	none	17
123[d]	0.28	4	13	0.28	none	17
258	0.37	42				16
411	0.57	none				18
331	0.30	none	17	0.34	none	15

[a] Model I, that is not too general. A later analysis[4] gives AAE = 0.50 for 694 compounds. [b] Model II, that is very general. A later analysis[4] gives AAE = 0.56 for 614 compounds. [c] Using a neural network with nine descriptors. [d] Using a linear model with nine descriptors.

In Tables 1 and 2 we summarize the methods that have been applied more generally; that is, to large sets of structurally diverse compounds. It is not always easy to compare different methods because some have been constructed using only a training set, others have used both a training set and a test set. In addition, various statistics have been used to describe the goodness of fit between observed and calculated log S_W values in a training set and between observed and predicted log S_W values in a test set. We prefer the SD, given by SD = $\sqrt{[\Sigma(\log S_W\text{obs} - \log S_W\text{calc})^2/(n - 1 - p)]}$ (where p is the number of parameters), but the average absolute error (AAE) is sometimes used. Defined as AAE = Σ(|log $S_W$obs − log $S_W$calc|)/n, the AAE is always much smaller than the corresponding SD value. In addition, some workers list outliers but other workers do not. Because the number of outliers can be very large (42 out of 300 for the general case in ref 16), care has to be taken in judging one model against another. Wherever possible in Tables 1 and 2, we have calculated the SD, as already described, to provide a uniform basis of comparison. Even then, comparisons of the various models is difficult. Bodor and Huang[15] obtain a very low SD value of 0.30 log units for a 331 compound data set, using 18 theoretically calculated descriptors, and Sutter and Jurs[17] find even lower SD values of 0.27 and 0.22 for a 123 compound data set. Myrdal and co-workers[10] however, find a much larger SD of 0.56 for an 873 compound training set using the AQUASOLVE model. However, the 331 training set[15] includes very few complicated molecules and the 123 compound data set[17] no complicated molecules at all, whereas the 873 training set[10] is much more diverse.
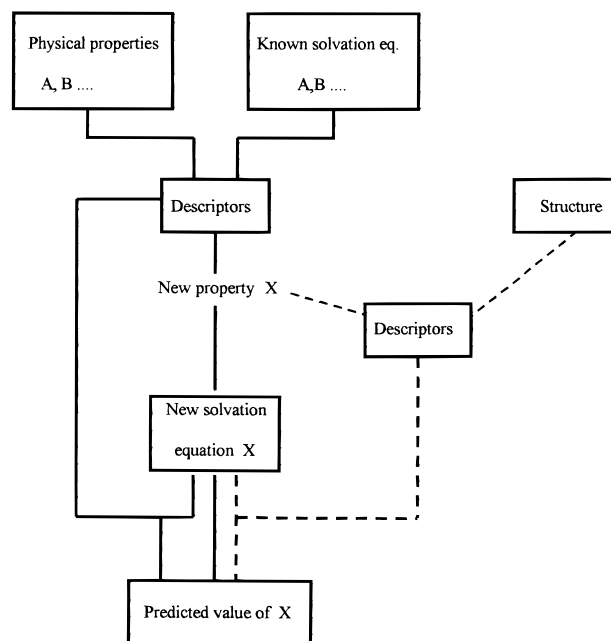

Figure 1—The application of the solvation equation. Full lines show the pathway using experimental descriptors and broken lines show the pathway using calculated descriptors.

Our conclusion is that for training sets that do not contain compounds of complicated structure, SD values as low as 0.30 log units may be obtained, but that for training sets that contain more varied compounds, SD values will not be lower than ~0.50 log units. Myrdal and co-workers[10] point out that the experimental solubilities themselves are a source of considerable error and note that recorded log S_W values for anthracene differ by 1.85 log units, and for fluoranthene by 1.15 log units. Hence, for training sets that contain a reasonable proportion of complicated structures, for many of which only one solubility determination has been made, experimental error probably precludes SD values less than around ~0.50 log units. What is also evident from Tables 1 and 2 is that there is no advantage, as regards SD values, of methods that require additional solute properties. Because there are very considerable advantages in methods that calculate log S_W from structure, especially in view of the importance of high throughput screening, our eventual aim is indeed to calculate aqueous solubility from structure.

Methodology

Our method starts with the following general solvation equation,[26]

$$\log SP = c + rR_2 + s\pi_2^H + a\Sigma\alpha_2^H + b\Sigma\beta_2^H + vV_x \quad (6)$$

Here, the dependent variable, log SP, is a property of a series of solutes in a given system, such as log P_{oct} or log S_W, and the independent variables are solute descriptors as follows:[26] R_2 is an excess molar refraction in units of (cm^3 mol^{-1})/10, π_2^H is the dipolarity/polarizability, $\Sigma\alpha_2^H$ is the overall or summation hydrogen-bond acidity, $\Sigma\beta_2^H$ is the overall or summation hydrogen-bond basicity, and V_x is the McGowan characteristic volume[27] in units of (cm^3 mol^{-1})/100. The coefficients in eq 6 are found by multiple linear regression analysis, using a set of solutes for which the descriptors are known. There are numerous applications of eq 6 to physicochemical properties, both by ourselves[28−47] and by other workers,[48−59] so that eq 6 can be regarded as a well-established general equation.

Table 3—Test Set of Compounds

compound	log S_W exp	log S_W calc	error	compound	log S_W exp	log S_W calc	error
2,2-dimethylbutane	−3.550	−3.293	0.257	diphenylmethane	−4.080	−4.347	−0.267
3-methylheptane	−5.160	−4.416	0.744	2,3-dimethylnaphthalene	−4.720	−4.790	−0.070
hexadecane	−8.400	−8.909	−0.509	anthracene	−6.350	−5.361	0.989
cycloheptane	−3.510	−3.697	−0.187	pyrene	−6.176	−6.108	0.068
2-methylbut-1-ene	−2.730	−2.324	0.406	benzo[ʲ]fluoranthene	−8.000	−7.513	0.487
buta-1,3-diene	−1.870	−1.547	0.323	1,4-difluorobenzene	−1.970	−1.972	−0.002
ethyne	0.290	−0.122	−0.412	1,2,3,5-tetrachlorobenzene	−4.630	−4.789	−0.159
trichloromethane	−1.170	−1.591	−0.421	3-chlorobiphenyl	−4.880	−4.951	−0.071
hexachloroethane	−3.670	−4.237	−0.567	2-bromonaphthalene	−4.400	−4.609	−0.209
1-chlorohexane	−3.120	−3.242	−0.122	4-chloroiodobenzene	−4.030	−4.417	−0.387
tribromomethane	−1.910	−2.478	−0.568	anthraquinone	−5.190	−3.709	1.481
1-bromoheptane	−4.430	−4.074	0.356	3-methylaniline	−0.850	−1.470	−0.620
bromodichloromethane	−1.540	−1.835	−0.295	N-ethylaniline	−1.700	−2.222	−0.522
methyl butyl ether	−0.990	−0.965	0.025	3-nitrotoluene	−2.440	−2.552	−0.112
tetrahydrofuran	1.150	0.167	−0.983	2,4-dinitrotoluene	−2.820	−2.200	0.620
2-ethylhexan-2-al	−2.460	−2.052	0.408	lidocaine	−1.710	−2.431	−0.721
heptan-2-one	−1.450	−1.354	0.096	2-methylbenzoic acid	−2.060	−1.761	0.299
propyl formate	−0.490	−0.497	−0.007	2-aminobenzoic acid	−1.520	−1.021	0.499
pentyl acetate	−1.890	−1.847	0.043	3,5-dimethylphenol	−1.400	−1.823	−0.423
ethyl heptylate	−2.740	−2.946	−0.206	2,4-dichlorophenol	−1.550	−2.266	−0.716
acetonitrile	0.260	0.803	0.543	2,3,4,6-tetrachlorophenol	−3.100	−3.506	−0.406
diethylamine	1.030	0.395	−0.635	4-hydroxybenzoic acid	−1.410	−1.110	0.300
acetamide	1.580	1.850	0.270	1-phenylethanol	−0.920	−1.087	−0.167
trichloroacetic acid	0.600	−0.023	−0.623	2,4-dimethylpyridine	0.380	−0.696	−1.076
pentan-2-ol	−0.290	−0.442	−0.152	morpholine	1.965	1.587	−0.378
2-methylpentan-1-ol	−1.110	−1.221	−0.111	codeine	−1.520	−2.318	−0.798
3,3-dimethylbutan-1-ol	−0.500	−1.222	−0.722	17a-methyltestosterone	−3.999	−4.318	−0.319
2,4-dimethylpentan-2-ol	−0.920	−1.483	−0.563	5-ethyl-5-(3-methylbutyl)barbital	−2.658	−2.580	0.078
decan-1-ol	−3.630	−3.423	0.207	5-allyl-5-phenylbarbital	−2.369	−3.309	−0.940
pent-4-ene-1-ol	−0.150	−0.315	−0.165	carbofuran	−2.800	−3.102	−0.302
diethyl disulfide	−2.420	−2.593	−0.173	fenoxycarb	−4.700	−4.821	−0.121
isopropylbenzene	−3.270	−3.576	−0.306	propoxur	−2.050	−1.631	0.419
1,4-diethylbenzene	−3.750	−4.090	−0.340				

The application of eq 6 is summarized in Figure 1 (full lines). Various sets of physical properties, already calibrated through known solvation equations, can be used to assign descriptors, exactly as detailed before.[60] In this way, a database of descriptors for some 3500 compounds has been established. If a new property, X, is to be investigated, the experimental database is used to obtain a correlation equation for the new property, through eq 6. Thus we have recently constructed[47] an equation for water−chloroform partition coefficients, log P_{chl}, for solutes as neutral species.

Once such an equation has been set up, more values of the dependent variable can be predicted from the experimental database of descriptors, as shown by the full lines in Figure 1. However, a further step is to calculate descriptors from structure, so that values of property X may be predicted from structure. This prediction is essential for any fast throughput screening of drugs, agrochemicals, etc. We have just completed a computer program, ABSOLVE, for the calculation of descriptors from structure;[61] this is then tantamount to a method for the prediction from structure of any property, X, for which we have solvation equations, as shown by the broken lines in Figure 1.

The aim of the present work is to obtain an equation for the correlation of log S_W values, without the need for a mp correction, using a large training set and also a reasonably large test set of compounds. Together with the program ABSOLVE, we will then be in a position to predict solubilities from structure.

Results and Discussion

Construction of an Equation for log S_W—We have used a number of databases to set out values of log S_W for 664 solids and liquids. We excluded five compounds from the 664 data set (cyclopropyl-5-spirobarbituric acid, uracil, chlorpheniramine, fentanyl, and adenine) because a preliminary analysis showed that these four were large outliers to all the equations we constructed. In addition we have not included any dicarboxylic acids, such as phthalic acid and succinic acid, partly because we have not yet finalised descriptors for these and partly because preliminary analyses suggest that calculated values of log S_W are always too positive. A total of 659 compounds were left for the final analysis. Every tenth compound in a random order was selected to form a test set, to give 594 compounds as a training set and 65 compounds as a test set. The total set of 664 compounds is given in the Appendix. The smaller test set of 65 compounds is in Table 3. Application of eq 6 to the 594 training set yielded eqs 7 and 8, with SD values of 0.56 and 0.63 log units, respectively; note that fewer compounds were used in eq 7, because of lack of mps. These SD values are not far short of the SD values obtained for models that have been applied to large data sets[3,18] (see Tables 1 and 2).

$$\log S_W = 0.579 - 0.576 R_2 + 0.980 \pi_2^H + 1.233 \Sigma \alpha_2^H + 3.389 \Sigma \beta_2^H - 4.079 V_x - 0.010(\text{mp} - 25)$$

$$n = 411, \text{SD} = 0.564, r^2 = 0.915, F = 724, \text{AAE} = 0.389 \quad (7)$$

$$\log S_W = 0.849 - 1.061 R_2 + 0.851 \pi_2^H + 0.646 \Sigma \alpha_2^H + 3.279 \Sigma \beta_2^H - 4.050 V_x$$

$$n = 594, \text{SD} = 0.630, r^2 = 0.895, F = 1004, \text{AAE} = 0.470 \quad (8)$$

It is somewhat surprising that eq 6 has led to the reasonable eqs 7 and 8, because eq 6 was not set up at all to

Table 4—Correlation Equations Without Compounds of Low and High Solubilities[a]

coefficients							statistics				condition[b]
r	s	a	b	k	v	c	r^2	SD	n	AAE	
−1.025	0.799	2.026	4.003	−2.953	−3.900	0.450	0.912	0.539	610	0.402	1
−0.891	0.693	2.160	4.231	−3.459	−3.913	0.451	0.901	0.534	636	0.388	2
−0.912	0.713	2.015	3.965	−3.047	−3.781	0.348	0.884	0.512	587	0.385	3
−1.020	0.813	2.124	4.187	−3.337	−3.986	0.510	0.918	0.562	594	0.409	4
−1.004	0.771	2.168	4.238	−3.362	−3.987	0.518	0.920	0.557	659	0.408	5

[a] All equations are without any mp correction term. [b] (1) Omit very soluble compounds; (2) omit very insoluble compounds; (3) omit both very soluble and very insoluble compounds; (4) eq 10; (5) eq 11.

correlate quantities such as log S_W. There is a fundamental difference between processes such as water−solvent partitions, to which eq 6 has previously been applied, and solubility in water. In the former processes, the thermodynamic standard states are those of unit molar concentration and unit activity in both the aqueous and the solvent phase. For solubility in water, the standard states are unit molar concentration and unit activity in the aqueous phase, but the pure liquid or solid (as the other phase). As pointed out before,[62] the standard state of pure liquid or pure solid is equivalent to a different standard state for each compound. Now eq 6 is constructed for processes in which different solutes have the same standard state in each phase. In chemical terms, this means that a solute in a given phase is surrounded by the phase molecules, whereas for the standard state of pure liquid or solid, the solute is surrounded by itself. Difficulties in application to aqueous solubility of equations similar to eq 6 have previously been encountered; for example, aliphatic and aromatic compounds give rise to quite different correlation equations.[63]

We can amend eq 6 to incorporate terms that reflect interactions in the pure liquid or solid. A term in $\Sigma\alpha_2^H \times \Sigma\beta_2^H$ will deal with hydrogen-bond interactions between acid and basic sites in the solid or liquid, and a term in $\pi_2^H \times \pi_2^H$ with dipole/dipole interactions. The best equations constructed on these lines are

$$\log S_W = 0.403 - 0.484R_2 + 0.814\pi_2^H + 1.956\Sigma\alpha_2^H + 4.018\Sigma\beta_2^H - 1.130\Sigma\alpha_2^H \times \Sigma\beta_2^H - 4.067V_x - 0.010(\text{mp} - 25)$$

$$n = 411,\ \text{SD} = 0.496,\ r^2 = 0.934,\ F = 819,\ \text{AAE} = 0.245\ \ (9)$$

$$\log S_W = 0.510 - 1.020R_2 + 0.813\pi_2^H + 2.124\Sigma\alpha_2^H + 4.187\Sigma\beta_2^H - 3.337\Sigma\alpha_2^H \times \Sigma\beta_2^H - 3.986V_x$$

$$n = 594,\ \text{SD} = 0.562,\ r^2 = 0.918,\ F = 1089,\ \text{AAE} = 0.409\ \ (10)$$

Inspection of eqs 7−10 shows that there is little to be gained by inclusion of the mp correction term (compare Tables 1 and 2). The equations with the cross-term are significantly better than those without this term, and eq 9 is better than eq 10. However, the practical advantages of eq 10 quite outweigh the better fit of eq 9; in any case, eq 10 compares well with the equations listed in Table 2 that cover a wide range of compound type.

We can probe the predictive capability of eq 10 through the test set of 65 compounds given in Table 3, where the observed and calculated log S_W values for eq 10 are given. The SD value for the 65 compound test set is 0.496 log units, AAE = 0.397, and av error = −0.122, which we can take as an estimate of the predictive power of eq 10.

Finally, we can combine the training set and test set and obtain eq 11 for the total of 659 compounds.

$$\log S_W = 0.518 - 1.004R_2 + 0.771\pi_2^H + 2.168\Sigma\alpha_2^H + 4.238\Sigma\beta_2^H - 3.362\Sigma\alpha_2^H\Sigma\beta_2^H - 3.987V_x$$

$$n = 659,\ \text{SD} = 0.557,\ r^2 = 0.920,\ F = 1256,\ \text{AAE} = 0.408\ \ (11)$$

We consider eq 11 to be the best equation we have constructed from the general solvation descriptors, and conclude that an amended version of eq 6, containing the extra $\Sigma\alpha_2^H \times \Sigma\beta_2^H$ term, can correlate and predict log S_W values to ~0.56 log units. The calculated values of log S_W from eq 11 are included in the Appendix.

There are particular experimental difficulties with regard to compounds that have very low solubilities. To ascertain if such compounds were exerting any undue influence on the regression, we re-ran the correlation leaving out the very insoluble compounds. We also left out separately the very soluble compounds, and finally omitted both the very insoluble and very soluble compounds. A summary of the resulting equations is given in Table 4, where the coefficient of the product term $\Sigma\alpha_2^H \times \Sigma\beta_2^H$ is denoted as 'k'. By comparison with eqs 10 and 11, changes in the regression coefficients are not very pronounced and so there is little disrupting effect of compounds with very low or very high solubilities.

More important effects probably arise when the solid in equilibrium with the saturated solution is a hydrate, because the solubility of the hydrate will not be the same as the unhydrated solid, to which all the correlation equations refer. In addition, solubilities may depend on the physical form of a solid, for example whether it is amorphous or crystalline.

The Factors that Influence Aqueous Solubility— Unlike most regression equations for log S_W, eqs 10 and 11 can be interpreted to show the physicochemical properties of the compound that influence aqueous solubility. We have already noted that most studies of aqueous solubility in which correlations are constructed without any correction term for solids do not discuss why a correction term is unnecessary. Neither eq 10 nor eq 11 include a solid correction term, and it seems obligatory to comment on this. The two main properties that lead to an increase in solubility are hydrogen-bond acidity and hydrogen-bond basicity; these properties no doubt reflect the strong hydrogen-bond basicity and strong hydrogen-bond acidity of water as a bulk solvent.[28,29] However, if the compound is itself both a hydrogen-bond acid and a hydrogen-bond base, then intermolecular hydrogen-bond interactions will lead to an increase in mp and to a decrease in solubility. Thus, the product term, $\Sigma\alpha_2^H \times \Sigma\beta_2^H$, takes the place (at least partly) of a solid correction term.

Table 5—Hydrogen-Bond Effects on Solubility (log S_W)

compound	2.124 $\Sigma\alpha_2^H$	4.187 $\Sigma\beta_2^H$	-3.337 $\Sigma\alpha_2^H \times \Sigma\beta_2^H$	resultant
acetic acid	1.30	1.84	−0.90	2.24
trichloroacetic acid	2.02	1.17	−0.89	2.30
benzoic acid	1.25	1.67	−0.79	2.14
phenol	1.27	1.26	−0.60	1.93
4-nitrophenol	1.75	1.09	−0.71	2.12
ethanol	0.79	2.01	−0.59	2.20
2,2,2-trifluoroethanol	1.21	1.05	−0.48	1.78
estratriol	2.97	5.11	−5.70	2.38
aniline	0.55	1.72	−0.36	1.91
benzamide	1.04	2.81	−1.10	2.75
pyrazole	1.15	1.88	−0.81	2.22
morpholine	0.13	3.81	−0.18	3.76
progesterone	0.00	4.77	0.00	4.77
trichloromethane	0.32	0.08	−0.01	0.39

Now all hydrogen-bond acids, with the exception of carbon acids, are also hydrogen-bond bases, so that the effect of hydrogen-bonding on solubility will be a resultant of the two single terms and the product term, as shown in Table 5 for some representative acids. It is quite clear that the net result of the presence of hydrogen-bond acid and hydrogen-bond base groups will increase solubility. The intermolecular acid−base interaction in a solid or liquid, given by the $\Sigma\alpha_2^H \times \Sigma\beta_2^H$ term, reduces the hydrogen-bond effect but still leaves a negative resultant. For the large number of compounds that are hydrogen-bond bases, but not acids, there is a straightforward effect of increased solubility (also shown in Table 5). As already mentioned, there are but few compounds that are hydrogen-bond acids and yet have no or very little hydrogen-bond basicity. Again, there will be virtually no cross-term, and all the effect of hydrogen-bond acidity will be toward an increase in solubility, as shown for trichloromethane. The single terms in the descriptors $\Sigma\alpha_2^H$ and $\Sigma\beta_2^H$ both lead to an increase in solubility. The other 'polar' term in eq 10 is $s.\pi_2^H$ that leads also to an increase in solubility. It might be expected that intermolecular interactions, such as dipole/dipole or dipole/induced dipole would lead to an increase in mp and, again, a decrease in solubility. However, the product term $\pi_2^H \times \pi_2^H$ is not significant, no doubt because it is very well correlated with π_2^H which leads to an increase in solubility. However the coefficient of π_2^H in eq 10 is very much less than for the solubility of gases and vapors, so that dipolar effects within the solid or liquid counteract to some extent the solute/water effects that lead to increased solubility.

Two other terms in eq 10, rR_2 and vV_x, both result in a decrease in solubility; the r and v coefficients in eq 10 are markedly more negative than in the solvation equation for gaseous solubility. The R_2 descriptor refers to the propensity of a solute to interact with surrounding σ and π electrons, the negative r coefficient suggesting that such interaction within the solid or liquid is much larger than the corresponding interaction between the solute and bulk water. Although the V_x descriptor refers to the size of the solute, the vV_x term for the solubility of gaseous solutes will be the resultant of two opposing effects, (i) a cavity effect that arises from the disruption of solvent−solvent interactions and leads to a negative coefficient, and (ii) a general solute−solvent dispersion interaction that leads to a positive coefficient. For the solubility of gaseous solutes in water, the v coefficient is negative (−0.869) so that the unfavorable cavity effect dominates.[29] In solids and liquids, part of the cohesive forces will be general dispersion interactions that help to hold the solid or liquid
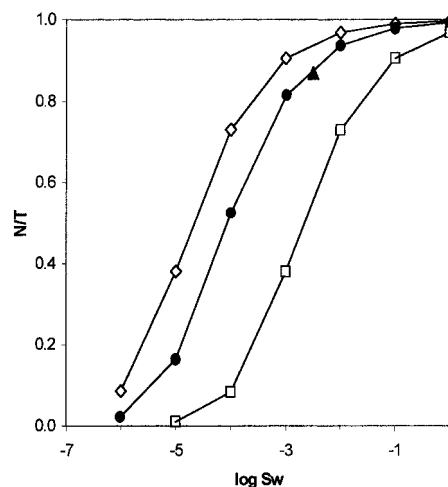
Figure 2—Values of N/T for Bronsted acids as a function of the total solubility, log S_W. Acids pK_a values are 3 (□), 4.37 (●), and 5 (◇), determined value (▲) for p-toluic acid.

together. These interactions within the solid or liquid will lead to an increase in mp, to a decrease in solubility, and to a much more negative v coefficient in eq 10, as observed (−3.986).

Thus, the sign and magnitude of the coefficients in eq 10 can be interpreted in terms of known chemical interactions, both between the compound and water, and between molecules of the compound itself. Such interpretation, in turn, leads to information about the physicochemical factors that influence the aqueous solubility of solids and liquids.

We note that in Mobile Order Theory[23−25] the terms in solute volume are interpreted as originating from a balance of entropic effects. The negative dependence of solubility on solute volume arises[24] from the mobile order entropy decrease of the hydrogen-bonded water molecules on introduction of the solute. We have used a cavity theory to interpret volume and other effects because we have used[29] this type of theory (i.e., scaled particle theory[64]) previously to obtain quantitative estimates of cavity terms for solution of gases and vapors in water.

The Solubility of Bronsted Acids and Bases—Grant and Higuchi[65] have noted the effect of pH on the solubility of Bronsted acids and bases and have given equations for the variation of solubility with pH.[65] Most studies on the correlation and prediction of solubility ignore this pH dependency; none of the studies in Tables 1 and 2 mention this problem at all.

If a Bronsted acid, such as a carboxylic acid, is dissolved in water, the pH of the resulting solution will depend both on the acid pK_a and on the total concentration of the acid in solution. For a given acid, the greater the concentration, the lower will be the pH and the larger will be the proportion of the neutral species. Hence, for acids with the same pK_a, the pH of the saturated solution will decrease as the solubility increases. For acids that are quite soluble, the proportion of neutral species will therefore be larger than for acids that are sparingly soluble. Our eq 11 and, indeed, all the other correlation equations in Tables 1 and 2, refer to the solubility of the neutral species, N, so that the predicted (neutral) solubility will be less than the observed solubility, T or S_W, with the difference depending on the acid pK_a value and the actual solubility.

In Figure 2 we give the calculated values of N/T, the fraction of the neutral species, for a series of acids of pK_a 3, 4.37, and 5 as a function of the observed total solubility, log S_W (log T). For an acid with a pK_a of 4.37 or 5, N/T is
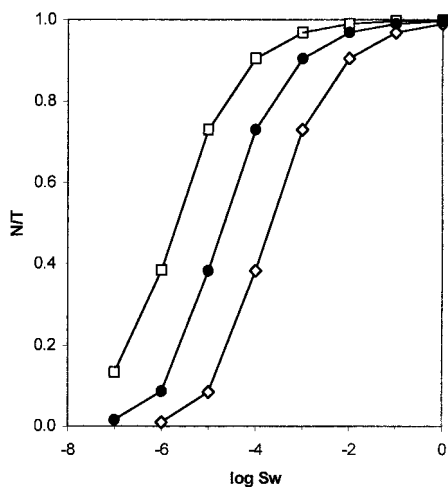
Figure 3—Values of N/T for Bronsted bases as a function of the total solubility, log S_W; pK_a values are 8 (□), 9 (●), and 10 (◇).

Table 6—Observed and Calculated log S_W Values for Strong Bronsted Acids ($pK_a < 4$)

acid	obs	calc	obs − calc	pK_a
trichloroacetic acid	0.600	−0.023	0.623	0.65
o-aminobenzoic acid	−1.520	−1.021	−0.499	2.11
chloroacetic acid	1.810	0.858	0.952	2.82
m-bromobenzoic acid	−2.276	−1.771	−0.505	2.85
o-chlorobenzoic acid	−1.890	−1.544	−0.346	2.94
salicylic acid	−1.820	−1.515	−0.305	2.98
p-nitrobenzoic acid	−2.800	−1.577	−1.223	3.42
m-nitrobenzoic acid	−1.680	−1.557	−0.123	3.49
m-chlorobenzoic acid	−2.590	−1.790	−0.800	3.87
o-toluic acid	−2.060	−1.761	−0.299	3.95
p-bromobenzoic acid	−3.539	−2.171	−1.368	3.97
p-chlorobenzoic acid	−3.310	−1.863	−1.447	3.98

larger than ∼0.5 even down to log S_W values of −4. Now an error of a factor of 0.5 (or 2.0) corresponds to an error of 0.3 log units and is not very important in the context of SD values of 0.5 log units. However, for very insoluble acids, with log S_W of −5 or −6, errors of one or two log units will arise if no consideration is given to ionization of Bronsted acids. For stronger acids with $pK_a = 3$, large errors will arise at log S_W values less than around −3.5 units.

We have chosen one of our pK_a values as 4.37 because this is the pK_a of p-toluic acid, studied in considerable detail by Strong and co-workers.[66] Their determined N/T value for p-toluic acid in the saturated solution at 25 °C is shown in Figure 2, and our calculated value is in excellent agreement.

A similar ionization phenomenon occurs in the solubility of Bronsted bases. Many drug molecules, of course, are strong Bronsted bases, with pK_a values of the conjugate acid ranging from 8 to 10. We give in Figure 3 plots of our calculated N/T values for three series of bases with $pK_a = 8$, 9, and 10 as a function of the observed solubility, log S_W. If we consider substantial errors in predicted values to arise when N/T is less than ∼0.5, these errors will be the case for bases with log $S_W < -4$ (pK_a 10), < -5 (pK_a 9) and < -6 (pK_a 8).

We can compare observed solubilities with those calculated from eq 10 for the neutral species for Bronsted acids and bases in our data set. In Table 6 are given values for strong Bronsted acids, that is, acids with pK_a values <4. Only in the case of trichloroacetic acid is the value of N/T

Table 7—Observed and Calculated log S_W Values for Strong Bronsted Bases ($pK_a > 10$)

base	obs	calc	obs − calc	pK_a
octylamine	−1.460	−2.151	0.691	10.57
hexylamine	−0.250	−1.038	0.788	10.64
butylamine	0.960	0.057	0.903	10.66
pentylamine	0.270	−0.491	0.761	10.64
heptylamine	−0.900	−1.600	0.700	10.66
propylamine	1.520	0.618	0.902	10.69
ethylamine	2.060	1.169	0.891	10.70
triethylamine	−0.140	−0.363	0.223	10.85
diethylamine	1.030	0.395	0.635	11.04
dibutylamine	−1.440	−1.804	0.364	11.25

<0.5 (0.480) for the saturated solution. The observed solubility would therefore be ∼0.32 log units more than the calculated solubility. For all the other acids in Table 6, the difference will be even less. Inspection of Table 6 shows that for three acids, the observed − calculated log S_W values are −1.2 to −1.5 units, so that other interfering factors are far more important than ionization, at least for the acids in Table 6.

A similar table can be constructed for the strong bases, those with $pK_a > 10$ (see Table 7). There is a general trend, with (observed − calculated) log S_W values always positive by ∼0.7 log units, on average. However, this trend cannot be accounted for by ionization; even the value of 0.80 for N/T for dibutylamine would make a difference of only 0.1 log units.

It seems, therefore, that only for very insoluble strong Bronsted acids and Bronsted bases will ionization lead to significant errors in calculation. However, it is worth pointing out that solubilities calculated with eqs 10 or 11, or by the methods summarized in Tables 1 and 2, refer to the solubility of the neutral species. For Bronsted acids and bases this will be the solubility in solutions of pH near to the compound pK_a (see Figures 2 and 3). The observed solubility is that at the pH of the saturated solution. This observed solubility does not refer to any specific pH, but to a pH that has to be calculated from the observed (total) solubility and the compound pK_a. If the solubility of a Bronsted acid or base is required at a given pH of 7 or 7.4, for example, then Figures 2 and 3 can be used to obtain the correction factor N/T, at least if the difference in pH between the saturated solution and the given pH is not too large.

Conclusions

An amended solvation equation can satisfactorily correlate and predict log S_W values to 0.56 log units. The descriptors used in the correlation equations are either calculated from structure (R_2 and V_x) or are obtained from experimental data (π_2^H, $\Sigma\alpha_2^H$, $\Sigma\beta_2^H$) as shown by the full lines in Figure 1. Now that the ABSOLVE program for the calculation of π_2^H, $\Sigma\alpha_2^H$, and $\Sigma\beta_2^H$ is in place, the correlation equations set up in this work will enable log S_W values to be predicted from structure in a high throughput manner, as shown by the broken lines in Figure 1. Although the correlation equations are for neutral species, corrections for Bronsted acids and Bronsted bases can be made from the predicted log S_W value and a (predicted) pK_a value.

ref	compound	log S_W obs	log S_W calc (eq 11)	ref	compound	log S_W obs	log S_W calc (eq 11)
62	methane	−0.900	−0.477	16	propyne	−0.410	−0.451
62	ethane	−1.360	−1.038	78	1-butyne	−1.240	−1.161
62	propane	−1.940	−1.600	4	1-pentyne	−1.640	−1.656
62	butane	−2.570	−2.162	4	1-hexyne	−2.360	−2.254
67	2-methylpropane	−2.550	−2.162	78	3-hexyne	−1.990	−2.300
4	pentane	−3.180	−2.723	78	1-heptyne	−3.010	−2.906
78	2-methylbutane	−3.180	−2.723	4	1-octyne	−3.660	−3.463
4	hexane	−3.840	−3.285	4	1-nonyne	−4.240	−4.020
4	2-methylpentane	−3.740	−3.285	78	dichloromethane	−0.630	−0.990
4	3-methylpentane	−3.680	−3.285	4	trichloromethane	−1.170	−1.590
4	2,2-dimethylbutane	−3.550	−3.285	4	tetrachloromethane	−2.310	−2.595
78	2,3-dimethylbutane	−3.650	−3.285	78	chloroethane	−1.060	−1.022
4	heptane	−4.530	−3.847	4	1,1-dichloroethane	−1.290	−1.353
78	2,2-dimethylpentane	−4.360	−3.847	4	1,2-dichloroethane	−1.060	−1.293
78	2,3-dimethylpentane	−4.280	−3.847	4	1,1,1-trichloroethane	−2.000	−2.175
78	2,4-dimethylpentane	−4.260	−3.847	4	1,1,2-trichloroethane	−1.480	−1.703
78	3,3-dimethylpentane	−4.230	−3.847	4	1,1,2,2-tetrachloroethane	−1.740	−2.211
78	2,2,3-trimethylbutane	−4.360	−3.847	4	1,1,1,2-tetrachloroethane	−2.180	−2.520
4	octane	−5.240	−4.408	78	pentachloroethane	−2.600	−3.032
78	2-methylheptane	−5.080	−4.408	4	hexachloroethane	−3.670	−4.225
78	3-methylheptane	−5.160	−4.408	4	1-chloropropane	−1.470	−1.573
78	4-methyloctane	−6.050	−4.408	4	2-chloropropane	−1.410	−1.487
4	2,2,4-trimethylpentane	−4.740	−4.408	78	1,2-dichloropropane	−1.600	−1.808
78	2,3,4-trimethylpentane	−4.800	−4.408	4	1,3-dichloropropane	−1.620	−1.839
78	nonane	−5.880	−4.970	4	1-chlorobutane	−2.030	−2.128
78	2,2,5-trimethylhexane	−5.050	−4.970	78	1-chloro-2-methylpropane	−2.000	−2.048
78	decane	−6.980	−5.532	4	2-chlorobutane	−1.960	−2.061
78	undecane	−7.590	−6.094	4	1-chloropentane	−2.730	−2.688
78	dodecane	−7.670	−6.655	4	2-chloro-2-methylbutane	−2.510	−2.851
78	tetradecane	−7.960	−7.779	78	1-chlorohexane	−3.120	−3.243
78	hexadecane	−8.400	−8.902	67	1-chloroheptane	−4.000	−3.797
4	cyclopentane	−2.640	−2.477	14	chloroethylene	−1.750	−1.109
4	methylcyclopentane	−3.300	−3.001	4	1,1-dichloroethylene	−1.640	−1.732
78	propylcyclopentane	−4.740	−4.124	4	cis-1,2-dichloroethylene	−1.300	−1.379
78	pentylcyclopentane	−6.080	−5.243	4	trichloroethylene	−1.960	−2.279
4	cyclohexane	−3.100	−3.081	4	tetrachloroethylene	−2.540	−3.121
4	methylcyclohexane	−3.850	−3.613	4	hexachloro-1,3-butadiene	−4.920	−5.116
78	cis-1,2-dimethylcyclohexane	−4.300	−4.181	78	bromomethane	−0.790	−0.820
78	trans-1,4-dimethylcyclohexane	−4.470	−4.096	4	dibromomethane	−1.170	−1.563
78	ethylcyclohexane	−4.250	−4.163	4	tribromomethane	−1.910	−2.474
78	cycloheptane	−3.510	−3.688	4	tetrabromomethane	−3.140	−3.738
78	cyclooctane	−4.150	−4.313	4	bromoethane	−1.090	−1.287
78	decalin	−5.190	−4.981	4	1,2-dibromoethane	−1.680	−1.716
62	ethylene	−0.400	−0.601	4	1-bromopropane	−1.730	−1.848
62	propylene	−1.080	−1.174	4	2-bromopropane	−1.590	−1.768
62	1-butene	−1.940	−1.732	4	1-bromobutane	−2.370	−2.404
15	2-methylpropene	−2.330	−1.732	4	1-bromo-2-methylpropane	−2.430	−2.404
4	1-pentene	−2.680	−2.287	67	1-bromopentane	−3.080	−2.962
67	cis-2-pentene	−2.540	−2.335	67	1-bromohexane	−3.810	−3.516
4	trans-2-pentene	−2.540	−2.320	67	1-bromoheptane	−4.430	−4.072
78	2-methyl-1-butene	−2.730	−2.314	67	1-bromooctane	−5.060	−4.630
78	3-methyl-1-butene	−2.730	−2.257	4	iodomethane	−1.000	−1.303
78	2-methy-2-butene	−2.560	−2.353	67	diiodomethane	−2.340	−2.419
4	1-hexene	−3.230	−2.834	78	iodoethane	−1.600	−1.766
78	2-methyl-1-pentene	−3.030	−2.846	4	1-iodopropane	−2.290	−2.322
4	1-heptene	−3.730	−3.409	4	2-iodopropane	−2.090	−2.264
4	trans-2-heptene	−3.820	−3.437	4	1-iodobutane	−2.960	−2.878
4	1-octene	−4.440	−3.973	67	1-iodoheptane	−4.810	−4.543
4	1-nonene	−5.050	−4.531	4	bromochloromethane	−0.890	−1.315
78	1-decene	−5.510	−5.096	4	bromodichloromethane	−1.540	−1.841
78	1,3-butadiene	−1.870	−1.539	4	chlorodibromethane	−1.900	−2.119
4	2-methyl-1,3-butadiene	−2.030	−2.094	4	1-chloro-2-bromoethane	−1.320	−1.691
78	2,3-dimethyl-1,3-butadiene	−2.400	−2.525	4	1,1,2-trichlorotrifluoroethane	−3.040	−2.624
4	1,4-pentadiene	−2.090	−1.988	78	1,2-dichlorotetrafluoroethane	−2.740	−2.410
4	1,5-hexadiene	−2.680	−2.556	4	diethyl ether	−0.090	−0.337
4	cyclopentene	−2.100	−1.874	4	dipropyl ether	−1.100	−1.427
78	cyclohexene	−2.590	−2.499	4	diisopropyl ether	−1.100	−1.635
4	1-methylcyclohexene	−3.270	−3.073	4	dibutyl ether	−1.850	−2.543
78	cycloheptene	−3.180	−3.065	4	methyl propyl ether	−0.390	−0.441
78	1,4-cyclohexadiene	−2.060	−2.022	4	methyl butyl ether	−0.990	−0.945
16	ethyne	0.290	−0.132	4	methyl tert-butyl ether	−0.240	−0.228

ref	compound	log S_W obs	log S_W calc (eq 11)	ref	compound	log S_W obs	log S_W calc (eq 11)
4	ethyl propyl ether	−0.660	−0.859	4	methyl acrylate	−0.220	−0.253
4	propyl isopropyl ether	−1.340	−1.444	4	glyceryl triacetate	−0.600	0.726
4	ethyl vinyl ether	−0.850	−0.391	4	malonic acid diethyl ester	−0.820	−0.534
4	dimethoxymethane	0.480	0.787	4	acetonitrile	0.260	0.795
4	1,1-diethoxyethane	−0.430	0.589	4	propionitrile	0.280	0.421
4	1,2-diethoxyethane	−0.770	0.149	4	acrylonitrile	0.150	0.347
4	1,2-propylene oxide	−0.590	0.701	14	ethylamine	2.060	1.200
68	tetrahydrofuran	0.490	0.182	14	propylamine	1.520	0.650
4	2-methyltetrahydrofuran	0.110	−0.151	14	butylamine	0.960	0.089
4	tetrahydropyran	−0.030	−0.369	14	pentylamine	0.270	−0.460
4	propionaldehyde	0.580	0.549	14	hexylamine	−0.250	−1.007
4	butyraldehyde	−0.010	−0.004	14	heptylamine	−0.900	−1.569
4	valeraldehyde	−0.850	−0.542	14	octylamine	−1.460	−2.121
4	caproaldehyde	−1.300	−1.086	14	diethylamine	1.030	0.429
14	2-ethylbutanal	−1.520	−1.103	4	dipropylamine	−0.460	−0.665
14	2-ethylhexanal	−2.130	−2.237	14	dibutylamine	−1.440	−1.771
4	tert-crotonaldehyde	0.320	0.294	14	trimethylamine	1.320	0.855
14	2-ethyl-2-hexanal	−2.460	−2.043	4	triethylamine	−0.140	−0.321
4	2-butanone	0.520	0.310	11	tripropylamine	−2.282	−1.853
4	2-pentanone	−0.190	−0.244	4	nitromethane	0.260	0.628
67	3-pentanone	−0.280	−0.271	4	nitroethane	−0.220	0.148
15	3-methyl-2-butanone	−0.120	−0.259	4	1-nitropropane	−0.800	−0.492
4	2-hexanone	−0.800	−0.799	4	2-nitropropane	−0.620	−0.404
15	3-hexanone	−0.830	−0.814	4	chloropicrin	−2.000	−2.069
15	3-methyl-2-pentanone	−0.670	−0.797	4	acetamide	1.580	1.859
4	4-methyl-2-pentanone	−0.740	−0.797	4	N,N-dimethylacetamide	1.110	1.344
15	3,3-dimethyl-2-butanone	−0.720	−0.815	4	urea	0.960	2.317
67	2-heptanone	−1.450	−1.348	4	o-ethyl carbamate	0.850	0.787
4	4-heptanone	−1.300	−1.353	4	acetic acid	2.000	1.184
15	2,4-dimethyl-3-pentanone	−1.300	−1.358	4	hexanoic acid	−1.060	−0.994
67	2-octanone	−2.050	−1.894	4	decanoic acid	−3.440	−3.191
4	2-nonanone	−2.580	−2.467	4	methacrylic acid	0.000	0.047
15	5-nonanone	−2.580	−2.466	4	chloroacetic acid	1.810	0.870
67	2-decanone	−3.300	−3.018	4	trichloroacetic acid	0.600	−0.014
4	cyclohexanone	−0.600	−0.284	15	methanol	1.560	1.594
24	carvone	−2.060	−2.704	15	ethanol	1.100	1.043
24	camphor	−1.960	−2.152	15	1-propanol	0.620	0.492
24	menthone	−2.350	−2.626	4	2-propanol	0.430	0.698
4	methyl formate	0.580	0.607	4	1-butanol	0.000	−0.058
4	ethyl formate	0.150	0.076	67	2-methylpropan-1-ol	0.100	−0.074
4	propyl formate	−0.490	−0.495	16	butan-2-ol	0.470	−0.074
15	isopropyl formate	−0.630	−0.392	4	1-pentanol	−0.600	−0.615
15	butyl acetate	−1.370	−1.046	4	2-pentanol	−0.290	−0.409
4	isobutyl formate	−1.010	−0.958	4	3-pentanol	−0.240	−0.432
15	isopentyl formate	−1.520	−1.517	4	2-methylbutanol	−0.470	−0.638
15	methyl acetate	0.460	0.361	16	3-methylbutan-1-ol	−0.510	−0.611
4	ethyl acetate	−0.040	−0.180	16	2-methylbutan-2-ol	0.150	−0.332
4	propyl acetate	−0.720	−0.743	69	3-methyl-2-butanol	−0.180	−0.431
4	isopropyl acetate	−0.550	−0.644	4	2,2-dimethylpropanol	−0.400	−0.512
4	isobutyl acetate	−1.210	−1.203	4	1-hexanol	−1.240	−1.167
4	pentyl acetate	−1.890	−1.841	4	2-hexanol	−0.890	−0.962
4	isopentyl acetate	−1.920	−1.764	4	3-hexanol	−0.800	−0.975
4	methyl propionate	−0.140	−0.217	69	2-methylpentanol	−1.110	−1.191
4	ethyl propionate	−0.660	−0.753	4	3-methyl-2-pentanol	−0.720	−1.191
4	methyl butyrate	−0.820	−0.757	69	4-methylpentanol	−1.140	−1.176
4	ethyl butyrate	−1.280	−1.296	69	2-methyl-2-pentanol	−0.490	−0.868
4	propyl butyrate	−1.920	−1.855	69	3-methyl-2-pentanol	−0.710	−0.968
4	methyl pentanoate	−1.360	−1.321	4	4-methyl-2-pentanol	−0.800	−0.965
4	ethyl pentanoate	−1.750	−1.839	4	2-methyl-3-pentanol	−0.700	−1.005
62	propyl propanoate	−1.340	−1.299	69	3-methyl-3-pentanol	−0.360	−0.910
62	pentyl propanoate	−2.250	−2.417	4	2-ethyl-1-butanol	−1.170	−1.211
62	methyl hexanoate	−1.870	−1.854	69	2,2-dimethyl-1-butanol	−1.040	−1.229
62	ethyl hexanoate	−2.350	−2.394	4	3,3-dimethyl-1-butanol	−0.500	−1.191
62	ethyl heptanoate	−2.740	−2.940	4	3,3-dimethyl-2-butanol	−0.620	−1.015
62	methyl octanoate	−3.170	−2.963	4	1-heptanol	−1.810	−1.730
62	ethyl octanoate	−3.390	−3.499	4	2-heptanol	−1.550	−1.525
62	methyl nonanoate	−3.380	−3.515	4	3-heptanol	−1.470	−1.515
62	ethyl nonanoate	−3.800	−4.058	4	4-heptanol	−1.400	−1.517
62	methyl decanoate	−4.690	−4.074	69	2-methyl-2-hexanol	−1.080	−1.424
62	ethyl decanoate	−4.100	−4.611	69	3-methyl-3-hexanol	−0.980	−1.458

ref	compound	log S_W obs	log S_W calc (eq 11)	ref	compound	log S_W obs	log S_W calc (eq 11)
4	3-ethyl-3-pentanol	−0.850	−1.495	4	2,3-dimethylnaphthalene	−4.720	−4.790
69	2,2-dimethylpentanol	−1.520	−1.785	4	2,6-dimethylnaphthalene	−4.890	−4.718
4	2,4-dimethyl-2-pentanol	−0.920	−1.445	4	1-ethylnaphthalene	−4.170	−4.784
4	2,4-dimethyl-3-pentanol	−1.220	−1.591	4	2-ethylnaphthalene	−4.290	−4.728
4	1-octanol	−2.390	−2.282	78	1,2,3,4-tetrahydronaphthalene	−4.370	−3.825
4	2-octanol	−2.090	−2.059	4	indan	−3.040	−3.309
69	3-octanol	−1.980	−2.077	4	acenaphthene	−4.630	−4.369
4	2-methyl-2-heptanol	−1.720	−1.994	4	acenaphthylene	−3.960	−4.105
4	3-methyl-3-heptanol	−1.600	−2.014	4	fluorene	−5.000	−4.608
4	2-ethyl-1-hexanol	−2.110	−2.315	4	1-methylfluorene	−5.220	−5.170
67	1-nonanol	−3.010	−2.835	4	anthracene	−6.350	−5.361
69	2-nonanol	−2.740	−2.628	4	2-methylanthracene	−6.960	−5.826
69	1-decanol	−3.630	−3.395	4	9-methylanthracene	−5.890	−5.869
69	2-undecanol	−2.940	−3.745	4	9,10-dimethylanthracene	−6.570	−6.345
69	1-dodecanol	−4.800	−4.502	4	phenanthrene	−5.260	−5.121
69	1-tetradecanol	−5.840	−5.614	78	1-methylphenanthrene	−5.850	−5.713
69	1-pentadecanol	−6.350	−6.170	78	2-methylphenanthrene	−5.840	−5.713
69	1-hexadecanol	−7.000	−6.726	4	fluoranthene	−6.000	−5.975
69	1-octadecanol	−8.400	−7.842	4	benzo[a]fluorene	−6.680	−6.794
4	cyclohexanol	−0.440	−0.636	4	benzo[b]fluorene	−8.040	−6.767
69	cycloheptanol	−0.880	−1.219	67	pyrene	−6.176	−6.115
69	cyclooctanol	−1.290	−1.834	4	7,12-dimethylbenz[a]anthracene	−7.020	−8.125
69	4-pentene-1-ol	−0.150	−0.284	4	naphthacene	−8.600	−7.071
69	1-hexene-3-ol	−0.590	−0.879	67	chrysene	−8.057	−6.932
4	2-butoxyethanol	−0.420	−0.239	78	5-methylchrysene	−6.590	−7.494
4	ethanethiol	−0.600	−0.797	78	6-methylchrysene	−6.570	−7.494
4	butanethiol	−2.180	−1.910	78	5,6-dimethylchrysene	−7.010	−8.056
70	dimethyl sulfide	−0.450	−0.574	67	triphenylene	−6.726	−6.666
70	diethyl sulfide	−1.340	−1.539	67	perylene	−8.804	−7.404
70	di-n-propyl sulfide	−2.580	−2.647	4	benzo[b]fluoranthene	−8.230	−7.380
70	diisopropyl sulfide	−2.240	−2.452	4	benzo[j]fluoranthene	−8.000	−7.519
70	dimethyl disulfide	−1.440	−1.514	4	benzo[k]fluoranthene	−8.490	−7.604
70	diethyl disulfide	−2.420	−2.581	78	cholanthrene	−7.850	−7.453
4	thiourea	0.320	1.139	4	3-methylcholanthrene	−7.920	−8.037
4	triethyl phosphate	0.430	0.226	71	benzo[a]pyrene	−8.699	−7.832
4	benzene	−1.640	−1.956	4	benzo[e]pyrene	−7.800	−7.917
4	toluene	−2.210	−2.509	67	benzo[ghi]perylene	−9.018	−8.509
4	ethylbenzene	−2.770	−3.048	78	picene	−7.870	−8.803
4	o-xylene	−2.800	−3.018	4	fluorobenzene	−1.800	−2.024
4	m-xylene	−2.820	−3.008	78	1,3-difluorobenzene	−2.000	−2.153
4	p-xylene	−2.770	−2.998	78	1,4-difluorobenzene	−1.970	−1.978
4	propylbenzene	−3.370	−3.609	78	benzyl trifluoride	−2.510	−2.500
4	isopropylbenzene	−3.270	−3.572	4	chlorobenzene	−2.380	−2.749
4	1,2,3-trimethylbenzene	−3.200	−3.479	4	1,2-dichlorobenzene	−3.050	−3.419
78	1,2,4-trimethylbenzene	−3.310	−3.466	4	1,3-dichlorobenzene	−3.040	−3.517
4	1,3,5-trimethylbenzene	−3.400	−3.469	4	1,4-dichlorobenzene	−3.270	−3.480
78	2-ethyltoluene	−3.210	−3.519	4	1,2,3-trichlorobenzene	−4.000	−4.173
78	4-ethyltoluene	−3.110	−3.500	78	1,2,4-trichlorobenzene	−3.590	−4.162
4	butylbenzene	−4.060	−4.159	4	1,3,5-trichlorobenzene	−4.480	−4.223
4	isobutylbenzene	−4.120	−4.169	4	1,2,3,4-tetrachlorobenzene	−4.570	−4.766
4	tert-butylbenzene	−3.660	−4.066	4	1,2,3,5-tetrachlorobenzene	−4.630	−4.800
78	1,2-diethylbenzene	−3.280	−4.097	4	1,2,4,5-tetrachlorobenzene	−5.560	−4.792
4	1,4-diethylbenzene	−3.750	−4.084	4	pentachlorobenzene	−5.650	−5.374
78	1,2,4,5-tetramethylbenzene	−4.590	−4.059	4	hexachlorobenzene	−7.680	−5.999
78	2-isopropyltoluene	−3.760	−4.043	78	2-chlorotoluene	−3.520	−3.355
78	4-isopropyltoluene	−3.770	−4.012	78	4-chlorotoluene	−3.080	−3.313
4	pentylbenzene	−4.640	−4.714	78	benzyl chloride	−2.390	−2.182
4	tert-pentylbenzene	−4.150	−4.713	4	1-chloronaphthalene	−3.930	−4.360
4	pentamethylbenzene	−4.000	−4.602	4	2-chloronaphthalene	−4.140	−4.363
4	hexylbenzene	−5.210	−5.281	69	2-chlorobiphenyl	−4.540	−4.991
78	hexamethylbenzene	−5.230	−4.921	69	3-chlorobiphenyl	−4.880	−4.960
4	styrene	−2.820	−2.964	4	bromobenzene	−2.550	−2.977
78	diphenylmethane	−4.080	−4.348	78	1,2-dibromobenzene	−3.500	−4.019
4	bibenzyl	−4.620	−4.910	78	1,3-dibromobenzene	−3.540	−4.061
67	biphenyl	−4.345	−4.262	4	1,4-dibromobenzene	−4.070	−4.056
78	4-methylbiphenyl	−4.620	−4.831	4	1,3,5-tribromobenzene	−5.600	−5.132
4	naphthalene	−3.600	−3.598	4	1,2,4,5-tetrabromobenzene	−6.980	−5.880
4	1-methylnaphthalene	−3.700	−4.164	78	2-bromotoluene	−2.230	−3.587
78	2-methylnaphthalene	−3.770	−4.124	78	4-bromotoluene	−3.190	−3.527
4	1,3-dimethylnaphthalene	−4.290	−4.769	78	1-bromonaphthalene	−4.350	−4.690
4	1,4-dimethylnaphthalene	−4.140	−4.790	4	2-bromonaphthalene	−4.400	−4.617
67	1,5-dimethylnaphthalene	−4.679	−4.789	67	iodobenzene	−3.010	−3.420

ref	compound	log S_W obs	log S_W calc (eq 11)	ref	compound	log S_W obs	log S_W calc (eq 11)
78	1-iodonaphthalene	−4.550	−5.156	24	p-fluoroacetanilide	−1.780	−1.113
14	o-fluorobromobenzene	−2.700	−3.285	24	p-chloroacetanilide	−2.843	−1.808
14	m-fluorobromobenzene	−2.670	−3.245	24	p-bromoacetanilide	−3.083	−2.085
20	o-chlorobromobenzene	−3.190	−3.898	11	4-nitroacetanilide	−2.692	−1.831
20	m-chlorobromobenzene	−3.210	−3.919	71	phenacetin	−2.350	−1.741
20	p-chlorobromobenzene	−3.630	−3.883	24	lidocaine	−1.710	−2.405
20	o-chloroiodobenzene	−3.540	−4.425	4	benzoic acid	−1.550	−1.055
20	m-chloroiodobenzene	−3.550	−4.417	4	o-toluic acid	−2.060	−1.742
20	p-chloroiodobenzene	−4.030	−4.427	4	m-toluic acid	−2.140	−1.662
14	p-bromoiodobenzene	−4.560	−4.639	4	p-toluic acid	−2.600	−1.609
4	anisole	−1.850	−2.038	4	o-chlorobenzoic acid	−1.890	−1.517
78	2-chloroanisole	−2.460	−2.916	4	m-chlorobenzoic acid	−2.590	−1.771
72	3-chloroanisole	−2.780	−2.896	4	p-chlorobenzoic acid	−3.310	−1.848
72	4-chloroanisole	−2.780	−2.782	67	2-bromobenzoic acid	−2.276	−1.739
4	diphenyl ether	−3.960	−4.579	67	4-bromobenzoic acid	−3.539	−2.157
4	benzaldehyde	−1.190	−1.362	4	m-nitrobenzoic acid	−1.680	−1.530
4	p-methoxybenzaldehyde	−1.490	−1.527	4	p-nitrobenzoic acid	−2.800	−1.554
4	acetophenone	−1.280	−1.533	4	o-aminobenzoic acid	−1.520	−1.020
4	benzophenone	−3.120	−3.564	4	aspirin	−1.720	−1.132
4	anthraquinone	−5.190	−3.728	4	phenylacetic acid	−0.890	−1.088
4	methyl benzoate	−1.850	−1.890	71	ibuprofen	−3.760	−3.927
4	ethyl benzoate	−2.320	−2.407	71	naproxen	−4.202	−3.892
4	dimethyl phthalate	−1.660	−1.315	4	phenol	0.000	−0.728
4	diethyl phthalate	−2.350	−2.310	4	2-methylphenol	−0.620	−1.440
4	di(2-ethylhexyl)phthalate	−6.960	−8.942	4	3-methylphenol	−0.680	−1.256
4	benzonitrile	−1.000	−1.446	67	p-cresol	−0.730	−1.331
4	phthalonitrile	−2.380	−1.103	4	2,4-dimethylphenol	−1.190	−1.819
4	aniline	−0.410	−1.012	67	2,6-dimethylphenol	−1.290	−1.963
4	o-toluidine	−2.210	−1.501	67	3,4-dimethylphenol	−1.380	−1.734
4	m-methylaniline	−0.850	−1.458	4	3,5-dimethylphenol	−1.400	−1.800
4	p-methylaniline	−1.210	−1.435	67	2,4,6-trimethylphenol	−2.050	−2.393
4	o-chloroaniline	−1.520	−1.959	4	p-tert-butylphenol	−2.410	−2.767
4	m-chloroaniline	−1.370	−1.816	4	thymol	−2.220	−2.813
4	p-chloroaniline	−1.660	−1.768	4	p-phenylphenol	−3.480	−3.182
4	o-nitroaniline	−1.960	−1.747	4	2-chlorophenol	−1.060	−1.564
4	m-nitroaniline	−2.190	−1.438	4	3-chlorophenol	−0.700	−1.372
4	p-nitroaniline	−2.370	−1.234	4	4-chlorophenol	−0.700	−1.297
4	ethyl-p-aminobenzoate	−2.100	−2.031	4	4-bromophenol	−1.090	−1.603
24	risocaine	−2.452	−2.598	67	2,3-dichlorophenol	−1.300	−2.222
24	butamben	−3.082	−3.173	4	2,4-dichlorophenol	−1.550	−2.249
4	N-methylaniline	−1.280	−1.611	67	2,6-dichlorophenol	−1.790	−2.224
4	N-ethylaniline	−1.700	−2.208	67	3,4-dichlorophenol	−1.250	−1.810
4	N,N-dimethylaniline	−1.920	−2.459	67	3,5-dichlorophenol	−1.340	−1.829
4	N,N-diethylaniline	−3.030	−3.586	67	2,3,4-trichlorophenol	−2.670	−2.570
4	1-naphthylamine	−1.920	−2.446	67	2,3,5-trichlorophenol	−2.670	−2.600
4	p,p′-biphenyldiamine	−2.700	−2.744	67	2,3,6-trichlorophenol	−2.640	−2.705
4	benzylamine	−1.540	−0.426	4	2,4,5-trichlorophenol	−2.210	−2.641
24	procaine	−1.780	−2.226	4	2,4,6-trichlorophenol	−2.340	−2.667
71	diphenylamine	−3.504	−3.632	67	2,3,4,5-tetrachlorophenol	−3.150	−3.258
4	azobenzene	−2.750	−4.283	67	2,3,4,6-tetrachlorophenol	−3.100	−3.491
4	nitrobenzene	−1.800	−1.865	67	2,3,5,6-tetrachlorophenol	−3.370	−3.387
4	o-nitrotoluene	−2.330	−2.422	4	pentachlorophenol	−4.280	−3.475
4	m-nitrotoluene	−2.440	−2.565	4	o-methoxyphenol	−1.960	−1.211
4	p-nitrotoluene	−2.490	−2.426	4	p-hydroxybenzaldehyde	−0.960	−0.671
4	o-chloronitrobenzene	−2.550	−2.614	4	o-aminophenol	−0.720	−0.471
4	m-chloronitrobenzene	−2.770	−2.587	4	p-aminophenol	−0.800	−0.129
4	p-chloronitrobenzene	−2.920	−2.748	4	o-nitrophenol	−1.740	−1.862
4	o-nitroanisole	−1.960	−2.157	4	m-nitrophenol	−1.010	−1.035
4	p-nitroanisole	−2.410	−2.113	4	p-nitrophenol	−0.740	−0.853
4	1,2-dinitrobenzene	−3.100	−1.981	4	salicylic acid	−1.820	−1.479
4	1,3-dinitrobenzene	−2.290	−1.657	4	p-hydroxybenzoic acid	−1.410	−1.074
4	1,4-dinitrobenzene	−3.390	−1.656	4	1,2-benzenediol	0.620	−0.424
4	2,4-dinitrotoluene	−2.820	−2.219	4	1,3-benzenediol	0.810	−0.274
4	2,6-dinitrotoluene	−3.000	−2.303	4	1,4-benzenediol	−0.170	−0.322
4	2,4,6-trinitrotoluene	−3.220	−2.116	24	methylparaben	−1.827	−1.481
4	1,3,5-trinitrobenzene	−2.890	−1.554	4	ethyl-p-hydroxybenzoate	−2.350	−2.018
4	1-nitronaphthalene	−3.540	−3.718	4	o-hydroxybenzamide	−1.820	−1.109
4	2,3-dichloronitrobenzene	−3.480	−3.374	4	p-hydroxyacetanilide	−1.030	−1.022
4	3,4-dichloronitrobenzene	−3.200	−3.436	4	1-naphthol	−2.220	−2.638
4	benzamide	−0.960	−0.401	4	2-naphthol	−2.280	−2.540
4	acetanilide	−1.330	−0.893	4	phenylmethanol	−0.400	−0.785

ref	compound	log S_W obs	log S_W calc (eq 11)	ref	compound	log S_W obs	log S_W calc (eq 11)
4	1-phenylethanol	−0.920	−1.062	75	5-ethyl-5-isopropylbarbituric acid	−2.148	−1.492
4	2-phenoxyethanol	−0.700	−0.882	76	butabarbital	−2.390	−1.974
24	ephedrine	−0.470	−0.781	76	pentobarbital	−2.390	−2.536
4	thiophenol	−2.120	−2.553	75	5-ethyl-5-(3-methylbutyl)barbital	−2.658	−2.536
4	phenylthiourea	−1.770	−1.023	75	5,5-diisopropylbarbital	−2.766	−1.997
4	p-toluenesulfonamide	−1.740	−1.057	75	5-allyl-5-methylbarbital	−1.160	−0.679
4	furane	−0.820	−1.031	75	5-allyl-5-ethylbarbital	−1.614	−1.241
4	furfural	−0.100	−0.148	75	5-allyl-5-isopropylbarbital	−1.708	−1.696
4	dibenzofurane	−4.600	−4.469	76	secobarbital	−2.356	−2.821
4	pyridine	0.760	0.043	75	5,5-diallylbarbital	−2.077	−1.733
4	2,3-dimethylpyridine	0.380	−0.737	75	5-(3-methyl-2-butenyl)-5-ethylbarbital	−2.253	−2.335
4	2,4-dimethylpyridine	0.380	−0.679	75	5-(3-methyl-2-butenyl)-5-isoprbarbital	−2.593	−2.868
4	2,6-dimethylpyridine	0.450	−0.698	75	5-ethyl-5-phenylbarbital	−2.322	−2.875
4	3,4-dimethylpyridine	0.360	−0.694	75	5-allyl-5-phenylbarbital	−2.369	−3.283
4	3,5-dimethylpyridine	0.380	−0.808	75	cyclobutyl-5-spirobarbituric acid	−1.655	−0.123
4	2-ethyl pyridine	0.510	−0.866	75	cyclopentyl-5-spirobarbituric acid	−2.349	−0.726
4	4-ethyl pyridine	0.830	−0.902	75	cyclohexyl-5-spirobarbituric acid	−3.060	−1.330
73	cocaine	−2.253	−2.167	75	cycloheptyl-5-spirobarbituric acid	−3.168	−1.937
74	atropine	−2.119	−2.532	75	cyclooctyl-5-spirobarbituric acid	−2.982	−2.562
63	quinoline	−1.300	−1.883	77	amitrole	0.522	0.864
63	isoquinoline	−1.450	−1.802	77	carbaryl	−3.224	−2.744
4	carbazole	−5.270	−3.871	77	carbofuran	−2.800	−1.774
4	antipyrene	0.720	0.440	13	chlorfenac	−3.078	−3.097
67	morpholine	1.965	1.615	77	coumatetralyl	−2.837	−4.767
4	theophylline	−1.390	−0.212	77	2,4-DB	−3.734	−3.602
67	caffeine	−0.876	0.446	4	DDT	−7.150	−7.457
4	morphine	−3.280	−2.382	77	desmedipham	−4.632	−5.103
4	codeine	−1.520	−2.278	77	dichlorophen	−3.953	−4.582
70	thiophene	−1.330	−1.653	77	dichlorprop	−2.827	−3.350
24	imipramine	−4.190	−4.153	4	diuron	−3.049	−2.975
4	progesterone	−4.420	−4.022	77	DNOC	−1.456	−2.397
4	testosterone	−4.020	−4.075	77	etofenprox	−8.600	−7.836
72	deoxycorticosterone	−3.450	−3.979	77	fenoxycarb	−4.700	−4.816
13	hydroxyprogesterone-17a	−3.817	−4.240	4	fenuron	−1.600	−1.637
72	corticosterone	−3.240	−4.043	4	fluometuron	−3.430	−2.111
4	cortisone	−3.110	−3.293	77	isoprocarb	−2.863	−2.213
4	hydrocortisone	−3.090	−4.929	11	isoproturon	−3.536	−3.000
13	17a-methyltestosterone	−3.999	−4.336	77	linuron	−3.592	−3.387
72	prednisolone	−3.180	−5.202	77	methyldymron	−3.350	−3.739
24	hydrocortisone 21-acetate	−4.880	−4.690	77	metolcarb	−1.803	−1.571
74	estrone	−3.955	−4.068	11	metoxuron	−2.564	−2.175
24	estradiol	−5.030	−4.389	4	monolinuron	−2.570	−2.358
72	dexamethasone	−3.590	−5.347	4	propoxur	−2.050	−1.617
75	5,5-dimethylbarbituric acid	−1.742	0.204	77	warfarin	−4.259	−4.074
75	5-methyl-5-ethylbarbituric acid	−1.228	−0.357	77	XMC	−2.581	−2.084
76	barbital	−2.400	−0.960				

Outliers

ref	compound	log S_W obs	log S_W calc (eq 11)	ref	compound	log S_W obs	log S_W calc (eq 11)
75	cyclopropyl-5-spirobarbituric acid	−1.886	0.358	74	fentanyl	−1.129	−4.224
4	uracil	−1.490	1.191	74	adenine	−2.432	0.185
74	chlorpheniramine	−0.235	−3.056				

References and Notes

1. Irmann, F. Eine einfache korrelation zwischen wasserloslichkeit und struktur von kohlenwasserstoffen und halogenkohlenwasserstoffen. *Chem. Ing. Tech.* **1965**, *37*, 789–798.

2. Wakita, K.; Yoshimoto, M.; Miyamoto, S.; Watanabe, H. A method for calculation of aqueous solubility of organic compounds using new fragment solubility constants. *Chem. Pharm. Bull.* **1986**, *34*, 4663–4681.

3. Klopman, G.; Wang, S.; Balthasar, D. M. Estimation of aqueous solubility of organic molecules by the group contribution approach. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 474–482.

4. Kuhne, R.; Ebert, R.-U.; Kleint, F.; Schmidt, G.; Schuurmann, G. Group contribution methods to estimate water solubility of organic compounds. *Chemosphere* **1995**, *30*, 2061–2077.

5. Nirmalakhanden, N. N.; Speece, R. E. Prediction of aqueous solubility of organic compounds based on molecular structure. 2. Application to PNAs, PCBs, PCDDs, etc. *Environ. Sci. Technol.* **1989**, *23*, 708–713.

6. Banerjee, S. Calculation of water solubility of organic compounds with UNIFAC-derived parameters. *Environ. Sci. Technol.* **1985**, *19*, 369–370.

7. Kan, A. T.; Tomson, M. B. UNIFAC prediction of aqueous solubilities of chemicals of environmental interest. *Environ. Sci. Technol.* **1996**, *30*, 1369–1376.

8. Myrdal, P. B.; Ward, G. H.; Dannenfelser, R.-M.; Mishra, D.; Yalkowsky, S. H. Aquafac 1: aqueous functional group activity coefficients; application to hydrocarbons. *Chemosphere* **1992**, *24*, 1047–1061.

9. Myrdal, P. B.; Ward, G. H.; Simamora, P.; Yalkowsky, S. H. Aquafac: aqueous functional group activity coefficients. *SAR QSAR Environ. Res.* **1993**, *1*, 53–61.

10. Myrdal, P. B.; Manka, A. M.; Yalkowsky, S. H. Aquafac 3: Aqueous functional group activity coefficients; application to the estimation of aqueous solubility. *Chemosphere* **1995**, *30*, 1619–1637.

11. Lee, Y.-C.; Myrdal, P. B.; Yalkowsky, S. M. Aqueous functional group activity coefficients (Aquafac) 4: applications to complex organic compounds. *Chemosphere* **1996**, *33*, 2129–2144.

12. Lee, Y.-C.; Pinsuwan, S.; Yalkowsky, S. M. A comparison of AQUAFAC group q-values to the corresponding CLOGP f-values. *Chemosphere* **1997**, *35*, 775−782.
13. Pinsuwan, S.; Myrdal, P. B.; Lee, Y.-C.; Yalkowsky, S. H. AQUAFAC 5: aqueous functional group activity coefficients; application to alcohols and acids. *Chemosphere* **1997**, *35*, 2503−2513.
14. Bodor, N.; Harget, A.; Huang, M.-J. Neural network studies. 1. Estimation of the aqueous solubility of organic compounds. *J. Am. Chem. Soc.* **1991**, *113*, 9480−9483.
15. Bodor, N.; Huang, M.-J. A new method for the estimation of aqueous solubility of organic compounds. *J. Pharm. Sci.* **1992**, *81*, 954−960.
16. Nelson, T. M.; Jurs, P. C. Prediction of aqueous solubility of organic compounds. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 601−609.
17. Sutter, J. M.; Jurs, P. C. Prediction of aqueous solubility for a diverse set of hereroatom- containing compounds using a quantitative structure−property relationship. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 100−107.
18. Katritzky, A. R.; Wang, Y.; Sild, S.; Tamm, T.; Karelson, M. QSPR studies on vapor pressure, aqueous solubility and the prediction of water−air partition coefficients. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 720−725.
19. Hansch, C.; Quinlan, J. E.; Lawrence, G. L. The linear free energy relationship between partition coefficients and the aqueous solubility of organic liquids. *J. Org. Chem.* **1968**, *33*, 347−350.
20. Yalkowsky, S. H.; Valvani, S. C. Solubility and partitioning. I: solubility of nonelectrolytes in water. *J. Pharm. Sci.* **1980**, *69*, 912−922.
21. Valvani, S. C.; Yalkowsky, S. H.; Roseman, T. J. Solubility and partitioning IV: solubility and octanol−water partition coefficients of liquid nonelectrolytes. *J. Pharm. Sci.* **1981**, *70*, 502−507.
22. Yalkowsky, S. H.; Valvani, S. C.; Roseman, T. J. Solubility and partitioning VI: octanol solubility and octanol−water partition coefficients. *J. Pharm. Sci.* **1983**, *72*, 866−870.
23. Ruelle, P.; Kesselring, U. W. The hydrophobic effect. 1. A consequence of the mobile order in H-bonded liquids. *J. Pharm. Sci.* **1998**, *87*, 987−997.
24. Ruelle, P.; Kesselring, U. W. The hydrophobic effect. 2. Relative importance of the hydrophobic effect on the solubility of hydrophobes and pharmaceuticals in H-bonded solvents. *J. Pharm. Sci.* **1998**, *87*, 998−1014.
25. Ruelle, P.; Kesselring, U. W. The hydrophobic effect. 3. A key ingredient in predicting n- octanol−water partition coefficients. *J. Pharm. Sci.* **1998**, *87*, 1015−1024.
26. Abraham, M. H. Scales of solute hydrogen-bonding: their construction and application to physicochemical and biochemical processes. *Chem. Soc. Rev.* **1993**, *22*, 73−83.
27. Abraham, M. H.; McGowan, J. C. The use of characteristic volumes to measure cavity effects in reversed phase liquid chromatography. *Chromatographia* **1987**, *23*, 243−246.
28. Abraham, M. H.; Chadha, H. S.; Whiting, G. S.; Mitchell, R. C. Hydrogen bonding. 32. An analysis of water-octanol and water-alkane partitioning and the ΔlogP parameter of Seiler. *J. Pharm. Sci.* **1994**, *83*, 1085−1100.
29. Abraham, M. H.; Andonian-Haftvan, J.; Whiting, G. S.; Leo, A.; Taft, R. W. Hydrogen bonding. Part 34: the factors that influence the solubility of gases and vapours in water at 298 K, and a new method for its determination. *J. Chem. Soc., Perkin Trans. 2* **1994**, 1777−1791.
30. Abraham, M. H.; Chadha, H. S.; Leo, A. J. Hydrogen bonding. 35. Relationship between high performance liquid chromatography capacity factors and water-octanol partition coefficients. *J. Chromatogr. A* **1994**, *685*, 203−211.
31. Abraham, M. H.; Roses, M. Hydrogen bonding. 38. Effect of solute structure and mobile phase on reversed phase high performance liquid chromatography capacity factors. *J. Phys. Org. Chem.* **1994**, *7*, 672−684.
32. Abraham, M. H.; Chadha, H. S.; Dixon, J. P.; Leo, A. J. Hydrogen bonding. 39. The partition of solutes between water and various alcohols. *J. Phys. Org. Chem.* **1994**, *7*, 712−716.
33. Abraham, M. H.; Chadha, H. S.; Dixon, J. P.; Rafols, C.; Treiner, C. Hydrogen bonding. Part 40. Factors that influence the distribution of solutes between water and sodium dodecyl sulfate micelles. *J. Chem. Soc., Perkin Trans. 2* **1995**, 887−894.
34. Abraham, M. H.; Treiner, C.; Roses, M.; Rafols, C.; Ishihama, Y. Linear free energy relationship analysis of microemulsion electrokinetic chromatographic determination of lipophilicity. *J. Chromatogr. A* **1996**, *752*, 243−249.
35. Siebert, D. S.; Poole, C. F.; Abraham, M. H. Retention properties of a spacer-bonded propanediol sorbent for reversed-phase liquid chromatography and solid-phase extraction. *Analyst* **1996**, *121*, 511−520.
36. Tan, L. C.; Carr, P. W.; Abraham, M. H. Study of retention in reversed phase liquid chromatography using linear solvation energy relationships. I. The stationary phase. *J. Chromatogr. A* **1996**, *752*, 1−18.
37. Abraham, M. H.; Poole, C. F.; Poole, S. K. Solute effects on reversed phase thin-layer chromatography, *J. Chromatogr. A* **1996**, *749*, 201−209.
38. Nasal, A.; Haber, P.; Kaliszan, R.; Forgacs, E.; Cserhati, T.; Abraham, M. H. Polyethylene coated silica and zirconia stationary phases in view of quantitative structure-retention relationships. *Chromatographia* **1996**, *43*, 484−490.
39. Abraham, M. H.; Chadha, H. S.; Dixon, J. P.; Rafols, C.; Treiner, C. Hydrogen bonding. 41. Factors that influence the distribution of solutes between water and hexadecylpyridinium chloride micelles. *J. Chem. Soc., Perkin Trans. 2* **1997**, 19−24.
40. Abraham, M. H.; Roses, M.; Poole, C. F.; Poole, S. K. Hydrogen bonding. 42. Characterization of reversed-phase high-performance liquid chromatographic C18 stationary phases. *J. Phys. Org. Chem.* **1997**, *10*, 358−368.
41. Abraham, M. H.; Chadha, H. S.; Leitao, R. A. E.; Mitchell, R. C.; Lambert, W. J.; Kaliszan, R.; Nasal, A.; Haber, P. Determination of solute lipophilicity as logP(octanol) and logP(alkane) using poly(styrene-divinylbenzene) and immobilised artificial membrane stationary phases in reversed-phase high-performance chromatography. *J. Chromatogr. A* **1997**, *766*, 35−47.
42. Pagliara, A.; Caron, G.; Lisa, G.; Fan, W.; Gaillard, P.; Carrupt, P.-A.; Testa, B.; Abraham, M. H. Solvatochromic analysis of di-*n*-butyl ether/water partition coefficients as compared to other systems. *J. Chem. Soc., Perkin Trans. 2* **1997**, 2639−2643.
43. Valko, K.; Plass, M.; Bevan, C.; Reynolds, D.; Abraham, M. H. Relationships between the chromatographic hydrophobicity indices and solute descriptors obtained by using several reversed phase diol, nitrile, cyclodextrin and immobilised artificial membrane-bonded high performance liquid chromatographic columns. *J. Chromatogr. A* **1998**, *797*, 41−55.
44. Plass, M.; Valko, K.; Abraham, M. H. Determination of descriptors of tripeptide derivatives based on high-throughput gradient high-performance liquid chromatographic retention data. *J. Chromatogr. A* **1998**, *803*, 51−60.
45. Du, C. M.; Valko, K.; Bevan, C.; Reynolds, D.; Abraham, M. H. Rapid gradient RP-HPLC method for lipophilicity determination: a solvation equation based comparison with isocratic methods. *Anal. Chem.* **1998**, *70*, 4228−4234.
46. Abraham, M. H.; Martins, F.; Mitchell, R. C.; Salter, C. J. Hydrogen bonding. 47. Characterization of the ethylene glycol−heptane partition system; hydrogen bond acidity and basicity of peptides. *J. Pharm. Sci.* **1999**, *88*, 241−247.
47. Abraham, M. H.; Platts, J. A.; Hersey, A.; Leo, A.; Taft, R. W. The correlation and estimation of gas-chloroform and water-chloroform partition coefficients by an LFER method. submitted for publication *J. Pharm. Sci.*, in press.
48. Seibert, D. S.; Poole, C. F. Retention properties of a cyanopropylsiloxane-bonded silica based sorbent for solid-phase extraction. *J. High Resol. Chromatogr.* **1995**, *18*, 226−230.
49. Poole, C. F.; Poole, S. K.; Seibert, D.; Chapman, C. M. Determination of kinetic and retention properties of cartridge and disk devices for solid-phase extraction. *J. Chromatogr. B* **1997**, *689*, 245−259.
50. Boillet, D.; Poole, C. F. Influence of solvent effects on retention for a porous polymer sorbent in reversed phase liquid chromatography. *Chromatographia* **1997**, *46*, 381−398.
51. Poole, C. F.; Poole, S. K. Interphase model for retention and selectivity in micellar electrokinetic chromatography. *J. Chromatogr. A* **1997**, *792*, 89−104.
52. Poole, S. K.; Poole, C. F. Retention of neutral organic compounds from solution on carbon adsorbents. *Anal. Comm.* **1997**, *34*, 247−251.
53. Boillet, D.; Poole, C. F. Influence of temperature on retention and selectivity in reversed phase liquid chromatography. *Analyst* **1998**, *123*, 295−299.
54. Kiridena, W.; Poole, C. F. Structure-driven retention model for solvent selection and optimization in reversed-phase thin-layer chromatography. *J. Chromatogr. A* **1998**, *802*, 335−347.
55. Kiridena, W.; Poole, C. F. Influence of solute size and site-specific surface interactions on the prediction of retention in liquid chromatography using the solvation parameter model. *Analyst* **1998**, *123*, 1265−1270.
56. Poole, C. F. Relationship between liquid−liquid distribution and liquid-micelle distribution systems. *J. Chromatogr. A* **1998**, *807*, 307−310.
57. Bolliet, D.; Poole, C. F.; Roses, M. Conjoint prediction of the retention of neutral and ionic compounds (phenols) in reversed-phase liquid chromatography using the solvation

parameter model. *Anal. Chim. Acta* **1998**, *368*, 129−140. Li, J.; Carr, P. W. Characterization of polybutadiene-coated zirconia and comparison to conventional bonded phases by use of linear solvation energy relationships. *Anal. Chim. Acta* **1996**, *334*, 239−250.

58. Jackson, P. T.; Schure, M. R.; Weber, T. P.; Carr, P. W. Intermolecular interactions involved in solute retention on carbon media in reversed-phase high-performance liquid chromatography. *Anal. Chem.* **1997**, *69*, 416−425.

59. Tan, L. C.; Carr, P. W. Study of retention in reversed-phase liquid chromatography using linear solvation energy relationships. *J. Chromatogr. A* **1998**, *799*, 1−19.

60. Abraham, M. H.; Kumarsingh, R.; Cometto-Muniz, J. E.; Cain, W. S.; Roses, M.; Bosch, E.; Diaz, M. L. The determination of solvation descriptors for terpenes, and the prediction of nasal pungency thresholds. *J. Chem. Soc., Perkin Trans. 2* **1998**, 2405−2411.

61. Platts, J. A.; Abraham, M. H.; Butina, D.; Hersey, A. *J. Chem. Inf. Comput. Sci.*, in press.

62. Abraham, M. H. Thermodynamics of solution of homologous series of solutes in water. *J. Chem. Soc., Faraday Trans. 1* **1984**, *80*, 153−181.

63. Kamlet, M. J.; Doherty, R. M.; Abraham, M. H.; Carr, P. W.; Doherty, R. F.; Taft, R. W. Linear solvation energy relationships. 41. Important differences between aqueous solubility relationships for aliphatic and aromatic solutes. *J. Phys. Chem.* **1987**, *91*, 1996−2004.

64. Pierotti, R. A. A scaled particle theory of aqueous and nonaqueous solutions. *Chem. Rev.* **1976**, *76*, 717−726.

65. Grant, D. G. W.; Higuchi, T. *Solubility Behaviour of Organic Compounds*; John Wiley & Sons: New York, 1990; pp 486−488.

66. Strong, L. E.; Neff, R. M.; Whitesel, I. Thermodynamics of dissolving and solvation for benzoic acid and the toluic acids in aqueous solution. *J. Soln. Chem.* **1989**, *18*, 101−114.

67. Brunas, J. P. The aqueous solubility of drugs and other compounds. University College London. Third Year Project, 1995.

68. Calculated from the gas−water partition coefficient and the saturated vapor concentration, this work.

69. Ruelle, P.; Kesselring, W. The hydrophobic propensity of water toward amphiprotic solutes: prediction and molecular origin of the aqueous solubility of aliphatic alcohols. *J. Pharm. Sci.* **1997**, *86*, 179−186.

70. Schwarzenbach, R. P.; Gschwend, P. M.; Imboden, D. M. *Environmental Organic Chemistry*; John Wiley-Interscience: New York, 1993; pp 618−625.

71. Pinsuwan, S.; Li, A.; Yalkowsky, S. Correlation of octanol/water solubility ratios and partition coefficients. *J. Chem. Eng. Data* **1995**, *40*, 623−626.

72. Huuskonen, J.; Salo, M.; Taskinen, J. Neural network modelling for estimation of the aqueous solubility of structurally related drugs. *J. Pharm. Sci.* **1997**, *86*, 450−454.

73. Siedell, A. *Solubilities of organic compounds: A compilation of quantitative solubility data from the periodical literature*, 3rd ed. (4th printing); D. Van Nostrand Company: 1941; Vol. 2, p 777.

74. *The Merck Index: An encyclopedia of chemicals, drugs, and biologicals*, 11th ed.; Merck & Company: Rahway, NJ, 1989.

75. Prankerd, R. J.; McKeown, R. H. Physico-chemical properties of barbituric acid derivatives: IV. Solubilities of 5,5-disubstituted barbituric acids in water. *Int. J. Pharm.* **1994**, *112*, 1−15.

76. Vaution, C.; Treiner, C.; Puisieux, F.; Cartensen, J. T. Solubility behaviour of barbituric acids in aqueous solution of sodium alkyl sulfonate as a function of concentration and temperature. *J. Pharm. Sci.* **1981**, *70*, 1238−1242.

77. *The Pesticide Manual*, 11th ed.; Tomlin, C. D. S., Ed.; British Crop Protection Council, 1997.

78. Ruelle, P.; Kesselring, W. Aqueous solubility prediction of environmentally important chemicals from the mobile order thermodynamics. *Chemosphere* **1997**, *34*, 275−298.

Acknowledgments