Pure & Appl. Chem., Vol. 65, No. 12, pp. 2503–2512, 1993.
Printed in Great Britain.

Application of solvation equations to chemical and biochemical processes

Michael H. Abraham
Department of Chemistry, University College London, 20 Gordon Street, London, WC1H OAJ, UK

Abstract - Two solvation equations that can be used either as LSERs or as QSARs have been applied to various processes that involve transfer of a series of solutes from the gas phase to a condensed phase or transfer of a series of solutes from one condensed phase to another. In the former class, the processes include gas-liquid chromatography, gas-solid chromatography, the solubility of gases and vapours in polymers and organic solvents, and upper respiratory tract irritation in mice. The latter class includes water-octanol and other partitions, the inhibition of firefly luciferase enzyme by aqueous nonelectrolytes, and general anaesthesia.

INTRODUCTION

There are a number of scales of hydrogen-bond strength currently available in the literature. The first such scales devised were those of Kamlet and Taft, who set out an α (or α_1) scale of solvent hydrogen-bond acidity [1] and a β (or β_1) scale of solvent hydrogen-bond basicity [2]. Since these scales were based on a solvatochromic comparison method, they are often known as Kamlet-Taft solvatochromic parameters. The α_1 and β_1 hydrogen-bond scales have been applied to all kinds of solvent effects, as summarised in a number of reviews [3-6].

Some years later, Abraham and co-workers [7,8] devised scales of solute hydrogen-bond acidity, α_2^H, and solute hydrogen-bond basicity, β_2^H, using equilibrium constants for 1:1 complex formation in tetrachloromethane:

$$\text{A-H} + \text{B} \overset{K}{\rightleftharpoons} \text{A-H}\cdots\text{B} \qquad (1)$$

The acid and base were present in dilute solution, so that the equilibrium constants refer to the simple monomeric solutes in equation 1. In the case of solute hydrogen-bond acidity [7], logK values were compiled for series of acids against 45 different reference bases, enabling 45 equations to be constructed of the form:

$$\log K \text{ (acids against reference base B)} = L_B.\log K_A^H + D_B \qquad (2)$$

where L_B and D_B characterise the reference base and $\log K_A^H$ now characterises the series of acids. All 45 equations were constrained to pass through a "magic point" at (-1.1, -1.1) which provides an automatic zero for the scale. A typical example of one of the 45 equations is that for acids against the reference base tetrahydrofuran (THF):

$$\log K \text{ (acids against THF)} = 0.8248\ \log K_A^H - 0.1970 \qquad (3)$$
$$n = 23 \quad \rho = 0.9960 \quad sd = 0.089 \quad F = 2609$$

Here, and elsewhere, n is the number of data points, ρ is the overall correlation coefficient, sd is the overall standard deviation and F is the Fisher F-statistic.

In a similar way [8], logK values for a series of bases against 34 reference acids led to 34 equations:

$$\log K \text{ (bases against reference acid A)} = L_A.\log K_B^H + D_A \qquad (4)$$

Now L_A and D_A characterise the reference acid, and $\log K_B^H$ characterises the series of bases. Again, all the equations were constrained to pass through the magic point (-1.1, -1.1). A typical example of the set of 34 equations is that for bases against the reference acid 4-chlorophenol:

$$\log K\text{(bases against 4-chlorophenol)} = 1.065\ \log K_B^H + 0.074 \qquad (5)$$
$$n = 38 \quad sd = 0.054$$

The scales of solute hydrogen-bonding, $\log K_A^H$ and $\log K_B^H$, have their origin at -1.1, but this can simply be moved to the more convenient origin of zero, and the scales compressed somewhat at the same time, through the defining equations [7,8]:

$$\alpha_2^H = (\log K_A^H + 1.1)/4.636 \quad (6)$$

$$\beta_2^H = (\log K_B^H + 1.1)/4.636 \quad (7)$$

The two scales can be combined to yield a simple equation for the correlation and estimation of logK values for equation 1 in tetrachloromethane at 298K [9]:

$$\log K = 7.354\ \alpha_2^H.\beta_2^H - 1.094 \quad (8)$$
$$n = 1312 \quad \rho = 0.9956 \quad sd = 0.09 \quad F = 148535$$

Raevsky and co-workers [10] have also used equilibrium constants in equation 1 to construct scales of solute hydrogen-bonding. Although their original equation suffered through lack of a constant term, this has now been rectified on their latest equation [11]:

$$\Delta G^O = 2.43 C_A.C_B + 5.70 \quad (9)$$

Here, ΔG^O is the standard Gibbs energy change for equation 1, in kJ mol^{-1}, C_A is the hydrogen-bond acidity of a solute, and C_B is the hydrogen-bond basicity of a solute. Equation 9 correlated 936 ΔG^O values for equation 1 with sd = 1.11 kJ mol^{-1}, equivalent to sd = 0.19 log units. Raevsky and co-workers [11] also provided a correlation equation and solute hydrogen-bond parameters in terms of the enthalpy change in equation 1:

$$\Delta H^O = 4.96\ E_A.E_B \quad (10)$$

where ΔH^O is in kJ mol^{-1} and E_A and E_B are the enthalpic solute hydrogen-bond parameters. Clearly equation 8 and equation 9 are similar, and would be expected to lead to similar hydrogen-bond scales.

However, the solvent scales of Kamlet and Taft are not the same as the solute scales of Abraham or of Raevsky, for two fundamental reasons. Firstly, the Kamlet-Taft scales refer to properties of bulk liquids, whereas the α_2^H and β_2^H scales refer to monomeric solutes in dilute solution in tetrachloromethane. Secondly, the Kamlet-Taft scales are mainly derived from spectroscopic measurements, and are not then related to any thermodynamic property, whereas α_2^H and β_2^H are rigorously Gibbs-energy related. Unfortunately, there is already confusion in the literature over solvent and solute scales, which have even been used interchangeably. It is already known [12] that β_1 and β_2^H are not well-related, even for nonassociated compounds.

We regard the solvent scales as quite separate and deal only with solute scales. There are available a reasonable number of solute α_2^H and β_2^H values, from the original work [8,9] and from subsequent work of Berthelot and Laurence and co-workers [13, 14] on solute hydrogen-bond basicity. Some values of α_2^H and β_2^H are given in Table 1. All the α_2^H and β_2^H values refer to 1:1 complexation between acids and bases and although they represent the most useful and extensive scales of solute hydrogen-bonding so far constructed, it is not obvious that the same scales can be used to express the hydrogen-bond strength of solutes when the latter are surrounded by an excess of solvent molecules. It is this situation that exists in processes such as the solubility of gases and vapours in liquids, or the partition of solutes between liquid phases. Hence α_2 and β_2 scales need to be devised that are appropriate to this situation.

In the event, it seems possible to use the original α_2^H and β_2^H scales to set up "effective" or "summation" scales of solute hydrogen-bond acidity or basicity [15,16]. The α_2^H and β_2^H descriptors are incorporated into linear solvation energy relationships, LSERs, and the summation descriptors, $\Sigma\alpha_2^H$ and $\Sigma\beta_2^H$ are back-calculated [16]. A comparison of the summation descriptor $\Sigma\beta_2^H$ and β_2^H is in Table 2 for a number of solutes, and a selection of $\Sigma\alpha_2^H$ and $\Sigma\beta_2^H$ values is in Table 3.

For rather simple solutes, there is quite good agreement between β_2^H and $\Sigma\beta_2^H$, showing that the former parameter, based on 1:1 equilibrium constants, can indeed be used to set up a $\Sigma\beta_2^H$ scale. With more complicated solutes, $\Sigma\beta_2^H$ is usually larger than β_2^H, and for activated aromatics such as phenol and aniline this is always the case. Of course, phenol and aniline when surrounded by an excess of solvent will act as though they have two basic sites - the functional group and the aromatic ring, so that an elevated $\Sigma\beta_2^H$ value is expected.

Table 1. Some values of the solute α_2^H and β_2^H descriptors[a]

Solute	α_2^H	β_2^H
n-Heptane	0.00	0.00
Hept-1-ene	0.00	0.07
Hept-1-yne	0.13	0.20
Dichloromethane	0.13	0.05
Trichloromethane	0.20	0.02
Tetrachloromethane	0.00	0.00
Diethyl ether	0.00	0.45
Propanone	0.04	0.50
Butanone	0.00	0.48
Ethyl acetate	0.00	0.45
Acetonitrile	0.09	0.44
Dimethylcyanamide	0.00	0.61[b]
Nitromethane	0.12	0.29[c]
N-Methylacetamide	0.38	0.71
Water	0.35	0.38
Ethanol	0.33	0.44
Propan-2-ol	0.32	0.47
t-Butyl alcohol	0.32	0.49
Trifluoroethanol	0.57	0.18
Hexafluoropropan-2-ol	0.77	0.03
Acetic acid	0.55	-
Benzene	0.00	0.15
Phenylethyne	0.12	-
Acetophenone	0.00	0.51
Nitrobenzene	0.00	0.34
Phenol	0.60	0.22
4-Flurophenol	0.63	0.21
Benzoic acid	0.59	-
n-Propylamine	-	0.70
Diethylamine	-	0.70
Triethylamine	0.00	0.67
Aniline	0.26	0.38
Pyridine	0.00	0.62
2,6-Di-t-butylpyridine	0.00	0.19
Pyrrole	0.41	0.34[d]
Indole	0.44	-
Diethylsulfide	0.00	0.28
Dimethylsulfoxide	0.00	0.78
Diphenylsulfoxide	0.00	0.67
Diphenylsulfone	0.00	0.51

[a] From refs. 7 and 8.
[b] Ref.14.
[c] From data by E.V.Titov, V.I. Shurpach, G.A.Belkina and N.P. Gonchar, *J. Mol. Structure*, 1990, **219**, 257.
[d] From data by M. Orban, A. Kiss and L. Barcza, *J.Chem.Soc.Perkin.Trans.*2, 1987, 1815.

Table 2. A comparison of β_2^H and $\Sigma\beta_2^H$

Solute	β_2^H	$\Sigma\beta_2^H$
n-Heptane	0.00	0.00
Hept-1-ene	0.07	0.07
Hept-1-yne	0.20	0.10
Diethyl ether	0.45	0.45
Butanone	0.48	0.51
Ethyl acetate	0.45	0.45
Acetonitrile	0.44	0.32
Nitromethane	0.29	0.31
Triethylamine	0.70	0.79
Water	0.38	0.35
Ethanol	0.44	0.48
Propan-2-ol	0.47	0.56
t-Butyl alcohol	0.49	0.60
Benzene	0.15	0.14
Acetophenone	0.51	0.48
Nitrobenzene	0.34	0.28
Pyridine	0.62	0.52
Diethylsulfide	0.28	0.32
Dimethylsulfoxide	0.78	0.88
N,N-Dimethylbenzenesulfonamide	0.53	0.86
Trimethylphosphate	0.76	1.00
Anisole	0.26	0.29
Phenol	0.22	0.30
p-Cresol	0.24	0.31
4-Fluorophenol	0.21	0.23
Benzyl alcohol	0.42	0.56
Aniline	0.38	0.41
p-Toluidine	0.42	0.45
4-Fluoroaniline	0.36	0.40
N,N-Dimethylaniline	0.35	0.41

Table 3. Some values of $\Sigma\alpha_2^H$ and $\Sigma\beta_2^H$

Solute	$\Sigma\alpha_2^H$	$\Sigma\beta_2^H$
Hept-1-yne	0.12	0.10
Dichloromethane	0.10	0.05
Trichloromethane	0.15	0.02
Acetamide	0.54	0.68
N,Methylacetamide	0.40	0.72
N,N-Dimethylacetamide	0.00	0.78
Acetic Acid	0.61	0.44
Trichloroacetic acid	0.95	0.28
Water	0.82	0.35
Ethanol	0.37	0.48
2,2,2-Trifluoroethanol	0.57	0.25
2-Methoxyethanol	0.30	0.84
Methyl benzoate	0.00	0.46
Dimethyl phthalate	0.00	0.88
Benzoic acid	0.59	0.40
Phenol	0.60	0.30
2-Chlorophenol	0.32	0.31
4-Chlorophenol	0.67	0.20
Benzenesulfonamide	0.55	0.80

Some of the $\Sigma\alpha_2^H$ and $\Sigma\beta_2^H$ values are interesting. For water, $\Sigma\alpha_2^H$ is more than double the α_2^H value of 0.35, and the $\Sigma\alpha_2^H$ and $\Sigma\beta_2^H$ values for the chlorophenols show clearly the effect of intramolecular hydrogen-bonding in the ortho isomer.

It should also be noted that for certain solutes, such as aniline, substituted anilines, pyridine and alkyl pyridines, and sulfoxides (but not sulfones) an alternative basicity descriptor, $\Sigma\beta_2^O$, is required in processes that involve transfer from water to rather aqueous solvent systems [15, 16].

The LSERs are themselves devised using a simple cavity theory of solvation, in which the process of dissolution of a gaseous solute in a solvent involves (i) the endoergic creation of a cavity in the solvent and (ii) incorporation of the solute into the cavity with consequent setting up of various exoergic solute-solvent interactions. The processes to be considered will all involve a series of solutes with a fixed solvent or solvents. Hence the properties of the solvent phase are constant, and the various interactions will be described by particular solute parameters. These can be set out as follows.

R_2 is an excess molar refraction that can be determined simply from a knowledge of the compound refractive index [17]. Since R_2 is almost an additive property, it is quite straightforward to deduce values for compounds that are gaseous or solid at room temperature. Several hundred R_2 values are at present available, and further values can be determined or estimated quite easily. The R_2 descriptor represents the tendency of a compound to interact with a solvent phase through π - or n - electron pairs.

π_2^H is the compound dipolarity/polarizability [18-20], it being not possible to devise descriptors for these properties separately. This descriptor can be obtained experimentally from gas-liquid chromatographic (GLC) data for solutes that are not too involatile, and from water-solvent partition coefficients for solutes in general. At present, several hundred values of π_2^H are known, and it is reasonably easy to obtain further values.

$\Sigma\alpha_2^H$ is the solute effective or summation hydrogen-bond acidity. For mono-acids, this descriptor was originally obtained directly from hydrogen - bond complexation constants, and in this way values were found for many types of solute such as carboxylic acids, alcohols, and phenols [7]. Now that the acid scale is established, further values can be obtained by chromatographic or partition measurements. In addition, all values for poly-acids such as glycols must be found by these methods.

$\Sigma\beta_2^H$ is the solute effective or summation hydrogen-bond basicity. Again, for mono-bases, this was first obtained from hydrogen-bond complexation constants [8]. Further values for mono-bases, and all values for poly-cases can be found from partition coefficients [16]. Several hundred $\Sigma\beta_2^H$ values have now been obtained [15, 16].

$\log L^{16}$ is a descriptor [21] based on the solute gas-liquid partition coefficient on hexadecane at 298K. A data base of several hundred such values is available [15, 17-19, 21], and additional values can easily be obtained by gas chromatography on a variety of nonpolar stationary phases. The $\log L^{16}$ descriptor is a measure of the lipophilicity of a solute.

V_x is the McGowan characteristic volume that can trivially be calculated for any solute simply from a knowledge of its molecular structure [22]. Calculation is aided by the algorithm for the number of bonds in a molecule, counting all bonds as equal, ie a double or a triple bond counts as one bond only:

$$B = N - 1 + R \tag{11}$$

Here B is the total number of bonds, N is the total number of atoms in the molecule, and R is the number of rings in the molecule [15].

These descriptors can be combined into two equations [15], either as linear solvation energy relationships LSERs or as quantitative structure-activity relationships QSARs;

$$\log SP = c + r.R_2 + s.\pi_2^H + a.\alpha_2^H + b.\beta_2^H + l.\log L^{16} \tag{12}$$

$$\log SP = c + r.R_2 + s.\pi_2^H + a.\alpha_2^H + b.\beta_2^H + v.V_x \tag{13}$$

Here, SP is a property for a series of solutes on a given phase. Thus, SP can be the gas-liquid partition coefficient for a number of solutes on a particular organic solvent, ie an LSER, or SP can be a biological property for a series of solutes, ie a QSAR. Equation (12,13) can be solved by the method of multiple linear regression analysis, MLRA, to yield the constants c, r, s, a, b and l (or v). Not every term in equation (12,13) may be significant, and each term is analyzed using students t-test. Usually, terms are retained only if the t-test shows >95% significance. A number of precautions are taken when using MLRA, in particular (i) the number of data points should not be less than five times the number of descriptors, and (ii) the descriptors must not be collinear. The constants obtained by MLRA are important in that they can be used to characterise the solvent phase (in LSERs) or receptor area (in QSARs) involved. In both cases, the r-constant gives the propensity of the phase to interact with solute π - and n - electron pairs, the s-constant is the phase-area dipolarity/polarizability, the a-constant is the phase-area basicity (because a basic phase will interact with acid solutes), similarly the b-constant is the phase-area acidity, and the l-constant or the v-constant is a measure of the phase-area lipophilicity: by definition l = 1.00 for hexadecane at 298K.

Because the constants in equation (12, 13) represent quite specific properties of the phase or receptor area, they must follow correct chemical principles. Thus for a completely nonacidic phase, the b-constant must be zero, within some reasonable experimental error. Thus equation (12, 13) are not simply some statistical fitting procedures, but are substantive equations expressing not only the effect of solutes on some particular process, but also the properties of the solvent phase or receptor area involved. LSERs or QSARs derived from equation (12, 13) have to be examined with regard to goodness-of-fit, as is the case for any LSER or QSAR, but also with regard to general chemical principles. This latter test is highly unusual in QSAR work, but is very important in that strict application of the test leads to QSARs that are chemically firmly based, and are not just fitting equations to a given data set.

APPLICATIONS OF EQUATION (12)

This equation is best applied to gas $\rightarrow$ condensed phase processes. One such process, for which considerable data exist, is that of gas-liquid chromatography (GLC). The usual physico-chemical quantity measured is V_G, the specific retention volume of a solute at the column temperature. This is related to the gas-liquid partition coefficient, K, or the Ostwald solubility coefficient, L, through equation (14), where ρ_1 is the density of the GLC stationary phase at the column temperature:

$$L \text{ or } K = V_G.\rho_1 \tag{14}$$

The definition of L or K is given by equation (15):

$$L \text{ or } K = \frac{\text{concentration of solute in solution}}{\text{concentration of solute in gas phase}} \tag{15}$$

A number of GLC stationary phases were examined by Poole and co-workers[23] who obtained logK values for a series of solutes at 394K, considerable care being taken to correct for any interfacial adsorption. Typical equations[24] for logK values on the stationary phases are those for Carbowax 20M, equation (16), and the liquid salt tetraethylammonium 4-toluene-sulfonate, equation (17):

$$\log K(\text{Carbowax 20M}) = -0.56 + 0.29R_2 + 1.29\pi_2^H + 1.80\Sigma\alpha_2^H + 0.450\ \log L^{16} \tag{16}$$
$$n = 39 \quad r = 0.9957 \quad sd = 0.059 \quad F = 982$$

$$\log K(Et_4NX) = -1.01 + 0.36R_2 + 2.06\pi_2^H + 3.61\Sigma\alpha_2^H + 0.34\ \log L^{16} \tag{17}$$
$$n = 29 \quad r = 0.9941 \quad sd = 0.076 \quad F = 81$$

In these equations, and elsewhere, n is the number of data points, r is the overall correlation coefficient, sd is the standard deviation, and F is the F-statistic. The equations show that both stationary phases are dipolar (s = 1.29 and 2.06) and are hydrogen-bond bases (a = 1.80 and 3.61), but have no hydrogen-bond basicity (the $b.\Sigma\beta_2^H$ term is not significant). Equations such as (16) and (17) have now been constructed for numerous GLC stationary phases at 393K, and provide a new method for the characterisation of such phases [24, see also 25].

Most of the GLC stationary phases normally used, have zero or rather low hydrogen-bond acidity [24, 25], as shown by the lack of the $b.\Sigma\beta_2^H$ term in equation (12). Recently, a number of acidic phases were synthesised [26], one of which seemed to have potential as a GLC stationary phase. This phase, bis(3-allyl-4-hydroxyphenyl)sulfone, known as H10, was tested at 394K and shown to have considerable hydrogen-bond acidity [27]:

$$\log K(H10) = -0.57 - 0.05R_2 + 1.32\pi_2^H + 1.27\Sigma\alpha_2^H + 1.46\Sigma\beta_2^H + 0.42\ \log L^{16} \tag{18}$$
$$n = 58 \quad r = 0.9940 \quad sd = 0.069 \quad F = 856$$

The phase, however, is not very selective because it also has pronounced hydrogen-bond basicity, with an a-constant of 1.27, compare a = 1.80 for the polyether, Carbowax 20M. Carr and co-workers [28] have also synthesised a phase that has a very high hydrogen-bond acidity with almost zero hydrogen-bond basicity, the fluoroalcohol 4-dodecyl-α, α-bis(trifluoromethyl)benzyl alcohol, BOH. A regression of relative partition coefficients according to equation (12) yields only a poor correlation at 353K, however [27]:

$$\log K'(BOH) = -1.60 - 0.22R_2 + 0.45\pi_2^H + 2.69\Sigma\beta_2^H + 0.68\ \log L^{16} \quad (19)$$
$$n = 143 \quad r = 0.9598 \quad sd = 0.265 \quad F = 403$$

It is not possible to compare the constants in equations (18) and (19) because they refer to different temperatures. In general, the constants s, a, b and l all decrease with increase in temperature. Even so, it is likely that the fluoralcohol BOH is more acidic than the phenol H10.

Analysis of the retention data obtained on H10 was also of interest in that it confirmed the previous finding [29] that the solvation theories of Abraham [15] and of Poole [30] are essentially equivalent.

The characterisation of phases is not restricted to temperatures usually encountered in GLC work, and has been applied to phases at 298K as well [26, 31]. Many of these were candidate coatings for chemical sensors, and hence had to be characterised at ambient temperature. Equation (12) can be used to analyze solute-solvent, or solute-phase, interactions term-by-term, and hence provides a logical rationale for the selection of coatings for chemical sensors and arrays [32].

The application of equation (12) is, of course, not restricted to materials used as GLC phases or coatings for chemical sensors. A very important application is in the study of compound-polymer interactions, using the polymer as a stationary phase in a GLC experiment; this technique is sometimes known as inverse GLC, because the properties of the stationary phase are under investigation. Thus logK values for 43 compounds on atatic polypropene at 273K, obtained by Munk and co-workers [33] yielded the regression equation [34]:

$$\log K = -0.28 + 0.16R_2 + 0.08\pi_2^H + 0.64\ \log L^{16} \quad (20)$$
$$n = 43 \quad r = 0.9994 \quad sd = 0.022 \quad F = 10824$$

The polymer is only slightly dipolar/polarisable and has no basicity or acidity at all. Not only can polymers and compound-polymer interactions be investigated using equation (12), but values of logK or of $\log V_G$ can be predicted for other compounds as well. Since important parameters such as the weight-fraction activity coefficient, Ω^∞, and the Flory-Huggins interaction coefficient, χ, can be obtained from V_G [35-37], this amounts to an indirect prediction of Ω^∞ and χ. Thus for poly(butadiene) at 363K $\log V_G$ is given [38] by:

$$\log V_G = -0.10 + 0.30R_2 + 0.33\pi_2^H + 0.39\Sigma\alpha_2^H + 0.61\ \log L^{16} \quad (21)$$
$$n = 24 \quad r = 0.9981 \quad sd = 0.041 \quad F = 1246$$

This enables $\log \Omega^\infty$ to be predicted for other compounds to around 0.04 units, and χ to be predicted to within 0.10 units [38].

In a similar way, the commercially important soybean oil can be characterised through equation (12), and values of $\log \Omega^\infty$ and χ predicted via a prediction of V_G. For soybean oil at 396K, the regression equation (22) was obtained [39]:

$$\log V_G = -0.43 + 0.58\pi_2^H + 0.90\Sigma\alpha_2^H + 0.61\ \log L^{16} \quad (22)$$
$$n = 21 \quad r = 0.9910 \quad sd = 0.060 \quad F = 311$$

Since V_G and K are connected by equation (14), only the constant c in equation (12) changes if $\log V_G$ is used instead of logK. Hence the characteristic constants in equation (12) can be compared with those obtained previously at the same temperature for other phases [24, 25]. As might be expected, the values of s, a, and l in equation (22) are quite close to those for esters such as di-2-ethylhexyladipate[39].

Not only can the technique of gas-chromatography, when combined with the solvation equation (12), be used to characterise liquids, and polymers above their glass transition temperature, but it can also be used to characterise solid phases as well. Pankow[40] attempted to relate $\log V_G$ values for solutes on the graphitised material Carbotrap, to their vapour pressure, but a much better relationship was obtained [41] using equation (12):

$$\log V_G = -4.73 - 2.27R_2 + 2.65\ \log L^{16} \quad (23)$$
$$n = 38 \quad r = 0.9737 \quad sd = 0.880 \quad F = 318$$

Although the V_G values listed by Pankow [40] refer to 293K, many were extrapolated from higher

temperatures, so that equation (23) is as good as could be expected. In equation (23) the terms in solute dipolarity/polarisability and solute hydrogen-bond properties are not significant, so that Carbotrap appears to be free from acidic or basic sites.

Simple organic solvents can also be investigated using equation (12). For nonvolatile solvents, the required K (orL) values can be obtained by the gas chromatographic method in which the solvent is the stationary phase. More generally the gas-liquid partition coefficients are obtained by other methods, such as headspace analysis or vapour-liquid equilibrium measurements. In the latter case, some extrapolation to infinite dilution has usually to be made. A number of solvents has been investigated at 298K using equation (12); as examples, methylene iodide and chloroform can be taken [42]:

$$\log K(CH_2I_2) = -0.84 + 0.32R_2 + 1.34\pi_2^H + 0.83\,\Sigma\alpha_2^H + 1.19\Sigma\beta_2^H + 0.87\,\log L^{16} \quad (24)$$
$$n = 37 \quad r = 0.9979 \quad sd = 0.089 \quad F = 1461$$

$$\log K(CHCl_3) = 0.10 - 0.35R_2 + 1.26\pi_2^H + 0.60\Sigma\alpha_2^H + 1.18\Sigma\beta_2^H + 0.99\log L^{16} \quad (25)$$
$$n = 35 \quad r = 0.9969 \quad sd = 0.153 \quad F = 754$$

Both of these solvents exhibit some dipolarity/polarizability, are weak bases, but are somewhat stronger hydrogen-bond acids. Other solvents investigated in this way [15] are N-formylmorpholine and tri(2-ethylhexylphosphate), both strong hydrogen-bond bases with a-constants 4.32 and 3.74 respectively.

Biological processes can also be examined; in these cases, equation (12) represents a quantitative structure-activity relationship (QSAR). A recalculation of an earlier equation [43] for the effect of airborne chemicals on the upper respiratory irritation in mice yields [15]:

$$-\log FRD_{50} = 0.96 + 0.81\pi_2^H + 2.55\Sigma\alpha_2^H + 0.72\,\log L^{16} \quad (26)$$
$$n = 39 \quad r = 0.9870 \quad sd = 0.12 \quad F = 440$$

Here FRD_{50} is the molar concentration in the gas phase that leads to a 50% decrease in respiratory rate. The use of equation (26) is not only that it can lead to predictions of FRD_{50}, but also that it provides some insight as to the receptor site or receptor area involved. This can be seen by examination of the characteristic constants in equations for gas-solvent partitions at 298K, see Table 4.

Table 4. Comparison of constants in equation (12) for various gas solvent partitions with those for upper respiratory tract irritation, as $-\log FRD_{50}$

	r	s	a	b	l
$-\log FRD_{50}$	-	0.81	2.55	-	0.72
Gas Water	0.82	2.74	3.90	4.81	-0.21
Gas $(CH_2Cl)_2$	-0.28	1.72	0.73	0.59	0.93
Gas 2-Ethylhexylphosphate	-0.26	0.91	3.74	-	0.96
Gas N-Formylmorpholine	-	2.57	4.32	-	0.73

It is quite clear that the receptor site/area cannot possibly resemble water, or the rather nonpolar solvent 1,2-dichloroethane. The amide N-formylmorpholine is rather too dipolar (s = 2.57) to be a suitable model, but 2-ethylhexylphosphate has about the same dipolarity (s = 0.91) and hydrogen-bond basicity (a = 3.74) as the receptor site/area, although the latter is still larger than the receptor site/area value (a = 2.55). It can be concluded that the receptor site/area is non acidic (b = 0.00), somewhat dipolar (s = 0.81) and quite basic (a = 2.55), although not quite as basic as the phosphate ester. Such an analysis illustrates one of the virtues of equation (12) as a QSAR, namely that it provides chemical information on the biological process, as well as being useful in a predictive way.

APPLICATIONS OF EQUATION (13)

The same solute descriptors are used in equation (13) as in equation (12), except that the characteristic volume, V_x, is used instead of $\log L^{16}$; the units of V_x in the following equations are $(cm^3\ mol^{-1})/100$. Equation (13) works best for processes within condensed phases, for example the partition of solutes between two liquid phases. For the very important water-octanol partition, equation (13) yields [15]:

$$\log P(oct) = 0.08 + 0.58R_2 - 1.09\pi_2^H + 0.03\Sigma\alpha_2^H - 3.40\Sigma\beta_2^H + 3.81V_x \quad (27)$$

where P is given by (concentration of solute in octanol)/(concentration of solute in water). Such an equation

can be used to deduce the relative solute-solvent effects, e.g., solute-octanol less solute-water interactions. In addition it gives information on the relative properties of water and octanol, or more correctly on octanol-saturated water and water-saturated octanol.

The latter may be important because although the solubility of octanol in water is only 0.0045 mol dm^{-3}, that of water in octanol is 2.33 mol dm^{-3} at 298K [44]. On equation (27), solute dipolarity/polarisability and especially solute hydrogen-bond basicity favour water, and solute size favours octanol. Conversely, it can be deduced that water is more dipolar and is a stronger hydrogen-bond acid than is octanol, but octanol is more lipophilic than is water. The almost zero a-constant in equation(27) implies that water and (wet) octanol have the same hydrogen-bond basicity. Numerous examples of the application of equation (13) to water-solvent partitions have been given [16]; conversely, equations such as equation (27) are very useful for the determination of $\Sigma\beta_2^H$ values [16].

There are numerous biological and toxicological processes in which aqueous solutes interact with a given system. One example is the work of Franks and Lieb [45, 46] on the inhibition of firefly luciferase activity by aqueous nonelectrolytes. A re-analysis of the data of Franks and Lieb yields equation (28), very similar to that recorded before [47]:

$$\log 1/EC_{50} = 0.58 + 0.80R_2 - 3.53\Sigma\beta_2^H + 3.78V_x \quad (28)$$
$$n = 43 \quad r = 0.9881 \quad sd = 0.340 \quad F = 535$$

As set out earlier [47] the major factors that determine the effect of solutes on the inhibition of firefly luciferase activity, are the solute hydrogen-bond basicity that decreases the solute effect, and the solute size that increases the effect. Furthermore, it can be deduced [47] that the target site on the enzyme has about the same dipolarity as water (since the s-constant is zero), the same hydrogen-bond basicity as water (since the a-constant is zero), but is much less acidic than water (b = -3.53) and is much more lipophilic than water (v = +3.78).

A very similar analysis can be carried out for general anaesthesia of tadpoles by aqueous solutes. The effective concentration, EC_{50}, is again in units of mol dm^{-3} and application of equation (13) leads to:

$$\log 1/EC_{50} = 0.69 + 0.66R_2 - 4.44\Sigma\beta_2^H + 4.16V_x \quad (29)$$
$$n = 28 \quad r = 0.9900 \quad sd = 0.224 \quad F = 392$$

Equation (29) is similar to the one described before [47] but contains no term in π_2^H. Inclusion of this descriptor gives:

$$\log 1/EC_{50} = 0.86 + 0.81R_2 - 0.54\pi_2^H - 4.17\Sigma\beta_2^H + 4.08V_x \quad (30)$$
$$n = 28 \quad r = 0.9915 \quad sd = 0.210 \quad F = 334$$

Judging by the F-statistic, equation (29) is possibly just preferable to equation (30), but there is not much to choose between them. In either case, the two main factors influencing general anaesthesia are solute hydrogen-bond basicity which decreases the effect, and solute size, which increases the effect, exactly as observed in the luciferase work. Since equations (28) and (29) are similar, the analysis of the anaesthetic target site or sites follows that of the target site on the luciferase enzyme.

As with equation (12), the use of equation (13) as a QSAR yields information about the target site/area in the process under consideration. How the two main factors in the luciferase inhibition and the tadpole ones compare with other processes can be seen from results in Table 5.

Table 5. Comparison of constants in equation (13) for various processes.

	r	s	a	b	v
Luciferase inhibition	0.80	-	-	-3.53	3.78
Tadpole anaesthesia	0.66	-	-	-4.44	4.16
Water/octanol	0.56	-1.05	-	-3.46	3.81
Water/decanol	0.48	-0.97	-	-3.80	3.95
Water/hexadecane	0.67	-1.62	-3.59	-4.87	4.43

As shown before [47] the target site or sites in the two biological processes resemble wet octanol, or wet decanol, in that all have the same hydrogen-bond basicity as water (a = 0.00), but are much less acidic (a= -3.5 to -4.4) and naturally much more lipophilic (v = 3.8 to 4.2). Both wet octanol and wet decanol are reasonable models for the biological process, although both are rather less dipolar (s = 1.05 or -0.97 as compared to 0.00). Hexadecane is a poor model because it is less dipolar and very much less basic (a = -3.59) than the target sites. As more water/solvent partitions are analyzed, solvent systems that are even better models than wet octanol or wet decanol might be found.

CONCLUSIONS

These results on luciferase enzyme and tadpole anaesthesia illustrate a particular advantage of equations such as equation (12) and equation (13) as LSERs and especially as QSARs. Most QSARs are set up in order to predict biochemical or biological effects of solutes in a given system. Indeed, this is usually their only function. It follows that such QSARs are unlikely to yield information on the system itself. But equations (12) and (13) have been constructed in such a way that they can be used for predictive purposes, and also can be used to obtain chemical information on the system. The given equations (29) and (30) lead to specific information on the dipolarity/polarizability, the hydrogen-bond acidity and basicity, and the lipophilicity of the target site or sites in the luciferase enzyme and the tadpole. As shown before [47] these quantities can be compared with those in physicochemical systems such as the water-octanol or the water-hexadecane system. As more and more systems are investigated using equation (13), so it will be possible to extend such comparisons. Nearly all applications of equation (13) are to systems at ambient temperature, so that temperature effects have not to be considered. This is not so for equation (12), which has been applied to systems at temperatures ranging from 298K to 423K and over, and care has to be taken to compare coefficients only at a common temperature. However, as more and more systems are investigated at 298K or (for many GLC phases) at 393K, so it will be possible to extend the comparison of coefficients in equation (12) as well.

Thus equations (12) and (13) represent two very general types of LSER and QSAR that can be applied to all kinds of physicochemical and biochemical phenomena. They can be used for predictive purposes and can also be used to obtain specific chemical information on the processes concerned.

REFERENCES

1. R.W. Taft and M.J., Kamlet, *J.Am Chem.Soc.*, 1976, **98**, 2886.
2. M.J. Kamlet and R.W. Taft, *J.Am.Chem.Soc.*, 1976, **98**, 377.
3. M.J. Kamlet, J.-L.M. Abboud and R.W.Taft, *Progr. Phys. Org.Chem.*, 1981, **13**, 485.
4. M.J. Kamlet, J.-L.M. Abboud, M.H.Abraham and R.W.Taft, *J.Org.Chem.*, 1983, **48**, 2877.
5. R.W. Taft, J.-L .M. Abboud, M.J. Kamlet and M.H. Abraham, *J. Soln. Chem.*, 1985, **14**, 153.
6. M.H. Abraham, P.L. Grellier, J.-L.M. Abboud, R.M. Doherty and R.W. Taft, *Can. J. Chem.*, 1988, **66**, 2673.
7. M.H. Abraham, P.L. Grellier, D.V. Prior, P.P. Duce, J.J. Morris and P.J. Taylor, *J.Chem.Soc. Perkin Trans. 2*, 1989, 699.
8. M.H. Abraham, P.L. Grellier, D.V. Prior, J.J. Morris and P.J. Taylor, *J.Chem.Soc. Perkin Trans. 2*, 1990, 521.
9. M.H. Abraham, P.L. Grellier, D.V. Prior, R.W. Taft, J.J. Morris, P.J. Taylor, C. Laurence, M. Berthelot, R.M. Doherty, M.J. Kamlet, J.-L. M. Abboud, K. Sraidi and G. Guiheneuf, *J.Am.Chem.Soc.*, 1988, **110**, 8534.
10. O.A. Raevsky and V.P.Novikov, *Khimico-farm. Z.*, 1982, **16**, 583.
11. O.A. Raevsky, V. Yu. Grigor'ev, D.B. Kireev and N.S. Zefirov, *Quant. Struct.-Act. Relat.*, 1992, **11**, 49.
12. M.H. Abraham, G.J. Buist, P.L. Grellier, R.A. McGill, D.V. Prior, S.Oliver, E. Turner, J.J.Morris, P.J. Taylor, P.Nicolet, P.-C. Maria, J.-F. Gal, J.-L.M. Abboud, R.M. Doherty, M.J. Kamlet, W.J. Shuely and R. W. Taft, *J. Phys.Org.Chem.*, 1989, **2**, 540.
13. C. Laurence, M. Berthelot, E. Raczynska, J.-Y. Le Questel, G. Dugay and P. Hudhomme, *J.Chem.Res.(s)*, 1990, 250.
14. M. Berthelot, M. Helbert, C. Laurence, J.-Y. Le Questel, F. Anvia and R. W. Taft, *J.Chem.Soc. Perkin Trans,2*, 1993, 625.
15. M.H. Abraham, *Chem.Soc. Revs.*, 1993, **22**, 73.
16. M.H. Abraham, *J.Phys.Org.Chem.*, in the press.
17. M.H.Abraham, G.S.Whiting, R.M. Doherty and W.J. Shuely, *J.Chem.Soc.Perkin Trans.2*, 1990, 1451.
18. M.H.Abraham, G.S.Whiting, R.M. Doherty and W.J. Shuely, *J.Chromatogr.*, 1991, **587**, 213.

19. M.H.Abraham and G.S. Whiting, *J.Chromatogr.*, 1992, **594**, 229.
20. M.H. Abraham, *J.Chromatogr.*, 1993, in the press.
21. M.H.Abraham, P.L Grellier and R.A. McGill, *J.Chem.Soc. Perkin Trans.2*, 1987, 797.
22. M.H. Abraham and J.C. McGowan, *Chromatographia*, 1987, **23**, 243.
23. B.R. Kersten, S.K. Poole and C.F. Poole, *J.Chromatogr.*, 1989, **468** 235.
24. M.H. Abraham, G.S. Whiting, R.M Doherty and W.J. Shuely, *J.Chromatogr.*, 1991, **587**, 229.
25. M.H. Abraham, G.S.Whiting, R.M. Doherty and W.J. Shuely, *J.Chromatogr.*, 1990, **518**, 329.
26. M.H. Abraham, I. Hamerton, J.B. Rose and J.W. Grate, *J.Chem.Soc.Perkin Trans 2*, 1991, 1417.
27. M.H.Abrahan, J. Andonian-Haftvan I. Hamerton, C.F. Poole and T.O. Kollie, *J. Chromatogr.*, in the press.
28. J. Li, Y. Zhang, H. Ouyang and P.W. Carr, *J.Am.Chem.Soc.*, 1992, **114**, 9813.
29. T.O. Kollie, C.F. Poole, M.H. Abraham and G.S. Whiting, *Anal.Chim.Acta*, 1992, **259**, 1.
30. C.F. Poole, T.O. Kollie and S.K. Poole, *Chromatographia* , 1992, **34** 281.
31. M.H. Abraham, G.S. Whiting, J. Andonian-Haftvan, J.W.Steed and J.W. Grate, *J. Chromatogr.*, 1991, **588**, 361.
32. J.W. Grate and M.H.Abraham, *Sensors and Actuators B*, 1991, **3**, 85.
33. Q. Du, P. Hattam and P. Munk, *J.Chem.Eng.Data*, 1990, **35**, 367.
34. M.H. Abraham, G.S. Whiting, R.M. Doherty, W.J. Shuely and P. Sakellariou, *Polymer*, 1992, **33**, 2162.
35. M.Galin, *Polymer*, 1983, **33**, 865.
36. A.J. Ashworth and G.J. Price, *Macromolecules*, 1986, **19**, 358.
37. I.H. Romdhane and R.P. Danner, *J.Chem.Eng.Data*, 1991, **36**, 15.
38. M.H. Abraham, C.M. Du, J.P. Osei-Owusu, P. Sakellariou and W.J. Shuely, *Eur.Polymer J.*, in the press.
39. M. H. Abraham and G.S. Whiting, *J.Am.Oil Chemists Assn.*, 1992, **69**, 1236.
40. J. F. Pankow, *J. Chromatogr.*, 1991, **547**, 488.
41. M.H. Abraham, and D.P. Walsh, *J. Chromatogr.*, 1992, **627**, 294.
42. M.H. Abraham, J. Andonian-Haftvan, J.P. Osei-Owusu, P. Sakellariou, J.S. Urieta, M.C. Lopez and R. Fuchs, *J.Chem.Soc.Perkin.Trans.2*, 1993, 299.
43. M. H. Abraham, G.S. Whiting, Y. Alarie, J.J. Morris, P.J. Taylor, R.M. Doherty, R.W. Taft and G.D. Nielsen, *Quant, Structure-Activity Relat.*, 1990, **9**, 6.
44. A. Leo and C. Hansch, *J.Org.Chem.*, 1971, **36**, 1539.
45. N.P. Franks and W.R. Lieb, *Nature*, 1984, **310**, 599.
46. N.P. Franks and W.R. Lieb, *Nature*, 1985, **316**, 349.
47. M. H. Abraham, W.R. Leib and N.P. Franks, *J.Pharm.Sci.*, 1991, **80**, 719.