Journal of Chromatography, 218 (1981) 409–425
Elsevier Scientific Publishing Company, Amsterdam — Printed in The Netherlands

CHROM. 14,098

ESTIMATION OF AQUEOUS SOLUBILITIES OF ORGANIC NON-ELECTROLYTES USING LIQUID CHROMATOGRAPHIC RETENTION DATA

T. L. HAFKENSCHEID and E. TOMLINSON*
Subfaculty of Pharmacy, University of Amsterdam, Plantage Muidergracht 24, 1018 TV Amsterdam (The Netherlands)

SUMMARY

The use of reversed-phase liquid chromatographic retention parameters in a modified Hildebrand–Scott equation to describe the aqueous solubilities of non-electrolyte liquids and solids has been studied. The results indicate that, for 32 model solutes of various characters, a semiempirical relationship between aqueous solubility, a theoretical capacity factor obtained by extrapolation of retention data in simple binary systems to a pure aqueous eluent and a function of solute melting points and entropies of fusion can be used to give good estimates of aqueous solubilities. Relationships are given between extrapolated capacity factors and liquid–liquid distribution coefficients or smoothed surface areas. It is suggested that the assumptions and intricacies often needed to calculate the last two parameters make the chromatographic parameter an appropriate candidate for describing solute non-ideality and for use in the modified Hildebrand–Scott equation.

INTRODUCTION

Aqueous solubility has long been recognised as a key factor in pharmaceutics and chemistry. The phenomenology of drug delivery, transport and distribution, the prediction of chemical environmental effects and development of analytical methods, etc., are dependent *inter alia* upon knowledge of aqueous solubilities. Aqueous solubilities are also of thermodynamic interest since they give information on the nature of non-ideal solutions.

The assessment of solubility can be extremely difficult, particularly for poorly soluble compounds. This may be due to very low equilibration rates, compound instability, the effect of impurities on solubility and (during a drug development programme) lack of sufficient compound or of a specific analytical technique. Discrepancies in reported solubilities are striking, for example Tulp and Hutzinger[1] have found that the aqueous solubility of DDT is between 0.2 and 1000 ppb (10^9). Thus there is a need to be able to make either *a priori* predictions of solubilities in simple and mixed systems or to form reliable estimates using readily obtainable parameters. A number of approaches[2–7] to this problem have been made, with frequent use being made of semiempirical relationships between solubility and physicochemical properties.

0021-9673/81/0000–0000/$02.50

Yalkowsky–Valvani equation

For a solid, taking as a good approximation that the latent heat of fusion, ΔH_f, is independent of temperature, and that the difference in heat capacities of the crystalline and molten forms of the solid is small, it can be shown[8,9] that the temperature variation of the mole fraction ideal solubility of a pure solid (A_s), $X^i_{A_s}$, is described by

$$\ln X^i_{A_s} = \frac{\Delta H_f}{R} \cdot \left(\frac{1}{T_m} - \frac{1}{T} \right) \tag{1}$$

where T_m and T are the melting point and temperature at which pure solid is in equilibrium with the solution of mole fraction X. For a liquid solute (A_l) the Gibbs free energy excess function, $G^E_{A_l}$, is

$$G^E_{A_l} = RT \ln \gamma_{A_l} = (G_{A_l} - G^i_{A_l}) \tag{2}$$

where γ is the activity coefficient. Thus, it follows that

$$\log X_{A_l} = \log X^i_{A_l} - \log \gamma_{A_l} \tag{3}$$

and, using a mole fraction scale:

$$\log X_{A_l} = -\log \gamma_{A_l} \tag{4}$$

Since the entropy of fusion, ΔS_f, at T_m is given by $\Delta H_f/T_m$, eqns. 1 and 4 can be combined to give:

$$\log X_{A_s} = \frac{\Delta S_f}{2.3\,R} \cdot \left(\frac{T_m - T}{T} \right) - \log \gamma_A \tag{5}$$

Eqn. 5 is similar to that derived by Yalkowsky and Valvani[10,11] using a phenomenological approach without invoking excess functions; from this relationship it is apparent that the estimation of aqueous solubilities of a solid requires knowledge of a term reflecting solid–solid interactions, and one describing solute–solvent and solute–solute interactions. Yalkowsky and Valvani[10] have suggested that, as ΔS_f is often constant, the first term be described by some function of T_m; and that, since most compounds of pharmaceutical (*sic*) interest have solubility parameters which do not differ greatly from that of octan-1-ol, then the second term in eqn. 5 can be approximated by the water–octan-1-ol liquid–liquid distribution coefficient[12], K_d (by implying that the solute activity coefficients in this oil are unity). This gives

$$\log X_A = -K \log K_{d_A} - K' T_{m_A} - K'' \tag{6}$$

which may be regarded as the Yalkowsky–Valvani equation[10]. Although this approach has been used by others[13], problems arise with the use of this K_d scale; in practice, it is extremely difficult to determine log K_d values greater than 4, solutes need to be pure and stable and there is a high solute consumption. If one resorts to calculation of K_d via a group or fragmental constant approach[14], problems arise with

complicated structures[15] and neighbourhood effects; also the data banks available are unable to predict K_d in mixed solvents (as would be the intention for estimating solute solubilities in the presence of a cosolvent).

The high precision afforded by modern liquid chromatography has prompted the use of this technique for estimating various physicochemical properties of small organic molecules. These include distribution coefficients[15–17]*, ionization constants[18,19], complex formation constants[20], diffusion coefficients[21] and critical micelle concentrations[22]. It has been demonstrated[17] that the use of reversed-phase high-performance liquid chromatographic (HPLC) retention parameters is suitable for indicating solute chemical potentials in polar solvents, and it appeared useful therefore to attempt to use the Yalkowsky–Valvani model to extend the applicability of HPLC to estimations of aqueous solubilities. Our findings are now described.

EXPERIMENTAL

Materials

Model solutes used were (Table I) benzene and mono- and 1,4-disubstituted benzenes (functional groups CH_3, Cl, NO_2, OH, NH_2 and COOH), naphthalene, azulene, anthracene, phenanthrene and isopropyl derivatives (with the above functional groups). These were of synthetic or analytical grade and were obtained from E. Merck (Darmstadt, G.F.R.), Fluka (Buchs, Switzerland), BDH (Poole, Great Britain), ICN Pharmaceuticals (Plainview, NY, U.S.A.) and Aldrich (Milwaukee, WI, U.S.A.). The stationary phase was Hypersil ODS (5 μm) (Shandon, Astmoor, Great Britain). Mobile phases were made up by volume using methanol (analytical grade; J. T. Baker, Phillipsburgh, NJ, U.S.A.) and either (i) water, once distilled from an all-glass still, (ii) 10 m*M* aqueous phosphate buffer (pH 2.15; ionic strength 0.01 *M*) or (iii) 2.5 m*M* aqueous hexadecyltrimethylammonium bromide solution (to give a final eluent concentration not exceeding 0.75 m*M*). Hexadecyltrimethylammonium bromide (HTAB) was of analytical grade (E. Merck).

Instruments and columns

Chromatographic equipment consisted of an Altex 110A single piston pump (Altex, Berkeley, CA, U.S.A.) with additional dampening, a Model 7125 injection valve (Rheodyne, Berkeley, CA, U.S.A.) and Waters 440 UV absorbance and R401 refractive index detectors (Waterx Assoc., Milford, MA, U.S.A.) arranged in tandem. Peak recording was achieved with a Kipp BD 41 flat-bed potentiometric recorder (Kipp & Zn., Delft, The Netherlands). The eluent reservoir and the column were kept at 20.00 $\pm$ 0.01°C by immersion in a Hetotherm 02PT 623 thermostat water-bath (Heto, Birkeröd, Denmark). The eluent reached the injection valve via a 1-ml coil immersed in the water-bath.

Procedures

Columns (100 × 3 mm and 50 × 4.6 mm) were of stainless steel (316), and were packed using a slurry of stationary phase (3%, w/v) in carbon tetrachloride–methanol (95:5, v/v), with methanol as packing liquid. Solutes were dissolved in the

* To date, 33 studies have been reported; these references will enable such studies to be identified.

eluent and retention times measured using a Model 310 microsplit stop watch (Heuer, Bienne, Switzerland), with corrections being made for dead volumes of connections. The capacity factor, k, was calculated from $(t \cdot t_0^{-1} - 1)$, where t and t_0 are the corrected solute retention time and the retention time of eluent slightly enriched with water. All k values were calculated from the mean of at least four t values, and had coefficients of variance of less that 0.5%.

Where necessary, aqueous solubility determinations were carried out by shaking/ultrasonicating (at 20°C) a supersaturated solution until equilibration, followed by ultracentrifugation, filtration and sampling.

Multiple linear regressions were carried out using a standard computer program.

RESULTS AND DISCUSSION

Using a fixed column and eluent at a given flow and temperature, the retention of a solute in the case of pure solvophobic chromatography is dependent upon the solvation of the solute[23], *i.e.*

$$\kappa = \text{constant} + \left(\frac{\Delta G_c + \Delta G_i}{2.3\,RT}\right) \tag{7}$$

where κ is the logarithmic form of k; the free energy of solvation is given by the excess free energy, G^E, and is the sum of the free energies required to create a cavity in the eluent for the solute, ΔG_c, and that gained due to interactive forces upon placing the solute in the eluent, ΔG_i. The excess free energy arising from placing a solute in the eluent is thus a measure of the deviation from ideal behaviour, such that $G^E = 0$ if the solute is "equivalent" to the eluent. Since[24]

$$G_A^E = RT \ln \gamma_{A_e} \tag{8}$$

where subscript e refers to eluent, and for which it is assumed[23] that stationary phase activity is constant, then the use of water as the eluent, combined with the knowledge that[25]

$$\ln \gamma_{A_e} = (A/RT)\, x_w^2 \tag{9}$$

where A is a constant for a given system (accounting for solute–solute, solute–solvent and solvent–solvent interactions) and x_w is the mole fraction of water in a saturated solution, results in an approximately constant value of $\ln \gamma_{A_e}$ if x_w approaches unity. Thus, at low solubility ($x_w \approx 1$), this gives (after combination of eqns. 7, 8 and 9), κ_A in a purely aqueous eluent, κ_{w_A}, as a measure of $\log \gamma_A$ in a saturated solution, *viz.*:

$$\log \gamma_{A_{saturated}} = \kappa_A - \text{constant} \tag{10}$$

For poorly soluble solutes the use of water as eluent in a reversed-phase LC system is made untenable by excessive retention and insufficient detector sensitivity. However, we can approach a value for the capacity ratio in water using extrapolation

techniques. One such method is to relate k to the organic modifier volume fraction, φ, in the eluent, as has been formalized by[26]

$$\kappa = \kappa_w + B\varphi \tag{11}$$

where κ_w refers to the logarithmic form of the solute capacity factor in pure water.

In this study we have examined the retention of 32 aliphatic and aromatic organic solutes in a reversed-phase mode (see Experimental) using three different eluents, *i.e.*, water–methanol, aqueous buffer (pH 2.15)–methanol and aqueous hexadecylammonium bromide–methanol, using a volume fraction of organic modifier of 0.3–0.90. The data obtained have been analysed using linear regression analysis, and the results are given in Table I for all eluents examined. It can be seen that eqn. 11 well describes retention of the studied volume fraction range, with correlation coefficients being in only one case less than 0.999. For some solutes, deviations from linearity occurred at high volume fractions; here, the resulting low retentions can lead to large inaccuracies in determining k. Moreover, the buffering capacity of the eluent for some carboxylic acids may be too low in this study (methanol increases eluent pH at $\varphi >$ 0.75, ref. 27). In these cases regression analysis was restricted to the linear portion of the κ *versus* φ plot.

For ionizable solutes, buffered eluents were used to obtain κ_w values, with the addition of buffer to the eluent being tested for other non-ionizable reference solutes (Table I). Thus, for eight reference solutes, the presence of buffer had none or negligible effect on either B or κ_w, which justifies (i) the use of buffering to obtain κ_w for non-ionized solutes and (ii) the inclusion of data from a buffered and a non-buffered system in the same data set.

Recently it has been suggested[28] that silanol groups at the surface of reversed stationary phases may be responsible for departure from the retention behaviour expected on the basis of solvophobic theory[23]. Being cognisant of this, we have added a silanol-masking compound (hexadecyltrimethylammonium bromide) to the eluents for three solutes expected to be silanophilic[29], and the results obtained are also given in Table I. Apart from expected ion-pair effects where appropriate, HTAB affected the retention of the reference compounds only very slightly at volume fractions of organic modifier less than 0.6. Apart from the solutes given in Table I, retention was studied for 4-aminophenol, 1,4-diaminobenzene, 2-aminopropane and trimethylamine. These solutes either did not have proper solvophobic retention (exclusion phenomena, irregular peak shape) or decomposed before or during chromatography.

Using Hypersil ODS and aqueous methanol eluents for the compounds studied, the value of B from the κ *versus* φ relationships (eqn. 11) ranges from -1.4 to -5.2, and has an approximate rank order with the intercept (κ_w) values. This is shown in Fig. 1, which may be described by

$$\kappa_w = -1.18\,B - 1.54 \qquad (n = 32,\ r = 0.930) \tag{12}$$

where n and r are the number of solutes and the correlation coefficient respectively. The variation in B is contrary to the comment of Snyder *et al.*[26] that B is constant for any one stationary phase–eluent system, but is consistent with other findings, *e.g.*, refs. 17, 30, 31, that eluent changes alter hydrophobic and polar groups selectivity, and

TABLE I

REGRESSION COEFFICIENTS ACCORDING TO EQN. 11

See Experimental for full details. n = Number of data points. When $n = 4$ the volume fractions are 0.30, 0.45, 0.60 and 0.75; when $n = 5$, $\varphi = 0.90$ is also included in the data set. r = Correlation coefficient.

Compound		*Eluent*											
		Water–methanol				*Buffer (pH 2.15)–methanol*				*Aqueous HTAB solution–methanol*			
		$-B$	κ_w	n	r	$-B$	κ_w	n	r	$-B$	κ_w	n	r
$R^1 \cdot C_6H_4 \cdot R^2$													
R^1	R^2												
H	H	2.79	2.03	5	0.9992	2.77	2.00	5	0.9996	2.92	2.12	5	1.000
	CH_3	3.27	2.59	5	0.9996								
	Cl	3.46	2.71	5	0.9998	3.46	2.71	5	0.9998				
	NO_2	2.97	1.93	5	0.9999								
	OH	2.66	1.37	5	0.9999	2.59	1.32	5	0.9999				
	NH_2									2.15	0.86	5	0.9998
	COOH					3.27	1.92	4	0.9995				
CH_3	CH_2	3.76	3.18	5	0.9998								
	Cl	3.98	3.31	5	0.9998	3.99	3.33	5	0.9998				
	NO_2	3.39	2.48	4	0.9998								
	OH	3.07	1.91	5	0.9997								

	NH_2									2.17	1.05	5	0.9994
	COOH					3.77	2.48	4	0.9990				
Cl	Cl	4.03	3.33	5	0.9998								
	NO_2	3.31	2.44	4	0.9999								
	OH	3.33	2.21	5	0.9998								
	NH_2	2.93	1.80	5	0.9992								
	COOH					3.86	2.70	4	0.9996				
NO_2	NO_2	2.87	1.72	5	0.9998	2.80	1.70	4	1.000				
	OH	3.09	1.77	5	0.9998								
	NH_2	2.77	1.31	5	0.9993								
	COOH					3.27	1.99	4	0.9999				
OH	OH	2.33	0.29	5	1.000								
	COOH					3.01	1.20	4	0.9990				
H_2	COOH					2.35	0.55	4	0.9999				
Naphthalene		4.07	3.32	5	0.9996								
Phenanthrene		5.06	4.42	4	0.9993								
Anthracene		5.20	4.58	4	0.9992								
$CH_3 \cdot CH(R^3)CH_3$													
R^3 =	OH	1.37	0.18	4	0.9980	1.35	0.17	4	0.9998				
	NO_2	1.98	0.90	4	0.9990	1.94	0.88	4	0.9999				
	Cl	2.70	1.92	4	0.9990	2.78	1.95	4	0.9980				
	COOH					2.16	0.95	4	0.9990				

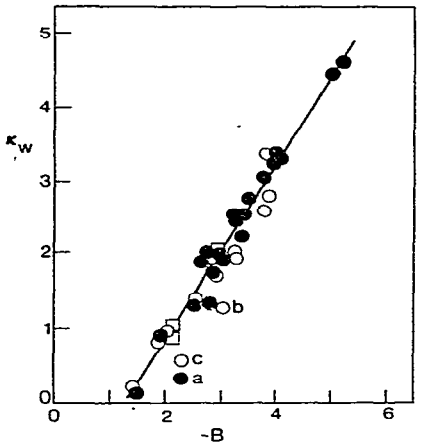


Fig. 1. Relationship between extrapolated log capacity factor, κ_w, and B derived using eqn. 11 for 32 non-electrolytes in a reversed-phase HPLC system. Eluents: ●, water–methanol; ○, buffer–methanol; □, aqueous HTAB solution–methanol. The straight line is the regression line according to eqn. 12. a = *p*-Hydroxyphenol; b = *p*-hydroxybenzoic acid; c = *p*-aminobenzoic acid.

thus appears to reflect not only solvent strength[26] but also specific interactions between solutes, stationary phase and eluent[15]. Finally, from Table I, it can be seen that, since the relationships given by eqn. 11 are linear over a fairly wide range of volume fractions of organic modifier, there is no need to introduce a parabolic function, *i.e.*, φ^2, as has been suggested[32].

Estimation of aqueous solubilities of liquids

Table II gives the physicochemical data used in this study which comprises entropies of fusion (calculated from enthalpies of fusion[33,34]), melting points[33,34], measured water–octan-1-ol distribution coefficients[14,35] and aqueous solubilities (this study and ref. 36). The validity of introducing a liquid chromatographic retention term in the Hildebrand–Scott equation (eqn. 5) can be tested first by regressing solubility data from Table II for solutes that are liquids at 20°C with κ_w values obtained using eqn. 11, since the term $\Delta S_f/R\,[T_m - T/T]$ does not play a rôle. This results in:

$$-\log X_w = 0.54 + 1.34\,\kappa_w \quad (n = 11,\ r = 0.959) \tag{13}$$

For the eleven liquids studied, both propan-2-ol and isobutyric acid are too soluble in water to obey the assumption that $\ln \gamma_{A_c}$ (eqn. 9) is constant, since x_w deviates significantly from unity. (For isobutyric acid this deviation is 3.9% and for propan-2-ol 100%.) Regression of the data without these two compounds gives (Fig. 2):

$$-\log X_w = 1.37 + 1.01\,\kappa_w \quad (n = 9,\ r = 0.993) \tag{14}$$

The excellent correlation coefficient obtained and the fact that the coefficient of κ_w approximates to unity is in accord with theory (eqn. 10), and appears to validate the

TABLE II

PHYSICOCHEMICAL PROPERTIES OF MODEL SOLUTES

ΔS_f = Entropy of fusion[33,34]; T_m = melting point[33,34]; X_w = observed aqueous mole fraction solubility[10,36]; SSA = smoothed surface area[38]; K_d = observed water–octan-1-ol distribution coefficient[14,35].

Compound		ΔS_f (J/K·mol)	T_m (°C)	$-\log X_w$	K_w*	*SSA* (A^2)	$\log K_d$	
							obs.	*calc.***
$R^1 \cdot C_6H_4 \cdot R^2$								
R^1	R^2							
H	H	***	***	3.36	2.05	89.7	2.13	2.07
	CH_3	***	***	4.00	2.59	107.0	2.74	2.61
	Cl	***	***	4.10	2.71	109.6	2.84	2.81
	NO_2	***	***	3.51	1.91		1.87	1.79
	OH	35.9	40.9	1.77	1.35		1.48	1.51
	NH_2	***	***	2.17	0.86		0.90	0.97
	COOH	43.8	122.4	3.29	1.92		1.87	1.86
CH_3	CH_3	***	***	4.54	3.19	124.2	3.15	3.09
	Cl	***	***	4.82 §	3.32	126.8	3.33	3.33
	NO_2	50.2	54.5	4.13	2.48		2.39	2.31
	OH	40.2	34.8	2.44	1.91		1.94	2.03
	NH_2	56.5	43.7	2.79	1.05		1.39	1.49
	COOH	51.9	182	4.29	2.48		2.27	2.39
Cl	Cl	56.1	53.1	5.00	3.35	129.5	3.39	3.57
	NO_2	49.0	85.6	4.59 §	2.44		2.40	2.55
	OH	46.4	43.5	2.73	2.23		2.39	2.27
	NH_2	57.7	72.5	3.41 §	1.80		1.83	1.73
	COOH	62.8	243	5.26	2.70		2.65	2.85
NO_2	NO_2	62.8	174	4.12	1.73		1.48	1.53
	OH	54.4	115	3.33	1.75		1.91	1.25
	NH_2	50.2	149	4.12	1.31		1.39	0.71
	COOH	72.0	242	4.80	1.99		1.89	1.61
OH	OH	61.1	174	1.91	0.29		0.55	0.97
	COOH	51.1	215	2.98	1.20		1.58	1.32
NH_2	COOH	45.6	189	2.96	0.55		0.68	0.79
Naphthalene		53.1	80.2	5.37	3.32	128.7	3.37	3.37
Phenanthrene		49.0	96.3	6.73	4.42	165.0	4.46	4.67
Anthracene		60.2	216.5	8.04	4.58	167.6	4.45	4.67
$CH_3 \cdot CH(R^3)CH_3$								
R^3 =	OH	***	***	0	0.18		0.05	−0.01
	NO_2	***	***	2.36 §	0.89		0.87	0.62
	Cl	***	***	3.18	1.93		1.90	1.70
	COOH	***	***	1.41	0.95		0.82	0.64

* Mean values from Table I.

** Using hydrophobic fragment constants[14].

*** Liquid at room temperature.

§ Determined in this study.

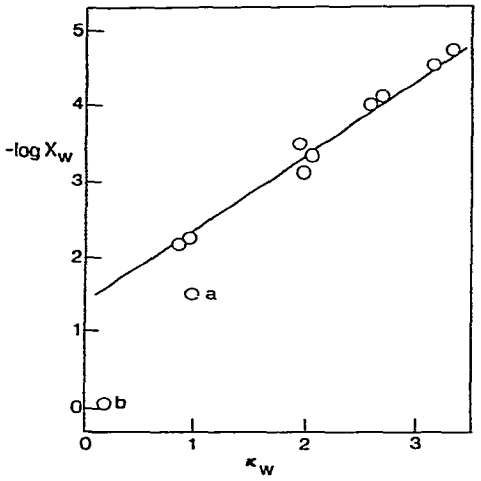


Fig. 2. Estimation of aqueous solubilities of liquids (at 20°C) from mean extrapolated log capacity factors, κ_w, using the general form of eqn. 13. The straight line is the regression line according to eqn. 14. Isobutyric acid (a) and propan-2-ol (b) are outliers since they do not obey the assumption given by eqn. 9.

substitution of log γ_A in eqn. 5 by (κ_w − constant). (Eqn. 14 implies that the product of κ_w and log X_w is constant for liquids, and so any structural modification that increases κ_w without altering the liquid state will result in an equivalent decrease in aqueous solubility.)

Estimation of aqueous solubilities of solids

Using the physicochemical constants given in Table II and the extrapolated κ_w values, the complete data set available for solids ($n = 21$) and for solids and liquids ($n = 32$) has been fitted by multiple linear regression to functions of κ_w and $\Delta S_f(T_m - 20)$, and (eqn. 6) κ_w and $(T_m - 20)$. These analyses gave

$$-\log X_w = 1.28\,\kappa_w + 0.0073\,(T_m - 20) + 0.42 \quad (15)$$
$(n = 21, R = 0.924, F = 55.4)$

$$-\log X_w = 1.27\,\kappa_w + 0.0070\,(T_m - 20) + 0.60 \quad (16)$$
$(n = 32, R = 0.956, F = 159.3)$

$$-\log X_w = 1.24\,\kappa_w + 1.23 \times 10^{-4}\,\Delta S_f(T_m - 20) + 0.55 \quad (17)$$
$(n = 21, R = 0.924, F = 55.4)$

$$-\log X_w = 1.26\,\kappa_w + 1.15 \times 10^{-4}\,\Delta S_f(T_m - 20) + 0.65 \quad (18)$$
$(n = 32, R = 0.955, F = 155.5)$

where R and F are the multiple correlation coefficient and the variance ratio. These results show that the aqueous solubilities for both liquids and solids can be estimated by use of the same semiempirical relationship, and that for this data set, the results are not improved by introduction of the ΔS_f term. This is in accord with previous suggestions[10] that the variation in ΔS_f can be neglected. Yalkowsky and Valvani[10] have shown that the ideal aqueous solubility of various classes of molecules can be calculated using the approximation

TABLE III

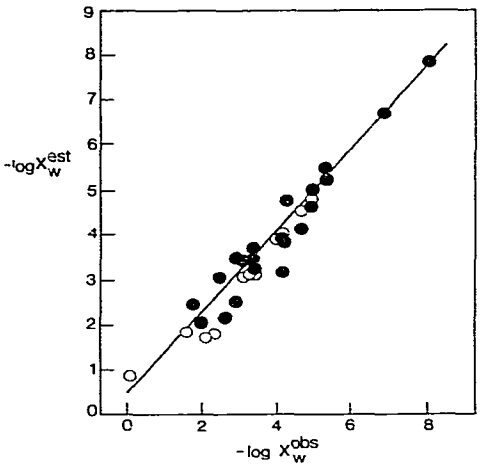
ESTIMATED AQUEOUS SOLUBILITIES USING HPLC RETENTION DATA TO DESCRIBE NON-IDEALITY

Compound		*Estimated aqueous solubility ($-\log X_w$)*	
		According to eqn. 16	*According to eqn. 18*
$R^1 \cdot C_6H_4 \cdot R^2$			
R^1	*R^2*		
H	H	3.20	3.24
	CH_3	3.88	3.92
	Cl	4.04	4.07
	NO_2	3.02	3.06
	OH	2.46	2.44
	NH_2	1.69	1.74
	COOH	3.74	3.59
CH_3	CH_3	4.64	4.67
	Cl	4.81	4.84
	NO_2	3.91	3.98
	OH	3.12	3.13
	NH_2	2.12	2.13
	COOH	4.87	4.74
Cl	Cl	5.08	5.09
	NO_2	4.14	4.08
	OH	3.59	3.59
	NH_2	3.25	3.27
	COOH	5.58	5.66
NO_2	NO_2	3.87	3.94
	OH	3.48	3.45
	NH_2	3.16	3.08
	COOH	4.67	5.00
OH	OH	2.04	2.10
	COOH	3.48	3.31
NH_2	COOH	2.47	2.23
Naphthalene		5.23	5.20
Phenanthrene		6.74	6.65
Anthracene		7.78	7.78
$CH_3 \cdot CH(R^3)CH_3$			
R^3 =	OH	0.83	0.88
	NO_2	1.73	1.77
	Cl	3.05	3.08
	COOH	1.80	1.85

$$-\log X^{i}_{A,} = 0.01\ (T_m - 25) \tag{19}$$

for rigid non-spherical molecules, and

$$-\log X^{i}_{A,} = [0.01 + 0.0018\ (n - 5)]\ (T_m - 25) \tag{20}$$



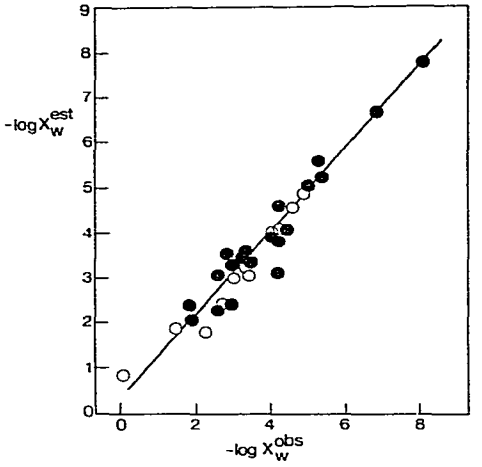
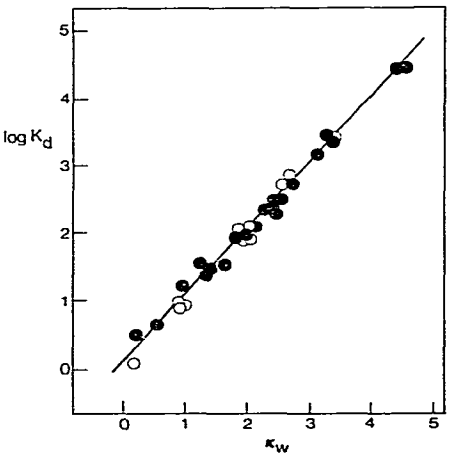


Fig. 3. Relationship between observed mole fraction aqueous solubilities for liquids (open points) and solids and solubilities estimated using the coefficients of eqn. 16.

Fig. 4. Relationship between observed mole fraction aqueous solubilities for liquids (open points) and solids and solubilities estimated using the coefficients of eqn. 18.

for partially flexible molecules, where n is the number of carbon and/or heteroatoms in a solute sidechain, and where measurements are at 20°C.

The good correlation found in this study using the approximations has prompted us to analyse the available data set with respect to structural groupings. Accordingly, eqns. 21–28 give the multiple linear regression analyses for four groupings according to the general forms of eqns. 16 and 18.



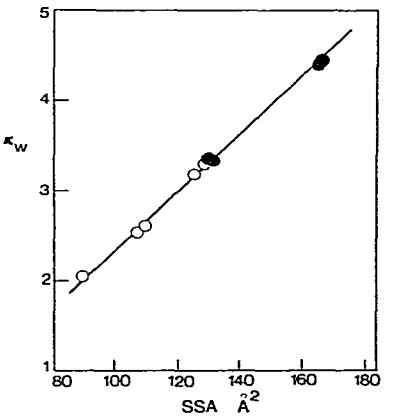


Fig. 5. Relationship between log extrapolated capacity factors, κ_w, according to eqn. 11 (Table II) and observed water–octan-1-ol distribution coefficients (Table II). The straight line is the regression line for all solutes according to eqn. 30. Key as in Fig. 3.

Fig. 6. Correlation between solute smoothed surface area, SSA[38], and κ_w. The straight line is the regression line according to eqn. 33. Key as in Fig. 3.

(i) Compounds with CH_3, Cl and NO_2 groups:

$$-\log X_w = 1.10\,\kappa_w + 0.0085\,(T_m - 20) + 1.18 \quad (21)$$
$$(n = 15,\ R = 0.989,\ F = 290.6)$$
$$-\log X_w = 1.13\,\kappa_w + 1.37 \times 10^{-4}\,\Delta S_f\,(T_m - 20) + 1.14 \quad (22)$$
$$(n = 15,\ R = 0.986,\ F = 227.3)$$

(ii) Compounds with OH groups:

$$-\log X_w = 1.22\,\kappa_w + 0.0093\,(T_m - 20) - 0.07 \quad (23)$$
$$(n = 7,\ R = 0.956,\ F = 26.5)$$
$$-\log X_w = 1.29\,\kappa_w + 1.71 \times 10^{-4}\,\Delta S_f\,(T_m - 20) - 0.14 \quad (24)$$
$$(n = 7,\ R = 0.968,\ F = 37.2)$$

(iii) Compounds with CO_2H groups:

$$-\log X_w = 1.28\,\kappa_w + 0.0097\,(T_m - 20) - 0.03 \quad (25)$$
$$(n = 5,\ R = 0.926,\ F = 9.02)$$
$$-\log X_w = 1.19\,\kappa_w + 1.21 \times 10^{-4}\,\Delta S_f\,(T_m - 20) + 0.37 \quad (26)$$
$$(n = 5,\ R = 0.988,\ F = 61.4)$$

(iv) Compounds with NH_2 groups:

$$-\log X_w = 0.53\,\kappa_w + 0.0127\,(T_m - 20) + 1.80 \quad (27)$$
$$(n = 4,\ R = 0.988,\ F = 40.9)$$
$$-\log X_w = 0.40\,\kappa_w + 2.63 \times 10^{-4}\,\Delta S_f\,(T_m - 20) + 1.91 \quad (28)$$
$$(n = 4,\ R = 0.991,\ F = 54.8)$$

From the variance ratios for these equations it is seen that the use of the melting point approximation is not always appropriate for closely related groups of compounds, and we conclude that further work is required to determine the validity of the approximation.

Table III gives the estimated aqueous solubilities found using eqns. 16 and 18, and Figs. 3 and 4 are plots of observed (Table II) and estimated aqueous solubilities (of liquids and solids) found using these relationships. It can be seen that use of the $(T_m - 20)$ or the $\Delta S_f(T_m - 20)$ term combined with κ_w gives estimates of aqueous solubilities ranging from reasonable to excellent, with better correlation being obtained for the more insoluble compounds. The study of compounds with even lower solubilities than those examined here should enable better estimates of the coefficients of eqn. 18, so that the equation can be used predictively.

Since, for the compounds studied, an excellent relationship (Fig. 5 and eqns. 9 and 30) is found between extrapolated κ_w values and experimentally measured water–octan-1-ol distribution coefficients, K_d (Table II)

$$\log K_d = -0.09 + 1.05\,\kappa_w \quad (n = 11,\ r = 0.997) \quad (29)$$
$$\log K_d = -0.06 + 1.02\,\kappa_w \quad (n = 32,\ r = 0.991) \quad (30)$$

where $n = 11$ refers to liquids only, it should follow that the use of K_d in estimating solubilities should give results comparable to those obtained with κ_w. Eqns. 31 and 32 show this to be the case:

$$-\log X_w = 1.29 \log K_d + 0.0071 (T_m - 20) + 0.50 \quad (31)$$
$(n = 32, R = 0.951, F = 141.9)$

$$-\log X_w = 1.29 \log K_d + 1.19 \times 10^{-4} \Delta S_f(T_m - 20) + 0.55 \quad (32)$$
$(n = 32, R = 0.952, F = 145.1)$

As discussed above (Introduction), liquid–liquid distribution coefficients are notoriously difficult to determine over a wide range, and use is often made of estimation procedures. Using a hydrophobic fragment approach and standard computational procedures[14], we have calculated K_d for some of the solutes studied here (Table II), and then used these values in multiple regression analysis of the data according to the general form of eqn. 18. For all compounds we obtain a R value of 0.930, $F = 96.0$, which compares with a value of 0.955, $F = 155.5$, by using κ_w values. Similarly, Yalkowsky and Morozowich (Table 14, ref. 36) found that use of calculated K_d values for complicated drug structures gives poor agreement ($n = 7$, $r = 0.825$) between observed solubilities and those estimated according to eqn. 6. Here lies the advantage in using liquid chromatography, since the experimentally accessible κ_w scale is far greater than that[17] for K_d and the ease and precision of determination obviates the need for calculations of K_d using approximation procedures. (It can be shown that, using reversed-phase LC to obtain κ_w values, aqueous solubilities on a log mole fraction scale of -1 to -11 can readily be accessed.)

Several relationships have been established between K_d[37–39], aqueous solubility[10,40–42], and some geometrical properties of molecules such as cavity surface area and relative surface area. From eqns. 18, 29 and 30, similar relationships should be expected between κ_w and such properties. Using the smoothed surface area (SSA) (Table II) recently calculated by Bultsma[38], which accounts only for exposed surfaces, we find (Fig. 6):

$$\kappa_w = 0.031 \text{ SSA} - 0.79 \quad (n = 9, r = 0.999) \quad (33)$$

The excellent fit of the data by this relationship is in accord with the use of solvophobic theory[23] for describing solute retention in reversed-phase liquid chromatographic systems[20], and further justifies our previous use of eqns. 7 and 8. As for calculation of K_d[14], SSA computations appear to be somewhat complicated for drug structures.

CONCLUSIONS

Locke[43] has described as "adequate" a relationship between relative retention data and the aqueous solubilities of nine aromatic fused ring molecules. In the present study the correlation between κ_w and log X_w is very poor ($n = 32$, $r = 0.452$); the variance between observed and estimated values of log X_w is only reduced to *ca.* 9% when the $\Delta S_f(T_m - 20)$ or $(T_m - 20)$ approximations are added.

This study has used theoretical and experimental findings to show that the use

TABLE IV
GROUP CONTRIBUTION TERMS, τ_w^X ACCORDING TO EQN. 34

R	*X*					
	CH_3	*Cl*	*NO_2*	*OH*	*NH_2*	*COOH*
H	0.54	0.66	−0.14	−0.70	−1.19	−0.13
CH_3	0.60	0.73	−0.11	−0.68	−1.54	−0.11
Cl	0.61	0.64	−0.27	−0.48	−0.91	−0.01
NO_2	0.57	0.53	−0.18	−0.16	−0.60	−0.08
OH	0.56	0.88	0.40	−1.06		−0.15
NH_2	0.19	0.94	0.45			−0.31
COOH	0.56	0.78	−0.07	−0.72	−1.37	
Mean τ_w	0.52	0.74	0.03	−0.63	−1.12	−0.11
Standard deviation	0.14	0.14	0.29	0.29	0.37	0.13

of reversed-phase retention data in the Hildebrand–Scott equation is justified both for solids and liquid non-electrolytes, although, as recognised by others[36], effects such as polymorphism and solid–solid phase interactions can perturb the approach.

Although we have measured κ_w for the non-benzenoid azulene, its aqueous solubility is unknown. Since its κ_w and T_m values are 3.17 and 100°C, using the coefficients of eqn. 5 we predict log X_w to be −5.19. It will be interesting to learn of the observed value at some future time.

The mono and 1,4-disubstituted benzenes studied (Table I) provide us with a data set from which estimation of functional group contributions to κ_w can be made, with five, six or seven independent estimates being possible (Table IV). Defining an extrapolated group contribution term, τ_w, as[17]

$$\tau_w^X = \kappa_{w_j} - \kappa_{w_i} \tag{34}$$

where j and i are solutes having substituents X and R and only R, respectively, it is found that

$$\tau_w^X = 0.93\,\pi + 0.07 \; (n = 6, r = 0.986) \tag{35}$$

where π is a hydrophobic group constant[14] obtained from water–octan-1-ol distribution coefficients (Table II). Tentatively, it is suggested that the use of τ_w values can be used to estimate liquid–liquid distribution (eqn. 30), aqueous solubilities (eqn. 18) and, in conjunction with equations of the form 12, solute retention using binary eluents.

The success of this use of liquid chromatographic retention data to help estimate aqueous solubilities has encouraged us to develop further this approach, and we are currently examining environmental effects (*i.e.*, temperature and ionic

strength) and the effects of co-solvents and co-solutes on the viability of the approach to predict the solubility of drug molecules.

A preliminary report of this work has been made[44].

ACKNOWLEDGEMENTS

We acknowledge initial discussions with Dr. J. C. Kraak, University of Amsterdam, technical assistance from F. Smedes and computational assistance from P. C. van Krimpen. The interest shown by Dr. S. H. Yalkowsky in this work has been encouraging to us.

REFERENCES

1 M. Th. M. Tulp and O. Hutzinger, *Chemosphere*, No. 10 (1978) 849.
2 F. Weimer and J. M. Prausnitz, *Hydrocarbon Process.*, 44 (1965) 237.
3 C. M. Hansen, *Ind. Eng. Chem., Prod. Res. Develop.*, 8 (1969) 2.
4 C. Tsonopoulos and J. M. Prausnitz, *Ind. Eng. Chem., Fundam.*, 10 (1971) 593.
5 P. J. Leinonen and D. Mackay, *Can. J. Chem. Eng.*, 51 (1973) 230.
6 R. A. Pierotti, *Chem. Rev.*, 76 (1976) 717.
7 T. Higuchi, F. M. Shih, T. Kimura and J. H. Rytting, *J. Pharm. Sci.*, 68 (1979) 1267.
8 J. Hildebrand and R. L. Scott, *The Solubility of Non-electrolytes*, Reinhold, New York, 1950.
9 W. J. Moore, *Physical Chemistry*, Longman, London, 5th ed., 1972, p. 249.
10 S. H. Yalkowsky and S. C. Valvani, *J. Chem. Eng. Data*, 24 (1979) 127.
11 S. H. Yalkowsky and S. C. Valvani, *J. Pharm. Sci.*, 69 (1980) 912.
12 C. Hansch and F. Helmer, *J. Polym. Sci.*, 6 (1968) 3295.
13 N. A. Armstrong, K. C. James and C. K. Wong, *J. Pharm. Pharmacol.*, 31 (1979) 627.
14 R. F. Rekker, *The Hydrophobic Fragmental Constant*, Elsevier, Amsterdam, 1977.
15 T. Braumann and L. H. Grimme, *J. Chromatogr.*, 206 (1981) 7.
16 R. Shaw, M. Rivetna and W. H. Elliott, *J. Chromatogr.*, 202 (1980) 347.
17 E. Tomlinson, H. Poppe and J. C. Kraak, *Int. J. Pharmaceutics*, 7 (1981) 225.
18 D. Palalikit and J. H. Block, *Anal. Chem.*, 52 (1980) 624.
19 C. M. Riley, E. Tomlinson and T. M. Jefferies, in A. M. Lawson, C. K. Lim and W. Richmond (Editors), *Current Developments in the Clinical Applications of HPLC, GC and MS*, Academic Press, London, 1980, Ch. 3, p. 35.
20 Cs. Horváth, W. Melander, I. Molnár and P. Molnár, *Anal. Chem.*, 49 (1977) 2295.
21 W. A. Wakeham, *J. Chem. Soc., Farad. Trans.*, (1981) in press.
22 J. A. Graham and L. B. Rogers, *J. Chromatogr. Sci.*, 18 (1980) 614.
23 Cs. Horváth, W. Melander and I. Molnár, *J. Chromatogr.*, 125 (1976) 129.
24 J. Novák, in Z. Deyl, K. Macek and J. Janák (Editors), *Liquid Column Chromatography —A Survey of Modern Techniques and Applications*, Elsevier, Amsterdam, Oxford, New York, 1975, p. 45.
25 K. Denbigh, *The Principles of Chemical Equilibrium*, Cambridge University Press, Cambridge, 1971, p. 432.
26 L. R. Snyder, J. W. Dolan and J. R. Gant, *J. Chromatogr.*, 165 (1979) 3.
27 S. H. Unger and T. F. Feuerman, *J. Chromatogr.*, 176 (1979) 426.
28 W. R. Melander, J. Stoveken and Cs. Horváth, *J. Chromatogr.*, 199 (1980) 35.
29 A. Nahum and Cs. Horváth, *J. Chromatogr.*, 203 (1981) 53.
30 N. Tanaka, H. Goodell and B. L. Karger, *J. Chromatogr.*, 158 (1978) 233.
31 C. M. Riley, E. Tomlinson and T. M. Jefferies, *J. Chromatogr.*, 185 (1979) 197.
32 P. J. Schoenmakers, H. A. H. Billiet and L. de Galan, *J. Chromatogr.*, 185 (1979) 179.
33 R. C. Weast (Editor), *Handbook of Chemistry and Physics*, CRC Press, Cleveland, OH, 60th ed., 1979/1980.
34 E. Martin, S. H. Yalkowsky and J. E. Wells, *J. Pharm. Sci.*, 68 (1979) 4273.
35 C. Hansch and A. Leo, *Substituent Constants for Correlation Analysis in Chemistry and Biology*, Wiley, New York, 1979.

36 S. H. Yalkowsky and W. Morozowich, in E. J. Ariens (Editor), *Drug Design*, Vol. IX, Academic Press, New York, 1980, pp. 122–183.
37 M. J. Harris, T. Higuchi and J. H. Rytting, *J. Phys. Chem.*, 77 (1973) 2694.
38 T. Bultsma, *Eur. J. Med. Chem.*, 15 (1980) 371.
39 R. Kühne, K. Bocek, P. Scharfenberg and R. Franke, *Eur. J. Med. Chem.*, 16 (1980) 7.
40 R. B. Hermann, *J. Phys. Chem.*, 76 (1972) 2754.
41 G. L. Amidon, S. H. Yalkowsky and S. Leung, *J. Pharm. Sci.*, 63 (1974) 1858.
42 S. C. Valvani, S. H. Yalkowsky and G. L. Amidon, *J. Phys. Chem.*, 80 (1976) 829.
43 D. C. Locke, *J. Chromatogr. Sci.*, 12 (1974) 433.
44 T. L. Hafkenscheid and E. Tomlinson, *Int. J. Pharmaceutics*, 8 (1981) 331.