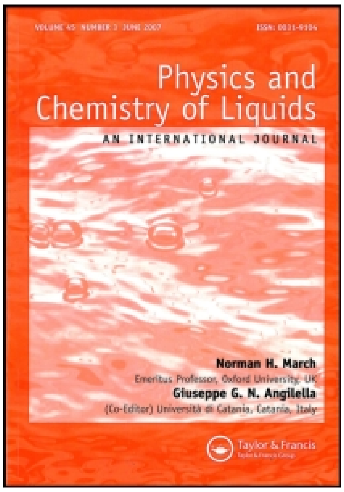

Physics and Chemistry of Liquids
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713646857

Correlation and Estimation of Aqueous Solubilities of Halogenated Benzenes

Ronald J. Baker[a]; Brian J. Donelan[a]; Lesa J. Peterson[a]; William E. Acree Jr[a]; Chun-Che Tsai[a]
[a] Department of Chemistry, Kent State University, Kent, Ohio, USA

To cite this Article Baker, Ronald J. , Donelan, Brian J. , Peterson, Lesa J. , Acree Jr, William E. and Tsai, Chun-Che(1987) 'Correlation and Estimation of Aqueous Solubilities of Halogenated Benzenes', Physics and Chemistry of Liquids, 16: 4, 279 — 292

To link to this Article: DOI: 10.1080/00319108708078530

URL: http://dx.doi.org/10.1080/00319108708078530

Phys. Chem. Liq., 1987, Vol. 16, pp. 279-292
Photocopying permitted by license only

Printed in the United Kingdom

Correlation and Estimation of Aqueous Solubilities of Halogenated Benzenes

RONALD J. BAKER, BRIAN J. DONELAN, LESA J. PETERSON, WILLIAM E. ACREE Jr and CHUN-CHE TSAI

Department of Chemistry, Kent State University, Kent, Ohio 44242, USA

(Received 2 October 1986)

Quantitative structure-activity relationships have been studied for the aqueous solubility of 42 halogenated benzenes (HBs) at 25°C. Multiple regression correlations of aqueous mole fraction solubility (X) with molecular size descriptors (total molecular surface area, total group surface area, molecular volume or molecular weight), melting point temperature (MP) and a combined molecular symmetry and structure descriptor ($R\cdot$Molecular size descriptor) are reported. A generalized correlation takes the form of

$$\log X = a_0 + a_1\,(\text{Molecular size descriptor})^n + a_2\,(\text{MP} - 25) + a_3\,(\text{R}\cdot\text{Molecular size descriptor})$$

in which $1 \leq n \leq 2$, a_i refers to the ith regression coefficient and R is a molecular redundancy index. These correlations enable the aqueous solubility to be estimated from melting point temperature and structural information. The best correlation found, from the standpoint of reduced standard deviation and calculational simplicity, does not contain a melting point term ($a_2 = 0$). Agreement between observed and estimated solubilities is quite good. More importantly, the correlations did provide very reasonable predictions for the 7 fluorobenzenes which were measured as part of this investigation, but not included in the regression analysis. The correlations may also be useful in screening experimental solubility for which serious discrepancies exist in the literature.

INTRODUCTION

Recently several workers have addressed the problem of correlating and predicting the solubilities of polycyclic aromatic hydrocarbons (PAHs), halogenated benzenes (HBs), polychlorobiphenyls (PCBs) and other nonelectrolytes in aqueous media[1-11]. Development of a model which provides a semi-quantitative means of predicting the solubility of a given aromatic hydrocarbon is of particular importance considering the well documented carcinogenic activity of this class of compounds. Our initial study[1] involving 25 polycyclic aromatic hydrocarbons documented thirty multiple regression correlations between aqueous mole fraction solubility (X) and molecular size descriptor (molecular surface area, molecular volume or molecular weight), melting point temperature (MP), and molecular symmetry and structure descriptor ($R \cdot r_d$). A general correlation takes the form of

$$\log X = a_0 + a_1 \,(\text{Molecular size descriptor}) + a_2\,(MP - 25) + a_3\,(R \cdot r_d)$$

in which a_i refers to the ith regression coefficient, R is a molecular redundancy index and r_d is the summation of interatomic distance within a molecule. These correlations enabled estimation of aqueous PAH solubility to within a factor of two in most cases employing only melting point temperatures and structural information for input data. More importantly, the correlations did provide very reasonable predictions for 10 PAHs not included in the original regression analysis.

Successful prediction of PAH solubilities suggests that the generalized correlation expression may be applicable to subsets of substituted aromatic hydrocarbons, with minor modifications needed to take into account the chemical nature of the substituent atoms or fragment molecular groups. To pursue this idea further, we extend our earlier considerations to include halogenated benzenes. This class of compounds was selected primarily because (1) the configurational arrangement of individual atoms within a molecule is fixed by the planar rigid structure (2) various structural isomers provide the opportunity to study the aqueous solubility as a function of molecular symmetry and (3) the substituent groups (which in this case are the halogen atoms) are nonpolar in nature and interactions between the dissolved solute and surrounding water molecules should not involve hydrogen bonding. Furthermore, experimental solubilities are available for 35 different HBs in the literature. The published data set unfortunately contains only four fluorobenzenes and a limited number of "mixed" halogenated benzenes. For this reason we have determined experimental solubilities for an additional seven fluorobenzenes and results from these measurements are used to test the predictive ability of our newly-derived correlations.

MATERIALS AND METHODS

1-Fluoro-4-iodobenzene, 1-chloro-3-fluorobenzene, 1-chloro-2-fluorobenzene, 1-bromo-2-fluorobenzene, 1-bromo-3-fluorobenzene, 1,2,3,5-tetrafluorobenzene, and 1,2,4,5-tetrafluorobenzene were used as received from commercial sources without further purification. Triply distilled water and the various solutes were placed in amber glass bottles and allowed to equilibrate in a constant temperature shaker bath at 25.00 ± 0.05°C for several days. Aliquots of the equilibrated solutions were transferred into a tared volumetric flask and then diluted quantitatively with methanol. Concentrations were determined spectrophotometrically in the UV region.

Mole fraction solubilities at 25°C for the remaining 35 halogenated benzenes considered in this study were obtained from the literature[11]. Total molecular surface areas (TSA) and total halogen surface areas (TSAX) were computed using the program SAREA[12]. Van der Waals molecular volumes (V_w) computed by the program MOLV[13], which uses an iterative summation of discrete increments within the confines of a given molecular unit. As it was desired to compare the viability of each descriptor, atomic coordinates used as input for TSA, TSAX and V_w calculations were derived, whenever possible, from x-ray or neutron diffraction studies[14–25]. Van der Waals radii employed in both surface and volume calculations were 1.70 Å for carbon, 1.20 Å for hydrogen, 1.80 Å for chlorine, 1.35 Å for fluorine, 1.95 Å for bromine and 2.15 Å for iodine.

As in an earlier paper[1], we compute the molecular redundancy index (R) from the following formula

$$R = \frac{\sum_{i=1}^{k} n_i \log n_i}{N \log N}$$

where n_i is the number of atoms in the ith atom set, i is the number of different atom sets, and $N = \sum_{i=1}^{k} n_i$ is the total number of atoms in the molecule. Considering as an example the 1-fluoro-4-iodobenzene molecule (Figure 1), the summation extends over the four types of non-equivalent carbon atoms (C_α, C_β, C_I, C_F), the fluorine atom, the iodine atom, and the two types of non-equivalent hydrogen atoms (H_α, H_β).

$$\begin{aligned} R &= (n_{C_\alpha} \log n_{C_\alpha} + \eta_{C_\beta} \log n_{C_\beta} + n_{C_I} \log n_{C_I} + n_{C_F} \log n_{C_F} + n_I \log n_I \\ &\quad + n_F \log n_F + n_{H_\alpha} \log n_{H_\alpha} + n_{H_\beta} \log n_{H_\beta})/(12 \log 12) \\ &= (2 \log 2 + 2 \log 2 + 1 \log 1 + 1 \log 1 + 1 \log 1 + 1 \log 1 \\ &\quad + 2 \log 2 + 2 \log 2)/(12 \log 12) \\ &= 0.1860 \end{aligned}$$

We also consider as part of this QSAR study, the molecular size discriptor based on the summation of interatomic distances, r_d. For a molecule composed of j atoms, r_d is computed *via*

$$r_d = \sum_{i} \sum_{j>i} r_{ij}$$

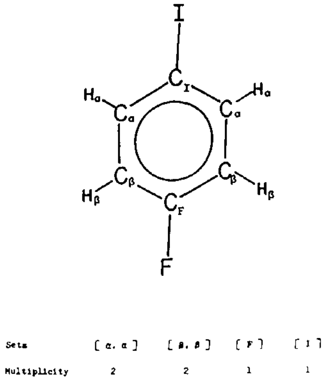

Figure 1 Decomposition of the 1-fluoro-4-iodobenzene molecule into sets of topologically equivalent atoms (C_α, C_β, C_F, C_I, H_α, H_β, F, I) and their multiplicities.

in which r_{ij} is the interatomic distances between any two atoms in the structure. Such a descriptor was found to be particularly useful in distinguishing structural isomers in our earlier paper dealing with the correlation and estimation of aqueous solubilities of polycyclic aromatic hydrocarbons.

Following the computation and tabulation of the molecular and thermodynamic input data (listed in Table 1) multiple regression analysis was performed using the program HANSCH3X[28]. In this study the logarithm of aqueous solubility ($\log X$) was examined in relation to various calculated molecular size descriptors, including the molecular weight (MW), TSA, TSAX, V_w and r_d. Additional terms considered included the thermodynamic (MP-25) and symmetry (R) correction terms along with various geometric combinations of the symmetry and molecular size descriptors. Regression equations were limited to contain no more than four terms.

RESULTS AND DISCUSSION

Correlations (Eqs 1–34) exhibiting significant statistical merit are presented in Table 2. The criteria employed in the decision to report a given correlation expression were: first the equations reported were required to exhibit favorable standard deviation(s) and correlation coefficients and second that the 95 percent confidence interval of the individual regression coefficients could not be inordinately large with respect to the coefficients themselves. In many instances the regression

Table 1 Parameters Used in the Derivation of Correlation Equations.

									log X	
Compound	log X[b]	MW	MP[d] (°C)	R	TSA[g] (nm^2)	TSAX (nm^2)	V_w (nm^3)	r_d (nm)	Eq. 18	Eq. 31
Benzene	−3.39	78.11	liq[e]	0.7211	1.050	0.000	0.0800	17.80	−3.34	−3.25
Fluorobenzene	−3.53	96.10	liq	0.1860	1.100	0.136	0.0829	18.07	−3.52	−3.45
Chlorobenzene	−4.09	112.56	liq	0.1860	1.240	0.307	0.0953	18.48	−4.02	−3.99
Bromobenzene	−4.38	157.01	liq	0.1860	1.355	0.360	0.1010	18.59	−4.45	−4.45
Iodobenzene	−4.69	204.01	liq	0.1860	1.438	0.456	0.1118	18.78	−4.81	−4.82
1,2-Difluorobenzene	−3.74	114.09	liq	0.2789	1.142	0.270	0.0859	18.33	−3.70	−3.67
Dichlorobenzene	−4.94	147.00	liq	0.2789	1.397	0.584	0.1111	19.09	−4.71	−4.77
Dibromobenzene	−5.24	235.91	liq	0.2789	1.505	0.704	0.1223	19.38	−5.19	−5.29
Diiodobenzene	−5.99	329.91	27.0	0.2789	1.659	0.872	0.1419	19.78	−5.93	−6.09
1,3-Difluorobenzene	−3.74	114.09	liq	0.1860	1.161	0.270	0.0876	18.58	−3.74	−3.70
Dichlorobenzene	−4.83	147.00	liq	0.1860	1.442	0.614	0.1121	19.40	−4.85	−4.88
Dibromobenzene	−5.12	235.91	liq	0.1860	1.555	0.746	0.1242	19.64	−5.36	−5.43
Diiodobenzene	−6.31	329.91	40.4	0.1860	1.731	0.941	0.1442	20.04	−6.28	−6.34
1,4-Difluorobenzene	−3.71	114.09	liq	0.4649	1.166	0.270	0.0877	18.63	−3.82	−3.83
Dichlorobenzene	−4.95	147.00	53.1	0.4649	1.424	0.616	0.1122	19.50	−5.04	−5.08
Dibromobenzene	−5.81	235.91	87.3	0.4649	1.559	0.745	0.1252	19.70	−5.82	−5.78
Diiodobenzene	−6.99	329.91	131	0.4649	1.737	0.942	0.1446	20.09	−6.93	−6.79
1,2,3-Trifluorobenzene[a]		132.08	liq	0.1860	1.202	0.404	0.0905	18.81	−3.90	−3.88
Trichlorobenzene	−5.50	181.45	53	0.1860	1.582	0.869	0.1276	20.00	−5.61	−5.59
Tribromobenzene[a]		314.80	87.8	0.1860	1.725	1.031	0.1443	20.38	−6.47	−6.34
Triiodobenzene[a]		455.80	116	0.1860	1.946	1.270	0.1737	20.96	−7.79	−7.61
1,2,4-Trifluorobenzene[a]		132.08	liq	0.0000	1.205	0.405	0.0900	18.85	−3.84	−3.77
Trichlorobenzene	−5.46	181.45	liq	0.0000	1.624	0.912	0.1286	20.15	−5.55	−5.53
Tribromobenzene	−6.24	314.80	44	0.0000	1.767	1.082	0.1462	20.43	−6.33	−6.25
Triiodobenzene[a]		455.80	91.5	0.0000	2.012	1.352	0.1795	21.93	−7.86	−7.64

Table 1 (*Continued*)

Compound	log X^b	MW	MP^d (°C)	R	TSA^g (nm^2)	TSAX (nm^2)	V_w (nm^3)	r_d (nm)	log X	
									Eq. 18	Eq. 31
1,3,5-Trifluorobenzene^a		132.08	liq	0.4421	1.204	0.404	0.0909	18.85	−4.00	−4.05
Trichlorobenzene	−6.18	181.45	63	0.4421	1.619	0.917	0.1279	20.03	−6.04	−6.15
Tribromobenzene	−7.34	314.80	121	0.4421	1.804	1.121	0.1466	20.54	−7.28	−7.24
Triiodobenzene^a		455.80	184.2	0.4421	1.947	1.355	0.1760	21.11	−8.41	−8.20
1,2,3,4-Tetrafluorobenzene^a		150.08	liq	0.2789	1.237	0.539	0.0917	19.01	−4.09	−4.13
Tetrachlorobenzene	−6.44	215.89	47.5	0.2789	1.724	1.149	0.1418	20.57	−6.41	−6.54
Tetrabromobenzene^a		393.70	62.5^f	0.2789	1.927	1.360	0.1650	21.10	−7.59	−7.73
Tetraiodobenzene^a		581.70	136	0.2789	2.214	1.671	0.2031	21.87	−9.68	−9.64
1,2,3,5-Tetrafluorobenzene^a	−4.05^c	150.08	liq	0.1860	1.237	0.539	0.0924	19.06	−4.04	−4.05
Tetrachlorobenzene	−6.53	215.89	54.5	0.1860	1.746	1.179	0.1422	20.39	−6.46	−6.49
Tetrabromobenzene^a		393.70	99	0.1860	1.939	1.401	0.1652	20.84	−7.70	−7.61
Tetraiodobenzene^a		581.70	148	0.1860	2.249	1.740	0.2042	21.61	−9.84	−9.64
1,2,4,5-Tetrafluorobenzene^a	−4.12^c	150.08	liq	0.4649	1.248	0.539	0.0922	19.06	−4.21	−4.33
Tetrachlorobenzene	−7.30	215.89	139	0.4649	1.797	1.209	0.1447	20.79	−7.38	−7.31
Tetrabromobenzene	−8.72	393.70	182	0.4649	1.973	1.418	0.1666	21.15	−8.60	−8.45
Tetraiodobenzene^a		581.70	254	0.4649	2.285	1.759	0.2057	21.93	−10.95	−10.67
Pentafluorobezene^a		168.07	liq	0.1860	1.262	0.674	0.0953	18.99	−4.16	−4.18
Pentachlorobenzene	−7.39	250.33	86	0.1860	1.885	1.431	0.1567	20.99	−7.35	−7.31
Pentabromobenzene^a		472.59	160	0.1860	2.114	1.690	0.1852	21.60	−9.03	−8.75
Pentaiodobenzene^a		707.60	172	0.1860	2.469	2.072	0.2326	22.56	−11.48	−11.28
Hexafluorobenzene^a		186.06	liq	0.7211	1.313	0.816	0.0983	19.30	−4.76	−5.11
Hexachlorobenzene	−9.50	284.78	230	0.7211	2.036	1.678	0.1709	21.85	−9.65	−9.70
Hexabromobenzene^a		551.49	327	0.7211	2.263	1.937	0.2042	22.50	−11.63	−11.45
Hexaiodobenzene^a		833.48	350^f	0.7211	2.745	2.424	0.2647	24.19	−15.51	−15.61

1-Fluoro-4-iodobenzene[a]	−4.87[c]	222.00	liq	0.1860	1.429	0.607	0.1148	19.07	−4.79	−4.83
1-Chloro-2-fluorobenzene[a]	−4.16[c]	130.55	liq	0.0000	1.276	0.429	0.0977	18.76	−4.10	−4.03
1-Chloro-3-fluorobenzene[a]	−4.09[c]	130.55	liq	0.0000	1.281	0.438	0.0989	18.76	−4.12	−4.05
1-Bromo-2-fluorobenzene[a]	−4.44[c]	175.01	liq	0.0000	1.333	0.500	0.1046	18.86	−4.31	−4.25
1-Bromo-3-fluorobenzene[a]	−4.41[c]	175.01	liq	0.0000	1.338	0.507	0.1039	18.86	−4.33	−4.27
1-Bromo-2-chlorobenzene[a]	−4.93	191.46	liq	0.0000	1.470	0.647	0.1187	19.46	−4.87	−4.83
1-Bromo-3-chlorobenzene[a]	−4.95	191.46	liq	0.0000	1.493	0.678	0.1182	19.47	−4.96	−4.93
1-Bromo-4-chlorobenzene[a]	−5.37	191.46	68.0	0.1860	1.505	0.680	0.1194	19.60	−5.31	−5.18
1-Bromo-4-iodobenzene[a]	−6.30	282.91	92.0	0.1860	1.648	0.843	0.1350	19.90	−6.08	−5.90
1-Chloro-2-iodobenzene[a]	−5.28	238.46	liq	0.0000	1.548	0.733	0.1280	19.65	−5.20	−5.17
1-Chloro-3-iodobenzene[a]	−5.29	238.46	liq	0.0000	1.581	0.776	0.1285	19.67	−5.35	−5.33
1-Chloro-4-iodobenzene[a]	−5.77	238.46	57.0	0.1860	1.594	0.779	0.1285	19.80	−5.67	−5.62

[a] Not employed in solubility correlations.

[b] Logarithm of aqueous mole fraction solubility. Taken, except as noted, from Reference [11].

[c] Measured as part of this investigation.

[d] Taken, except where noted, from Reference [26].

[e] Compound is liquid at 25°C, thus (MP − 25) = 0 in correlation expression.

[f] Reference [27].

[g] Structural information from References [14]–[25].

Table 2 Correlation Equations (Equations 1-34).

Eqn	MW $\times 10^2$	MW^2 $\times 10^4$	TSA	TSA^2	V_w	V_w^2	TSAX	MT $\times 10^2$	$R \cdot$MW $\times 10^2$	$R \cdot$TSA	$R \cdot V_w$	$R \cdot r_d$	$R \cdot$TSAX	CONST.	s	r^2	F (P, Q)
(1)	−1.450 (0.427)													−2.50 (0.97)	0.913	0.667	52 (1,26)
(2)	−0.865 (0.310)							−1.69 (0.49)						−3.24 (0.60)	0.534	0.890	102 (2,25)
(3)	−0.290 (0.109)						−2.21 (0.26)	−0.940 (0.161)						−2.99 (0.17)	0.147	0.992	1000 (3,24)
(4)	−0.953 (0.296)							−2.44 (0.79)	0.962 (0.839)					−3.42 (0.58)	0.491	0.911	82 (3,24)
(5)	−0.747 (0.336)							−2.11 (0.71)				0.0671 (0.0852)		−3.74 (0.87)	0.517	0.901	73 (3,24)
(6)		−0.309 (0.103)												−4.08 (0.65)	1.004	0.597	38 (1,26)
(7)		−0.161 (0.794)						−1.80 (0.57)						−4.20 (0.41)	0.625	0.850	71 (2,25)
(8)		−0.056 (0.022)					−2.34 (0.24)	−0.893 (0.166)						−3.23 (0.14)	0.150	0.992	962 (3,24)
(9)		−0.185 (0.079)						−2.57 (0.94)	1.006 (1.023)					−4.46 (0.47)	0.589	0.872	54 (3,24)
(10)		−0.130 (0.081)						−2.37 (0.78)				0.0933 (0.0940)		−4.72 (0.65)	0.588	0.872	55 (3,24)
(11)			−5.55 (0.60)											2.92 (0.94)	0.411	0.932	359 (1,26)
(12)			−4.10 (0.32)					−1.01 (0.16)						1.04 (0.46)	0.151	0.991	1424 (2,25)
(13)			−4.07 (0.29)					−0.689 (0.292)					−0.741 (0.587)	1.06 (0.41)	0.136	0.993	1171 (3,24)
(14)			−4.26 (0.34)					−0.784 (0.277)		−0.347 (0.358)				1.37 (0.55)	0.142	0.992	1065 (3,24)
(15)			−4.29 (0.38)					−0.841 (0.253)				−0.0221 (0.0262)		1.40 (0.61)	0.145	0.992	1027 (3,24)
(16)				−1.85 (0.15)										−1.11 (0.39)	0.314	0.960	632 (1,26)

(17)		−1.45 (0.10)			−0.805 (0.152)					−1.81 (0.21)	0.134	0.993	1806 (2,25)
(18)		−1.44 (0.08)			−0.422 (0.233)				−0.883 (0.458)	−1.75 (0.17)	0.107	0.996	1921 (3,24)
(19)		−1.52 (0.97)			−0.504 (0.237)	−0.445 (0.292)				−1.53 (0.26)	0.115	0.995	1636 (3,24)
(20)		−1.53 (0.11)			−0.591 (0.227)			−0.0264 (0.0222)		−1.54 (0.30)	0.122	0.994	1447 (3,24)
(21)			−57.9 (6.4)							1.53 (0.81)	0.420	0.930	343 (1,26)
(22)			−42.6 (3.1)		−1.03 (0.15)					−0.01 (0.35)	0.141	0.992	1614 (2,25)
(23)			−42.2 (3.7)		−0.743 (0.276)				−0.672 (0.558)	−0.02 (0.32)	0.129	0.994	1301 (3,24)
(24)			−43.2 (3.4)		−0.924 (0.272)		−2.09 (4.36)			0.11 (0.43)	0.142	0.993	1075 (3,24)
(25)			−43.3 (3.8)		−0.969 (0.240)			−0.0086 (0.0252)		0.11 (0.49)	0.143	0.993	1054 (3,24)
(26)				−239 (21)						−1.83 (0.36)	0.335	0.955	553 (1,26)
(27)				−185 (15)	−0.833 (0.178)					−2.39 (0.21)	0.157	0.991	1304 (2,25)
(28)				−184 (13)	−0.474 (0.303)				−0.830 (0.600)	−2.34 (0.19)	0.139	0.993	1121 (3,24)
(29)				−189 (16)	−0.685 (0.315)		−2.75 (4.84)			−2.28 (0.29)	0.156	0.991	883 (3,24)
(30)				−189 (18)	−0.759 (0.284)			−0.0095 (0.0281)		−2.30 (0.33)	0.159	0.991	851 (3,24)
(31)		−1.49 (0.09)							−1.59 (0.29)	−1.61 (0.19)	0.131	0.993	1891 (2,25)
(32)				−191 (15)					−1.63 (0.36)	−2.20 (0.19)	0.163	0.990	1218 (2,25)
(33)	−4.26 (0.38)								−1.94 (0.40)	1.41 (0.53)	0.187	0.987	915 (2,25)
(34)			−44.3 (4.0)						−1.98 (0.40)	0.32 (0.44)	0.191	0.986	883 (2,25)

statistics could be favourably affected by augmentation of the data set to include more compounds. However, in the interests of evaluating the effectiveness of the molecular and thermodynamic descriptors as predictive tools we elected to determine regression coefficients based on only 28 of the HBs appearing in Table 1. These linear and multilinear correlations were in turn utilized to predict the solubilities of the remaining 14 compounds. Comparisons between the experimental and calculated values are listed in the last two columns of Table 1 for Eqs 18 and 31. For the majority of compounds agreement between experimental and predicted solubility is quite good; and in all instances well within one order of magnitude and, in most instances, differing by no more than a factor of two. Furthermore, the experimental uncertainty and reproducibility associated with solubility measurements of slightly soluble compounds is comparable in magnitude to the standard deviations of many of our correlations. For example, the experimental values of Miller *et al.*[10] for several chlorobenzenes differ from those used in this study by as much as a factor of 3 or 0.477 when expressed as log X. The solubility data set of Miller *et al.*[10] was not included in the regression analysis because we hoped to eliminate discrepancies in experimental conditions as much as possible by taking solubility data from a single laboratory. Out of curiosity, however, we did rerun the regression analysis using both sets of solubility data and the second set of correlations was nearly identical to the ones presented in Table 2.

Examination of Table 2 reveals that the better two parameter correlations combine the melting point contribution with either the molecular surface area

$$\log X = 1.04 - 4.10(\mathrm{TSA}) - 0.0101(\mathrm{MP\text{-}25}) \qquad \text{(Eq. 12)}$$
$$s = 0.151 \quad r^2 = 0.991 \quad F(2{,}25) = 1424$$

$$\log X = -1.81 - 1.45(\mathrm{TSA})^2 - 0.00805(\mathrm{MP\text{-}25}) \qquad \text{(Eq. 17)}$$
$$s = 0.134 \quad r^2 = 0.993 \quad F(2{,}25) = 1806$$

or molecular volume descriptor.

$$\log X = -0.01 - 42.6(V_w) - 0.0103(\mathrm{MP\text{-}25}) \qquad \text{(Eq. 22)}$$
$$s = 0.141 \quad r^2 = 0.992 \quad F(2{,}25) = 1614$$

$$\log X = -2.39 - 185(V_w)^2 - 0.00833(\mathrm{MP\text{-}25}) \qquad \text{(Eq. 27)}$$
$$s = 0.157 \quad r^2 = 0.991 \quad F(2{,}25) = 1304$$

An interesting discovery which was not originally anticipated based on our earlier studies[1] is the dissimilar results obtained for molecular weight correlations. Molecular surface area (Eq. 12) and molecular volume (Eq. 22) correlations are comparable in predictive accuracy with standard deviations of 0.151 and 0.141, respectively, and both are far superior to the corresponding 2 parameter molecular weight expression (Eq. 2) which has a standard deviation of 0.534.

Molecular weight correlations could be significantly improved by introducing a second molecular size descriptor

$$\log X = -2.99 - 0.00290(\mathrm{MW}) - 0.00940(\mathrm{MP}{-}25) - 2.21(\mathrm{TSAX}) \qquad \text{(Eq. 3)}$$

$$s = 0.147 \quad r^2 = 0.992 \quad F(3,24) = 1000$$

Despite the reduced standard deviation, the usefullness of Eq. 3 must be questioned since comparable, and in many cases better, predictions can be made with either of the three term molecular surface area or molecular volume correlations given in Table 2. In addition, computational time is not saved as Eq. 3 requires that the halogen group molecular surface area be calculated. Similarly, standard deviations for the three parameter molecular surface area and molecular volume correlations are in general lower than their two parameter counterparts. Reduced standard deviations, however, must far outweigh the additional time required for calculation if the expression is to offer practical advantages.

The best overall correlation we found contains three parameters.

$$\log X = -1.75 - 1.44(\mathrm{TSA})^2 - 0.00422(\mathrm{MP}{-}25) - 0.883(R \cdot \mathrm{TSAX}) \qquad \text{(Eq. 18)}$$

$$s = 0.107 \quad r^2 = 0.996 \quad F(3,24) = 1921$$

Very good predictions could still be obtained if the melting point term was removed.

$$\log X = -1.61 - 1.49(\mathrm{TSA})^2 - 1.59(R \cdot \mathrm{TSAX}) \qquad \text{(Eq. 31)}$$

$$s = 0.131 \quad r^2 = 0.993 \quad F(2,25) = 1891$$

It is difficult based on our regressional data set to conclude if the removal of the melting point term is significant since only 14 out of the 28 compounds are crystalline at 25°C. From a thermodynamic point of view, the melting point term is required to properly describe the solubility of crystalline compounds. Elimination of this term should not be construed as a complete departure from thermodynamics, but rather an attempt to predict a compound's melting point temperature from structural information, thereby replacing a_2 (MP–25) in our generalized correlation expression with a linear combination of molecular descriptors. As shown in Table 3, there is a high correlation between

Table 3 Squared Correlation Coefficients

	MW	MW^2	TSA	TSA^2	V_w	V_w^2	TSAX	MT	$R \cdot$ MW	$R \cdot$ TSA	$R \cdot V_w$	$R \cdot$ TSAX	$R \cdot r_d$
MW	1.00												
MW^2	0.97	1.00											
TSA	0.73	0.62	1.00										
TSA^2	0.72	0.63	0.99	1.00									
V_w	0.77	0.67	0.99	0.99	1.00								
V_w^2	0.77	0.68	0.98	0.99	0.99	1.00							
TSAX	0.60	0.49	0.96	0.97	0.95	0.95	1.00						
MT	0.35	0.35	0.52	0.58	0.51	0.57	0.54	1.00					
$R \cdot$ MW	0.39	0.41	0.35	0.39	0.37	0.41	0.33	0.79	1.00				
$R \cdot$ TSA	0.11	0.12	0.17	0.21	0.17	0.21	0.19	0.68	0.84	1.00			
$R \cdot V_w$	0.14	0.15	0.21	0.25	0.21	0.26	0.22	0.72	0.87	0.99	1.00		
$R \cdot$ TSAX	0.28	0.26	0.48	0.53	0.48	0.53	0.53	0.87	0.82	0.79	0.83	1.00	
$R \cdot r_d$	0.02	0.03	0.04	0.06	0.04	0.06	0.05	0.46	0.67	0.93	0.90	0.55	1.00

MP-25 and $R\cdot$TSAX (squared correlation coefficient of 0.87). Predictive expressions requiring *a priori* knowledge of the compound's melting point temperature are of no use if the required information is not readily available. Equation 31, on the other hand, requires only structural information which can be estimated, if need be, assuming a rigid and planar structure.

Results from this study indicate that our generalized correlation expression should take the form

$$\log X = a_0 + a_1\,(\text{Molecular size descriptor})^n + a_2\,(\text{MP-25}) + a_3\,(R\cdot\text{Molecular size descriptor})$$

with the total group molecular surface area being another very viable molecular size descriptor and $1 \leq n \leq 2$. The 34 correlations listed in Table 2 are specific one, two, and three-parameter expressions derived from our generalized form. As one might surmise from this study, the selection of predictive expression depends, to a large extent, on the predictive accuracy desired and the effort that one is willing to spend.

Acknowledgement

The authors thank Cheryl Judy and Ferdinand Fojas for their assistance. Support from the National Institutes of Health, US Environmental Protection Agency, and The Donors of the Petroleum Research Fund (administered by the American Chemical Society) is gratefully acknowledged.

References

1. R. J. Baker, W. E. Acree, Jr. and C.-c. Tsai, *Quant. Struct.-Act. Relat.*, **3**, 10 (1984).
2. D. Mackay and W. Y. Shiu, *J. Chem. Eng. Data.*, **22**, 399 (1977).
3. S. C. Valvani, S. H. Yalkowsky and G. L. Amidon, *J. Phys. Chem.*, **80**, 829 (1976).
4. S. H. Yalkowsky and S. C. Valvani, *J. Chem. Eng. Data*, **24**, 127 (1979).
5. S. H. Yalkowsky and S. C. Valvani, *J. Pharm. Sci.*, **69**, 912 (1980).
6. G. L. Amidon and S. T. Anik, *J. Chem. Eng. Data*, **26**, 28 (1981).
7. R. P. Eganhouse and J. P. Calder, *Geochim. Cosmochim. Acta*, **40**, 555 (1976).
8. S. T. Anik, PhD Dissertation, University of Wisconsin, Madison (1978).
9. D. Mackay, R. Mascarenhaus, W. Y. Shiu, S. C. Valvani and S. H. Yalkowsky, *Chemosphere*, **9**, 257 (1980).
10. M. M. Miller, S. Ghodbane, S. P. Wasik, Y. B. Tewari and D. E. Martire, *J. Chem. Eng. Data*, **29**, 184 (1984).
11. S. Yalkowsky, R. J. Orr and S. C. Valvani, *Ind. Eng. Chem. Fundam.*, **18**, 351 (1979).
12. R. S. Pearlmann, QCPE Program No. 413, Indiana University.
13. R. J. Baker and C.-c. Tsai, unpublished.
14. E. C. Cox, D. W. J. Cruickshank and J. A. S. Smith, *Proc. Roy. Soc.* (London) Ser. A, **247**, 1 (1958).
15. P. D. Andre, R. Fourme and M. Renaud, *Acta Cryst.*, **B27**, 2371 (1971).
16. G. L. Wheeler and S. D. Colson, *Acta Cryst.*, **B31**, 911 (1975).
17. R. G. Hazell, M. S. Lehmann and G. S. Pawley, *Acta Cryst.*, **B28**, 1388 (1972).
18. H. J. Milledge and L. M. Pant, *Acta Cryst.*, **13**, 285 (1960).

19. P. Marsh and D. E. Williams, *Acta Cryst.*, **B37**, 279 (1981).
20. E. B. Halac, E. M. Burgos, H. Bonadeo and E. A. D'Alessio, *Acta Cryst.*, **A33**, 86 (1977).
21. G. Gafner and F. H. Herbstein, *Acta Cryst.*, **17**, 982 (1964).
22. P. Marsh and D. E. Williams, *Acta Cryst.*, **B37**, 705 (1981).
23. N. Boden, P. P. Davis, C. H. Stam and G. A. Wesselink, *Mol. Phys.*, **25**, 81 (1973).
24. G. M. Brown and O. A. W. Strydom, *Acta Cryst.*, **B30**, 801 (1974).
25. E. Baharie and G. S. Pawley, *Acta Cryst.*, **A35**, 233 (1979).
26. R. C. Weast (ed.), *CRC Handbook of Chemistry and Physics*, 64th edition, CRC Press Inc., Boca Raton, FL. (1983).
27. J. Buckingham (ed.), *Dictionary of Organic Compounds*, 5th edition, Chapman and Hall, New York, NY (1982).
28. C. Hansch, A. Leo, D. A. Elkins and N. D. Dreyer, private communication.