

CHROM. 16,417

HIGH-PERFORMANCE LIQUID CHROMATOGRAPHY OF ALKYL BENZENES

RELATIONSHIP WITH LIPOPHILICITIES AS DETERMINED FROM OCTANOL-WATER PARTITION COEFFICIENTS OR CALCULATED FROM HYDROPHOBIC FRAGMENTAL DATA AND CONNECTIVITY INDICES; LIPOPHILICITY PREDICTIONS FOR POLYAROMATICS

RUDY E. KOOPMANS and ROELOF F. REKKER*

Department of Pharmacochemistry, Vrije Universiteit, De Boelelaan 1083, 1081 HV Amsterdam (The Netherlands)

(First received October 14th, 1983; revised manuscript received November 3rd, 1983)

SUMMARY

High-performance liquid chromatography (HPLC) was used to study the relationships between the retention and hydrophobicity of a series of alkylbenzenes. One of the aims was a critical comparison of the use of the logarithms of experimentally determined partition coefficients ($\log P$) (octanol-water system), calculated $\log P$ (applying the Rekker fragment system), calculated $\log P$ (applying the Leo-Hansch fragment system) or Kier's connectivity indices. The conclusions are that the Rekker system gives a good description (both relative and absolute) of HPLC hydrophobicities, that the Leo-Hansch system lacks absolute prediction in a number of instances and that the Kier indices can be regarded as reasonable substitutes in the description of the hydrophobicity of alkylbenzenes. Lipophilicities of polyaromatic structures can be calculated easily and with surprisingly high accuracy from Rekker's f constants.

INTRODUCTION

In recent years there has been an increasing tendency to apply reversed-phase thin-layer chromatography (RPTLC) and high-performance liquid chromatography (HPLC) in studies of the lipophilic properties of organic structures. Workers concerned with both pharmaceutical quantitative structure-activity relationships (QSAR) and environmental problems have become increasingly interested in these techniques.

The advantages of RPTLC and HPLC over the shaking-flask method, mostly applied to solvent partition, are evident: they are rapid, accurate and there is no objection to the use of impure substances; the disadvantages are the limited lipophilicity range that can be covered by one elution composition and the frequently

overlooked demand for structural congenericity in an investigated series of compounds. The concept "congeneric" series is not so easy to define correctly; literally it means "of the same kind, class or stock". In many instances non-congenericity may be hidden in a series of compounds consisting of outwardly congeneric-looking structures with an inconsistent partition behaviour towards solvent systems with dissimilar phase polarities. Many investigators hold the view that RPTLC on silica gel with acetone-water or methanol-water as the eluent and HPLC on a C₁₈ column with methanol-water as the eluent offers sufficient guarantees to accept a satisfactory congenericity, *i.e.*, phase non-polarity can be completely neglected.

In this investigation we chose an HPLC study on a small series of alkylbenzenes that are completely congeneric. Some earlier studies on a range of alkylbenzenes with differing substitution patterns and with varying alkyl groups are available. Sleight¹ correlated log (capacity factor) with the total number of aliphatic carbon atoms. Colin and Guiochon² obtained close correlations between log (capacity factor) and the number of carbon atoms for *n*-alkane, *n*-alkylbenzene and normal fatty acid methyl ester homologous series. In a fairly extensive study, Smith³ compared the separations of a number of alkylbenzenes on three different reversed-phase columns (SAS-Hypersil, ODS-Hypersil and C₂₂-Magnusil) and correlated log (capacity factors) with the number of carbon atoms, Kier's connectivity index, hydrophobic surface areas and experimentally determined log *P* values. Jinno and Kawasaki⁴ correlated retention data of isomeric alkylbenzenes with log *P*, van der Waals volume and connectivity index. In general, all of these experiments give an indication that reasonable correlations can be found between log *k'* values and a suitable physical parameter indicating size, topological shape or hydrophobicity of the solute molecules. Experimental log *P* values commonly seem to be the least suitable for obtaining high-quality correlations. Thus, Jinno and Kawasaki⁴ obtained for 19 alkylbenzenes with log *P* as the independent parameter a correlation coefficient not higher than 0.919 on a Jasco FineSil C₁₈ column.

As the main purpose of this work was the acquisition of rigid information on the appropriate parametrization of HPLC data, we decided to correlate log *k'* values with the following:

- (a) Experimentally determined log *P*_{o/w} values from the literature;
- (b) Calculated log *P* values, making use of Rekker's hydrophobic fragmental system⁵, the basis for which is the equation

$$\log P = \sum_1^n a_n f_n \quad (1)$$

where *f* is the hydrophobic fragmental constant, the lipophilicity contribution of a constituent part of a structure to the total lipophilicity, and *a* is a numerical factor indicating the incidence of a given fragment in the structure. In 1979 the system was extended to a series of about 100 *f* values, based on a data file consisting of about 1000 log *P* values from the octanol-water system⁶. Since the available data set represents a good arbitrary cross-section of organic compounds and assuming that both the fragmentation procedure and fragment computation are performed correctly, the *f* values obtained will be suitable for log *P* predictions with comparably accurate

results, *i.e.*, with contributions to the correlation coefficient as high as 0.996 (the *r* finally attained in the fragment computation).

(c) Calculated log *P* values making use of Leo and Hansch's hydrophobic fragmental system^{7,8}. If one defines Rekker's fragmental system as a reduction approach in deriving an *f* system from a large data set accompanied by strong statistical criteria, then the Leo-Hansch approach should be regarded as a constructional approach: it starts from a few carefully measured log *P* values and construes the system step by step, duly applying correction factors as soon as they seem necessary. Basic fragmental *f* values are those of H and C (0.23 and 0.20, respectively) and the first correction presents itself as soon as a carbon chain is formed: this corrective "bond factor", *F*_b, amounts to -0.12 per bond, neglecting, however, the first bond.

(d) Kier connectivity indices; Kier and Hall's connectivity method⁹ may be considered as a *de novo* approach as it expresses connections between atoms in topological terms and not in forces, properties or other physico-chemical type quantities. The parameter χ includes information on size, branchings, cyclization, unsaturated character and presence of hetero atoms. Since its introduction, Kier's method has found wide application in different types of QSAR work¹⁰⁻¹².

EXPERIMENTAL

The compounds investigated were of several origins from laboratory stock. They were all of sufficient purity with one main peak in the chromatograms.

Methanol (absolute), acetonitrile-water (15:85) and triethylamine-water (1%) were investigated with regard to their elution properties, but we finally selected methanol-water (70:30) prepared from analysed-reagent grade methanol obtained from J. T. Baker (Deventer, The Netherlands) and doubly distilled water. The eluent was degassed before use by drawing a water-pump vacuum for 5 min with stirring.

Retention times were determined on a Waters Assoc. (Milford, MA, U.S.A.) HPLC apparatus equipped with a Waters Assoc. Model 440 UV absorbance detector (254 nm). The column used throughout was RP type (μ Bondapak C₁₈; Waters Assoc.). Sample sizes were chosen so as to give a good detection signal from 10 μ l of injected sample (variations between 10 and 100 ppm).

Retention times were expressed in terms of log (capacity factors, *k'*) by

$$\log k' = \log (t_R - t_0)/t_R$$

where *t_R* represents the retention time of the compound and *t₀* the elution time of an unretained peak generated by formamide (J. T. Baker, analysed quality, concentration 100 μ l/ml in methanol). The recorder chart speed was 10 mm/min and the flow-rate of the eluent was 1.5 ml/min (2000 p.s.i.) or 2.3 ml/min (3000 p.s.i.).

RESULTS AND DISCUSSION

The experimental log *P*_{oct} values (from the literature), *t_R*, log *k'* and log *P*_{calc} values calculated according to Rekker (Σf_R) are collected in Table I for ten methyl-substituted benzene derivatives, including benzene. The *t_R* values are averages from at least three (sometimes five or six) measurements, mostly performed on different days. The eluent was methanol-water (70:30), the flow-rate 2.3 ml/min and *t₀* = 1.30 min.

TABLE I
LIPOPHILICITIES AND HPLC DATA FOR METHYLBENZENES

Substituents	Log P_{obs} *	t_R	Log k'	Σf_R	Σf_R^{est}	Δ
H	2.13	2.30	-0.114	2.02	1.93	0.09
Methyl	2.69	2.93	0.099	2.54	2.51	0.03
1,2-dimethyl	3.12	3.80	0.284	3.06	3.01	0.05
1,3-dimethyl	3.20	4.01	0.319	3.06	3.11	-0.05
1,4-dimethyl	3.15	4.07	0.328	3.06	3.13	-0.07
1,2,3-trimethyl	n.m.**	5.13	0.470	3.58	3.52	0.06
1,2,4-trimethyl	n.m.	5.54	0.513	3.58	3.64	-0.06
1,3,5-trimethyl	3.44	5.88	0.547	3.58	3.73	-0.15
1,2,4,5-tetramethyl	4.00	7.89	0.705	4.10	4.16	-0.06
Pentamethyl	n.m.	10.78	0.863	4.62	4.59	0.03
Hexamethyl	4.31***	15.00	1.023	5.14	5.02	0.12

* Selected values; other values available from the literature: benzene 1.56¹³, 2.15^{14,15}; toluene 2.73¹⁶, 2.11¹³, 2.80¹⁴; 1,2-dimethylbenzene 2.77¹⁴.

** n.m., not measured.

*** Apparent experimental outlier.

The relationship between Σf_R and $\log k'$ can be expressed by

$$\Sigma f_R = 2.715 (\pm 0.149)^* \log k' + 2.242 (\pm 0.083)^* \quad (2)$$

$n = 11; r = 0.996; s = 0.085; F = 1089$

The statistics are of excellent quality, better than usually observed for correlations between $\log P$ values and HPLC data; it should be borne in mind, however, that the series studied is completely congeneric.

A comparison of the estimates for Σf_R (Σf_R^{est} in Table I) with the $\log P_{obs}$ values clearly gives an experimental value of 4.31 for hexamethylbenzene as an outlier. The results provide confirmation of what is steadily being realized about solvent partition, that $\log P$ values can be measured in practice with good accuracy only when they remain below 4.0-4.5, possibly because non-centrifugable octanol droplets in the aqueous phase of the partition system will introduce a persistent unreal aqueous concentration in the partitioned solute. This implies that we should reject the concept of Hansch and Leo⁸ that a so-called "pot effect" should be held responsible for a much too low lipophilicity of a number of structures. The term "pot effect" was proposed in "honour" of Δ^9 -tetrahydrocannabinol, which was the first structure for which grossly anomalous partition behaviour was observed; it is explained by holding an aryl -OH or -O- group in combination with with an alkyl chain responsible in that they can easily be solubilized by 10^{-3} M octanol in the aqueous phase at saturation.

Table II gives results for a set of congeneric structures, but differing from those in Table I in that for all of the compounds experimental $\log P$ values are available and that a branched-chain alkyl group (isopropyl) and a slightly longer group (n -

* 95% confidence levels in parentheses.

butyl) have been added. Four regression equations were derived, connecting the $\log k'$ data with $\log P_{\text{obs}}$, Σf_{LH} , Σf_{R} and χ , respectively:

$$\log P_{\text{obs}} = 2.077 (\pm 0.293) \log k' + 2.633 (\pm 0.140) \quad (3)$$

$n = 10; r = 0.977; s = 0.156; F = 169$

$$\Sigma f_{\text{LH}} = 3.299 (\pm 0.458) \log k' + 2.638 (\pm 0.218) \quad (4)$$

$n = 10; r = 0.978; s = 0.243; F = 175$

$$\Sigma f_{\text{R}} = 2.706 (\pm 0.208) \log k' + 2.439 (\pm 0.054) \quad (5)$$

$n = 10; r = 0.993; s = 0.111; F = 568$

$$\chi = 2.275 (\pm 0.182) \log k' + 2.331 (\pm 0.087) \quad (6)$$

$n = 10; r = 0.992; s = 0.097; F = 523$

The best correlations are obtained with Σf_{R} and χ , the correlations with $\log P_{\text{obs}}$ and Σf_{LH} being of significantly poorer quality. With regard to the inadequacy of the $\log P_{\text{obs}}/\log k'$ equation (eqn. 3), *n*-butylbenzene and hexamethylbenzene are responsible; the same is true for the $\Sigma f_{\text{LH}}/\log k'$ equation (eqn. 4) but this has the remarkable feature that when the number of carbon fragments in the alkyl substituent increases, the actual differences with respect to the observed $\log P$ values increase. This observation might be indicative of some failure in the conceptional approach of the Leo-Hansch fragmentation; in all respects the π -system of Hansch seems more suitable for correlating $\log k'$ values than their fragmental system, as is clearly demonstrated by the equation

$$\Sigma \pi = 2.871 (\pm 0.273) \log k' + 2.571 (\pm 0.130) \quad (7)$$

$n = 10; r = 0.989; s = 0.145; F = 372$

In addition, we draw attention to the regression equations 8-13, which in combination with eqns. 3-7 make it clear that:

(a) there is a good parallel between the fragmental approach of Rekker and that of Leo and Hansch, although the remarks made above with respect to the observed differences when using the Leo-Hansch system remain valid and may be more closely indicated as follows: between $\log P$ values of 1.00 and 2.50 the differences between the two fragmentation procedures can be neglected; below $\log P = 1.00$ the Leo-Hansch estimates remain increasingly behind the Rekker estimates and above $\log P = 2.50$ the reverse is true;

(b) experimental $\log P$ values are inferior to their analogues that make use of an additive approach for calculation;

(c) Kier's connectivity indices correlate correctly with Rekker's fragmental values.

$$\Sigma f_{\text{R}} = 0.800 (\pm 0.071) \Sigma f_{\text{LH}} + 0.352 (\pm 0.283) \quad (8)$$

$n = 10; r = 0.991; s = 0.128; F = 424$

$$\log P = 0.747 (\pm 0.146) \Sigma f_{\text{R}} + 0.836 (\pm 0.514) \quad (9)$$

$n = 10; r = 0.958; s = 0.211; F = 88$

TABLE II
LIPOPHILICITIES, HPLC DATA AND CONNECTIVITY INDICES FOR A SERIES OF ALKYL BENZENES

Substituents	t_R	$\text{Log } k'$	$\text{Log } P_{\text{obs}}^*$	$\text{Log } P_{\text{est}}$	Δ_1	Σf_{LH}	$\Sigma f_{\text{LH}}^{\text{est}}$	Δ_2	Σf_R	Σf_R^{est}	Δ_3	χ	χ_{est}	Δ_4
H	2.15	-0.186	2.13	2.25	-0.12	2.13	2.02	0.11	2.02	1.94	0.08	2.000	1.909	0.081
Methyl	2.70	0.031	2.69	2.70	-0.01	2.79	2.74	0.05	2.54	2.52	0.02	2.411	2.402	0.009
1,2-dimethyl	3.40	0.208	3.12	3.07	0.05	3.45	3.32	0.13	3.06	3.00	0.06	2.827	2.805	0.022
1,3-dimethyl	3.60	0.246	3.20	3.15	0.05	3.45	3.45	0.00	3.06	3.10	-0.04	2.821	2.891	-0.070
1,4-dimethyl	3.60	0.250	3.15	3.16	-0.01	3.45	3.46	-0.01	3.06	3.12	-0.06	2.821	2.901	-0.080
Isopropyl	4.65	0.410	3.66	3.49	0.17	3.74	3.99	-0.25	3.58	3.55	0.03	3.354	3.264	0.090
1,3,5-trimethyl	5.10	0.466	3.44	3.61	-0.17	4.11	4.18	-0.07	3.58	3.70	-0.12	3.232	3.392	-0.160
<i>n</i> -Butyl	7.45	0.674	4.26	4.04	0.22	4.41	4.86	-0.45	4.10	4.26	-0.16	3.971	3.865	0.106
1,2,4,5-tetramethyl	6.70	0.616	4.00	3.92	0.08	4.77	4.67	0.10	4.10	4.11	-0.01	3.655	3.733	-0.078
Hexamethyl	12.20	0.924	4.31**	4.56	-0.25	6.09	5.69	0.40	5.14	4.94	0.20	4.500	4.433	0.067

* Selected values; other values available from the literature: benzene 1.56¹³, 2.15^{14,15}; toluene 2.73¹⁶, 2.11¹³, 2.80¹⁴; 1,2-dimethylbenzene 2.77¹⁴.

** Apparent experimental outlier.

$$\log P = 0.581 (\pm 0.158) \Sigma f_{\text{LH}} + 1.167 (\pm 0.628) \quad (10)$$

$$n = 10; r = 0.922; s = 0.284; F = 45$$

$$\log P = 0.905 (\pm 0.131) \chi + 0.537 (\pm 0.423) \quad (11)$$

$$n = 10; r = 0.976; s = 0.159; F = 161$$

$$\log P = 0.693 (\pm 0.155) \Sigma \pi + 0.891 (\pm 0.576) \quad (12)$$

$$n = 10; r = 0.949; s = 0.229; F = 73$$

$$\Sigma f_{\text{R}} = 1.179 (\pm 0.101) \chi - 0.299 (\pm 0.327) \quad (13)$$

$$n = 10; r = 0.991; s = 0.123; F = 459$$

The extent to which careful observance of the requirements for congenericity in a series of structures, indicated in the introductory remarks given above, is necessary may be demonstrated by an investigation performed by D'Amboise and Hanai¹⁷. They studied the relationship between hydrophobicity and retention time (HPLC) of a number of compounds belonging to various structural classes: (a) polyaromatics, (b) alkylbenzenes, (c) benzoates, (d) substituted benzenes, (e) aliphatic alcohols and (f) aliphatic carboxylic acids.

In illustration of our approach, we abstract two equations from the work of D'Amboise and Hanai:

$$\log k' = 0.241 \log P - 0.182 \quad (14)$$

$$n = 23; r = 0.980$$

$$\log k' = 0.197 \log P - 0.763 \quad (15)$$

$$n = 7; r = 0.999$$

Eqn. 14 includes the classes a-e and eqn. 15 applies to class f; the experimental conditions for both are similar using acetonitrile-water (50:50) as the eluent. Log P values are f summation taken from ref. 5. In our opinion, the correlation coefficient of eqn. 14 is too low, leaving 4% unaccounted for, and the difference in slopes of eqns. 14 and 15 is unacceptably high, indicating significant non-congenericity in the data set belonging to eqn. 14. In order to elucidate this suspected non-congenericity, we decided to re-evaluate part of D'Amboise and Hanai's material that had led to their eqn. 14. However, they had omitted from their paper a detailed tabulation of the observed capacity factors, so that we had to abstract them from the graphical representation. A fully correct reconstruction of the data was impossible; two data numberings are missing (chlorobenzene and bromobenzene), one data point is missing and three data points seem to have been mutually transposed (phenol for aniline, benzoic acid for phenol, pentyl alcohol for butyl alcohol). Table III summarizes a number of our abstracted data; Nos. 1-7 could be correlated by the following equation, in which correct fragment summations were applied⁶, thereby replacing some erroneously calculated items as used by D'Amboise and Hanai:

$$\log k' = 0.182 (\pm 0.011) \Sigma f + 0.002 (\pm 0.040) \quad (16)$$

$$n = 7; r = 0.998; s = 0.014; F = 1208$$

TABLE III

PARTIAL RE-EVALUATION OF D'AMBOISE AND HANAI'S HPLC DATA¹⁷ ON POLYAROMATICS, ALKYL BENZENES, BENZOATES, SUBSTITUTED BENZENES AND FATTY ACIDS

Type	No.	Compound	Log P*	Log P**	Log k'***	Log k' (eqn. 16)
Polyaromatics	1	Benzene	2.16	2.02	0.37	0.37
	2	Naphthalene	3.18	3.38	0.60	0.62
	3	Anthracene	4.20	4.57	0.83	0.83
	4	Pyrene	4.50	5.17	0.94	0.94
Alkylbenzenes	5	Toluene	2.59	2.54	0.45	0.46
	6	Ethylbenzene	3.12	3.06	0.58	0.56
	7	Isopropylbenzene	3.52	3.58	0.66	0.65
Benzoates	8	Methyl benzoate	2.15	2.16	0.35	0.39
	9	Isopropyl benzoate	3.09	3.10	0.55	0.58
	10	Butyl benzoate	3.74	3.71	0.70	0.68
	11	Isopentyl benzoate	4.15	4.23	0.79	0.77
Substituted benzenes	12	Aniline	1.03	1.00	0.20	0.18
	13	Phenol	1.54	1.53	0.06	0.28
Fatty acids	14	Hexanoic acid	1.87	1.84	-0.35	0.34
	15	Octanoic acid	2.93	2.88	-0.16	0.53
	16	Decanoic acid	3.99	3.92	0.05	0.72
	17	Dodecanoic acid	5.05	4.95	0.25	0.90
	18	Tetradecanoic acid	6.11	5.99	0.45	1.09
	19	Hexadecanoic acid	7.17	7.03	0.65	1.28
	20	Octadecanoic acid	8.23	8.07	0.87	1.47

* D'Amboise and Hanai's calculations applying Rekker's fragmental values from ref. 5 with erroneously calculated log *P* values for Nos. 1, 2, 3 and 4.

** Calculated according to ref. 6.

*** Derived from D'Amboise and Hanai's graphical representation of log *k'* versus log *P*¹⁷.

TABLE IV

RELATIONSHIP BETWEEN NUMBER OF CARBON ATOMS AND LIPOPHILICITY FOR AROMATIC HYDROCARBONS

Adapted from ref. 1.

Compound	n_c^*	Δn_c^{**}	$\Delta(C_xH_y)^{***}$	$\Sigma f[\Delta(C_xH_y)]$	$\Sigma f/\Delta n_c$
Benzene	6	—	—	—	—
Naphthalene	10	4	C ₄ H ₂	0.98 + 0.29 = 1.27	0.32
Diphenyl	12	6	C ₆ H ₄	1.66 + 0.29 = 1.95	0.33
Phenanthrene	14	8	C ₈ H ₄	1.97 + 0.58 = 2.55	0.32
Anthracene	14	8	C ₈ H ₄	1.97 + 0.58 = 2.55	0.32
Fluoranthene	16	10	C ₁₀ H ₄	2.17 + 0.87 = 3.04	0.30
Pyrene	16	10	C ₁₀ H ₄	2.17 + 0.87 = 3.04	0.30
Chrysene	18	12	C ₁₂ H ₆	2.95 + 0.87 = 3.82	0.32
Naphthacene	18	12	C ₁₂ H ₆	2.95 + 0.87 = 3.82	0.32
Perylene	20	14	C ₁₄ H ₆	3.26 + 1.16 = 4.42	0.32
3,4-Benzopyrene	20	14	C ₁₄ H ₆	3.26 + 1.16 = 4.42	0.32

* n_c = number of carbon atoms.

** Δn_c = difference relative to benzene.

*** $\Delta(C_xH_y)$ = CH difference relative to benzene.

The slope of eqn. 16 is no longer different from that of eqn. 15, and its statistics, in contrast to those of eqn. 14, are fully acceptable. Eqn. 16 is suitable for predictive purposes, as can be seen from the benzoate data in Table III. The excellent agreement between estimates and observed $\log k'$ values indicates that the benzoates are fully congeneric with the series of polyaromatics and alkylbenzenes. The same is true for aniline (predicted $\log k' = 0.18$ compared with an observed value of 0.20), but the non-congenericity of phenol in the envisaged series is certain (predicted $\log k' = 0.28$ compared with an observed value of 0.06). The seven fatty acids covered by eqn. 15, although congeneric in themselves, are completely non-congeneric with regard to polyaromatics and alkylbenzenes and they all pass through the column with $\log k'$ values 0.76 lower than predicted. When we apply eqn. 16 they pass through the column with $\log k'$ values 0.65 lower than predicted.

The problem, indicated by several groups^{1,3}, that the total number of carbon atoms in an aromatic compound is not in line with the lipophilicity expressed as \log (capacity factors) deserves some closer attention.

Table IV gives results for a series of aromatic hydrocarbons (taken from ref. 1), which enable us to compare the increase in the number of carbon atoms with the corresponding increase in lipophilicity. This increase, which is very constant, is 0.32 for each extra carbon atom. This implies a perfect correlation between $\log k'$ and the number of carbon atoms (n_C):

$$\log k' = 0.109 (\pm 0.006) n_C - 1.739 (\pm 0.989) \quad (17)$$

$n = 11; r = 0.997; s = 0.039; F = 1441$

Table V gives the number of carbon atoms and the corresponding increase in lipophilicity for compounds 1-7 from Table III. With respect to this increase the series is inhomogeneous and the incompatibility of alkylbenzenes and polyaromatics is evident:

$$\log k' = 0.053 (\pm 0.014) n_C + 0.107 (\pm 0.150) \quad (18)$$

$n = 7; r = 0.971; s = 0.053; F = 82$

$$\log k' = 0.057 (\pm 0.001) n_C + 0.028 (\pm 0.011) \quad (19)$$

$n = 4$ (polyaromatics); $r = 1.000; s = 0.002; F = 37850$

TABLE V

NUMBERS OF CARBON ATOMS IN AND LIPOPHILICITIES OF POLYAROMATICS AND ALKYL BENZENES

Adapted from ref. 17. Symbols as in Table IV.

No.	Compound	n_C	Δn_C	$\Delta(C_xH_y)$	$\Sigma f[\Delta(C_xH_y)]$	$\Sigma f[\Delta n_C]$
1	Benzene	6	—	—	—	—
2	Naphthalene	10	4	C ₄ H ₂	0.98 + 0.29 = 1.27	0.32
3	Anthracene	14	8	C ₈ H ₄	1.97 + 0.58 = 2.55	0.32
4	Pyrene	16	10	C ₁₀ H ₄	2.17 + 0.87 = 3.04	0.30
5	Toluene	7	1	CH ₂	0.52	0.52
6	Ethylbenzene	8	2	C ₂ H ₄	1.04	0.52
7	Isopropylbenzene	9	3	C ₃ H ₆	1.56	0.52

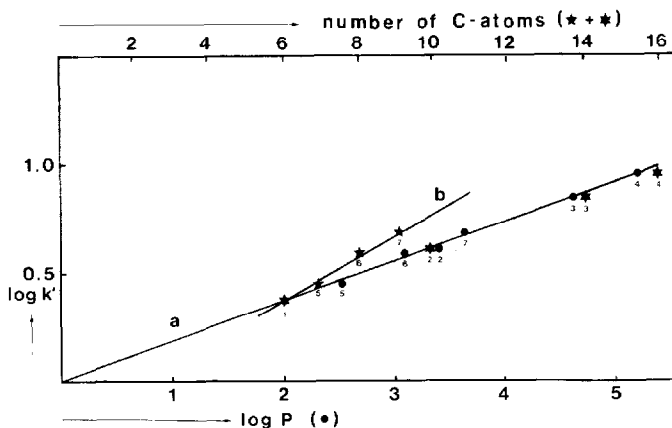


Fig. 1. Log (capacity factors) of alkylbenzenes and polyaromatics *versus* calculated log P values (Rekker method). (a) (filled circles), represents eqn. 16 and applies to the bottom horizontal axis; (b) (\star), represents eqn. 20 and applies the top horizontal axis; (c) (\ast) is not drawn; it would represent eqn. 19.

$$\log k' = 0.100 (\pm 0.023) n_c - 0.235 (\pm 0.171) \quad (20)$$

$n = 4$ (alkylbenzenes); $r = 0.995$; $s = 0.016$; $F = 200$

Eqn. 20 predicts $\log k' = 1.36$ for pyrene, instead of the observed value of 0.94. A graphical representation of some of the above equations is shown in Fig. 1.

Smith³ noticed that biphenyl and naphthalene, when included in a series of alkylbenzenes, were eluted much earlier than would be expected from their carbon numbers, and remarked that the additional aromatic ring in each compound apparently makes it more polar than a corresponding alkyl chain. Our re-evaluation of part of D'Amboise and Hanai's work clearly refutes Smith's view. The lower elution times of biphenyl and naphthalene are connected with a lower fragmental lipophilicity value of the extra CH units in comparison with values in alkylbenzenes.

Finally, we shall show how log P estimates obtained from Rekker's fragmental constants present a parameterization which is advantageous in practice. For this purpose a small series of soil sorption coefficients (K_{oc}) of polycyclic aromatics published by Sabljic and Protic¹⁸ were re-correlated. The soil sorption coefficient of a compound is its concentration sorbed by the soil or by a sediment divided by its concentration in the soil water:

$$K_{oc} = c_{soil}/c_{water} \quad (21)$$

The organic carbon material of the soil is its main dissolving constituent, which means that under steady-state conditions eqn. 21 represents the normal expression for a partition coefficient of a compound partitioned in a solvent system consisting of an organic phase and water.

Sabljić and Protic obtained an excellent correlation between log (soil sorption coefficient) and Kier's second-order valence molecular connectivity index, ${}^2\chi^v$, in the following Collander-type equation:

TABLE VI

LIPOPHILICITIES, CONNECTIVITY INDICES AND SOIL SORPTION COEFFICIENTS FOR A SERIES OF POLYAROMATICS

Adapted from ref. 18.

Compound	Σf	c.c.*	$\Sigma f(c)**$	${}^2\chi^v$	Log K_{oc}		
					Obs	Eqn. 23	Eqn. 22
Benzene	2.02	—	2.02	1.155	1.919	1.93	1.95
Naphthalene (I)	3.01	0.29	3.30	2.347	3.114	3.18	3.18
2-Methylnaphthalene	3.53	0.29	3.82	2.851	3.929	3.69	3.70
Phenanthrene (IV)	3.99	0.58	4.57	3.508	4.362	4.43	4.38
Anthracene (III)	3.99	0.58	4.57	3.547	4.415	4.43	4.42
9-Methylanthracene	4.51	0.58	5.09	3.944	4.813	4.94	4.83
Pyrene (VII)	4.30	0.87	5.17	4.290	4.924	5.02	5.18
Tetracene (V)	4.97	0.87	5.84	4.746	5.813	5.67	5.65
<i>Predictions:</i>							
Fluorene (II)	3.83	0.29	4.12			3.99	
Fluoranthene (VI)	4.30	0.87	5.17			5.02	
Chrysene (VIII)	4.97	0.87	5.84			5.68	
Benzo[a]pyrene (IX)	5.28	1.16	6.44			6.26	
Benzo[ghi]perylene (X)	5.57	1.45	7.02			6.83	

* c.c. = corrective constant.

** $\Sigma f(c)$ = corrected fragmental sum.

$$\log K_{oc} = 1.030 (\pm 0.127) {}^2\chi^v + 0.764 (\pm 0.442) \quad (22)$$

$$n = 8; r = 0.992; s = 0.159; F = 392$$

Our recalculation is given in eqn. 23, and makes use of log P values calculated from fragmental values incorporating correctly derived multiples of the factor 0.289. Because experience indicates that the required procedure is often not properly understood, we refer to Fig. 2, which gives a representative set of examples.

$$\log K_{oc} = 0.980 (\pm 0.103) \Sigma f - 0.048 (\pm 0.456) \quad (23)$$

$$n = 8; r = 0.995; s = 0.135; F = 545$$

Table VI gives all relevant data incorporated in eqn. 23; the excellent quality of eqn. 23 fully justifies a recommendation for the use of calculated log P values in soil sorption studies of aromatic hydrocarbons. In addition to the correctness of the correlation, the absence of a significant intercept and a slope not significantly different from unity are notable features. This means that the soil-water system can be indicated as iso-discriminative towards octanol-water¹⁹ *i.e.* that the soil-water system is not different from the octanol-water system in discriminating between a set of compounds offered for partitioning (gives an equal "spread"), and that in a first evaluation of soil sorption coefficients for polyaromatics and related compounds the fragmental lipophilicities are very useful substitutes. A few predictions of K_{oc} on the basis of eqn. 23 are included in the bottom part of Table VI.

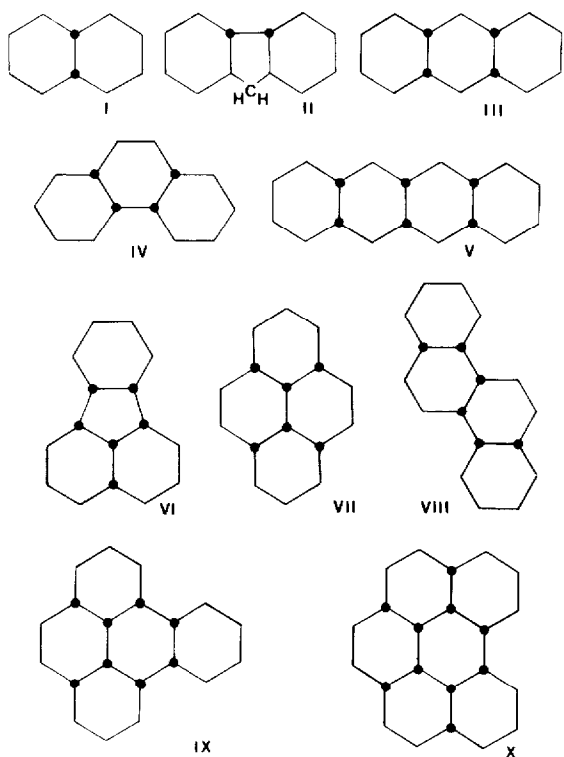


Fig. 2. Polyaromatics with condensation points indicated by Clarendon dots. I, Naphthalene; II, fluorene; III, anthracene; IV, phenanthrene; V, tetracene; VI, fluoranthene; VII, pyrene; VIII, chrysene; IX, benzo[a]pyrene; X, benzo[ghi]perylene.

CONCLUSIONS

Retention of aromatics and their alkyl-substituted derivatives in reversed-phase HPLC is simply correlated with Rekker's calculated $\log P$ values and there is no necessity for any analysis of aromatic hydrocarbon in terms of aromatic ring types and of alkyl substituents, provided that due attention is given to condensation corrections in polyaromatic structures.

Experimentally obtained $\log P$ values, especially those higher than 4.0, are far less reliable for correlative purposes and the same is true for $\log P$ values calculated by means of the fragmental method of Hansch and Leo.

Kier's connectivity indices, on the other hand, offer a good parametrization mode for the lipophilic behaviour of alkylbenzenes and polyaromatics.

ACKNOWLEDGEMENT

The authors acknowledge the valuable assistance of Mrs. G. J. Bijloo with the computations.

REFERENCES

- 1 R. B. Sleight, *J. Chromatogr.*, 83 (1973) 31.
- 2 H. Colin and G. Guiochon, *J. Chromatogr. Sci.*, 18 (1980) 54.
- 3 R. M. Smith, *J. Chromatogr.*, 209 (1981) 1.
- 4 K. Jinno and K. Kawasaki, *Chromatographia*, 17 (1983) 337.
- 5 R. F. Rekker, in W. Th. Nauta and R. F. Rekker (Editors), *The Hydrophobic Fragmental Constant; Its Derivatives and Application; a Means of Characterizing Membrane Systems*, Pharmacochimistry Library Vol. 1, Elsevier, Amsterdam, 1977.
- 6 R. F. Rekker and H. M. de Kort, *Eur. J. Med. Chem.*, 14 (1979) 479.
- 7 K. Rogers and A. Cammarata, *J. Med. Chem.*, 12 (1969) 692.
- 8 M. Tichy and K. Bocek, personal communication to C. Hansch; see ref. 16.
- 9 L. B. Kier and L. H. Hall, *Molecular Connectivity in Chemistry and Drug Research*, Academic Press, New York, 1976.
- 10 L. B. Kier and L. H. Hall, *J. Pharm. Sci.*, 70 (1981) 583.
- 11 T. W. Schultz, L. B. Kier and L. H. Hall, *Bull. Environ. Contam. Toxicol.*, 28 (1982) 373.
- 12 W. J. Murray, L. H. Hall and L. B. Kier, *J. Pharm. Sci.*, 61 (1975) 1978.
- 13 P. Jow and C. Hansch, unpublished results; see ref. 16.
- 14 K. Buchel and W. Draber, *Progr. Photosynth. Res.*, 3 (1969) 1777.
- 15 A. Leo, P. Y. C. Jow, C. Silipo and C. Hansch, *J. Med. Chem.*, 18 (1975) 865.
- 16 C. Hansch and A. Leo, *Substituent Constants for Correlation Analysis in Chemistry and Biology*, Wiley, New York, 1979.
- 17 M. D'Amboise and T. Hanai, *J. Liq. Chromatogr.*, 5 (1982) 229.
- 18 A. Sabljic and M. Protic, *Bull. Environ. Contam. Toxicol.*, 28 (1982) 162.
- 19 R. F. Rekker, *J. Chromatogr.*, 300 (1984) 109.