A Surface Area Approach to Determination of Partition Coefficients

Patrick Camilleri,* Simon A. Watts, and Joe A. Boraston
Shell Research Limited, Sittingbourne Research Centre, Sittingbourne, Kent ME9 8AG

A suite of computer programs (AMBER, MS, SURFACE, and GENSTAT) has been used to calculate the contribution to surface area by the component atoms of a large number of benzene derivatives containing a variety of substituents such as alkyl, hydroxy, alkoxy, amino, and carbonyl functions (ester, ketone, and aldehyde). We have also considered a wide range of polyaromatic compounds. Component surface areas were related to the measured n-octanol–water partition coefficients (*P*) of the molecules under consideration, using linear regression analysis. This surface area group contribution approach was then used to estimate the partition coefficient of other molecules, the structures of which could be defined in terms of the components that have been used in the current model for predicting log *P*.

The partition coefficient (*P*) of a physiologically active compound is considered to play an important role in establishing the level of observed biological activity.[1] Both the translocation of a substrate to an active site and its binding at this site are thought to be related to its inherent hydrophobicity. Unfortunately, it is virtually impossible to determine experimentally the partition coefficient of a chemical in a realistic biological medium, so that partitioning between octanol and water has been widely used as a model for correlation with biological data and has been extensively applied[1,2] to quantitative structure–activity relationships used in predicting biological activity.

Although partition coefficients in octanol–water measured by the 'shake-flask' method have been most useful, experimental difficulties, such as the very low solubility of some compounds in the aqueous phase, have at times led to grossly inaccurate values. Correlation of octanol–water partition coefficients with chromatographic measurements either from thin-layer chromatography[3] (t.l.c.) or high-pressure liquid chromatography[4] (h.p.l.c.) has overcome some of these experimental difficulties in the measurement of high partition coefficients ($P > 10^5$) and has accelerated the measurement process. However, these indirect methods appear to be most successful only when structurally related molecules are being considered.

The empirical prediction of the partition coefficient of chemicals from a consideration of their molecular structure has been an important goal in the design of molecules with distribution properties that make them biologically active. The group contribution approach developed by Hansch[5] has been used extensively to estimate the partition coefficient of organic molecules in the octanol–water system. The Hansch hydrophobicity parameter π_X is defined in equation (1), where P_X and P_H are the partition coefficients of C_6H_5X and benzene, respectively. From the determination of a number of partition coefficients, Hansch and Leo developed a hydrophobicity scale, which can be used to predict the partition coefficient of a compound from a knowledge of its structure. π_X is an additive

$$\pi_X = \log P_X - \log P_H \quad (1)$$

quantity. The use of this parameter to estimate log *P* is only limited to closely related congeners. This limitation primarily arises from the complexity of the mechanism by which solute distributes itself between the organic and aqueous phases.

More recently Rekker[6] introduced the hydrophobic fragment constant, *f*, which is defined by equation (2). f_n is the

$$\log P = \sum_1^n a_n f_n \quad (2)$$

lipophilicity contribution from a fragment of a chemical structure to the total lipophilicity, and a_a is a numerical factor, representing the number of times a particular fragment occurs in the structure. *f* values were derived for a large number of fragments using regression and statistical analysis of experimentally determined log *P* values.

The calculation of log *P* derived by Hansch and Leo[7] also utilises fragment constants. However, the method used in this case has a different theoretical foundation to the Rekker method. Hansch and Leo have derived fragment constants (*f*) for the simplest constituents of a structure and log *P* is calculated by combining these fragment constants with other factors *F*, such as branch and bond factors [see equation (3)].

$$\log P = \sum_1^n a_n f_n + \sum_1^n b_m F_m \quad (3)$$

This process of estimation of log *P* together with some elements of the methodology used by Rekker has been computerised in recent years and is available commercially as the Med-Chem—C log *P* program.† Initially this method could only be used to calculate the log *P* values of simple molecules. Recent upgrading by the introduction of an array of new factors has allowed the estimation of log *P* of more complicated molecules. Despite these refinements, the Med-Chem—C log *P* program fails to calculate log *P* consistently. Moreover, the nature of the various factors that have been added to the calculation have either not been disclosed or are empirical so that it is difficult to understand the physicochemical reasons for variations in log *P* with chemical structure. Such information can be imperative in the rational design of compounds with the correct transport and/or binding properties for enhanced biological activity.

It has been known[8] for some time that for a homologous series of compounds, molecular surface area is linearly related to log *P*, thus allowing calculations of the partition coefficient of further compounds in a series. Surface area has been estimated by several approaches, including glueing styrofoam balls, representing the solvent, to a CPK model of a solute molecule.[9] This method is rather tedious so that algorithms have now been

† The Med-Chem program is available from Pomana College, Claremont, California.

Table 1. H.p.l.c. retention time (min) and n-octanol–water partition coefficients (*P*) of the hydrocarbons used in estimating log P for paracyclophane (**222**)

Compound	Measured log *P*	Retention time (min)	log (Retention time)
Toluene (**2**)	2.73	3.61	0.56
Xylene (**8**)	3.12	4.42	0.65
Ethylbenzene (**3**)	3.15	4.51	0.65
Naphthalene (**20**)	3.30	4.27	0.63
Phenanthrene (**32**)	4.46	7.40	0.87
Benzanthracene (**33**)	5.90	15.30	1.18
Paracyclophane (**222**)		8.25	0.92

Table 2. Components used in the current model for predicting log *P*

(1) Aromatic hydrocarbon, *e.g.* benzene ring
(2) Saturated hydrocarbon chains that do not qualify under (3), (6), (10), or (12), *e.g.* CH_3 group in toluene
(3) Single saturated carbon atom plus attached hydrogens directly attaching a non-hydrocarbon group to a hydrocarbon chain or ring, *e.g.* CH_2 group in benzyl alcohol
(4) OH group, *e.g.* as in phenol
(5) Oxygen atom of OR group, *e.g.* as in anisole, that is not type (11)
(6) Hydrocarbon part of OR group, *e.g.* CH_3 group in anisole, that is not type (12)
(7) Cl atom, *e.g.* as in chlorobenzene
(8) NH_2 or NH group, *e.g* NH_2 in aniline.
(9) C(=O)H or C(=O) group, *e.g.* in benzaldehyde
(10) Hydrocarbon chain part of C(=O)R group, *e.g.* CH_3 in $C(=O)CH_3$
(11) Oxygen atom of OR group in C(=O)OR
(12) Hydrocarbon part of OR group in C(=O)OR

developed[10] for computing the solvent-accessible molecular surface, defined by Richards[11] as the area traced out by a sphere, representing a solvent molecule, as it is rolled over the surface of a solute (sometimes referred to as the *contact surface*). Surface areas (*S*) are linearly related[9] to log *P* by equations such as (4), where α and β are constants. In this way a measure of

$$\log P = \alpha S - \beta \quad (4)$$

log *P* and a knowledge of the corresponding surface area for a number of molecules should allow an estimation of the log *P* for other molecules of related structure, if the surface areas of the latter have been calculated.

This surface area approach has advantages over other methods of estimating log *P* in that no correction factors are necessary for vicinal effects, such as branching and cyclisation. These features are automatically taken into account in determining the molecular surface area of a solute. Thus the surface area approach can be used to explain differences in the lipophilicity of stereoisomers. Moreover, as the surface area of a solute depends on its molecular conformation, the latter can be determined from log *P* using a combination of experiment and calculation. Therefore the surface area method can, in principle, provide a conceptual basis for understanding how the molecular structure of a compound can affect its partitioning.

Using an approach similar to the one in ref. 8 we have combined a number of available computer programs and have developed an algorithm which allows the estimation of the solvent-accessible surface area of the component atoms of a series of molecules. These surface area parameters have been calculated for a number of substituted benzene derivatives with known measured partition coefficients and containing constituents such as alkyl, chloroalkyl, alkoxy, hydroxy, amino, and carbonyl functions (mainly ester, ketone, and aldehyde). Several polyaromatic hydrocarbons containing one or more of these substituents have also been included. Regression analysis of these measured partition coefficients and the corresponding surface areas of the components has given an equation which can be used effectively to estimate the partition coefficient of other molecules. This method also estimates the standard errors of the calculated logarithm of the partition coefficient.

Experimental

Materials.—Hydrocarbons used in the determination of the partition coefficient of paracyclophane were purchased from Aldrich Chemical Co. and were used without further purification. Water was ion-exchanged and double-distilled.

Determination of the Partition Coefficient of Paracyclophane.—(a) *'Shake-flask' method.* Paracyclophane was dissolved in water-saturated n-octanol. This solution was added to water previously saturated with octanol and shaken for *ca.* 2 h. Separation of the octanol and water layers was achieved by centrifugation. The concentration of paracyclophane in the octanol layer was analysed directly by h.p.l.c. while the concentration in the water layer was analysed by the same method after extraction with ethyl acetate and concentration to a smaller volume.

(b) *H.p.l.c. retention time method.* Six compounds of known log *P* (see Table 1) were analysed on an HP 1090 machine using a C-18 column (20 cm in length) and a water–acetonitrile (35:65 v/v) solvent mixture as eluant (flow rate 2 ml min^{-1}). Retention times are given in Table 1. Regression analysis of a plot of measured log *P versus* log (retention time) gave equation (5) (correlation coefficient *r* 0.995). From equation (5) log *P* for paracyclophane was estimated from its h.p.l.c. retention time (8.25 min) as 4.61. This value is very close (within experimental error) to the value of 4.33 determined by the 'shake-flask' method.

$$\log P = 5.147 \log (\text{retention time}) - 0.125 \quad (5)$$

A Procedure for estimating log P *from Surface Area.*—The first step involves the 'sketching' of a molecule on the Evans and Sutherland PS 300 graphics system. The co-ordinates (bond distance and angles) of this molecule are then established by an energy-minimisation procedure. The program AMBER,[12] which is based on molecular mechanics, is utilised for this purpose. Each atom or group of atoms in the molecule under consideration is assigned to a component forming a component's file. Table 2 lists the 12 components that can be considered by the model developed so far.

In the next step, surfaces of the molecule are calculated as solvent-accessible surfaces, using the program MS.[10] This computes the area traced out by the edge of a sphere 1.4 Å in radius (this radius approximates the dimensions of an entire water molecule)[13] as it is rolled over the surface of the molecule. The 'contact surface' of a molecule and its derivation is illustrated in Figure 1. The density of points computed by the MS program is 10 points per $Å^2$.

Each atom within the molecule has associated with it surface points calculated from the Connolly surface[10] of the molecule. These points contribute to a component's surface area. Points resting on a concave or 'saddle-shaped' surface are normally shared out between two or three atoms, respectively. In our procedure to determine surface area we divided these so called 're-entrant points' by two or three depending on the number of atoms involved. The 'corrected' or average number of component points was then divided by ten in order to obtain the surface area ($Å^2$) for each component. This part of the

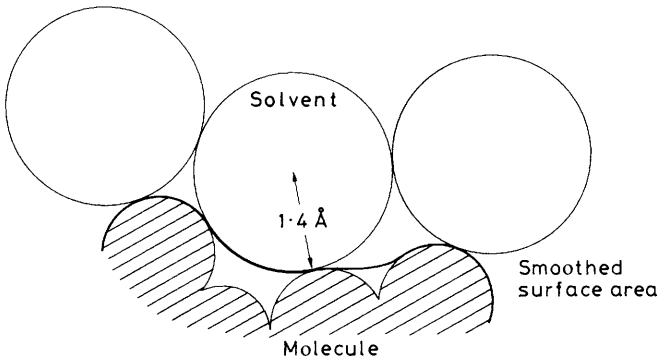


Figure 1. Diagrammatic illustration of the area traced out by the edge of a test sphere (1.4 Å radius), representing water, as it is rolled over the surface of a molecule

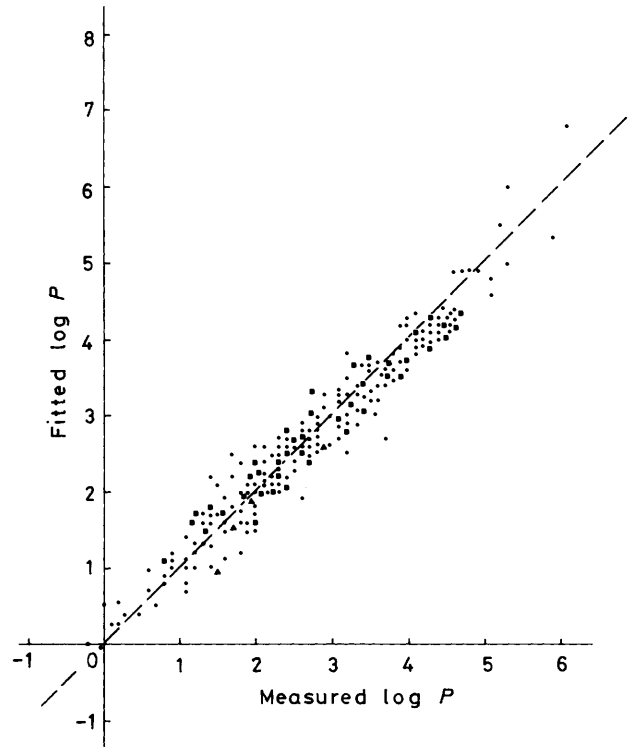


Figure 2. A plot of measured log P against log P from surface area measurements: (●), (■), and (▲) indicate one point, two overlapping points, and three overlapping points, respectively

procedure has been computerised and utilises a program which we have written and called SURFACE.

The above calculation of surface areas of the components of a molecule was carried out for >200 benzene derivatives (Table 3) which have one or more of the components in Table 2. These surface area descriptors, together with measured octanol–water log P values, were stored in a data file. The GENSTAT program* was finally used to perform multiple linear regression, using a model of the form (6) where the a_n values are

$$\log P = a_0 + a_1A_1 + a_2A_2 + \ldots a_nA_n \qquad (6)$$

the coefficients of the regression model and the A_n values are the surface areas of the various components of a molecule. a_0 is the intercept term which largely takes into account the contribution to surface area by the benzene moiety of the compounds considered. Coefficients obtained by performing regression on a large number of molecules (>200) with known log P were used to predict the unknown log P of other molecules (see Table 5), provided the latter had surface areas that could be defined in terms of the 12 components handled by the model.

Computations.—The computations described in the previous section were performed with DEC hardware. A computer in the VAX cluster with operating system VMS 4.3 was used. Programs were written in FORTRAN 77 and statistical analysis was performed using GENSTAT 4.04.

Results and Discussion

Table 3 lists the 217 compounds that have been considered in this study. This Table includes measured[14] octanol–water log P values together with two calculated log P values, one from the present surface area group contribution approach and another from the Med-Chem—C log P program. The sum of squares of residuals obtained from the two methods were calculated as 17.40 and 18.56, respectively. This result provides increased confidence in the surface area approach in that the estimation of log P is comparable to that of Med-Chem. Figure 2 shows a plot of experimental log P *versus* log P from surface area measurements; the line drawn is of unit slope. A plot (not shown) of 'residual' against calculated log P indicates that the scatter in this plot is not due to any systematic error in estimating log P from surface area considerations. Table 4 shows the individual regression coefficients estimated for the 12 components, defined and listed in Table 2. The percentage variance accounted for between measured and calculated log P values, using the surface area model, was 94.9.

The coefficients in Table 4 were used to compute the log P value for a number of compounds. Representative examples are given in Table 5. Also included in this Table is the standard error (which is a measure of the limits of confidence) for every calculated log P and the corresponding log P value obtained from the Med-Chem group contribution approach. The variety of structures of compounds (**218**)—(**230**) demonstrates the wide and useful application of the present model, which allows not only consideration of simple substituted benzene derivatives, such as compounds (**218**)—(**221**), but also more complicated cyclic aliphatic and heterocyclic structures.

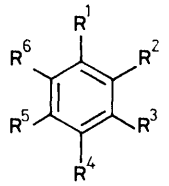
The case of paracyclophane (**222**) is of particular interest. The experimental log P value reported[14] for this compound is 2.33. Using the Med-Chem program a value of 5.79 was obtained (approximately twice the value of *p*-xylene minus the contribution of four hydrogen atoms). In contrast a value of 4.83 was calculated from the surface area of this molecule, that is, about an order of magnitude lower than the Med-Chem value. This large discrepancy between experimental and calculated log P values led us to remeasure the octanol–water partition coefficient of (**222**) by two methods, namely the 'shake-flask' method and h.p.l.c. (using appropriate standards, listed in Table 1). The values obtained by these two methods were 4.33 and 4.61, respectively. These values are remarkably close to the value of 4.83 calculated using the regression coefficients in Table 4.

The experimental log P for 9,10-dihydroanthracene (**223**) has been reported[6] as 4.25. This is again very similar in magnitude to the value of 4.55 obtained by the surface area approach, especially when the standard error of 0.05 for calculating this value is taken into account.

A comparison of the log P values obtained by the surface area approach and by the Med-Chem—C log P program shows that, for the majority of compounds, values are of the same order of

* The GENSTAT statistical program is available from Rothamsted Experimental Station, Harpenden, Hertfordshire and was written by N. H. Alvery, *et al.*

Table 3. Measured and fitted log *P* values for a number of benzene derivatives



Number	R^1	R^2	R^3	R^4	R^5	R^6	log *P* experimental	log *P* Calc.[a]	log *P* Residual	log *P* Calc.[b]	log *P* Residual
(1)	H	H	H	H	H	H	2.13	2.34	−0.21	2.14	−0.01
(2)	CH_3	H	H	H	H	H	2.73	2.79	−0.06	2.79	−0.06
(3)	C_2H_5	H	H	H	H	H	3.15	3.25	−0.10	3.32	−0.17
(4)	C_3H_7	H	H	H	H	H	3.72	3.63	0.09	3.85	−0.13
(5)	C_4H_9	H	H	H	H	H	4.26	4.20	0.06	4.38	−0.12
(6)	$CH(CH_3)_2$	H	H	H	H	H	3.66	3.60	0.06	3.72	−0.06
(7)	$C(CH_3)_3$	H	H	H	H	H	4.11	3.85	0.26	4.12	−0.01
(8)	CH_3	CH_3	H	H	H	H	3.12	3.16	−0.04	3.44	−0.32
(9)	CH_3	H	CH_3	H	H	H	3.20	3.31	−0.11	3.44	−0.24
(10)	CH_3	H	H	CH_3	H	H	3.15	3.29	−0.14	3.44	−0.29
(11)	C_2H_5	CH_3	H	H	H	H	3.53	3.62	−0.09	3.97	−0.44
(12)	CH_3	CH_3	CH_3	H	H	H	3.66	3.42	0.24	4.09	−0.43
(13)	CH_3	CH_3	H	CH_3	H	H	3.78	3.66	0.12	4.09	−0.31
(14)	CH_3	H	CH_3	H	CH_3	H	3.42	3.76	−0.34	4.09	−0.67
(15)	CH_3	CH_3	CH_3	CH_3	H	H	4.11	4.01	0.10	4.74	−0.63
(16)	CH_3	CH_3	CH_3	H	CH_3	H	4.17	4.07	0.10	4.74	−0.57
(17)	CH_3	CH_3	H	CH_3	CH_3	H	4.00	4.08	−0.08	4.74	−0.74
(18)	CH_3	CH_3	CH_3	CH_3	CH_3	H	4.56	4.49	0.07	5.39	−0.83
(19)	CH_3	CH_3	CH_3	CH_3	CH_3	CH_3	5.11	4.94	0.17	6.04	−0.93
(20)	Naphthalene						3.30	3.18	0.12	3.32	−0.02
(21)	2-Methylnaphalene						3.86	3.73	0.13	3.97	−0.11
(22)	2-Ethylnaphthalene						4.38	4.39	−0.01	4.49	−0.11
(23)	1,2-Dimethylnaphthalene						4.31	4.07	0.24	4.61	−0.30
(24)	1,3-Dimethylnaphthalene						4.42	4.19	0.23	4.61	−0.19
(25)	1,4-Dimethylnaphthalene						4.37	4.08	0.29	4.61	−0.24
(26)	1,5-Dimethylnaphthalene						4.38	4.03	0.35	4.61	−0.23
(27)	1,7-Dimethylnaphthalene						4.44	4.21	0.23	4.61	−0.17
(28)	1,8-Dimethylnaphthalene						4.26	3.98	0.28	4.61	−0.35
(29)	2,3-Dimethylnaphthalene						4.40	4.18	0.22	4.61	−0.21
(30)	2,7-Dimethylnaphthalene						4.31	4.30	0.01	4.61	−0.30
(31)	Anthracene						4.45	4.33	0.12	4.49	−0.04
(32)	Phenanthrene						4.46	4.14	0.32	4.49	−0.03
(33)	2,3-Benzanthracene						5.90	5.99	−0.09	5.66	0.24
(34)	Biphenyl						3.89	3.90	−0.01	4.03	−0.14
(35)	Diphenylmethane						4.14	4.24	−0.10	4.36	−0.22
(36)	1,2-Diphenylethane						4.79	4.83	−0.04	4.89	−0.10
(37)	OH	H	H	H	H	H	1.46	1.57	−0.11	1.48	−0.02
(38)	OH	CH_3	H	H	H	H	1.95	2.03	−0.30	2.12	−0.17
(39)	OH	H	CH_3	H	H	H	1.96	2.12	−0.16	2.12	−0.16
(40)	OH	H	H	CH_3	H	H	1.94	2.12	−0.18	2.12	−0.18
(41)	OH	C_2H_5	H	H	H	H	2.47	2.54	−0.07	2.65	−0.18
(42)	OH	H	C_2H_5	H	H	H	2.40	2.60	−0.20	2.65	−0.25
(43)	OH	H	H	C_2H_5	H	H	2.58	2.70	−0.12	2.65	−0.07
(44)	OH	CH_3	H	CH_3	H	H	2.30	2.65	−0.35	2.77	−0.47
(45)	OH	CH_3	H	H	CH_3	H	2.33	2.59	−0.26	2.77	−0.44
(46)	OH	CH_3	H	H	H	CH_3	2.36	2.56	−0.20	2.77	−0.41
(47)	OH	H	CH_3	CH_3	H	H	2.23	2.49	−0.26	2.77	−0.54
(48)	OH	H	CH_3	H	CH_3	H	2.35	2.45	−0.10	2.77	−0.42
(49)	CH_2OH	H	H	H	H	H	1.10	1.49	−0.39	1.10	0.00
(50)	C_2H_4OH	H	H	H	H	H	1.42	2.11	−0.69	1.41	0.01
(51)	$CH(OH)CH_3$	H	H	H	H	H	1.36	1.56	−0.20	1.33	0.03
(52)	C_3H_6OH	H	H	H	H	H	1.88	1.89	−0.01	1.86	0.02
(53)	CH_2OH	H	CH_3	H	H	H	1.60	1.64	−0.04	1.75	−0.15
(54)	CH_2OH	H	H	CH_3	H	H	1.58	1.65	−0.07	1.75	−0.17
(55)	OH	OH	H	H	H	H	0.88	0.84	0.04	0.81	0.07
(56)	OH	H	OH	H	H	H	0.80	0.79	0.01	0.81	−0.01
(57)	OH	H	H	OH	H	H	0.59	0.84	−0.25	0.81	−0.22
(58)	CH_2OH	OH	H	H	H	H	0.73	0.34	0.39	0.44	0.29
(59)	CH_2OH	H	OH	H	H	H	0.49	0.30	0.19	0.44	0.05
(60)	CH_2OH	H	H	OH	H	H	0.25	0.38	−0.13	0.44	−0.19
(61)	OH	H	OH	H	OH	H	0.16	0.09	0.07	0.14	0.02

Table 3 (*continued*)

(62)	1-Naphthol						2.84	2.48	0.36	2.65	0.19
(63)	2-Naphthol						2.70	2.57	0.13	2.65	0.05
(64)	1,3-Dihydroxynaphthalene						1.97	1.91	0.06	1.98	−0.01
(65)	1,5-Dihydroxynaphthalene						1.82	1.87	−0.05	1.98	−0.16
(66)	1,7-Dihydroxynaphthalene						1.94	1.91	0.03	1.98	−0.04
(67)	3-Hydroxybiphenyl						3.23	3.15	0.08	3.36	−0.13
(68)	4-Hydroxybiphenyl						3.20	3.42	−0.22	3.36	−0.16
(69)	Benzhydrol						2.67	2.59	0.08	2.45	0.22
(70)	Cl	H	H	H	H	H	2.84	3.05	−0.21	2.86	−0.02
(71)	Cl	CH_3	H	H	H	H	3.42	3.40	0.02	3.50	−0.08
(72)	Cl	H	CH_3	H	H	H	3.28	3.43	−0.15	3.50	−0.22
(73)	Cl	H	H	CH_3	H	H	3.33	3.41	−0.08	3.50	−0.17
(74)	Cl	Cl	H	H	H	H	3.98	3.73	0.25	4.28	−0.30
(75)	Cl	H	Cl	H	H	H	3.60	3.78	−0.18	3.57	0.03
(76)	Cl	H	H	Cl	H	H	3.52	3.74	−0.22	3.57	−0.05
(77)	CH_3	Cl	H	Cl	H	H	4.24	4.05	0.19	4.22	0.02
(78)	CH_3	Cl	H	H	H	Cl	4.29	4.06	0.23	4.22	0.07
(79)	Cl	H	CH_3	CH_3	H	H	3.82	3.84	−0.02	4.15	−0.33
(80)	Cl	Cl	Cl	H	H	H	4.05	4.41	−0.36	4.28	−0.23
(81)	Cl	Cl	H	Cl	H	H	3.98	4.30	−0.32	4.28	−0.30
(82)	Cl	H	Cl	H	Cl	H	4.49	4.38	0.11	4.28	0.21
(83)	Cl	Cl	Cl	Cl	H	H	4.64	5.00	−0.36	4.99	−0.35
(84)	Cl	Cl	Cl	H	Cl	H	4.92	4.96	−0.04	4.99	−0.07
(85)	Cl	Cl	H	Cl	Cl	H	4.82	4.96	−0.14	4.99	−0.17
(86)	Cl	Cl	Cl	Cl	Cl	H	5.17	5.53	−0.36	5.71	−0.54
(87)	Cl	Cl	Cl	Cl	Cl	Cl	5.31	6.00	−0.69	6.42	−1.11
(88)	CH_2Cl	H	H	H	H	H	2.30	2.56	−0.26	2.70	−0.40
(89)	C_2H_4Cl	H	H	H	H	H	2.95	2.92	0.03	3.03	−0.08
(90)	C_3H_6Cl	H	H	H	H	H	3.55	3.37	0.18	3.56	−0.01
(91)	2-Chlorobiphenyl						4.38	4.39	−0.01	4.74	−0.36
(92)	3-Chlorobiphenyl						4.58	4.45	0.13	4.74	−0.16
(93)	4-Chlorobiphenyl						4.61	4.38	0.23	4.74	−0.13
(94)	4,4′-Bichlorobiphenyl						5.33	4.54	0.79	5.46	−0.13
(95)	2,2′,4,5,5′-Pentachlorobiphenyl						6.11	6.36	−0.25	7.59	−1.48
(96)	2,2′,3,3′,6,6′-Hexachlorobiphenyl						6.51	7.28	−0.77	8.31	−1.80
(97)	OCH_3	H	H	H	H	H	2.08	2.23	−0.15	2.06	0.02
(98)	OC_2H_5	H	H	H	H	H	2.51	2.41	0.10	2.59	−0.08
(99)	OC_3H_7	H	H	H	H	H	3.18	2.60	0.58	3.12	0.06
(100)	OCH_3	CH_3	H	H	H	H	2.74	2.67	0.07	2.71	0.03
(101)	OCH_3	H	CH_3	H	H	H	2.66	2.52	0.14	2.71	−0.05
(102)	OCH_3	H	H	CH_3	H	H	2.81	2.73	0.08	2.71	0.10
(103)	OCH_3	CH_3	H	H	H	CH_3	2.92	3.10	−0.18	3.36	−0.44
(104)	OCH_3	OCH_3	H	H	H	H	2.21	2.18	0.03	1.59	0.62
(105)	OCH_3	H	OCH_3	H	H	H	2.21	2.13	0.08	2.15	0.06
(106)	OCH_3	H	OCH_3	H	OCH_3	H	1.53	2.09	−0.56	1.08	0.45
(107)	CH_2OCH_3	H	H	H	H	H	1.34	1.68	−0.34	1.35	−0.01
(108)	$C_3H_6OCH_3$	H	H	H	H	H	2.70	2.72	−0.02	2.44	0.26
(109)	1,3-Benzodioxole						2.08	1.73	0.35	1.50	0.58
(110)	1,4-Benzodioxane						2.01	1.98	0.03	1.85	0.16
(111)	Diphenyl ether						4.21	4.04	0.17	4.24	−0.03
(112)	Phenoxytoluene						3.79	4.16	−0.37	3.83	−0.04
(113)	OCH_3	OH	H	H	H	H	1.32	1.49	−0.17	1.29	0.03
(114)	OCH_3	H	OH	H	H	H	1.58	1.54	0.04	1.57	0.01
(115)	OCH_3	H	H	OH	H	H	1.34	1.59	−0.25	1.57	−0.23
(116)	OC_2H_5	OH	H	H	H	H	1.68	1.67	0.01	1.82	−0.14
(117)	OC_2H_5	H	OH	H	H	H	1.98	1.73	0.25	2.10	−0.12
(118)	OC_2H_5	H	H	OH	H	H	1.81	1.78	0.03	2.10	−0.29
(119)	OH	H	OCH_3	H	OCH_3	H	1.64	1.43	0.21	1.67	−0.03
(120)	CH_2OH	OCH_3	H	H	H	H	1.13	1.10	0.03	1.02	0.11
(121)	CH_2OH	H	OCH_3	H	H	H	1.10	1.19	−0.09	1.02	0.08
(122)	CH_2OH	H	H	OCH_3	H	H	1.05	0.77	0.28	0.46	0.59
(123)	OH	Cl	H	H	H	H	2.15	2.25	−0.10	2.21	−0.06
(124)	OH	H	Cl	H	H	H	2.50	2.30	0.20	2.49	0.01
(125)	OH	H	H	Cl	H	H	2.39	2.26	0.13	2.49	−0.10
(126)	OH	Cl	Cl	H	H	H	3.15	2.99	0.16	3.07	0.08
(127)	OH	Cl	H	Cl	H	H	2.92	2.97	−0.05	3.07	−0.15
(128)	OH	Cl	H	H	Cl	H	3.06	2.87	0.19	3.07	−0.01
(129)	OH	Cl	H	H	H	Cl	2.64	2.91	−0.27	2.79	−0.15
(130)	OH	H	Cl	Cl	H	H	3.33	3.03	0.30	3.35	−0.02
(131)	OH	H	Cl	H	Cl	H	3.62	2.97	0.65	3.35	0.27
(132)	OH	Cl	H	Cl	Cl	H	3.72	3.61	0.11	3.85	−0.13

Table 3 (*continued*)

(133)	OH	Cl	H	Cl	Cl	H	3.69	3.61	0.08	3.57	0.12
(134)	OH	Cl	Cl	Cl	Cl	H	4.21	4.13	0.08	4.60	−0.39
(135)	OH	Cl	Cl	Cl	H	Cl	4.45	4.23	0.22	4.32	0.13
(136)	OH	Cl	Cl	H	Cl	Cl	3.88	4.18	−0.30	4.32	−0.44
(137)	OH	Cl	Cl	Cl	Cl	Cl	5.12	4.66	0.46	5.06	0.06
(138)	OH	H	CH_3	Cl	H	H	3.10	2.73	0.37	3.13	−0.03
(139)	OH	Cl	H	OH	H	H	1.40	1.59	−0.19	1.69	−0.29
(140)	CH_2OH	H	Cl	H	H	H	1.94	1.84	0.10	1.82	0.12
(141)	CH_2OH	H	H	Cl	H	H	1.96	1.76	0.20	1.82	0.14
(142)	OCH_3	H	H	Cl	H	H	2.78	2.86	−0.08	2.91	−0.13
(143)	OCH_3	H	Cl	H	Cl	H	3.80	3.56	0.24	3.70	0.10
(144)	NH_2	H	H	H	H	H	0.90	1.20	−0.30	0.92	−0.02
(145)	NH_2	NH_2	H	H	H	H	0.15	0.30	−0.15	−0.31	0.46
(146)	NH_2	CH_3	H	H	H	H	1.32	1.70	−0.38	1.56	−0.24
(147)	NH_2	H	CH_3	H	H	H	1.40	1.64	−0.24	1.56	−0.16
(148)	NH_2	H	H	CH_3	H	H	1.39	1.65	−0.26	1.56	−0.17
(149)	NH_2	C_2H_5	H	H	H	H	1.74	2.19	−0.45	2.09	−0.35
(150)	NH_2	H	C_2H_5	H	H	H	1.96	2.08	−0.12	2.09	−0.13
(151)	2-Aminonaphthalene						2.28	1.97	0.31	2.09	0.19
(152)	2-Aminobiphenyl						2.84	3.02	−0.18	2.80	0.04
(153)	Diphenylamine						3.50	3.55	−0.05	3.62	−0.12
(154)	1,2-Diphenylhydrazine						2.94	3.41	−0.47	2.97	−0.03
(155)	*N*-Phenylbenzylamine						3.13	3.25	−0.12	3.34	−0.21
(156)	CH_2NH_2	H	H	H	H	H	1.09	0.63	0.46	1.09	0.00
(157)	$C_2H_4NH_2$	H	H	H	H	H	1.41	1.01	0.40	1.43	−0.02
(158)	NH_2	OH	H	H	H	H	0.62	0.58	0.04	0.65	−0.03
(159)	NH_2	H	OH	H	H	H	0.17	0.48	−0.31	0.25	−0.08
(160)	NH_2	H	H	OH	H	H	0.04	0.46	−0.42	0.25	−0.21
(161)	CH_2OH	H	NH_2	H	H	H	−0.05	−0.04	−0.01	−0.12	0.07
(162)	CH_2OH	H	H	NH_2	H	H	−0.22	−0.08	−0.14	−0.12	−0.10
(163)	NH_2	Cl	H	H	H	H	1.90	2.02	−0.12	1.93	−0.03
(164)	NH_2	H	Cl	H	H	H	1.88	1.89	−0.01	1.93	−0.05
(165)	NH_2	H	H	NH_2	H	H	1.83	1.88	−0.05	1.93	−0.10
(166)	NH_2	Cl	Cl	H	H	H	2.78	2.58	0.20	2.80	−0.02
(167)	NH_2	Cl	H	Cl	H	H	2.78	2.61	0.17	2.80	−0.02
(168)	NH_2	Cl	H	H	Cl	H	2.75	2.58	0.17	2.80	−0.05
(169)	NH_2	Cl	H	H	H	Cl	2.71	2.67	0.04	2.80	−0.09
(170)	NH_2	H	Cl	Cl	H	H	2.69	2.52	0.17	2.80	−0.11
(171)	NH_2	H	Cl	H	Cl	H	2.90	2.55	0.35	2.80	0.10
(172)	NH_2	Cl	Cl	Cl	H	H	3.33	3.23	0.10	3.58	−0.25
(173)	NH_2	Cl	H	Cl	Cl	H	3.45	3.20	0.25	3.58	−0.13
(174)	NH_2	Cl	H	Cl	H	Cl	3.52	3.28	0.24	3.58	−0.06
(175)	NH_2	H	Cl	Cl	Cl	H	3.32	3.14	0.18	3.58	−0.26
(176)	NH_2	Cl	Cl	Cl	Cl	H	3.94	3.75	0.19	4.33	−0.39
(177)	NH_2	Cl	Cl	H	Cl	Cl	4.10	3.77	0.33	4.33	−0.23
(178)	NH_2	Cl	Cl	Cl	Cl	Cl	4.59	4.27	0.32	5.07	−0.48
(179)	NH_2	OCH_3	H	H	H	H	1.18	1.18	0.00	1.02	0.16
(180)	NH_2	H	OCH_3	H	H	H	0.93	1.05	−0.12	1.02	−0.09
(181)	NH_2	H	H	OCH_3	H	H	0.95	1.01	−0.06	1.02	−0.07
(182)	NH_2	H	H	OC_2H_5	H	H	1.24	1.32	−0.08	1.55	−0.31
(183)	NH_2	OCH_3	H	H	H	OCH_3	1.20	1.05	0.15	1.11	0.09
(184)	CH(=O)	H	H	H	H	H	1.48	1.63	−0.15	1.50	−0.02
(185)	$C(=O)CH_3$	H	H	H	H	H	1.58	2.38	−0.80	1.58	0.00
(186)	$C(=O)C_2H_5$	H	H	H	H	H	2.19	2.90	−0.71	2.11	0.08
(187)	CH(=O)	CH_3	H	H	H	H	2.26	2.03	0.23	2.14	0.12
(188)	$C(=O)CH_3$	H	H	CH_3	H	H	2.10	2.34	−0.24	2.23	−0.13
(189)	$C(=O)C_6H_5$	H	H	H	H	H	3.18	3.77	−0.59	3.21	−0.03
(190)	Benzil						3.38	3.73	−0.35	3.41	−0.03
(191)	$C(=O)CH_2C_6H_5$	H	H	H	H	H	3.18	3.47	−0.29	3.35	−0.17
(192)	1,3-Phenylindan-1,3-dione						2.90	3.34	−0.44	2.78	0.12
(193)	CH(=O)	OH	H	H	H	H	1.81	0.96	0.85	2.07	−0.26
(194)	CH(=O)	H	OH	H	H	H	1.38	0.99	0.39	1.44	−0.06
(195)	CH(=O)	H	H	OH	H	H	1.35	1.08	0.27	1.44	−0.09
(196)	$C(=O)CH_3$	OH	H	H	H	H	1.92	1.45	0.47	2.08	−0.16
(197)	$C(=O)CH_3$	H	OH	H	H	H	1.39	1.25	0.14	1.45	−0.06
(198)	$C(=O)CH_3$	H	H	OH	H	H	1.30	1.27	0.03	1.45	−0.15
(199)	CH(=O)	Cl	H	H	H	H	2.33	2.31	0.02	2.33	0.00
(200)	CH(=O)	H	Cl	H	H	H	2.10	2.51	−0.41	2.33	−0.23
(201)	$C(=O)CH_3$	Cl	H	H	H	H	2.09	2.56	−0.47	2.09	0.00
(202)	$C(=O)CH_3$	H	Cl	H	H	H	2.51	2.55	−0.04	2.37	−0.05
(203)	$C(=O)CH_3$	H	H	Cl	H	H	2.32	2.52	−0.20	2.37	−0.05

Table 3 (*continued*)

(204)	CH(=O)	H	H	Cl	H	H	1.76	1.95	−0.19	1.78	−0.02
(205)	$C(=O)OCH_3$	H	H	H	H	H	2.12	2.05	0.07	2.11	0.01
(206)	$C(=O)OC_2H_5$	H	H	H	H	H	2.64	2.81	−0.17	2.64	0.00
(207)	$C(=O)OCH_3$	CH_3	H	H	H	H	2.75	2.82	−0.07	2.76	−0.01
(208)	$C(=O)OCH_3$	OH	H	H	H	H	2.55	1.87	0.68	2.61	−0.06
(209)	$C(=O)OCH_3$	H	OH	H	H	H	1.89	1.56	0.33	1.99	−0.10
(210)	$C(=O)OCH_3$	H	H	OH	H	H	1.96	1.46	0.50	1.99	−0.03
(211)	$C(=O)OCH_3$	Cl	H	H	H	H	2.38	2.67	−0.29	2.62	−0.24
(212)	$C(=O)OCH_3$	H	H	OCH_3	H	H	2.27	2.03	0.22	2.33	−0.06
(213)	$C(=O)CH_3$	H	OCH_3	H	H	H	1.84	2.39	−0.55	1.80	0.04
(214)	$C(=O)CH_3$	H	H	OCH_3	H	H	1.74	2.46	−0.72	1.80	−0.06
(215)	$C(=O)C_2H_5$	H	H	H	H	H	3.59	3.19	0.20	3.62	−0.03
(216)	$C(=O)CH_3$	NH_2	H	H	H	H	1.63	1.05	0.58	1.54	0.09
(217)	$C(=O)CH_3$	H	H	NH_2	H	H	0.83	0.78	0.05	0.90	−0.07

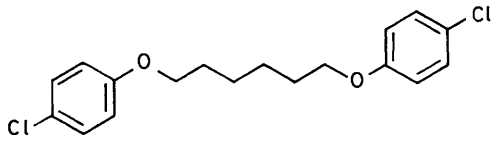
[a] Calculated by equations (6). [b] Calculated using the Med-Chem—C log *P* program.

Table 4. Regression coefficients, a_i [equation (3)] for the 12 components described in Table 2

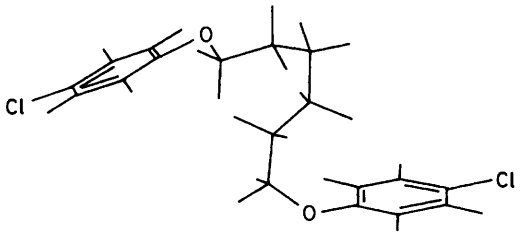
Regression coefficient	Estimates	Standard error	*T*
a_0	−0.239	0.165	−1.45
a_1	0.024 90	0.001 02	24.37
a_2	0.027 31	0.001 35	20.21
a_3	−0.022 37	0.003 59	−6.24
a_4	−0.018 09	0.002 47	−7.32
a_5	−0.000 42	0.009 32	−0.04
a_6	0.009 63	0.002 90	3.32
a_7	0.036 34	0.001 14	31.08
a_8	−0.031 97	0.002 81	−11.37
a_9	−0.007 12	0.003 02	−2.35
a_{10}	0.006 97	0.003 63	1.92
a_{11}	−0.0854	0.0295	−2.90
a_{12}	0.035 26	0.007 79	4.53

magnitude. As already discussed, exceptions are paracyclophane (**222**) and hexasubstituted benzene derivatives such as hexachloro- (**87**) and hexamethyl-benzene (**19**) where the surface area method appears to give log *P* values nearer to the experimentally determined values than the Med-Chem approach. Currently, a contrary situation exists in the case of some 'carbonyl'-containing compounds, such as acetophenone (**185**) and *o*-hydroxybenzaldehyde (**193**), although residual differences appear to be random. As it was mentioned earlier the residual sum of squares obtained for the 217 compounds are 17.4 and 18.6 for the surface area approach and the Med-Chem method. Excluding all 'carbonyl'-containing compounds (**184**)—(**212**) the residual sum of squares becomes 12.6 and 18.2 for the two methods, respectively. Thus in the case of ketones, aldehydes, and esters, a better correlation between the experimental log *P* and the corresponding value obtained by surface area calculation could perhaps be achieved (a) by considering a bigger data set, (b) by considering other minimum-energy conformations using AMBER, and (c) by the inclusion of solvation energy terms to equation (3). It might also be worthwhile remeasuring some of the experimentally determined log *P* values for these compounds.

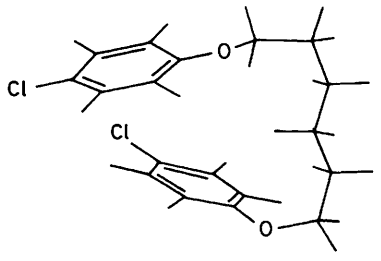
The overall success of calculating log *P* by combining a number of computer programs to obtain the surface area of twelve well defined (see Table 2) components in substituted benzene derivatives has encouraged us to automate this method fully and to apply it in the future to calculate the partition coefficient of more complicated chemical structures, *e.g.* other heterocyclic aromatic molecules. Moreover, since the conformation of a molecule may have a considerable influence on its lipophilic nature[13,15] the surface area approach can be applied to analyse the conformation of a flexible molecule such as the ether (**231**). As a preliminary test we have calculated the log *P* values for the two minimum-energy conformations (**232**) and (**233**) of this compound. These were found to be 7.36 and 6.40, respectively. This simple example shows that correlating partition coefficient to conformation *via* surface area could prove to be valuable in structure–activity correlations of biologically active molecules and could also be most useful in the identification of an 'active' conformation.



(231)

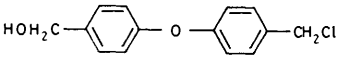
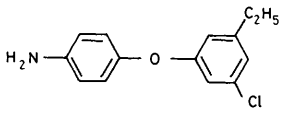
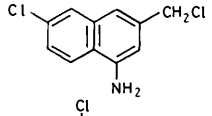
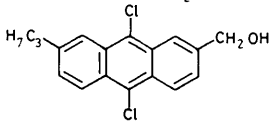
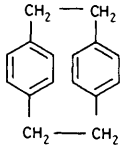
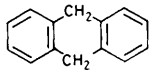
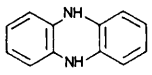
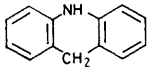
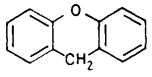
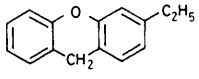
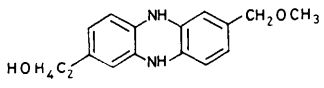
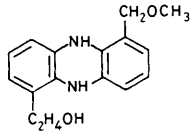
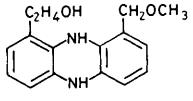


(231)



(233)

Table 5. Predicted log *P* values by (a) the surface area and (b) the group contribution (using the Med-Chem—C log *P* program) approaches

Compound number	Structure	Predicted log *P* (a)	Predicted log *P* (b)
(218)		3.04 (0.14)*	3.55
(219)		4.51 (0.14)	5.23
(220)		3.03 (0.09)	3.52
(221)		5.87 (0.17)	6.81
(222)		4.83 (0.06)	5.79
(223)		4.55 (0.05)	4.67
(224)		2.72 (0.07)	3.20
(225)		3.71 (0.06)	3.94
(226)		4.11 (0.10)	4.56
(227)		5.66 (0.12)	6.59
(228)		1.66 (0.19)	1.59
(229)		1.82 (0.18)	1.59
(230)		1.70 (0.17)	1.59

* Numbers in parentheses are standard errors.

Acknowledgements

We thank Drs. L. J. Mulheirn and A. J. Kirby for useful discussions.

References

1 C. Hansch and J. M. Clayton, *J. Pharm. Sci.*, 1973, **62**, 1.
2 Y. Connolly, 'Quantitative Drug Design,' Marcel Dekker, New York, 1978.
3 E. Tomlinson, *J. Chromatogr.*, 1975, **113**, 1.
4 P. J. Twitchett and A. C. Moffat, *J. Chromatogr.*, 1975, **111**, 149.
5 T. Fiyita, I. Junkichi, and C. Hansch, *J. Am. Chem. Soc.*, 1964, **86**, 5175.
6 R. F. Rekker, 'The Hydrophobic Fragment Constant,' Elsevier, Amsterdam, 1977.
7 C. Hansch and A. Leo, 'Substituent Constants for Correlation Analysis in Chemistry and Biology,' Wiley, New York, 1979.
8 S. H. Yalkowsky and S. C. Valvani, *J. Med. Chem.*, 1976, **19**, 727.
9 N. Furasaki, S. Hada, and S. Neya, *J. Phys. Chem.*, 1985, **89**, 3046.
10 M. L. Connolly, *Science*, 1983, **221**, 709.
11 F. M. Richards, *Annu. Rev. Biophys. Bioengin.*, 1977, **6**, 151.
12 P. A. Kollman and P. H. Weinder, *J. Comput. Chem.*, 1981, **2**, 287.
13 A. J. H. Hoppinger and R. D. Battershall, *J. Med. Chem.*, 1976, **19**, 569.
14 C. Hansch, A. Leo, S. H. Unger, K. H. Kim, D. Nickaitani, and E. J. Lien, *J. Med. Chem.*, 1973, **16**, 1207.
15 L. R. Pratt and D. Chandler, *J. Chem. Phys.*, 1977, **67**, 3683.

Received 7th October 1987; *Paper* 7/1789