Computer-Assisted Computation of Partition Coefficients from Molecular Structures Using Fragment Constants

J. T. CHOU and PETER C. JURS*

Department of Chemistry, 152 Davey Laboratory, The Pennsylvania State University, University Park, Pennsylvania 16802

Received January 17, 1979

Log P, the logarithm of the partition coefficient between an aqueous phase and a model lipid phase (usually 1-octanol), has been shown to be highly correlated with various types of biological activities, drug–receptor interactions, drug–macromolecule interactions, and drug metabolism. A software package for calculating log P values from connection tables has been implemented and tested. The routine, named CLOGP, is based on the fragment method development by Leo and Hansch. During the development of the algorithm some revisions and extensions of the methodology have been made. Comparisons of experimental log P values and those calculated by CLOGP show good agreement. The program can be used to develop new fragment constants from observed log P values for members of a congeneric series. The program is particularly useful for calculation of log P's for large sets of compounds for structure–activity studies.

It is well known that log P, the logarithm of the partition coefficient between an aqueous phase and a model lipid phase, is highly correlated with various types of biological activities, such as many beneficial effects of drugs, toxicity, pesticidal activity, etc.[1-3] It is evident that log P, as an operational definition of lipophilicity or hydrophobic bonding, plays a significant role in the interaction between chemicals and macromolecules or receptors.[2] It also has been shown that the lipophilic character of drugs is an important factor in drug metabolism.[4,5]

Even though experimental values for partition coefficients are preferred over calculated ones in quantitative structure–activity studies,[6] there is an ever-increasing need for reliable estimations for heterogeneous sets of chemicals or drugs for which either the experimental values are not available or may be difficult to measure. For example, the log P's for polycyclic aromatic hydrocarbons are difficult to measure,[7] and the time required to measure values for a large set of these compounds (a few hundred) for a study of structure–carcinogenicity relations would be prohibitive. Therefore, a dependable calculation of log P is desired.

The calculation of log P can be done either by the π-addition method based on eq 1[7,8] or by summing f values as shown in eq 2.[6,9-12]

$$\log P_{R-X} = \log P_{R-H} + \pi_X \quad (1)$$

$$\log P = \sum_{1}^{n} a_i f_i \quad (2)$$

Nys and Rekker derived "hydrophobic fragment constants" by a statistical "reductionist approach". However, Leo, Hansch, and co-workers have developed the fragment addition method by a "constructionist approach".[6,12] This approach starts with a set of "fundamental fragments" whose values (f values) are summed with the appropriate weightings, that is, the number of times each fragment appears. Corrections are added to refine the calculation if necessary.[12] This method can be described explicitly by eq 3, where f_i is the fragment

$$\log P = \sum_{i=1}^{n} a_i f_i + \sum_{j=1}^{m} c_j \quad (3)$$

constant for the ith fragment, a_i is the number of occurrences of the ith fragment, and c_j is the jth correction factor.

The fundamental assumption of this "constructionist" approach is that hydrophobicity is an additive-constitutive property of molecules. The log P value calculated for a solute is equal to the summation of the hydrophobic contributions of each constitutive fragment. That is, a fragment contributes the same degree of hydrophobicity to all molecules. The f value for a fragment is constant regardless of the molecule which contains this fragment.

The f values (fragment constants) were derived from the partition coefficients of compounds without "surprise interactions".[6,12] For example, the f value for a hydrogen atom was derived from the carefully measured log P of hydrogen gas, and the f value for CH_3 was calculated from accurately measured log P values for ethane and methane.[6] As molecules increase in length, the flexibility decreases the log P relative to what it would be expected to be. Also, branching leads to the necessity for a smaller cavity for accomodation of the solute in the solvent, leading to a reduction of log P from the expected value. When there is an unsaturated bond, the localized polarity of this bond lowers the hydrophobicity. These are called bond and branching factors.[12] As multipolar groups, either H-polar or S-polar, are added to a molecule, intramolecular hydrogen bonding, proximity interactions, or hydrophobic shielding effects can occur.[12] To account for these interactions, correction values are used to further refine the calculation. As shown in eq 3, these correction factors are included in the summation to estimate log P.

The calculation of log P using eq 3 has been heretofore a manual procedure. In view of the fact that log P is a very important predictor for a variety of biological activities, and in order to facilitate its calculation for large sets of compounds, the procedure for estimation of log P has been automated. During the development of the algorithm implementing the "constructionist approach", some revisions and extentions of the methodology have been made.

DESIGN AND STRUCTURE OF CLOGP

The algorithm to automate the 1-octanol/water log P calculation was developed based on Leo and Hansch's fragment constants addition method (eq 3). The program, CLOGP, to calculate log P has been written in Fortran IV.

The structure and flowchart of CLOGP program is depicted in Figure 1. The details are explained in the following.

1. Molecular Structure Input. The hydrogen-suppressed topological representation of a molecular structure can be input either from a card reader or from a Tektronix graphics terminal through the ADAPT system.[13] After the structure has been entered, the structural information is stored in a con-

0095-2338/79/1619-0172$01.00/0

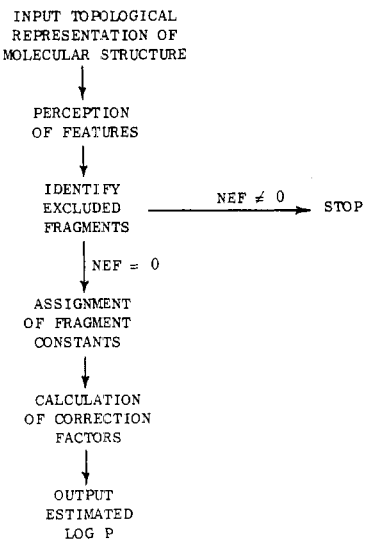


Figure 1. Logical flow of CLOGP program. NEF is the number of excluded fragments.

nection table form which can be easily accessed for further examination of the structure.

2. Perception of Features. The features needed for the log *P* calculation are perceived from the stored connection table. They are:

(a) general structural information, which includes number of each atom type, number of each bond type, number of hydrogens attached to each nonhydrogen atom, and an index for connections characteristic of each atom.

(b) functional groups or basic polar fragments, which include carbonyl (C=O), alcohol (OH), ether (-O-), sulfoxide (S=O), sulfonyl (SO_2), sulfhydryl (SH), thioether (-S-), amines ($-NH_2$, -NH-, -N<), nitroso (N=O), nitro (NO_2), cyano (CN), thiocyanate (SCN), phosphoryl (P=O), and the following groups: A=B, where A stands for C, N, S, P, and B stands for N, S, P; O-X, where X stands for O, N, S, P, halogen; N-Y, where Y stands for N, S, P, halogen; S-Z, where Z stands for S, P, halogen; X-Y, where X stands for P or halogen and Y stands for P or halogen.

(c) ring information. This information consists of the number of rings, number of each ring atom type, number of each ring bond type, characteristic of each ring atom (aromatic or nonaromatic, or fused aromatic ring atom), and character of each ring type (aromatic or nonaromatic).

The "fundamental fragments" as defined by Leo and Hansch[12] are perceived in the subsequent sections of the algorithm from the above information.

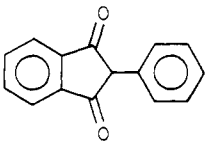
3. Identification of Excluded Fragments. The fundamental fragments for which fragment constants are not available are perceived, since some fragment constants have not been developed. If the presence of an unavailable fragment constant is detected (i.e., NEF $\neq$ 0, as in Figure 1), the calculation of log *P* will not be performed, and the structure of this excluded fragment will be output.

4. Assignment of Fragment Constants. If there is no excluded fragment present, the assignments of constants to all fundamental fragments will proceed. The values for most of the fragment constants have been taken from ref 12. The others have been derived using this program (see Table III).

5. Calculation of Correction Factors. Corrections are performed for bonding, branching, multiple halogenation, and proximity effects, if any. Two additional correction factors that cannot be perceived or calculated by this program are ring cluster and intramolecular hydrogen bonding corrections.

6. Output of the Estimated Log *P*. After fragment constant values are assigned to all fundamental fragments, and all corrections are accounted for, then the summation is formed and the estimated log *P* value is output.

Example 1



2-phenyl-1,3-indandione

fundamental fragment[a]	a_i	f_i	$a_i f_i$
	9	0.355	3.195
	3	0.130	0.390
	2	−0.830[b]	−1.660
	1	0.430	0.430
			$\Sigma a_i f_i = 2.357$

correction factor	c_j
ring bonds (4-1)(−0.090)	−0.270
alkyl chain branch	−0.130
polar fragment proximity effects	
F_{P-1}[c]: $-0.320(f_1 + f_2) = -0.320(-0.830-0.830)$	0.531
F_{P-2}[c]: $-0.200(f_1 + f_2) = -0.200(-0.830-0.830)$	0.332
	$\Sigma c_j = 0.463$

$\log P_{calcd} = \Sigma a_i f_i + \Sigma c_j = 2.818 \approx 2.82$
$\log P_{obsd} = 2.90$

[a] (⋯) is an aromatic bond type. [b] Enhanced by the other carbonyl group ($\delta_I \geq 0.30$). [c] $P-1$ and $P-2$ denote proximity effects between two polar fragments (two carbonyl fragments in this case) which are separated by one carbon atom and two carbon atoms, respectively.

Table I. Percentage of Compounds for Which Log *P* Could Be Calculated

no. of compds	class	percentage calcd
209	heterogeneous set of mutagens and carcinogens	80.4
200	polycyclic aromatic hydrocarbons	97.5
155	*N*-nitroso compounds	75.5
90	random selection from the 3052 compounds of ref 8	84.4
		av 84.5

CAPABILITIES OF THE CLOGP ALGORITHM

(1) Ability to Calculate Log *P*. An example of the sequence of calculations for a specific compound, 2-phenyl-1,3-indandione, is shown in Example 1. Four fundamental fragments are perceived, and the weighted summation of their fragment constants is formed to give 2.357. A ring bond correction is made for the four single bonds in the five-membered ring. A correction is made for the presence of a single alkyl chain branch. The two carbonyl groups are separated by one carbon atom and by two carbon atoms (tracing the five-membered ring in opposite directions), leading to two proximity effect corrections. The sum of the correction factors is 0.463. The log *P* calculated for 2-phenyl-1,3-indandione is 2.82, in quite good agreement with the experimentally observed value of 2.90.

The algorithm has been applied to several large sets of organic compounds currently under investigation in several structure–activity relations studies. The percentages of the data sets for which CLOGP could calculate log *P* values are shown in Table I. These range from a high of 97.5% for 200 polycyclic aromatic hydrocarbons to a low of 75.5% for a set

Table II. Comparison of Observed and Calculated Log P Values

index	name	log P_{obsd} [a]	log P_{calcd}
1	methylacetylene	0.94	0.90
2	fluoroform	0.64	0.64
3	isobutylene	2.34	2.18
4	ethanol	−0.31	−0.21
5	dimethyl ether	0.10	0.12
6	cyclohexane	3.44	3.51
7	propane	2.36	2.32
8	2-propanol	0.05	0.11
9	*tert*-butylamine	0.40	0.53
10	2-phenylethylamine	1.41	1.44
11	*N*-phenylacetamide	1.16	1.17
12	halothane	2.30	2.46
13	benzimidazole	1.34	1.51
14	*p*-nitrophenol	1.91	1.97
15	cyclohexene	2.86	2.96
16	1,2-dichlorotetrafluoroethane	2.82	2.86
17	hexachlorophene	3.93	3.89
18	1,2-methylenedioxybenzene	2.08	2.10
19	2-phenyl-1,3-indandione	2.90	2.82
20	carbon tetrachloride	2.83	2.96
21	dioxane	−0.42	0.01
22	2-bromoacetic acid	0.41	0.48
23	2-chloroethanol	0.03	0.03
24	indene	2.92	2.97
25	fluorene	4.12	4.03
26	anthracene	4.45	4.45
27	pyrene	4.88	4.90
28	quinoxaline	1.08[b]	1.13
29	carbazole	3.51[c]	3.50
30	menadione	2.20	1.75
31	chloramphenicol	1.14	0.56
32	2-hydroxy-1,4-naphthoquinone	1.46[d]	0.52
33	2-methyl-3-hydroxy-1,4-naphthoquinone	1.20	1.18
34	2-methoxy-1,4-naphthoquinone	1.35	1.26
35	benzothiazole	2.01	2.01
36	*o*-phenanthroline	1.83	1.93
37	thiazole	0.44	0.42
38	piperazine	−1.17	−1.25
39	morpholine	−1.08	−0.99
40	salicyclic acid	2.24[e]	1.97
41	imidazole	−0.08	−0.08
42	cyclohexanol	1.23	1.42
43	*o*-phenyleneurea	1.12	1.39
44	tripropylamine	2.79	2.85
45	di-*n*-propylamine	1.62[f]	1.67
46	coumarin	1.39	1.44
47	trifluoromethylbenzene	2.90[g]	2.20
48	trifluoromethylsulfonanilide	3.05	3.06
49	1,3-indandione	0.61	1.27
50	9-fluorenone	3.58	2.87
51	phenazine	2.84	2.72
52	morphine	0.83[h]	1.18
53	2,2,2-trifluoroethanol	0.41	0.41
54	2,2,2-trifluoroacetamide	0.12	0.12
55	2,2,2-trichloroethanol	1.35	1.39
56	2,2,2-trichloroacetamide	1.04	1.04
57	pyrimidine	−0.40	−0.46
58	glucose	−3.24[i]	−3.39
59	cyclohexylamine	1.49	1.52
60	neopentane	3.11	3.14
61	2-methylpropane	2.76	2.73
62	crotonic acid	0.72	0.85
63	cinnamonitrile	1.96	1.92
64	cinnamic acid	2.13	2.18
65	cinnamamide	1.41	1.00
66	methyl cinnamate	2.62	2.47
67	phenyl vinyl ketone	1.88	1.58
68	styrene	2.95	2.92
69	1-phenyl-3-hydroxypropene	1.95	1.47
70	methyl styryl ketone	2.07	1.98
71	1,1,2-trichloroethylene	2.29	2.28
72	2-methoxyanisole	2.08	2.00
73	ethyl vinyl ether	1.04	0.98
74	pyrazole	0.13	0.24
75	1,1-difluoroethylene	1.24	1.12
76	1,2,3,4-tetrahydroquinoline	2.29	2.58

[a] All the observed log P values reported in this paper are from Pomona College Medicinal Chemistry Data Bank, unless specified. [b] Average of 1.32 and 0.84. [c] Average of 3.72 and 3.29. [d] Average of 1.55 and 1.38. [e] Average of 2.26 and 2.21. [f] Average of 1.73, 1.46, and 1.67. [g] Average of 2.79 and 3.01. [h] Average of 1.03, 0.70, and 0.76. [i] From ref 12.

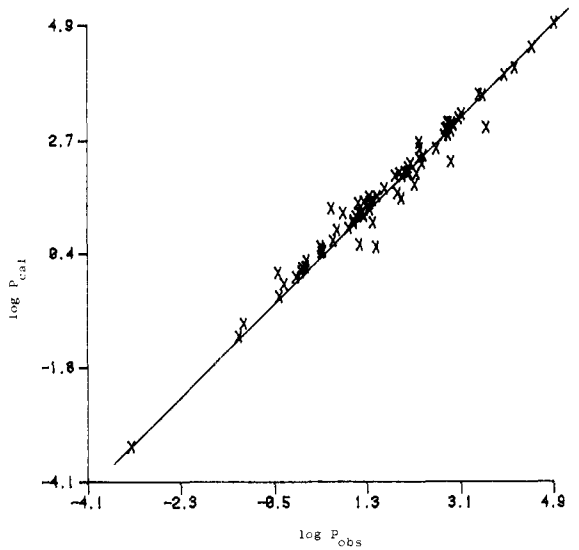
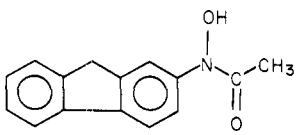


Figure 2. Plot of log P_{obsd} vs. log P_{calcd} for a test set of 76 compounds.

Example 2



N-hydroxy-2-acetylaminofluorene

No fragment constant available for (-CONR, -O(H)-)

of 155 *N*-nitroso compounds. For a large set of randomly selected compounds the figure was 84.4%, and this is the average of all sets reported in Table I as well.

The comparison of observed and calculated log *P* values for a test set of 76 compounds is shown in Table II. The calculated values are in good agreement with the observed values. This correlation can be demonstrated by the regression equation (eq 4) and its graphical appearance (see Figure 2).

$$\log P_{obsd} = 1.001 \log P_{calcd} + 0.028 \quad (4)$$

$$N = 76,\ r = 0.985,\ r^2 = 0.971,\ \text{S.E.} = 0.239$$

(2) Detection of Fundamental Fragments for Which Fragment Constants Are Not Available. This program is designed for estimation of log *P*. However, not all possible fragment constants have been developed, and therefore the perception of these fragments is necessary. After the molecule is decomposed into "fundamental fragments", the detection of fragments whose constants are not available is performed. If such fragments are present, the calculation of log *P* will not proceed and the structures of these excluded fragments will be printed out. As shown in Example 2, the polar fragment (-CONR, -OH) is perceived and its fragment constant is found to be not available. The message with the structure of this fragment is output. By obtaining this information one is able to pinpoint what type of fragments or which specific fragment has no fragment constant.

(3) Choice of Calculation or Checking. Another feature of this program is that the user has the choice of the calculation mode or checking mode. If the checking mode is chosen, only the detection of unavailable fragment constants will be performed without further calculation. If the calculation mode is selected, both checking and calculation will be performed, provided that every fragment constant for the molecule is available. This feature may be useful to find out whether a particular molecule of interest can be calculated or not, or what type of compounds in your work list of a large set of data cannot be handled. The execution time for the checking mode is less than half of that for calculation mode.

Table III. Fragment Constants Derived with CLOGP

fragment	f^a	$f^{\phi\,b}$	attachment to right
$-OSO_2-$	-2.11	-	alkyl
-NHCHO	-	-0.69	none
-CONHOH	-	-1.64	none
-NHOH	-	-1.11	none
-COO-	-1.18 (est)	-	aromatic
>NCHO	-2.67	-	alkyl
-N=N-	-	0.34	aromatic
$-OCONH_2$	-1.67	-	none
-ONHCO-	-	-1.02	alkyl
$-CSNH_2$	-1.30[c]	-0.42	none
>N-N=O	-2.38[d]	-	alkyl

[a] Attachment to left is an alkyl group. [b] Attachment to left is an aromatic group. [c] The log *P* in 1-octanol/water was calculated from the log *P* in $CHCl_3$/water using $\log P_{o/w} = 0.576 + 0.866 \log P_{c/w}$. [d] The observed log *P* values were obtained from ref 14. Only 21 compounds without proximity effects were used.

Table IV. Derivation of the Fragment Constant for NH_2COOR

R	log P_{obsd}	log P_{inc}	derived fragment constant
CH_3	-0.66	0.89	-1.55
CH_3CH_2	-0.15	1.43	-1.58
$CH_3(CH_2)_2$	0.36	1.97	-1.61
$CH_3(CH_2)_3$	0.85	2.51	-1.66
$CH_3CH_2(CH_3)CH$	0.65	2.29	-1.64
$(CH_3)_3C$	0.48	2.07	-1.59
$CH_3(CH_2)_4$	1.35	3.05	-1.70
$CH_3(CH_3)_2C$	0.94	2.61	-1.67
$CH_3(CH_2)_5$	1.85	3.59	-1.74
$CH_3CH_2(CH_3CH_2)(CH_3)C$	1.45	3.15	-1.70
$CH_3(CH_2)_6$	2.36	4.13	-1.77
$CH_3(CH_2)_7$	2.84	4.67	-1.83
			av -1.67 ± 0.08

(4) Development of New Fragment Constants. The program can be used to acquire information about what types of fragments are present in a set of compounds for which fragment constants are not available. If the development of a new fragment constant is desirable, one can do so using the program. Of course, the experimental values for congeners with this fragment must be available, and they must be sufficiently simple structures so that proximity effects are not present. Table III shows some fragment constants developed using CLOGP. An example of the procedure used is given in Table IV. Initially the program was not able to deal with compounds containing the fragment NH_2COO-. A set of experimental log *P* values was available so the derivation of the fragment constant was done. The estimated log *P* using all other fragments in each molecule was computed and is listed as log P_{inc} in Table IV. Then the difference between the experimentally observed log *P* and log P_{inc} was taken. The mean value for these 12 compounds is -1.67 ± 0.08. The trend for the fragment constant values to change smoothly from -1.55 to -1.83 as the chain length increases shows a folding effect may be playing a role. Leo has derived the value of -1.58 for the fragment by a different approach.

In order to utilize this feature of the program effectively, one must have some knowledge of the organization of the program, because some changes must be made in several subroutines.

DISCUSSION

As demonstrated in Table I, the partition coefficients of most compounds in a large set of data can be calculated by the fragment method or the CLOGP program. The fraction of calculable compounds varies depending on the class of

Table V. Comparison of Three Different Calculation Methods

		π addition method		SCAP method[a]		f addition (CLOGP)	
compound[a]	log P_{obsd}[a]	log P_{calcd}[a]	% rel error[a,b]	log P_{calcd}[a]	% rel error[a,b]	log P_{calcd}	% rel error[b]
benzene	2.13	2.13	0.0	2.23	−4.7	2.13	0.0
aniline	0.90	0.90	0.0	0.92	−2.2	0.91	−1.1
propylbenzene	3.68	3.63	+1.4	3.52	+4.4	3.88	−5.4
2-butanone	0.29	0.29	0.0	0.24	+17.2	0.30	−3.5
cyclohexanol	1.23	1.07	+13.0	1.22	+0.8	1.42	−15.5
2,2-dimethyl-propanol	1.36	0.94	+30.9	1.43	−5.1	1.15	+15.4
2-butanol	0.61	0.61	0.0	0.50	+18.0	0.65	−6.6
ethyl acetate	0.73	0.73	0.0	0.59	+19.2	0.71	+2.8
chloroform	1.97	1.67	+15.2	2.11	−7.1	1.96	+0.5
chlorobenzene	2.84	2.84	0.0	2.82	+0.7	2.85	−0.4
2-methyl-2-butanol	0.89	0.91	−2.2	0.75	+15.7	0.97	−9.0
propionitrile	0.16	0.16	0.0	0.19	−18.8	0.16	0.0
1-pentyne	1.98	1.98	0.0	1.96	+1.0	1.98	0.0
benzyl alcohol	1.10	1.47	−33.6	1.37	−24.5	1.11	−0.9
chlorobutane	2.39	2.39	0.0	2.17	+9.2	2.57	−7.5
toluene	2.69	2.63	+2.2	2.62	+2.6	2.80	−4.1
ethylbenzene	3.15	3.21	−1.9	3.01	+4.4	3.34	−6.0
fluorobenzene	2.27	2.27	0.0	2.28	−0.4	2.28	−0.4
nitrobenzene	1.85	1.85	0.0	2.08	−12.4	1.88	−1.6
pentane	3.39[c]	2.50	+26.3	2.17	+36.0	3.40	−0.3
average absolute error			6.3		10.2		4.1

[a] Taken from ref 15. [b] % rel error = [(log P_{calcd} − P_{obsd})/log P_{obsd}] × 100. [c] The original reported value was 2.50. This value was obtained from the Pomona College Medicinal Chemistry Data Bank.

compounds in the data set. To achieve a higher percentage will require the development of currently unavailable fragment constants.

There is nearly perfect agreement between observed and calculated log P values for some data sets as shown by eq 4. The slope is nearly equal to unity (1.001), the intercept is negligibly small (0.028), and 97% of the variance can be explained by this regression equation. Considering the variety of structural types, the wide range of log P values (−3.24 to 4.88), and the relatively small standard deviation (0.239), the correlation derived from these 76 test compounds indicates a reasonable dependability for estimating log P values of these compounds.

Table V shows a comparison of three alternative methods for the calculation of log P values for a set of 20 compounds. The values reported for the π-addition method and the SCAP method are taken from ref 15. The π-addition method should produce good agreement with observed values because this method is more closely tied to observed values than either the SCAP or CLOGP approaches. The "constructionist approach" as implemented in CLOGP has the smallest average absolute error for this data set. This may be an additional indication of the usefulness of the method.

The correlation between observed and calculated log P values was calculated for a randomly selected set of compounds. For Table I a set of 90 compounds was randomly selected from the tabulation of 3052 compounds by Leo, Hansch, and Elkins.[8] Only 40 of the 90 compounds had measured log P values in 1-octanol/water reported and can be calculated by CLOGP. Equation 5 shows the correlation

$$\log P_{obsd} = 0.737 \log P_{calcd} + 0.567 \quad (5)$$

$$n = 40,\ R = 0.846,\ R^2 = 0.715,\ \text{S.E.} = 0.769$$

obtained for these 40 compounds. The set includes compounds with structures as simple as methanol and as complex as amitryptyline ($C_{20}H_{23}N$) and has a wide range of observed log P values from −0.92 to 5.19. This regression equation shows that further refinement of the methodology is necessary for compounds with more complex structures. The poorest results were obtained for compounds with a degree of molecular complexity that would have suggested that the results would be suspect. These are compounds with many polar groups in close proximity such as a tetracycline with three carbonyls, five hydroxyls, and an amide group.

There are a number of rules currently implemented in the program that differ from those of Leo and Hansch;[12] most of them are due to ease of programming.

(1) Pragmatic Rule of Aromaticity. The rings bonded to the aromatic rings in the structures listed in Table VI are considered to be aromatic rings in the definition of Leo and Hansch.[12] However, they are not recognized as aromatic rings by CLOGP. The algorithm to define aromaticity in CLOGP follows the Hückel formalism. Because different rules are used, different log P values are obtained for such compounds. The values generated by both approaches are shown in Table VI. The log P values estimated by either approach are not perfect, with only three out of the six compounds in good agreement with observation.

(2) Proximity Effect of Calculations. (a) One type of proximity effects concerns those in a nonaromatic ring containing an unsaturated bond with aromatic properties such as the fused bond in compound 3 in Table VI. Leo and Hansch have suggested that the choice of type of proximity effect between two polar fragments in a nonaromatic ring depends on the type of bond intervening between these two groups.[12] However, the ring type (either aromatic or nonaromatic) rather than the bond type is taken into consideration by CLOGP. For example, as shown in Table VII, the aromatic type of the proximity effect correction $P - 2$ is considered by Leo and Hansch because the two carbonyl groups are separated by an aromatic bond. However, the alicyclic correction $P - 2$ is used by CLOGP because the ring in which the two polar groups reside is not aromatic. The extent of interaction between these carbonyl groups should not be the same as that if they were in an aromatic ring. Moreover the second ring from the left in structures 1 and 3 in Table VII are very similar in terms of bond lengths and conformation, although the first rings are quite different one from the other. The extent of proximity effects for $P - 2$ for both compounds should be very close if not identical.

(b) Another type of proximity effects is applicable to a pair of polar fragments, where one is an alicyclic ring and the other

Table VI. Comparison of Calculated log P Using Different Pragmatic Rules of Aromaticity

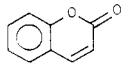
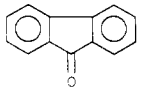
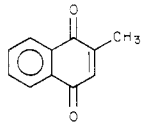
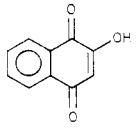
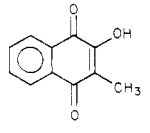
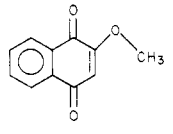
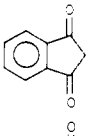
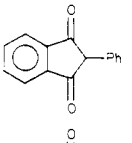
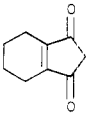
no.	structure	compound	log P_{obsd}	log P_{calcd} Leo and Hansch	log P_{calcd} CLOGP
1		coumarin	1.39	1.40	1.44
2		9-fluorenone	3.58	3.58	2.87
3		menadione	2.20	2.68	1.75
4		2-hydroxy-1,4-naphthaquinone	1.46	1.35	0.52
5		2-hydroxy-3-methyl-1,4-naphthaquinone	1.20	2.02	1.18
6		2-methoxy-1,4-naphthaquinone	1.35	1.95	1.26

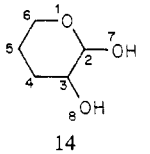
Table VII. Comparison of Calculated Log P Using Different Rules for the Proximity Effects

no.	compound		log P_{obsd}	Leo and Hansch log P^b	Leo and Hansch $P-2^a$	CLOGP log P	CLOGP $P-2^a$
1		1,3-indandione	0.61	0.76	A[c]	1.27	N[c]
2		2-phenyl-1,3-indandione	2.90	2.31	A[c]	2.82	N[c]
3		4,5,6,7-tetrahydro-1,3-indandione			N[c]		N[c]

[a] Proximity effect between two carbonyl groups which are separated by two carbon atoms. [b] From ref 12. [c] A: aromatic type of proximity effect is considered. N: nonaromatic or alicyclic type of proximity effect is considered.

is not in the ring. Leo and Hansch suggest regarding both polar groups as if they were in the ring. CLOGP takes the average of the proximity effect corrections calculated for the case where both are in the ring and both are not in a ring.

(c) There are two $P-2$ proximity effects: 7-OH and 8-OH, 1-O and 8-OH, involved for the following structure.



14

It has been proposed that only the proximity interaction of 7-OH (more polar than 1-O) and 8-OH should be considered, since 2-C and 3-C are involved in both effects.[12] Instead, the average of these two proximity effects will be calculated by the program.

(d) The extent of $P-2$ proximity effect between halogen and H-polar group has not been suggested by Leo and Hansch.[12] The value of +0.46 is used in CLOGP for this type interaction tentatively as proposed by Rekker.[11]

(3) The Fragment Constant for Aromatic Fused Carbon Atom. It is suggested that the value of 0.44 should be assigned to a fused aromatic carbon atom which attaches to a polar group or heteroatom, only when both rings are "aromatic".[12] However, this value will be used by the program whenever this carbon atom attaches to a ring heteroatom. Examples of calculated values using these rules are shown in Table VIII. Actually the second rule was also initially proposed by Leo. Naturally, these empirical rules are subject to further changes, if there are sufficient data to show a preference for one choice over the other. Additionally, new rules may be generated.

The fragment method, just like other methods, is based on assumptions: here, the additive-constitutive property of molecular hydrophobicity.[12] If this basic hypothesis is not

Table VIII. Comparison of Calculated Log P Values Using Different Rules for Assigning Fused Aromatic Carbon Atom

compound		log P_{obsd}	log P_{calcd} Leo and Hansch	log P_{calcd} CLOGP
	1,2-methylenedioxybenzene	2.08	1.34[a]	2.10[b]
	1,2,3,4-tetrahydroquinoline	2.29	2.27	2.58

[a] From ref 12. [b] The additional difference between these two calculations is $P - 2$.

Table IX. Discrepancy between Trends Observed for Log P Values Observed in Several Solvent Systems and Calculated for Homologs of $HOCH_2CH_2R$

no.	R	log P_{obsd} in oct/w	ether/w	oil/w	log P_{calcd} in oct/w
11	H	−0.31	−0.58	−1.52	−0.21
2	OCH_2CH_3	−0.54	−0.70	−1.72	−0.17
3	$O(CH_2)_2OCH_2CH_3$	−0.87[a]	−1.19	−2.22	−0.08

[a] This value was calculated from $\log P_{ether/water}$ using the equation $\log P_{oct/w} = 0.160 + 0.864 \log P_{ether/w}$ from the Pomona College Medicinal Chemistry Data Bank.

valid, or "pot anomaly" occurs,[12] or the "folding effect" of long alkyl chains appears, then exceptions or outliers may result. Leo and Hansch have demonstrated some outliers due to violations of this fundamental assumption.[12]

One specific discrepancy was noticed while using this program. As shown in Table IX, the trend of observed log P values in three different solvent pairs are in decreasing order from compound 1 (ethanol) to compound 3 (carbitol), whereas the estimated ones are in increasing order. That is, the calculated log P for the higher congener is obtained such that a positive value is added to the estimated value of next lower congener rather than a negative value as it should be. This positive value is due to the proximity interactions of polar fragments. This is also true for dioxane (compound 21 in Table II: obsd, −0.42; calcd, 0.01). The extent of proximity effect between OH and OH, OH and O, or O and O should be less than that for other polar groups; i.e., the lower coefficient may be preferably used. The coefficient of 0.15 was derived for this type of proximity effect. The new calculated values using this new coefficient for compounds with $O-(CH_2)_2-O$ type of proximity interaction are shown in Table X. The estimated values are in good agreement with the observed ones. This example indicates that when a discrepancy occurs, it may be due to anomalous solvation behavior of the solute, or it might indicate the necessity for a "fine-tuned" correction factor. In this case a better coefficient for a proximity effect was needed.

Table X. New Calculated Values for Log P for Compounds Involving an $O(CH_2)_2O$ Type of Proximity Effect

compound	log P_{obsd}	log P_{calcd}[a]
$HOCH_2CH_3$	−0.31	−0.21
$HOCH_2CH_2OCH_2CH_3$	−0.54	−0.55
$HOCH_2CH_2OCH_2CH_2OCH_2CH_3$	−0.87[b]	−0.86
	−0.42	−0.36
$HOCH_2CH_2OH$	−1.93	−1.83
$HOCH_2CH(OH)CH_2OH$	−2.49[b]	−2.67
$CH_3CH(OH)CH(OH)CH_3$	−0.92	−1.07

[a] The coefficient of 0.15 is used for calculation of the proximity effect. [b] See footnote of Table IX.

There are some limitations to the use of this program. First, it cannot handle any ionic, inorganic, or organometallic compounds. Only nonionic, organic compounds containing C, H, O, H, S, P, or halogens can be handled. Secondly, two of the correction factors, i.e., ring cluster and intramolecular hydrogen bonding, cannot be perceived and calculated.

It is intended to continue in the improvement and generalization of the routine in the future in order to expand the set of compounds for which it will reliably estimate log P values. However, the program in its current form will already make feasible structure–activity investigations that would have been difficult or impossible in the absence of the ability to estimate log P values.

ACKNOWLEDGMENT

This work was supported by the National Cancer Institute through Contract No. N01 CP 75926. The computer used was purchased with partial financial support of the National Science Foundation. The authors wish to express their gratitude to Albert J. Leo for his manuscript, his invaluable discussion on calculation method, and advice. The authors also wish to thank A. J. Stuper for his helpful comments.

REFERENCES AND NOTES

(1) W. Van Valkenburg, Ed., "Biological Correlations - The Hansch Approach", American Chemical Society, Washington, D.C., 1972.
(2) W. J. Dunn III and C. Hansch, *Chem.-Biol. Interact.*, **9**, 75 (1974).
(3) C. Hansch and J. M. Clayton, *J. Pharm. Sci.*, **62**, 1 (1973).
(4) C. Hansch, *Drug Metab. Rev.*, **1**, 1 (1972).
(5) K. A. S. Al-Gailany, J. B. Houston, and J. B. Bridges, *Biochem. Pharmacol.*, **27**, 783 (1978).
(6) A. Leo, P. Y. C. Jow, C. Silipo, and C. Hansch, *J. Med. Chem.*, **18**, 865 (1975).
(7) C. Hansch and T. Fujita, *J. Am. Chem. Soc.*, **86**, 1616 (1964).
(8) A. Leo, C. Hansch, and D. Elkins, *Chem. Rev.*, **71**, 525 (1971).
(9) G. G. Nys and R. F. Rekker, *Chim. Ther.*, **8**, 521 (1973).
(10) G. G. Nys and R. F. Rekker, *Chim. Ther.*, **9**, 361 (1974).
(11) R. F. Rekker, "The Hydrophobic Fragmental Constant", Pharmacochemistry Library, Vol. 1, Elsevier, New York, 1977.
(12) C. Hansch and A. Leo in "Substituent Constants for Correlation Analysis in Chemistry and Biology", Wiley, New York, in press.
(13) A. J. Stuper and P. C. Jurs, *J. Chem. Inf. Comput. Sci.*, **16**, 99 (1976).
(14) G. M. Singer, H. W. Taylor, and W. Lijinsky, *Chem.-Biol. Interact.*, **19**, 133 (1977).
(15) A. J. Hopfinger and R. D. Battershell, *J. Med. Chem.*, **19**, 569 (1976).