Atomic Physicochemical Parameters for Three-Dimensional Structure-Directed Quantitative Structure-Activity Relationships I. Partition Coefficients as a Measure of Hydrophobicity

Arup K. Ghose and Gordon M. Crippen

College of Pharmacy, University of Michigan, Ann Arbor, Michigan 48109

Received 4 November 1985; accepted 6 February 1986

Earlier we showed (A. K. Ghose and G. M. Crippen, *J. Med. Chem.*, 28, 333, 1985) the necessity of atomic physicochemical parameters in three-dimensional receptor mapping. Here we derive more refined and widely applicable hydrophobicity parameters. Carbon, hydrogen, oxygen, nitrogen, sulfur, and halogens are classified into 110 atom types. Among these, the hydrophobic contributions of 90 atom types have been evaluated from the log $P_{(\text{water-octanol})}$ values of 494 molecules, using the additive model and least-squares technique. It gave a standard deviation of 0.347, a correlation coefficient of 0.962, and an explained variance of 0.908. These atomic values were used to predict the log P values of 69 compounds. The predicted values showed a standard deviation of 0.404 and a correlation coefficient of 0.896. This work has been compared with more conventional approaches.

INTRODUCTION

There are many approaches to explaining the activity of drugs given their chemical structures and experimentally determined affinities.[1] The *ab initio* treatment of the problem is almost impossible due to the complexity of the biological system. The simplest biological data, namely, the binding energy of the drugs with the purified receptor, are hard to explain even after crystallographic determination of the structure of the receptor and the receptor-ligand complex.[2] The main source of error is the effect of the solvent and the various entropic contributions. The molecular mechanics based intermolecular energy calculation evaluates the van der Waals, electrostatic and other interatomic interactions as a function of their distances. However, this approach cannot be used in most ligand-receptor interaction studies, since either the explicit structure of the receptor is not known or the active site has not been identified even when the structure is known. In order to overcome this difficulty, in the distance geometry based 3-D QSAR,[3-5] a three-dimensional model of the active site cavity is developed. The cavity is divided into smaller regions or pockets either by some spheres or some planes[6] to study the nature of interaction at various regions. The ligand-receptor binding consists of the occupancy of the site pockets by some ligand atoms. The boundaries between the site pockets are not physical barriers, but only signal changes in the nature of interaction. When an atom enters a site pocket it interacts with it. The interaction energy may be modeled as a function of one or more physicochemical parameters of the atom. However, not all physicochemical properties can be dissected into atomic values. A proper understanding of the property is necessary before the dissection.

The advantages of atomic physicochemical parameters are: (i) A large substituent occupies a large space. Because of the differences in the nature of the site atoms surrounding such a substituent, it may experience different types of interactions at different regions. Such differences may be accounted for, if the physicochemical properties of the subfragments are used; (ii) The

Journal of Computational Chemistry, Vol. 7, No. 4, 565–577 (1986)

CCC 0192-8651/86/040565-13$04.00

relative orientation or conformation of a group may affect its interaction with the receptor, and such changes are difficult to study with overall group physicochemical properties.

To overcome those difficulties we used atomic physicochemical properties in an earlier work.[4] The present objective is to develop more refined and widely applicable hydrophobicity parameters to be used in developing the quantitative expression for ligand site interaction. Unfortunately, although hydrophobicity is one of the most important factors in the biochemical processes, it is poorly understood.[7] The partition coefficient of a molecule in the water-octanol system is assumed as the measure of its hydrophobicity.

Earlier Works

Fujita, Iwasa, and Hansch[8] first showed that partition coefficients of the molecules can be reliably estimated from their chemical structures. Later Rekker *et al.*[9,10] gave some fragmental values for calculating the partition coefficients in the water-octanol system. Finally Hansch and Leo[11] gave more elaborate tables for fragment values. Their method is based on the definitions of "isolating" and "nonisolating" carbons, "fundamental fragments" and some correction factors. The reason for defining carbons in these two types probably came from two problems: (i) the carbons which are multiply bonded to hetero atoms have very different chemical as well as electronic behavior, and (ii) since such carbons often exist uniquely in a group, linear dependency does not allow their individual evaluation and, therefore, they are "nonisolating". The fundamental fragment has been defined as a fragment involving one or more atoms so that its unsatisfied bonds lead to isolating carbons. Such restriction comes from the fact that delocalized systems having heteroatoms interact strongly to redistribute the electrons, thereby causing large shifts in their physicochemical behavior. The properties of the isolated atoms or groups cannot, therefore, be used in these delocalized systems. Even such restrictions do not allow one to evaluate the partition coefficients just by adding the fragment values. One should add various correction factors to cover the influence of one structure over another. The problem with this approach is that many fragment values have not been defined. Also, due to the correction factors, one does not know how the hydrophobicity has been distributed over the entire structure. This method may work well for evaluating the hydrophobicity of the molecule, but it is not suitable for modeling the interaction of the drug molecules with the receptor. Broto, Moreau, and Vandycke[12] tried to improve the situation by removing the correction factors. However, their method often used fragments containing multiple atoms to evaluate the partition coefficients, since they did not explicitly include hydrogen. On the other hand in an earlier work[4] we used pure atomic fragments to evaluate the partition coefficients of the molecules. Since the partition coefficient is not a simple additive property, the constitutive feature was covered by classifying each element into a number of atom types according to structural environment. The partition coefficient according to our approach can, therefore, be expressed by eq. (1).

$$\log P_{(\text{water-octanol})} = \Sigma n_i a_i \qquad (1)$$

where n_i is the number of atoms of type i and a_i is the contribution for atom type i. The most obvious problem of such an approach is the linear dependency of columns resulting from the unique classification of atom types in some structural environment. For example, if nitrogen and oxygen are given different classes uniquely in nitro group, their separate values cannot be evaluated. Linear dependency can be removed if at least one atom is defined in another structural environment. Klopman and Iroff[13] on the other hand, assumed that the structural changes alter the charge densities. Consideration of charge density to calculate the partition coefficient will take care of the constitutive factors. Their more recent work[14] showed that even neglect of charge density can fit the partition coefficient of a large number of molecules.

COMPUTATION

Classification of the Atoms

Since the partition coefficient in the water-octanol system is assumed as a measure of the hydrophobicity of a molecule, its origin is,

therefore, the difference in free energy of the solute in two phases. The free energy change is the manifestation of three factors: (i) enthalpy of interaction between the solute and solvent molecules, (ii) enthalpy of interaction between the solvent molecules, and (iii) the entropy changes. The entropy change comes mainly from the change in the structure of the solvent phase around the solute molecules.[7] The change in conformational distribution of the solute molecules may also contribute in the entropy factor. The interaction of any atom of the solute molecule with the solvent is a function of the various molecular forces as well as the approachability of the solvent. The classification of the atoms is, therefore, made to differentiate (at least partly): (i) the electron distribution around the atom, and (ii) the approachability of the solvent to the atom. The nature of the atoms attached to the atom concerned as well as to its nearest neighbors influences these factors. A carbon atom may be attached to as many as four atoms, so consideration of the nature of these atoms as well as the atoms attached to its neighbors makes the carbon into a large number of atom types. We have not considered the nature of the atoms attached to the nearest neighbors, assuming that such effects are small, in order to keep the number of atom types to a minimum. Only compounds containing carbon, hydrogen, oxygen, nitrogen, halogens, and sulfur are considered. These elements are classified into 110 atom types as shown in Table I. The classification is complete to cover most neural organic molecules. However, it may not be final, since subclassification is always possible. The effect of substitution of even the distant neighbors may not be negligible in conjugated and aromatic systems. In order to understand the positional effect of substitution, we compiled the log P values of some of these isomers (Table II). From this collection it is obvious that the changes in partition coefficient in different isomers are often within the range of experimental error and although some regularities exist in some cases, they are not uniform. For example, halophenols show the same type of changes ongoing from ortho to meta to para. However, when the hydroxy group is replaced by a similar electron releasing amino group, the changes are quite different. A thorough analysis is not possible due to the unavailability of the data. The consideration of the substitution in the neighbors is, therefore, dropped during the atom classification in the aromatic system. Some flaws in the atomic hydrophobic values evaluated that way will be rectified if the atomic charge density is considered during the correlation with the biological data.

Preparation of Data

There are two steps: (i) getting the partition coefficients of the compounds having the atom types defined in Table I, and (ii) classifying the atoms from the structure of the molecules to generate the n_i values of eq. (1). With a large number of atom types one should have a large number of compounds in the regressional analysis to get statistically significant values. The first step has been simplified by the work of Hansch and Leo.[11] The classification of the atoms with a large number of atom types has been found to be extremely error prone, if performed manually. Therefore, this step was automated by writing two sets of computer programs. The first program, CHEMSTRUC, generates the chemical structures. It can be used interactively, and its commands are comparable to CAS ONLINE substructure generation. Using interactive commands one can initiate with a ring or a chain, add a chain or a ring, fuse a ring, and alter atom types or bond types. During or after the generation of the structure, one can check the correctness of the structure by computer graphics. At present it works on Televideo 950 enhanced with Retro-Graphics RG1000 and MASSCOMP Raster Graphic systems. The program allocates the number of hydrogens attached to any atom, but one can overwrite the number of hydrogens attached to any atom to cover special cases, such as ammonium salts. The program then evaluates the total number of atoms thus generated and compares it with the input value and reports any discrepancy. This check is found to be very helpful, since even visual aid sometimes fails to detect the discrepancies in bond or atom types. The default bond type is single bond and the default atom type is carbon.

The topology of the molecule is kept in a format comparable to that used in the Cambridge Crystallographic Data File. For each

Table I. Classification of atoms, their contributions to hydrophobicity and the statistics of the study.

	Type Description[a]	Hydrophobic[b] Contribution	No. of Compounds	Frequency of Use	*t*-test
	C in:				
1	:CH3R, CH4	−0.6327	159	234	100.00
2	:CH2R2	−0.3998	80	152	100.00
3	:CHR3	−0.2793	9	10	98.78
4	:CR4	0.2202	2	2	58.18
5	:CH3X	−1.1461	52	68	100.00
6	:CH2RX	−0.9481	132	214	100.00
7	:CH2X2	0.2394	5	5	83.57
8	:CHR2X	−0.9463	25	33	100.00
9	:CHRX2	0.5822	7	7	99.99
10	:CHX3	0.7245	4	4	99.98
11	:CR3X	−1.0777	8	8	100.00
12	:CR2X2	1.1220	2	2	100.00
13	:CRX3	0.6278	21	23	100.00
14	:CX4	1.2558	3	3	100.00
15	:=CH2	−0.2633	15	19	99.96
16	:=CHR	−0.0460	24	37	69.95
17	:=CR2	0.3496	4	4	93.04
18	:=CHX	−0.3053	15	15	99.78
19	:=CRX	−0.4451	6	6	99.52
20	:=CX2	−0.1915	4	5	81.20
21	:≡CH	0.1785	3	4	74.41
22	:≡CR, R=C=R	0.1541	4	5	71.07
23	:≡CX	—	0	0	—
24	:R——CH——R	−0.0548	226	826	100.00
25	:R——CR——R	0.3345	77	86	100.00
26	:R——CX——R	−0.1153	176	312	100.00
27	:R——CH——X	0.0219	42	70	49.39
28	:R——CR——X	0.2093	6	6	81.71
29	:R——CX——X	−0.1378	28	35	98.76
30	:X——CH——X	−0.2686	4	4	83.71
31	:X——CR——X	0.7376	4	4	99.99
32	:X——CX——X	0.0339	5	5	15.63
33	:R——CH...X	0.0230	11	14	21.05
34	:R——CR...X	0.2455	9	9	94.40
35	:R——CX...X	−0.1883	6	6	76.93
36	:Al—CH=X	0.7853	1	1	95.84
37	:Ar—CH=X	0.1682	5	5	67.18
38	:Al—C(=X)—Al	−0.4349	5	6	99.85
39	:Ar—C(=X)—R	−0.2392	6	6	87.18
40	:R—C(=X)—X R—C≡X, X=C=X	−0.1703	114	124	100.00
41	:X—C(=X)—X	0.0340	44	47	47.97
42	:X——CH...X	−0.7231	6	6	100.00
43	:X——CR...X	0.2256	1	1	44.25
44	:X——CX...X	−0.2692	9	12	99.69
45	unused	—	0	0	—
	H attached to[d]				
46	:C^0_{sp3}	0.4307	152	869	100.00
47	:C^1_{sp3}, C^0_{sp2}	0.3722	387	1517	100.00
48	:C^2_{sp3}, C^1_{sp2} C^0_{sp}	0.0065	40	48	10.78
49	:C^3_{sp3}, C^2_{sp2} C^3_{sp2}, C^3_{sp}	−0.2232	62	96	100.00
50	:Heteroatom	−0.3703	292	472	100.00
51	:α-C	0.2421	72	199	100.00
52–55	unused	—	0	0	—
	O in:				
56	:alcohol	−0.0517	47	54	75.23
57	:phenol, enol carboxyl OH	0.5212	90	104	100.00
58	:=O	−0.1729	174	223	100.00
59	:Al—O—Al	0.0407	18	21	41.72
60	:Al—O—Ar, Ar_2O R...O...R, R—O—C=X	0.3410	42	42	100.00

Table I *(continued)*

61[e]	:——O	1.8020	44	106	100.00
62–65	unused	—	0	0	—
	N in				
66	:Al—NH_2	0.2658	13	13	98.70
67	:Al_2NH	0.2817	15	16	99.80
68	:Al_3N	0.3990	6	6	98.86
69	:Ar—NH_2, X—NH_2	0.4442	39	41	100.00
70	:Ar—NH—Al	1.0841	3	3	100.00
71	:Ar—NAl_2	0.6632	2	2	98.49
72	:RCO—N<, >N—X=X	0.1414	110	148	100.00
73	:Ar_2NH, Ar_3N Ar_2N—Al, R...N...R[f]	0.3493	17	17	99.98
74	:R≡N, R=N—	−0.1201	19	20	85.65
75	:R——N——R[g], R——N——X	0.1757	61	84	100.00
76	:Ar—NO_2, R——N(——R)——O[h] $RO—NO_2$	−3.1516	38	49	100.00
77	:Al—NO_2	−3.3332	6	6	100.00
78	:Ar—N=X, X—N=X	0.1709	26	28	98.77
79–80	unused				
	F attached to				
81	:C^1_{sp3}	0.4649	4	4	98.40
82	:C^2_{sp3}	−0.1701	6	10	93.89
83	:C^3_{sp3}	0.1172	20	61	100.00
84	:C^1_{sp2}	0.6035	14	23	100.00
85	:C^{2-4}_{sp2}, C^1_{sp} C^4_{sp}, X	0.4752	1	2	98.62
	Cl attached to				
86	:C^1_{sp3}	1.0723	15	20	100.00
87	:C^2_{sp3}	0.3027	5	8	99.66
88	:C^3_{sp3}	0.4108	12	29	100.00
89	:C^1_{sp2}	1.0278	48	73	100.00
90	:C^{2-4}_{sp2}, C^1_{sp} C^4_{sp}, X	0.6972	15	30	100.00
	Br attached to				
91	:C^1_{sp3}	1.0966	10	11	100.00
92	:C^2_{sp3}	0.4292	1	1	73.53
93	:C^3_{sp3}	—	—	—	—
94	:C^1_{sp2}	1.3224	24	30	100.00
95	:C^{2-4}_{sp2}, C^1_{sp} C^4_{sp}, X	0.9987	4	8	100.00
	I attached to				
96	:C^1_{sp3}	1.4334	3	3	100.00
97	:C^2_{sp3}	—	—	—	—
98	:C^3_{sp3}	—	—	—	—
99	:C^1_{sp2}	1.8282	12	12	100.00
100	:C^{2-4}_{sp2}, C^1_{sp} C^4_{sp}, X	1.0735	1	3	100.00
101–105	unused halogens				
	S in				
106	:R—SH	1.0152	6	6	100.00
107	:R_2S, RS—SR	1.0339	12	13	100.00
108	:R=S	0.0727	6	6	35.66
109	:R—SO—R	−0.3332	2	2	77.92
110	:R—SO_2—R	−0.1005	15	15	68.83

[a]R represents any group linked through carbon; X represents any heteroatom (O, N, S, and halogens), Al and Ar represent the aliphatic and aromatic groups respectively; = represents double bond; ≡ represents triple bond; —— represents aromatic bonds as in benzene or delocalized bond as the N—O bond in nitro group; ... represents aromatic "single" bond, as the C—N bond in pyrrole.

[b]Hydrophobic contribution of only one atom.

[c]Level of significance of each contribution.

[d]The subscript represents hybridization and the superscript its formal oxidation number.

[e]As in nitro, =N—Oxides.

[f]Pyrrole type structure.

[g]Pyridine type structure.

[h]Pyridine—N—oxide type.

Table II. Positional effect of the aromatic disubstituted compounds on water-octanol partition coefficients.

Compound	log P_{w-o} o—	m—	p—	comment
Bromophenol	2.35	2.63	2.65	o < m ≃ p > o
Chlorophenol	2.17	2.50	2.39	o < m > p > o
fluorophenol	1.71	1.93	1.81	o < m > p > o
Iodophenol	2.65	3.00	2.93	o < m > p > o
Nitrophenol	1.77	2.00	1.90	o < m > p > o
Hydroxybenzaldehyde	1.70	1.38	1.35	o > m ≃ p < o
Hydroxybenzoic acid	2.21	1.50	1.57	o > m < p < o
Dihydroxy benzene	0.88	0.80	0.59	o > m > p < o
Aminophenol	0.62	0.17	0.04	o > m > p < o
Bromoaniline	2.29	2.10	2.26	o > m < p ≃ o
Chloroaniline	1.90	1.88	1.83	o ≃ m > p < o
Nitroaniline	1.44	1.37	1.39	o > m ≃ p < o
Aminobenzoic acid	1.21		0.46	p < o

nonhydrogen atom, a number is kept which tells the number of nonhydrogen and hydrogen atoms attached to it. For all nonhydrogen atoms involving bonds, one number is kept which tells the atoms linked by the bond and the nature of the bond. The cyclic bond type is kept as *minus* bond order and the acyclic bonds are kept as *plus* bond order. The aromatic bonds are classified into two types: all benzene type bonds are represented as −5 and the heteroatom-carbon bond in pyrrole-type structure is represented as −6. The delocalized bonds like the N—O bonds in the nitro group or pyridine-N-oxide are represented as +7.

The above structural information is sent to the second program CLASIF to classify the atom types according to Table I. It then evaluates the n_i values of the regression eq. (1). The least-squares program first checks so that no two columns are linearly dependent, if so it removes the second column. However, in the final data structure there were no linearly dependent columns. It also removes any unused column from the data set. A generalized inverse matrix[15] program has been used for the inversion of the least-squares matrix, since it guarantees the solution even for the ill-conditioned matrices. Except for the matrix inverse routine, the program was written by the authors. The statistical test routines were written according to Zelen and Severo.[16]

RESULTS AND DISCUSSION

The water-octanol partition coefficients of 494 compounds[11] were used to evaluate 90 atomic contribution values as shown in Table I. Some atom types in Table I are currently undefined for future use. Although we have defined certain atom types, partition coefficients for molecules containing these types are not available to us. On the other hand since some atom types occur too frequently in the organic structure, their frequency of use in the data set ought to be high compared with the others. Initially, when fewer molecules were incorporated, the solution for the contributions was very unstable. Often introduction of a few more compounds changed the atomic contribution values greatly. We continued to introduce more compounds until a stable solution was attained.

The most unusual feature of the present calculation is the hydrophobicity values of the carbons. According to Hansch and Leo the hydrophobicity is distributed almost equally among the carbons and hydrogens, so far as the base value is concerned. However, they did not specify the region undergoing change in hydrophobic behavior due to branching. On the other hand, the present method suggests that the hydrophobicity comes from hydrogen. Although that seems to be very unusual, one should remember that saturated carbon is always shielded from the solvent by the atoms attached to it. Therefore, its behavior may be different from the surface atoms. Once the base value has been accepted, it is easy to understand the trend in the change of their values. Carbon substitution increases the hydrophobicity in general, while one heteroatom decreases its value. More than one heteroatom increases hydrophobicity. However, the contributions having more than one het-

eroatom attached to a carbon have been determined from a limited number of compounds due to their scarcity, and also their statistical test of significance often is not very good. This makes it difficult to comment on these values. Most hydrogens in the present classification are strongly hydrophobic. The basis of the present classification was that the charge density over the hydrogen is determined by either the formal charge over the carbon or its hybridization state. The greater the electron density over the hydrogen, the higher will be its hydrophobicity values. This interpretation is also reflected in the observed values. The halogens were classified along similar lines, however, they do not have any monotonic trend in their values. It seems that other factors are also important here, like approachability of the solvent. All divalent sulfurs are hydrophobic in nature; sulfuryl and sulfonyl sulfurs are weakly hydrophilic.

The oxygen and nitrogen classification should be done with care since most drug molecules contain these atoms in various structural environments, and often they are responsible for the biological interaction. In the present work only six types of oxygen and 13 types of nitrogen have been used. Oxygens are in general weakly hydrophobic to weakly hydrophilic, with one exception, namely oxygen with a delocalized bond, as in nitro and aromatic heterocyclic *N* oxides, is strongly hydrophilic. It may or may not be true. The problem here is that the present data need the aromatic nitro groups to be weakly hydrophobic, while the *N*-oxide group is considerably hydrophilic. However, their nitrogens and oxygens have the same types, differing only in oxygen number. This is the only possible solution without giving the problem more degrees of freedom by making the nitrogens of different type. The nitrogens, otherwise, are weakly hydrophobic. The hydrophilic nature of the compounds containing nitrogen seems to come from the hydrogens attached. It is interesting to remember here that Klopman, Namboodiri, and Schochet[14] found that neglecting carbon atoms from their equation fits the partition coefficients almost equally well.

Instead of presenting the entire list of compounds used in the original data set and for prediction, we will compare our method with earlier ones. The comparison is made separately for different classes of compounds in order to understand the problem of any method in a particular class.

Saturated Hydrocarbon

According to the present approach the carbons in the saturated hydrocarbons are not all of the same type, their type and contribution toward partition coefficient changing with the extent of substitution. The number of carbon atom types used for saturated hydrocarbons is four. The hydrogens are assumed to be of the same type. Hansch and Leo,[11] on the other hand, assumed that the carbons and hydrogens are all of the same type, and instead used some correction factors due to chain bonds, ring bonds, and branching. Broto *et al.*[12] on the other hand, did not separate the hydrogen contributions, used different contribution for mono-, di-, tri-, and tetra- substituted carbons, and did not use any correction factors. Comparisons of the three methods are illustrated in Table III. The calculated values suggest that the present method as well as the method of Broto *et al.* consistently underestimate the partition coefficient values for saturated hydrocarbons. On the other hand, the Hansch approach is much better here. The reason for such discrepancy in the present method is obvious. The contribution values of these carbons came mainly from compounds containing heteroatoms. Such atoms, due to polarization, decrease the hydrophobicity of carbons and hydrogens beyond its point of attachment. However, in the present classification we approximated it by assuming that the effect is negligible if the heteroatom is not directly attached to it. Broto *et al.* have not even differentiated the heteroatoms from the carbons during the classification of carbon. Since the drug molecules, in general, contain heteroatoms, the hydrophobic contribution reported here can be used safely for the correlation.

Unsaturated and Aromatic Hydrocarbons

Unsaturated and aromatic carbons in the present approach have different types depending on the extent of substitution. Hydrogens are also differentiated according to their point of attachment. If attached to a saturated

Table III. Calculation of the partition coefficient of saturated hydrocarbon.

Compound	Method[a]	Expression Used[b]	log P_{calc}	log P_{obs}	Δ[c]
Methane	I	1[1] + 4[46]	1.09	1.09	0.00
	II	1[C] + 4[H]	1.12		0.03
	III	d			
Propane	I	2[1] + 1[2] + 8[46]	1.78	2.36	−0.58
	II	3[C] + 8[H] + $1F_b$	2.36		0.0
	III	2[3] + 1[2]	1.718		−0.642
Neopentane	I	4[1] + 1[4] + 12[46]	2.858	3.11	−0.252
	II	5[C] + 12[H] + $3F_b$ + $2F_{cBr}$	3.14		0.03
	III	1[1] + 4[3]	2.722		−0.388
Cyclopentane	I	5[2] + 10[46]	2.308	3.00	−0.692
	II	5[C] + 10[H] + $4F_b$	2.94		−0.06
	III	5[2]	2.28		−0.72

[a]I, present; II, Hansch *et al.*; III; Broto *et al.*
[b]For the notations see the original references.
[c]Calculated-observed.
[d]Since the hydrogens are not separated in this approach, a carbon attached to four hydrogen occurs uniquely in methane, Broto *et al.*, therefore did not define carbon of this type.

carbon it has one type, whereas if attached to an unsaturated or aromatic carbon it has another type. Hansch and Leo considered unsaturated hydrocarbons as saturated and introduced some correction factors to evaluate their partition coefficient. Aromatic carbons can also be given different types, while no differentiation is made on hydrogen. The carbon-hydrogen combination in unsaturated aliphatic and aromatic systems differs in the method of Broto *et al.* The partition coefficients of some unsaturated and aromatic hydrocarbons are evaluated in Table IV.

Monofunctional Compounds

In order to bypass the linearity problem, the effect of substitution by heteroatoms on a carbon is assumed to be independent of the nature of the atom. In addition, it is necessary to define some atom types in more than one functional group, which is often very difficult. The carbon and nitrogen in a cyanide group, for example, have a very unique environment which is hard to find in any other structural unit. In the present work the carbon atom has been equated with the carbon in carboxylic acids and derivatives, and with carbon which is double bonded to two hetero atoms. Such approximation may be poor. In the first the number of atoms bonded to it and the overall geometry is different although the formal oxidation number is the same. In the second, the geometry is the same, as is also the number of attachments, but the formal oxidation number and the polarization of the bonds may be different. Although the nitrogen was equated with the imino nitrogen, it was not necessary from the linearity point of view. In the approach of Hansch and Leo the fragmental value of most functional groups are given, and these values differ depending on whether they are attached to aliphatic or aromatic system. Their method of calculation still needs bond correction for the bonds beyond the functional group and for branching. Broto *et al.* differentiate the carbons on the basis of hybridization and nonhydrogen atom attachment and do not differentiate between carbon and heteroatom while considering the nonhydrogen atoms. Heteroatoms form polarized bonds with carbon and affect its electron distribution. Such effects are not reflected in their classification. The total change goes to the heteroatom. The partition coefficients of some monofunctional compounds are evaluated in Table V.

Polyfunctional Compounds

The parameters which can explain the partition coefficients of the monofunctional groups will also explain that of the polyfunctional compounds if the electronic interaction between the groups or the shielding of the groups from the solvent due to their mutual existence is negligible. The present approach uses different atom types when the substitution is on the same atom and assumes that the effect of the substituents beyond that

Table IV. Calculation of the partition coefficient of unsturated hydrocarbon.[a]

Compound	Method	Expression Used	log P_{calc}	log P_{obs}	Δ
I-Butene	I	1[1] + 1[2] + 1[15] + 1[16] + 5[46] + 3[47]	1.928	2.40	−0.472
	II	2[CH3] + 2[CH2] + $2F_b$ + $F_=$	2.31		−0.08
	III	1[3] + 1[2] + 1[9] + 1[10]	1.929		−0.471
t-Butyl-benzene	I	3[1] + 1[4] + 5[24] + 1[25] + 9[46] + 5[47]	4.175	4.11	0.065
	II	1[C6H5] + 3[CH3] + 1[C] $3F_b$ + $2F_{cBr}$	3.95		−0.16
	III	1[5] + 5[6] + 3[3] + [1]	3.741		−0.369
1,3-Buta-diene	I	2[15] + 2[16] + 6[47]	1.615	1.99	−0.375
	II	2[CH3] + 2[CH2] + $2F_b$ + $F_=$	1.76		−0.23
	III	2[9] + 2[10] + 2[222]	2.224		0.234
Anthracene	I	10[24] + 4[25] + 10[47]	4.512	4.45	0.062
	II	10[*CH*] + 4[*C*]	4.38		−0.07
	III	10[6] + 4[11]	4.106		−0.344
Piperidine	I	3[2] + 2[6] + 6[46] + 4[47] + 1[50] + 1[67]	0.889	0.85	0.039
	II	5[CH2] + 1[NH] + $5F_b$	0.70		−0.15
	III	5[2] + 1[91]	0.759		−0.091
2,5-Dihydro-furan	I	2[6] + 2[16] + 6[47] + 1[59]	0.286	0.46	−0.174
	II	4[CH2] + 1[O] + $4F_b$ + $1F_=$	−0.09		−0.55
	III	2[2] + 2[9] + 1[60]	0.116		−0.344
Acetophenone	I	1[1] + 5[24] + 1[25]+ 1[39] + 5[47] + 3[51] + 1[58]	1.603	1.58	0.023
	II	1[C6H5] + 1[CO] + 1[CH3] + $1F_b$	1.58		0.00
	III	5[6] + 1[5] + 1[14] + 1[13] + 1[3] + 2[222]	1.592		0.012
Crotonic	I	1[1] + 2[16] + 1[40] + 3[46] + 2[47] + 1[50] + 1[57] + 1[58]	1.119	0.72	0.399
	II	1[CH3] + 2[CH2] + 1[CO2H] + $2F_b$ + $1F_=$	0.31		−0.41
	III	1[3] + 2[9] + 1[13] + 1[14] 1[15] + 2[222]	0.672		−0.048

[a]See the footnote of Table III for some explanations on the columns.

is negligible. The substitution process can be viewed in two distinct ways, namely, the point of substitution and the incoming group. The point of substitution when carbon, is classified according to the hybridization and number of carbon and heteroatom attachments. These factors mostly govern the approachability of the solvent and the extent of interaction with the solvent. The effect of conjugation has been largely overlooked in the classification except that some groups have been classified according to their attachment to aromatic and nonaromatic system. During the classification of halogens the formal oxidation number of the carbon to which it is attached is considered, since the charge density on carbon governs the polarity of the carbon-halogen bond. Hansch and Leo used some correction factors to cover these effects. The problem in their approach is that many heterocyclic and conjugated systems involving heteroatoms cannot be fragmented down to simpler systems using the criterion of fundamental fragments and, therefore, cannot be used to estimate their partition coefficient values. Further complications arise when this rule is violated in the examples worked out. A saturated carbon attached to more than two heteroatoms is not an isolating carbon and hence cannot be considered as a fundamental fragment, yet in some of their examples it has been considered as a fundamental fragment. The method becomes worse in complex heterocyclic aromatic systems, since in order to decide the fundamental fragments and isolating carbons, it is necessary to consider all the resonating structures in which there is no charge separation. The cor-

Table V. Calculation of the partition coefficient of monofunctional compounds[a].

Compound	Method	Expression Used	log P_{calc}	log P_{obs}	Δ
Ethylbromide	I	1[1] + 1[6] + 3[46]+ 2[47] + 1[91]	1.552	1.61	0.058
	II	1[CH3] + 1[CH2] + 1[Br] + $1F_b$	1.56		− 0.05
	III	1[3] + 1[2] + 1[44]	1.701		0.097
t-Butyl-alcohol	I	3[1] + 1[11] + 9[46]+ 1[50] + 1[56]	0.47	0.35	0.128
	II	3[CH3] + 1[C] + 1[OH]+ $3F_b$ + $2F_{gBr}$	0.43		0.08
	III	1[1] + 3[3] + 1[19]	0.597		0.247
n-Butyl amine	I	1[1] + 2[2] + 1[6] + 7[46] 2[47] + 2[50] + 1[66]	0.904	0.97	− 0.066
	II	1[CH3] + 3[CH2] + 1[NH2]+ $3F_b$	0.97		0.00
	III	1[3] + 3[2] + 1[22]	0.913		− 0.057
sec-Butyl amine	I	2[1] + 1[2] + 8[46]+ 1[47] + 2[50] + 1[66]	0.731	0.74	− 0.009
	II	2[CH3] + 1[CH2] + 1[CH] 1[NH2] + $3F_b$ + $1F_{gBr}$	0.75		0.01
	III	2[3] + 1[2] + 1[4] + 1[23]	0.86		0.12
Acetanilide	I	1[1] + 5[24] + 1[26] + 1[40]+ 5[47] + 1[50] + 3[51] + 1[58] + 1[72]	0.993	1.16	− 0.167
	II	1[C6H5] + 1[CONH]+ [CH3] + F_b	1.16		0.0
	III	5[6] + 1[5] + 1[78] + 1[13] + 1[14] + 1[3]	1.282		0.122

[a]See the footnotes of Table III for some explanations on the columns.

rection factors make the approach so complex that they forgot to note that some of them exist uniquely in a single compound. For example, the $F_{H/S}$ factor between $-CONH_2$ and three α chlorine exists in only one compound. No wonder such compounds fit exactly, since the factor does not have any predictive feature.* The approach of Broto *et al.* is comparable to the present method, and the interaction between the groups is considered by defining the atoms in different structural environments. Since the halogen atoms are classified on the basis of nonhydrogen atoms attached to the adjacent carbon without differentiating between carbon and heteroatom, it does not consider the effect of one polarized bond over another. It only considers the approachability of the solvent molecules toward the halogen. The examples in Table VI compares the three methods in estimating the partition coefficient values.

A survey of the performance of the present method is given in Table VII. A survey like this may help in future classification of the atoms. If a particular class of compounds has consistently positive deviation, it is obvious that one or more atom types defined in this class of compounds have lower values when occurring in other classes. In other words, the classification is not perfect. If certain compounds in a particular class have large positive or negative deviations, those compounds should be checked for some special structural environment that has been neglected during atom classification.

The selection of the compounds used for the prediction was unbiased, except that they all had more carbon atoms than the 494 compounds used in the original data set. This implies that the prediction is made in the extrapolated region, so far as the carbon content is concerned. The predicted values gave a correlation coefficient of 0.896 and a standard deviation of 0.404.

*In a private communication Hansch and Leo gave a more concise definition of isolating carbon. "An I.C. is one which is not doubly or triply bonded to a hetero atom." This clarifies the situation with regard to the I.C. bonded to as many as four chlorines, but it changes the definition with regard to an aromatic carbon between two aromatic ring nitrogens; such a carbon is isolating since an aromatic bond is neither double nor triple. They have also designated the *H/S* polar factors as either *XCY* or *XCCY*, where *X* stands for a halogen, *Y* for an H-bonding polar fragment, and the number of I.C.'s between them shown with *C*'s. These have been represented in their CLOGP-3.33 computer program.

Table VI. Calculation of the partition coefficient of polyfunctional compounds[a].

Compound	Method	Expression Used	log P_{calc}	log P_{obs}	Δ
Adenine	I	1[29] + 1[30] + 1[31] + 1[34] 1[42] + 2[49] + 3[50]+ 1[69] + 1[73] + 3[75]	− 0.383	− 0.09	− 0.29
	II[b]	1[N=CH—N=] + 1[*C] + 2[*C] 1[NH—CH=N] + 1[NH2] + $2F_{p1}$ + $2F_{p2}$	− 0.946		
	III	—	—	—	
Hexachloro-benzene	I	6[26] + 6[84]	5.475	4.13	1.345
	II	6[C] + 6[Cl]	6.42		2.29
	III	6[30]	5.01		0.88
Acetyl-acetone[c]	I	2[1] + 1[16] + 1[19] + 3[46] + 3[46] + 1[47] + 1[50] + 3[51] + 1[57] + 1[58]	0.177	0.34	− 0.163
	II	2[CH3] + 1[CH] + 1[CH2]+ 1[CO] + 1[OH] + $4F_b$ + $1F_{p2}$ + $1F_=$ + $1F_H$	− 1.62		− 1.96
	III	2[3] + 1[15] + 2[13] + 1[9] 1[14]	1.212		0.872
Acetyl-acetone[c]	I	2[1] + 1[2] + 2[38]+ 8[51] + 2[58]	− 0.771	0.34	− 1.111
	II	2[CH3] + 2[CO] + 2[CH2] + 1[CO]	− 0.224		− 0.564
	III	1[2] + 2[3] + 2[13] + 2[14]	− 1.369		− 1.709
3,5-Dibromo-1,2,4-triazole	I	2[44] + 1[50] + 1[73] + 2[75] + 2[95]	1.789	2.24	− 0.451
	II	[d]			
	III	2[83] + 1[107] + 2[13] + 2[40]	1.646		− 0.594
2-Amino-1,3,4—thiadiazole-sulfonamide	I	2[44] + 4[50] + 2[58] + 1[69] + 1[72] + 2[75] + 1[107] + 1[110]	− 0.495	− 0.90	0.405
	II	[d]			
	III	2[83] + 1[126] + 2[13] + 1[27]+ 1[212] + 2[33] + 1[26] + 2[222]	− 0.775		0.125
N-nitroso-morpholine	I	4[6] + 8[47] + 1[58] + 1[59] + 1[72] + 1[78]	− 0.635	− 0.44	− 0.195
	II	[O] + 4[CH2] + [NNO]+ $5F_b$ + $2F_{p2}$	− 1.186		− 0.746
	III	4[2] + 1[60] + 1[166] + 1[82] + 1[33]	− 1.224		− 0.784
1,1,3-Trimethyl-3-nitroso-urea		3[5] + 1[41] + 9[47] + 2[58] 2[72] + 1[78]	0.053	0.36	− 0.307
	II	[e]			
	III	3[3] + 1[174] + 1[13] + 1[14] + 1[202] + 1[82] + 1[33]	0.804		0.444

[a]See the notes of Table III for some explanations of the columns.
[b]Four H-polar proximity effects are considered, two for the amino substituent and two within the ring. A fifth F_{p2} within the five-membered ring is doubtful and hence not considered.
[c]First on the basis of enol form since in aqueous phase enol form constitutes 76% of the total, second on the basis of diketo form.
[d]This molecule cannot be fragmented using the definition of fundamental fragment.
[e]The value of one fundamental fragment is not available.

It will be unwise to make a concrete comment on the overall performance of the methods on the basis of the few compounds that have been tested. However, we found that the approach of Hansch works very well for simple molecules and often very poorly for complex molecules. The present method as well as that of Broto *et al.* works fairly well for most molecules. The present method gives a total atomic distribution of the partition coefficient. The method of Broto *et al.* gives a partial distribution, since it does not include hydrogens explicitly. The Hansch approach is important for getting the overall hydrophobicity and does not give a good picture of its distribution, since the correction factors often have large values and do not point out the atom or group undergoing changes.

It may be interesting to consider the methods in terms of the number of parameters

Table VII. Nature of deviation between the calculated and observed log P values in different classes of compounds.

Class	No. of Compounds	Range of $\Delta_{calc-obs}$	Comment
Saturated hydrocarbons	8	−0.687(n-pentane) 0.000(methane)	The value of Δ decreases with the increase of C. Differences are all on the negative side.
Unsaturated hydrocarbons	12	−0.472(1-Butene) 0.596(1, 4-pentadiene)	Δ's are mostly on the negative side.
Monofunctional compounds			
Halides	17	−0.244(o-$CH_3C_6H_4Cl$) 0.133(CH_3Cl)	Δ's are slightly inclined on the negative side.
Alcohols and phenols	17	−0.147(3-methyl-2-butanol) 0.335(ethanol)	Δ's are inclined more on the positive side.
Amines	14	−0.124(n-pentylamine) 0.275(aniline)	Δ's are inclined slightly towards the positive side.
Carboxylic acid	6	−0.141(benzoic acid) 0.399(crotonic acid)	All aliphatic acids have positive Δ values.

used to evaluate the partition coefficients. Hansch and Leo used more than 200 fragments and correction factors, yet they cannot cover many structural units containing heteroatoms. The problem comes from the restrictions on the fundamental fragments. For urea derivatives, for example, one needs five different fragments (−HNCONH2, >NCONH2, −NHCONH−, >NCONH−, and >NCON<) to cover the structures in which the substituents are attached by "isolating carbon". If the attachment is not through isolating carbon, fragmentation is not possible, and hence even these fragment values become invalid. Broto *et al.* used 222 fragment or atomic values, but can cover most of the molecular structures. The present method used only 90 atom types. Four defined atom types could not be used due to the unavailability of the partition coefficient values of the required molecules. It should, however, be admitted that it is very difficult to select a molecule having a particular atom type from a long list, especially when the atom type is rare. These 90 parameters fit and predict the partition coefficient values remarkably well. However, in the future the heteroatoms may be subdivided according to conjugation effects to get even more reliable parameters. Phosphorus has not been considered, although the importance of phosphorus compounds in agricultural chemistry demands its inclusion in the near future.

CONCLUSION

The present work clearly suggests that a proper classification of the atoms can represent the partition coefficient of a large number of molecules simply as an additive property.* It also encourages one to improve and develop some other atomic physicochemical properties that can handle other kinds of molecular forces, such as electrostatic, and van der Waals.

The greatest advantage of the present method is the ease of documentation. The dissection of the molecule and the classification of the atoms are unambiguous and simple to computerize. The method of Broto *et al.* also has the same advantage. The computerization of the approach of Hansch and Leo is difficult. Although Chou and Jurs[17] did that, many values computed by their program varies considerably from the values calculated by Hansch and Leo.

*The constitutive factor here is kept hidden in the atom classification.

This work was supported by a grant from the National Institute of Health (5-R01-GM30561-02), the National Science Foundation (PCM-8314998), and the Robert A. Welch Foundation. Acknowledgment is also made to the Donors of the Petroleum Research Fund, administered by the American Chemical Society, for partial support of this research. One of the authors (A.K.G.) thanks Dr. G. Moreau for sending some worked out examples of their method, which removed some confusion. Thanks are also due to Drs. Hansch and Leo for informing us of their latest developments in the calculation of water-octanol partition coefficients.

References

1. (a) Y.C. Martin, *Quantitative Drug Design: A Critical Introduction,* Marcel Dekker, New York, 1978. (b) R. Franke, *Theoretical Drug Design Methods,* Akademie-Verlag, Berlin, 1984.
2. J.M. Blaney, P.K. Weiner, A. Dearing, P.A. Kollman, E.C. Jorgensen, S.T. Oatley, J.M. Burridge, and C.C.F. Blake, *J. Am. Chem. Soc.,* **104,** 6424 (1985).
3. A.K. Ghose and G.M. Crippen, *J. Med. Chem.,* **27,** 901 (1984).
4. A.K. Ghose and G.M. Crippen, *J. Med. Chem.,* **28,** 333 (1985).
5. A.K. Ghose and G.M. Crippen, *J. Comput. Chem.,* **6,** 350 (1985).
6. G.M. Crippen, *N.Y. Acad. Sci.,* **439,** 1 (1985).
7. G. Nemethy, *Angew. Chem. Int. Ed.,* **6,** 195 (1967).
8. T. Fujita, J. Iwasa, and C. Hansch, *J. Am. Chem. Soc.,* **86,** 5175 (1964).
9. G.C. Nys and R.F. Rekker, *Chim. Therap.,* **8,** 521 (1973).
10. R.F. Rekker, *The Hydrophobic Fragmental Constants,* Elsevier, New York, 1977.
11. C. Hansch and A. Leo, *Substituent Constants for Correlation Analysis in Chemistry and Biology,* Wiley, New York, 1979.
12. P. Broto, G. Moreau, and C. Vandycke, *Eur. J. Med. Chem.-Chim. Ther.,* **19,** 71 (1984).
13. G. Klopman and L. Iroff, *J. Comput. Chem.,* **2,** 157 (1981).
14. G. Klopman, K. Namboodiri, and M. Schochet, *J. Comput. Chem.,* **6,** 28 (1985).
15. E.V. Krishnamurthy and S.K. Sen, *Computer Based Numerical Algorithms,* Affiliated East-West Press, New Delhi, 1976, p. 187.
16. M. Zelen and N.C. Severo, in *Handbook of Mathematical Functions,* M. Abramowitz and I.A. Stegun, Eds., Dover Publications, New York, 1972, pp. 925–964.
17. J.T. Chou and P.C. Jurs, in *Physical Chemical Properties of Drugs,* S.H. Yalkowsky, A.A. Sinkula, and S.C. Valvani, Eds., Marcel Dekker, Inc., New York, 1980, p. 163.