Atomic Physicochemical Parameters for Three Dimensional Structure Directed Quantitative Structure-Activity Relationships III: Modeling Hydrophobic Interactions

Arup K. Ghose[*][#], **Avis Pritchett**[†], **and Gordon M. Crippen**[*]

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

In an earlier article[8] the need was demonstrated for atomic physicochemical properties for three dimensional structure directed quantitative structure-activity relationships, and it was shown how atomic parameters can be developed for successfully evaluating the molecular octanol-water partition coefficient, which is a measure of hydrophobicity. In this work we report more refined atomic values of octanol-water partition coefficients derived from nearly twice the number of compounds. Carbon, hydrogen, oxygen, nitrogen, sulfur and halogens are divided into 110 atom types of which 94 atomic values are evaluated from 830 molecules by least squares. These values gave a standard deviation of 0.470 and a correlation coefficient of 0.931. These parameters predicted the octanol-water partition coefficient of 125 compounds with a standard deviation of 0.520 and a correlation coefficient of 0.870. There is only a correlation coefficient of 0.432 between the atomic octanol-water partition coefficients and the atomic contributions to molar refractivity over the 93 atom types used for both the properties. This suggests that both parameters can be used simultaneously to model intermolecular interactions. We evaluated the CNDO/2 gross atomic charge distribution over several molecules to check the validity of our classification. We found that the charge density on the heteroatoms in conjugated systems is strongly affected by the presence of similar atoms in the conjugation which suggests it should be incorporated as a separate parameter in evaluating the partition coefficient.

INTRODUCTION

In the process of drug design, medicinal chemists determine the biological activity of some closely related compounds. When the biological system is complex, the biological activity is a consequence of a large number of very complex steps. Modeling such a system is extremely difficult unless the steps are examined individually. One very important step in biological activity is the binding of the ligand (drug) with a biological receptor, which depends on the (generally unknown) three dimensional structure of the receptor. The ultimate objective of any quantitative structure-activity relationships is to deduce the complementary features of the receptor by correlating the structural and physico-chemical properties of the ligands with their biological activities. Explanation of the simplest biological data, namely the binding energy of the ligand with the purified receptor involves: (i) the three dimensional structure of the biological receptor, (ii) knowledge of the active site,[1] (iii) the interaction of the biophase with the ligand/receptor, and most important (iv) the interaction of the ligand with the receptor.[2] Each process has its energetic (enthalpic) and entropic contribution. The energetic contribution often is easier to model than the entropic part. Entropy is related to the flexibility of the ligand and the receptor as well as the structural randomness of the biophase around the ligand and the receptor before and after binding. The complexity of these processes leads to very slow development of the rigorous approach and suggests finding some method that can give us a rough model of the active site.

Most QSAR approaches are accepted as a working tool where the three dimensional structure of the receptor is unknown or the

---

*To whom correspondence should be sent.
#Present address: Department of Molecular Modeling and Computers, Nucleic Acid Research Institute, Costa Mesa, CA 92626.
†Department of Pharmacy, Mercer University, 345 Boulevard NE, Atlanta, GA 30312.

active site has not been identified. Where such information is known, QSAR may still be done to check the validity of the conclusions drawn from such calculations.[3] However, a more appropriate calculation in such a situation is molecular mechanics[4] or molecular dynamics.[5] The inherent hypothesis in QSAR approaches is that we know the relative location of one or more groups in the ligands at the active site. If the nature of the groups is the only difference in the ligands, the correlation between binding and the physicochemical properties of the groups may give insight into the nature of the complementary receptor site, since different physicochemical properties are responsible for different types of intermolecular interaction. There are three major forces that are important in biochemical ligand binding: hydrophobic, dispersive, and electrostatic interactions. Molar refractivity is related to dispersive forces, and the molecular orbital charge distribution or the electrostatic potential at the van der Waals radius may be used for modeling the electrostatic interaction. However, the hydrophobic interaction, although probably the most important factor for biochemical interaction, is least understood.[6] The term *hydrophobic interaction*[7] refers to the force or the corresponding energy that operates between two or more nonpolar solutes in liquid water. Although the theoretical work on hydrophobic interactions led to a clear understanding of the molecular structure of aqueous solution, it has hardly begun to build a satisfactory theoretical description of the process that has a wide range of practical applicability. In such a situation, medicinal chemists try to model this interaction using a physicochemical property which closely parallels hydrophobicity, namely the partition coefficient of the ligand molecules between water and a nonpolar solvent (usually *n*-octanol). This property, in fact, represents nonregiospecific dispersive and electrostatic forces and the consequent entropic factor. However, the objective of this article is not to consider the theoretical aspect of hydrophobicity, but *rather to develop a different (atomic) form of an already established parameter that has been successfully used to represent the hydrophobic interaction.* This will not only help to understand the distribution of the property throughout the molecule,

but it will help to differentiate the interaction between the ligand and receptor at different regions. The advantages of atomic physicochemical parameters over more conventional group parameters have already been discussed.[8]

METHODS

Three important factors[9] affect the solubility of a molecule in a solvent: (1) the energy necessary to form a molecular sized cavity in the solvent against the intermolecular force within the solvent, (2) dispersive interaction between the solvent and the solute and (3) the electrostatic interaction between the solute and the solvent. The partition coefficient of a solute in two solvents may be approximated as the ratio of their solubilities.[10] The logarithm of partition coefficient, on the other hand, is directly related to the change in free energy during the transfer of the solute from one solvent to the other. One way of understanding the octanol-water partition coefficient is to correlate it with more fundamental physicochemical properties, like molar volume, formal charge density, and polarizability. With some modification, the atomic values of these fundamental properties can then be used to get the atomic contributions to partition coefficient. Although this approach is scientifically attractive, there are several problems with it, particularly, the conformational dependency of these fundamental properties for conformationally flexible molecules. The alternate approach is to express the octanol-water partition coefficient in terms of the chemical structure of the ligand. Rekker et al.[11] first gave some fragmental values for calculating the partition coefficient from the chemical structure of the molecules. Hansch and Leo[12] followed the same direction and gave a thorough list of fragmental values and also a large number of correction factors to account for various intramolecular interactions. It is due to these correction factors that regional contributions towards the partition coefficient are difficult to evaluate. Our approach to avoiding correction factors was to evaluate the hydrophobicity on an individual atom basis,[8] accounting for the undeniable intramolecular interactions by employing a large number of atom types. Unfortunately, intramolecular inter-

action is a continuous function, and its discretization in terms of atom types should be considered an approximation. Therefore, our formula for estimating the octanol-water partition coefficient is:

$$\log P_{ow} = \sum n_i a_i \quad (1)$$

where n_i is the number of atoms of type i, and a_i is the contribution of the corresponding atom type.

The various types of atoms defined for this study are indicated in Table I. The factors that led us to this classification are: (i) valence geometry (hybridization) of the atom, (ii) formal charge density on the atom, (iii) approachability of the solvent molecule towards the atom, and (iv) the linear independency of the columns of the data matrix arising from Eq. (1).

The number of a particular atom type in a molecule can be evaluated from its chemical

Table I. Classification of atoms, and their contributions to octanol-water partition coefficient which is a measure of hydrophobicity.

Type	Description[a]	Hydrophobic[b] Contribution	No. of Compounds	Frequency of Use	T-test	Molar Refraction[c]
	C in:					
1	:CH$_3$R, CH$_4$	−0.6037	360	548	100.00	2.3000
2	:CH$_2$R$_2$	−0.4295	216	454	100.00	2.3071
3	:CHR$_3$	−0.3426	45	50	100.00	2.4926
4	:CR$_4$	−0.1155	24	24	74.32	2.3000
5	:CH$_3$X	−1.0578	157	224	100.00	3.4006
6	:CH$_2$RX	−0.8188	257	402	100.00	3.2624
7	:CH$_2$X$_2$	−0.1540	5	5	51.00	3.6770
8	:CHR$_2$X	−0.5995	73	118	100.00	3.0137
9	:CHRX$_2$	0.0095	27	27	7.85	3.225
10	:CHX$_3$	0.5134	4	4	96.02	3.2401
11	:CR$_3$X	−0.4807	14	14	99.97	2.6140
12	:CR$_2$X$_2$	0.2853	2	2	58.14	3.1488
13	:CRX$_3$	0.5335	34	36	100.00	2.3010
14	:CX$_4$	1.1114	6	6	100.00	3.3559
15	:=CH$_2$	−0.1654	25	31	97.01	3.5071
16	:=CHR	−0.1033	48	70	97.97	4.4814
17	:=CR$_2$	−0.2330	9	10	89.41	3.7781
18	:=CHX	−0.0649	22	23	48.45	3.6211
19	:=CRX	−0.7814	22	23	100.00	4.4310
20	:=CX$_2$	0.1734	4	5	64.21	3.2000
21	:≡CH	0.0859	3	4	32.70	3.4161
22	:≡CR, R=C=R	0.1335	4	5	52.10	4.3043
23	:≡CX	—	—	—	—	3.4905
24	:R—CH—R	−0.0220	522	1988	99.99	3.4127
25	:R—CR—R	0.1596	282	340	100.00	4.3725
26	:R—CX—R	−0.0064	381	629	35.81	3.8182
27	:R—CH—X	0.0245	92	142	56.42	2.5001
28	:R—CR—X	0.1114	62	69	95.79	2.5000
29	:R—CX—X	0.2378	62	74	100.00	2.7967
30	:X—CH—X	0.2921	18	18	98.68	2.5000
31	:X—CR—X	0.8471	6	6	100.00	—
32	:X—CX—X	0.3002	13	13	96.97	2.5000
33	:R—CH...X	0.0183	15	18	14.26	3.4372
34	:R—CR...X	−0.2625	37	42	99.98	3.4494
35	:R—CX...X	−0.2959	7	8	93.90	3.1048
36	:Al-CH=X	−0.1243	5	5	42.27	3.8251
37	:Ar-CH=X	0.3310	5	5	86.19	4.5401
38	:Al-C(=X)-Al	0.5353	25	29	100.00	3.7529
39	:Ar-C(=X)-R	−0.2182	22	28	99.53	4.1288
40	:R-C(=X)-X R-C≡X, X=C=X	0.0278	270	333	77.26	2.7938
41	X-C(=X)-X	0.3514	104	109	100.00	2.4165
42	:X—CH...X	−0.3040	17	17	98.78	3.0606
43	:X—CR...X	0.0102	29	29	8.75	2.5001
44	:X—CX...X	0.1746	10	13	87.27	2.5001
45	unused					

Table I. *(continued)*

Type	Description[a]	Hydrophobic[b] Contribution	No. of Compounds	Frequency of Use	T-test	Molar Refraction[c]
	H attached to[d]					
46	:$C^0_{sp3'}$ having no X attached to next C	0.4234	263	1487	100.00	1.1461
47	:C^1_{sp3}, C^0_{sp2}	0.3610	711	3605	100.00	0.8000
48	:C^2_{sp3}, C^1_{sp2} C^0_{sp}	0.1183	63	75	98.15	0.8006
49	:C^3_{sp3}, C^2_{sp2} C^3_{sp2}, C^3_{sp}	−0.1573	133	199	100.00	0.8001
50	:Heteroatom	−0.2106	542	944	100.00	0.8000
51	:α-C	0.1869	195	516	100.00	1.0026
52	C^0_{sp3}, having 1 X attached to next carbon	0.3546	194	690	100.00	1.1461[e]
53	C^0_{sp3}, having 2 X attached to next carbon	0.2676	8	19	99.99	1.1461[e]
54	C^0_{sp3}, having 3 X attached to next carbon	0.3528	3	7	99.64	1.1461[e]
55	C^0_{sp3}, having 4 or more X attached to next carbon	—	—	—	—	1.1461[e]
	O in:					
56	:alcohol	−0.0876	87	130	99.44	1.4430
57	:phenol, enol carboxyl OH	0.1665	134	148	100.00	1.4090
58	:=O	−0.2473	408	572	100.00	1.6506
59	:Al-O-Al	0.0380	32	36	38.64	1.2000
60	:Al-O-Ar, Ar_2O R...O...R, R-O-C=X	0.1938	159	180	100.00	1.8434
61[f]	:—O	1.4968	69	155	100.00	1.6001
62–65	unused					
	N in:					
66	:Al-NH_2	−0.2577	21	21	98.19	2.5001
67	:Al_2NH	0.0266	21	23	21.79	2.5001
68	:Al_3N	0.1680	18	19	87.69	2.5377
69	:Ar-NH_2, X-NH_2	−0.0362	78	83	51.65	3.6195
70	:Ar-NH-Al	0.0274	6	6	10.72	2.9832
71	:Ar-NAl_2	0.5799	17	17	100.00	3.9733
72	:RCO-N<, >N-X=X	−0.2736	274	361	100.00	3.0059
73	:Ar_2NH, Ar_3N Ar_2N-Al, R...N...R[g]	0.3323	80	82	100.00	2.6295
74	:R$\equiv$N, R=N-	0.1849	57	82	100.00	3.1464
75	:R—N—R[h], R—N—X	−0.0545	162	233	97.27	4.5123
76	:Ar-NO_2, R—N(—R)—O[i] RO-NO_2	−2.6143	63	75	100.00	4.7725
77	:Al-NO_2	−2.6455	6	6	100.00	3.0389
78	:Ar-N=X, X-N=X	0.5466	40	53	100.00	3.6838
79–80	unused					
	F attached to					
81	:C^1_{sp3}	0.4093	5	5	93.32	0.8060
82	:C^2_{sp3}	0.1590	8	14	89.56	0.8000
83	:C^3_{sp3}	0.1890	34	103	100.00	1.3484
84	:C^1_{sp2}	0.5035	14	23	100.00	0.8000
85	:C^{2-4}_{sp2}, C^1_{sp} C^4_{sp}, X	0.2550	1	2	69.33	1.6440
	Cl attached to					
86	:C^1_{sp3}	0.9282	18	24	100.00	5.3647
87	:C^2_{sp3}	0.4659	8	14	100.00	5.6484
88	:C^3_{sp3}	0.4381	13	32	100.00	5.6858
89	:C^1_{sp2}	0.9036	78	110	100.00	5.0000
90	:C^{2-4}_{sp2}, C^1_{sp} C^4_{sp}, X	0.5302	17	32	100.00	5.9312
	Br attached to					
91	:C^1_{sp3}	1.0474	12	13	100.00	8.3379

Table I. *(continued)*

Type	Description[a]	Hydrophobic[b] Contribution	No. of Compounds	Frequency of Use	T-test	Molar Refraction[c]
92	$:C_{sp3}^{2}$	0.5809	2	3	99.06	8.5393
93	$:C_{sp3}^{3}$	0.5407	1	3	99.88	8.8635
94	$:C_{sp2}^{1}$	1.1743	32	39	100.00	8.0866
95	$:C_{sp2}^{2-4}, C_{sp}^{1}$					
	C_{sp}^{4}, X	0.8656	4	8	100.00	9.0569
	I attached to					
96	$:C_{sp3}^{1}$	1.4378	4	4	100.00	13.7535
97	$:C_{sp3}^{2}$	—	—	—	100.00	13.6306
98	$:C_{sp3}^{3}$	—	—	—	100.00	13.4586
99	$:C_{sp2}^{1}$	1.7028	13	13	100.00	12.8876
100	$:C_{sp2}^{2-4}, C_{sp}^{1}$					
	C_{sp}^{4}, X	0.8654	1	3	100.00	13.5530
101–105	unused halogens					
	S in:					
106	:R-SH	0.7412	9	9	100.00	7.7751
107	:R₂S, RS-SR	0.7598	26	27	100.00	7.3151
108	:R=S	0.2968	9	9	92.55	9.2916
109	:R-SO-R	−0.2515	2	2	52.41	5.3957
110	:R-SO₂-R	0.0425	57	61	52.11	5.4662

[a]R represents any group linked through carbon; X represents any heteroatom (O, N, S and halogens); Al and Ar represent aliphatic and aromatic groups respectively; = represents double bond; ≡ represents triple bond; — represents aromatic bonds as in benzene or delocalized bonds such as the N–O bond in nitro group; ... represents aromatic *single* bonds as the C–N bond in pyrrole.
[b]Atomic hydrophobicity in the unit of log P(octanol-water).
[c]Atomic refractivity from reference 16.
[d]The subscript represents hybridization and the superscript its formal oxidation number.
[e]Extrapolated value from atom type 46.
[f]As in nitro, =N-Oxides.
[g]Pyrrole type structure.
[h]Pyridine type structure.
[i]Pyridine-N-oxide type.

structure and the atom type definition, and log P_{ow} is taken from the compilation of Hansch and Leo.[12] The contribution of the various atoms, a_i, was evaluated using least squares.

RESULTS AND DISCUSSION

Although the method of calculation in this approach is very straightforward, preparation of error free input data is by far the greatest challenge for this type of calculation. In order to do that, we wrote several computer programs to automate the various steps.[8] The program CHEMSTRUC, for example, allows one to generate the topology of a molecule using simple interactive commands. The correctness of the structure can be checked by graphics. It creates the topology file in the format of the Cambridge Crystallographic Data File. Another program, CLASIF, uses this topographical information to classify the

atoms according to Tab. I, and then generates the data file corresponding to Eq. (1) for a least squares program.

We generated 830 compounds from the compilation of Hansch and Leo[12] to evaluate the atomic physicochemical parameters. The structures of some of the model compounds are shown in Figure 1, and their atom classification is shown in Table II along with their observed and calculated log P_{ow}. The atomic values obtained in this study correspond very well with our earlier values. The atom classification was basically the same, with the minor modification that hydrogens attached to a saturated carbon having zero formal oxidation number were given different types to take into account the effect of somewhat distant heteroatoms (types 52–55). The statistics of the fitting is shown in Table III. The standard deviation of 0.47 is not a large value, if we remember that the experimental error[13] in the estimation of log P_{ow} is often 0.4 (and
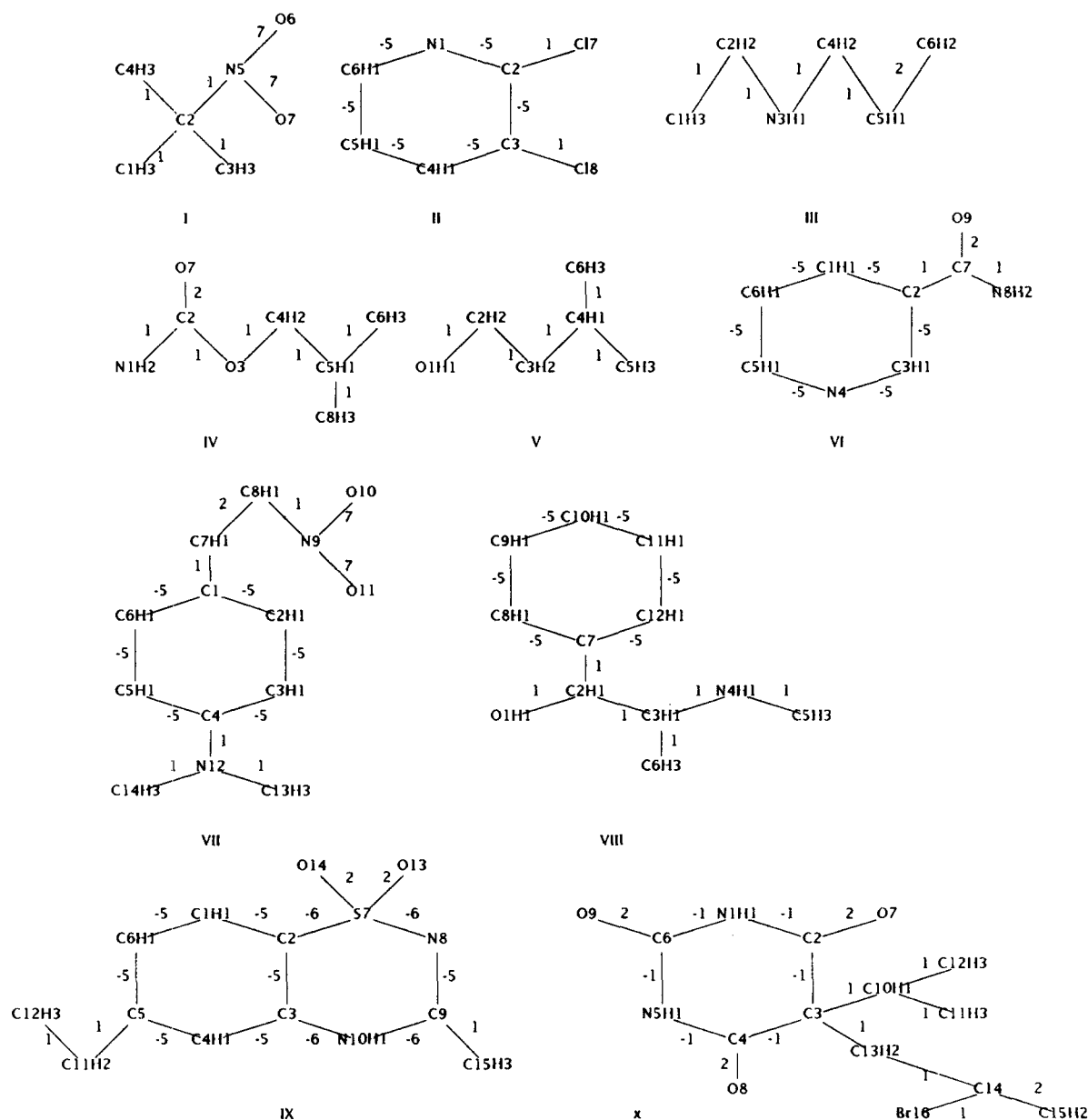
Figure 1. Schematic representation of the structures of some representative molecules. The number after the nonhydrogen atom indicates the atom label, and that after hydrogen indicates the quantity. The label number for hydrogen atoms can be obtained from their point of attachment and the number of nonhydrogen atoms. Start numbering from lower to higher nonhydrogen atoms, and then the first hydrogen should be numbered one plus the number of nonhydrogen atom. The number at the middle of the bond indicates the bond type as is defined in the Cambridge Crystallographic Data File with some modification. If a heteroatom in an aromatic ring system is attached by single bond, it is given the type -6. In general, acyclic bonds are represented by plus bond order and the cyclic bonds are represented by minus bond order. A delocalized bond, as in a nitro group, is represented as 7.

generally not given). The atomic contributions thus obtained were subjected to predict the log P_{ow} values of 125 compounds. The predicted values were equally good (see statistics in Table IV). There were 65 of the compounds used in the fitting, approximately 7.5% of the total, having deviation greater than 0.8 ≈ 1.5σ. (See Table V.)

Although it is difficult to identify specific structural properties responsible for these outliers, many fell into five classes, not mutually exclusive: (i) conjugated multiheteroatomic systems, (ii) compounds having intramolecular hydrogen bonds, (iii) tautomeric compounds, (iv) dipolar compounds and (v) compounds having hydrophobic carbon chains. Although the first type is, by far, the most frequently occurring one, the difficulty in the other classes cannot be underestimated. Dipolar compounds, such as amino

Table II. Classification of atoms in selected molecules and their calculated and observed values.

Molecule Id	Structure[a]	Atom Type (Atom List)	Obs	Calc
1508	I	1 (1,3,4), 11 (2), 52 (8-16), 61 (6,7) 77 (5)	1.01	1.25
1760	II	24 (4,5), 26 (3), 27 (6), 29 (2) 47 (9,10), 49 (11), 75 (1), 89 (8), 90 (7)	2.11	2.16
2136	III	1 (1), 6 (2,4), 15 (6), 16 (5), 47 (10,11,13-17) 50 (12), 52 (7-9), 67 (3)	0.81	0.90
2158	IV	1 (6,8), 3 (5), 6 (4), 41 (2) 46 (14-19), 47 (11,12), 50 (9,10) 52 (13), 58 (7), 60 (3), 72 (1)	0.65	0.85
2243	V	1 (5,6), 2 (3), 3 (4), 6 (2), 46 (12-18) 47 (8,9), 50 (7), 52 (10,11), 56 (1)	1.42	1.30
3109	VI	24 (1,6), 25 (2), 27 (3,5), 40 (7) 47 (10,13), 49 (11,12), 50 (14,15), 58 (9) 72 (8), 75 (4)	−0.37	−0.40
8249	VII	5 (13,14), 16 (7), 18 (8), 24 (2,3,5,6) 25 (1), 26 (4), 47 (15-19, 21-26), 48 (20) 61 (10,11), 71 (12), 76 (9)	2.67	2.83
8629	VIII	1 (6), 5 (2,3), 24 (8-12) 25 (7), 47 (14,15,17-19,23-27), 50 (13,16) 52 (20-22), 56 (1), 67 (4)	0.93	1.38
8251	IX	1 (12,15), 2 (11), 24 (1,4,6), 25 (5), 34 (2,3) 43 (9), 46 (20-24), 47 (16-18), 50 (19) 51 (25-27), 58 (13,14), 73 (10), 75 (8), 110 (7)	1.25	1.32
8378	X	1 (11,12), 2 (13), 3 (10), 4 (3), 15 (15) 19 (14), 40 (2,4), 41 (6), 46 (19-25), 47 (28,29) 50 (17,18), 52 (26,27), 58 (7-9), 72 (1,5), 94 (16)	1.47	1.22

[a]See Fig. 1 for the chemical structure of the molecules and their atom numbering.

Table III. Statistics of fit of the atomic partition coefficient

No. of Comp.	No. of Adjustable Parameters	Standard Deviation	Correl. Coeff.	Explained Variance
830	94	0.470	0.931	0.850

Table IV. Statistics of predictive power of the atomic parameters.

No. of Comp.	No. of Parameters Used	Standard Deviation	Correl. Coeff.
125	68	0.520	0.870

acids, have largely been underestimated in the evaluation of log P_{ow}, since they do not really contain the free amino or carboxylic acid groups attributed to them by the classification program. The problem in the compounds containing hydrophobic hydrocarbon chain are of two types: (i) in the small hydrocarbons the value has been underestimated, e.g., n-pentane; and (ii) the compounds having long hydrocarbon chain have been exaggerated, e.g., decanol. The reason is that the values of the contribution of the atom types present in these compounds were derived mainly from compounds containing one or more heteroatoms. Only withdrawing electrons from carbon or hydrogen makes them more hydrophilic. Since our atom classification does not distinguish heteroatomic substitution even in the second atoms the value obtained in our method is somewhat less than their normal value. However, the overestimation of hydrophobicity in compounds having long hydrocarbon chains is apparently not due to chain folding.[14] A conceivable explanation would be experimentally unsuspected aggregation of such compounds in the aqueous phase yielding an erroneously low value for hydrophobicity. In-

Table V. List of the compounds having a deviation of 0.8 or more from the observed log P_{ow}.

No.	ID[a]	Compound	Obs.	Calc.	Dev.
1	360	1,2,3-triazole-4,5-dibromo	2.44	1.08	−1.36
2	577	2-fluoroethanol	−0.92	−0.08	0.84
3	683	ethyleneglycol	−1.93	−0.79	1.14
4	711	imidazole-2,4,5-tribromo	1.96	2.78	0.82
5	723	acrylonitrile	−0.92	1.03	1.95
6	731	5-azauracil	−1.87	−0.70	1.17
7	1156	pyridazine	−0.72	0.30	1.02
8	1172	8-azaguanine	−0.71	−1.64	−0.93
9	1211	cytosine	−1.73	−0.55	1.18
10	1240	butan-2,3-dione	−1.34	0.49	1.83
11	1367	2-methyl-2-imidazoline	0.52	−0.51	−1.03
12	1754	purine-2,6,8-trichloro	3.90	2.85	−1.05
13	1785	2-nitropyridine	0.48	1.45	0.97
14	1869	pyridine-N-oxide	−1.69	−0.37	1.32
15	2005	δ-valerolactone	−0.35	0.52	0.87
16	2047	cyclopentane	3.00	2.09	−0.91
17	2178	n-pentane	3.39	2.59	−0.80
18	2308	hexachlorobenzene	4.13	5.38	1.25
19	2319	pentafluorophenol	3.23	2.43	−0.80
20	2380	1,3,5-trinitrobenzene	1.18	2.14	0.96
21	3532	4-aminophenol	0.04	0.84	0.80
22	3557	2-amino-5-carboxamide pyridine	0.70	−0.48	−1.18
23	4137	5-bromosalicylic acid	2.87	2.03	−0.84
24	4279	5-nitrosalicylic acid	2.34	1.24	−1.10
25	4474	o-hydroxybenzoic acid	2.21	1.20	−1.01
26	5872	p-methoxybenzoic acid	2.33	1.46	−0.87
27	5891	methylsalicylate	2.46	1.46	−1.00
28	8153	indole-1,2,-dimethyl	2.82	1.64	−1.18
29	8168	3-allyloxy-4-aminobenzoic acid	0.42	1.35	0.93
30	8193	1-(3-carboxypropyl)benzotriazole	0.10	0.95	0.85
31	8209	N-cyano-2(3,3-dimethyl-1-triazeno)benzamide	0.80	1.89	1.09
32	8419	fuscaric acid/5-Butylpicolinic acid/pK$_a$ = 5.9/	0.68	2.42	1.74
33	8468	1,2,4-benzothiadiazine-1,1-dioxide-3-Me-7-(N,N-dimethylsulfonamido)	0.32	−0.61	−0.93
34	8528	ω-phenylpropionaldehyde,N-methylhydrazone	1.00	2.27	1.27
35	8558	N-methyl-3-N,N-Dimethylsulfamoylphenylcarbamate	2.27	0.51	−1.76
36	8605	p-amylpyridine	3.75	2.85	−0.90
37	8706	N-pentanoylcyclobutanecarboxamide	0.75	1.62	0.87
38	8707	N-iso-pentanoylcyclobutanecarboxamide	0.48	1.53	1.05
39	8708	N-tert-pentanoylcyclobutanecarboxamide	0.53	2.06	1.53
40	8715	1-(2-chloroethyl)-3(4-methylcyclohexyl)-1-nitrosourea	3.30	2.28	−1.02
41	8733	2,4-diamino-6-dipropylamino-pyrimidine-3-oxide	2.16	1.31	−0.85
42	8747	decanoic acid	4.09	3.27	−0.82
43	8751	glucopyranoside,4-tert-butyl(Beta)	1.18	−0.73	−1.91
44	8761	decanol	1.70	3.47	1.77
45	8769	decylamine	1.92	3.09	1.17
46	8851	3-p-anisoyl-3-bromoacrylic acid	−1.15	1.65	2.80
47	8869	2,4—(NO$_2$)$_2$—C$_6$H$_3$NHN=C(CN)COOEt	4.14	3.15	−0.99
48	8880	N1-(6-Cl-3-pyridyl)sulfanilamide	2.24	1.20	−1.04
49	8883	3-acetaminoquinoline	1.92	1.03	−0.89
50	8893	4-acetyloxyaminoquinoline	0.97	1.83	0.86
51	8899	2—SO$_2$Me—4—NO$_2$—C$_6$H$_3$NHN=C(CN)COOMe	3.21	1.48	−1.73
52	8922	2-(p-hydroxybenzyl)-5-methylthiazole	2.62	1.69	−0.93
53	8936	sulfapyridine	−0.02	1.02	1.04
54	8993	antipyrene	0.28	1.42	1.14
55	9055	sulfamethoxypyridazine	0.40	1.21	0.81
56	9119	1,2,4-benzothiadiazine-1,1-dioxide-3-t-butyl-7-Cl	2.40	3.27	0.87
57	9157	1-(4-carboxybutyl)benzotriazole	0.49	1.36	0.87
58	9182	sulfathiazole,N,N-dimethyltriazene	1.04	2.02	0.98
59	9232	p-hydroxybenzoic acid,butylester	3.57	2.71	−0.86
60	9261	N,N-diethylbenzamide	1.25	2.09	0.84
61	9269	tert-valeranilide	1.99	2.87	0.88
62	9301	valeric acid,2-amino-5-phenyl	−0.36	1.61	1.97
63	9344	3-t-butylhydrochlorothiazide	0.36	1.30	0.94
64	9392	6-(N-propionylamino)penicillanic acid	−2.21	−0.35	1.86
65	9410	p-hexylpyridine	4.35	3.27	−1.08

[a] The compound number in the compilation of Hansch and Leo, ref. 11.

tramolecular hydrogen bonding is expected to increase the hydrophobicity, and although it is observed in many aromatic systems, it may not be true for aliphatic systems. For example, the log P_{ow} of 2-fluoroethanol and ethyleneglycol were underestimated. Decreased hydrogen bonding due to free rotation around single bonds may be partly responsible. Since most of these causes are ultimately related to electron redistribution, we examined the theoretical charge distribution in a large number of compounds, using CNDO/2.[15] Flexible molecules were given totally extended conformations during charge calculation. If the atom classification scheme is valid, one would expect little variation of charge density within an atom class. Table VI

shows that even for a methyl carbon attached to a carbon (type 1) the charge density varies considerably. Although all heteroatoms count equally during the classification of the carbon atom, the charge density on the carbon has been found to be greatly affected by the electronegativity of the atom. The variation in charge density in the same type is even more pronounced among heteroatoms, especially when they are in conjugation with other heteroatoms. This result in turn suggests another approach to get even more refined atomic contributions to the partition coefficient. In this approach the atoms would be classified according to the nonhydrogen atom substitution in the immediate and neighboring atoms in order to distinguish roughly

Table VI. The CNDO/2 gross atomic charge distribution on various atom types.

Atom Type[a]	Description	Formal Charge in a.u. (atom designation[b])
1	n-Hexane	−0.031 (C-1)
1	1-Acetic Acid	−0.108 (C-2)
2	n-Hexane	0.006 (C-2), 0.001 (C-3)
2	1,1-Difluorohexane	−0.061 (C-2), 0.010 (C-3)
6	1-Fluorohexane	0.206 (C-1)
6	1-Hexanol	0.143 (C-1)
24	Pyridine	0.033 (C-4)
24	Pyrrole	−0.074 (C-3)
26	Phenol	0.189 (C-1)
26	Nitrophenol	0.032 (C-1)
27	Pyrimidine	0.134 (C-4)
27	Pyridine-N-oxide	−0.001 (C-2)
33	Pyrazole	0.066 (C-5)
33	Histidine	0.004 (C-5)
38	Acetylacetone (enol)	0.271 (carbonyl C)
38	Acetone	0.268
39	Acetophenone	0.248
39	Benzophenone	0.239
40	Benzoic Acid	0.379
40	Acetic Acid	0.400
57	Hydroxylamine	−0.182
57	Benzoic acid	−0.275
57	Acetylacetone (enol)	−0.233
58	Acetic acid	−0.322
58	Acetophenone	−0.270
58	Acetaldehyde	−0.229
61	Nitrobenzene	−0.344
61	Nitroethane	−0.338
61	Pyridine-N-oxide	−0.444
69	Aniline	−0.265
69	Hydroxylamine	−0.166
73	Diphenylamine	−0.219
73	Pyrrole	−0.055
75	Pyrimidine	−0.177
75	Pyridine	−0.145
75	Pyridazine	−0.072
76	Nitrobenzene	0.481
76	Pyridine-N-oxide	0.274

[a]See Tab. I for the definition of the atom types.

[b]When the molecule has more than one of the indicated type of carbon atom, the IUPAC number is indicated within parenthesis to avoid confusion.

steric effects. The electronic effect may be modeled in terms of the charge density of the atom. Work along this line is in progress and will be reported in the near future.

MODELING THE DRUG-RECEPTOR INTERACTION

The drug-receptor free energy of binding may be divided into four parts: (i) hydrophobic interaction, (ii) dispersive interaction, (iii) electrostatic interaction, and (iv) steric repulsion. For drug design purposes, we would like to be able to at least estimate each of these factors in the course of correlating chemical structure with biological activity and predicting the binding of novel compounds. Using this approximate free energy would be easier if it consisted of a linear combination of atomic terms, and indeed this goal is within reach for each of the four factors. First, although the hydrophobic interaction is poorly understood, its definition suggests that it should be strongly correlated with the octanol-water partition coefficient. As we have already seen, that in turn can be estimated by a linear function of atomic contributions. The dispersive interaction has two limiting situations.[16] It can be either proportional to MR or $(MR \times N)^{1/2}$, where N is an empirical parameter known as the effective number of electrons. It may be a good idea to represent dispersive interactions as a linear combination of both of these quantities. Electrostatic interaction may also be considered to be a linear function of the atomic charges, provided the induction of the ligand on the receptor is minimal, and they maintain approximately the same relative position in different ligands. Lastly, if the surface of the receptor at a particular region is relatively rigid, repulsive interactions may play a major role in the binding process. This interaction depends not only on the nature of the receptor atoms and their location, but also on the variability of their positions. So far, the most detailed treatment of steric effects in binding site models are done with distance geometry techniques.[17]

Using all of these parameters is reasonable only if they are linearly independent. We found a correlation coefficient between atomic refractivity[16] and atomic partition coefficient of only 0.395, suggesting that they can be simultaneously used in a linear expression for the ligand receptor interaction. The correlation of atomic charge density with any of these properties has not been checked in general. Charge depends on the conformation of interest and therefore should be checked for the problem concerned. Similarly, the correlation coefficient between the octanol-water partition coefficient and atomic refractivity may not be low in a particular set of molecules due to the absence of many atom types or due to the difference of their frequency of occurrence.

CONCLUSION

The present study shows that the octanol-water partition coefficient of a large number of compounds can be represented as an additive function of the atom contribution. Since octanol-water partition coefficient is not a simple additive function, our method of atom classification tries to hide the constitutional factors. The predicted partition coefficients for a large number compounds were also within the experimental error. The method may give even better results if we take into account the effect of the ionization and tautomerization. However, the large difference in the charge density among various atoms of the same class suggests that a better approach may be to classify the atoms to differentiate the nature of nonhydrogen substitution in the neighboring atom only and use one parameter for charge density for each atom type. Each atom type may need a separate parameter for its charge distribution in order to account for its contribution to the partition coefficient, since the effect of charge density will also be determined by the approachability of the solvent towards the atom concerned.

References

1. D. E. Tronrud, H. M. Holden, and B. W. Matthews, Science, 235, 571–574, (1987).
2. P. A. Bash, U. C. Singh, F. K. Brown, R. Langridge, and P. A. Kollman, Science, 235, 574–576, (1987).

3. A. K. Ghose and G. M. Crippen, *J. Med. Chem.,* **28,** 333, (1985).

4. S. J. Weiner, P. A. Kollman, D. T. Nguyen and D. A. Case, *J. Comput. Chem.,* **7,** 230–252, (1986).

5. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, *J. Comput. Chem.,* **4,** 187–217, (1983).

6. A. Ben-Naim, in *Hydrophobic Interactions,* Plenum, New York, 1980.

7. L. R. Pratt, *Ann. Rev. Phys. Chem.,* **36,** 433, (1985).

8. A. K. Ghose and G. M. Crippen, *J. Comput. Chem.,* **7,** 565, (1986).

9. R. W. Taft, M. H. Abraham, G. R. Famini, R. M. Doherty, J. M. Abboud, and M. J. Kamlet, *J. Pharm. Sci.,* **74,** 807–814, (1985).

10. S. Glasstone, in *Textbook of Physical Chemistry,* p. 543, Macmillan, London, 1948.

11. G. C. Nys and R. F. Rekker, *Chim. Therap.,* **8,** 521, (1973).

12. C. Hansch and A. Leo, in *Substituent Constants for Correlation Analysis in Chemistry and Biology,* John Wiley, New York, 1979.

13. P. Broto, G. Moreau, C. Vandycke, *Eur. J. Med. Chem.-Chim. Ther.,* **19,** 61–65, (1984).

14. C. Tanford, *The Hydrophobic Effect,* 2nd ed., Wiley, New York, 1980, p. 16.

15. J. A. Pople and D. L. Beveridge, in *Approximate Molecular Orbital Theory,* McGraw-Hill, New York, 1970.

16. A. K. Ghose and G. M. Crippen, *J. Chem. Inf. Comput. Sci.,* **27,** 21–35, (1987).

17. A. K. Ghose and G. M. Crippen, *J. Med. Chem.,* **27,** 901–914, (1984).