HYDROGEN BONDING CONTRIBUTION TO LIPOPHILICITY PARAMETERS. HYDROGEN ACCEPTOR AND HYDROGEN ACCEPTOR DONOR PARAMETERS

Marvin CHARTON[1,*] and Barbara I. CHARTON[2]

Chemical Department, Pratt Institute, Brooklyn, NY 11205, U.S.A.;
e-mail: [1] mcharton@pratt.edu, [2] bc33@nyu.edu

Received July 12, 2004
Accepted December 1, 2004

Dedicated to Professor Otto Exner on the occasion of his 80th birthday in recognition of his outstanding contributions to physical organic chemistry.

In our analysis of the composition of lipophilicity parameters by the intermolecular force (IMF) model we have made use of n_H, the number of OH and/or NH bonds, as a measure of the hydrogen donor capacity of a substituent; and n_n, the number of lone pairs on O and/or N atoms in the substituent, as a measure of the hydrogen acceptor capacity of the substituent. The basis of this method is the reasonable assumption that in 55.6 molar water hydrogen bonding is maximized. The method does not account however for differences in the energy of different types of hydrogen bonds, but further assumes that these differences are to a first approximation negligible. In order to improve the model we have defined a scale of group hydrogen bonding acceptor parameters, $\acute{\eta}_{\mathrm{XHA}}$, and overall hydrogen bond parameters $\acute{\eta}_{\mathrm{XHAD}}$ from the water/1-octanol partition coefficients of AkX where Ak is alkyl. These parameters should account for both the extent of hydrogen bonding in water and for the difference in hydrogen bond energies of the various types of hydrogen bonds encountered. Correlations of log P values for $\mathrm{Ph(CH_2)}_n\mathrm{X}$, $\mathrm{X^1(CH_2)X^2}$, and substituted amino acids Xaa with the IMF equation using the $\acute{\eta}_{\mathrm{XHA}}$ and $\acute{\eta}_{\mathrm{XHAD}}$ parameters demonstrated their use. Correlation of log P values for PhX suggested that for many groups separate sets of $\acute{\eta}_{\mathrm{XHA}}$ and $\acute{\eta}_{\mathrm{XHAD}}$ values are required when they are bonded to sp^2 hybridized carbon rather than sp^3 hybridized carbon.

Keywords: Hydrogen bonds; Lipophilicity; Intermolecular force model; Partition coefficient; Substituent effects.

There are many phenomena which are dependent on the difference in intermolecular forces between an initial and a final state. They include partition, distribution, solubility, phase changes such as melting point and boiling point, and chromatographic properties such as retention times in

doi:10.1135/cccc20042147

gas chromatography, relative flow rates in paper and thin layer chromatography, and in capacity factors in high performance liquid chromatography.

The Intermolecular Force (IMF) Equation

We have been interested in the development of a quantitative model for the description of structural effects on these phenomena based on the parametrization of intermolecular forces. Then when Q_X is the quantity of interest, E'' is the energy due to the intermolecular forces, X denotes the variable structural feature, and i and f indicate the initial and final states, respectively, we obtain

$$Q_X = E''_f - E''_i = \Delta E'' . \qquad (1)$$

The intermolecular forces and the factors on which they depend are summarized in Table I [1].

Parametrization of the intermolecular forces described in Table I gives the inter/intramolecular force (IMF) equation which in its most general form[2–9] was written as

TABLE I
Intermolecular forces and the quantities upon which they depend

Intermolecular force	Quantity
Molecule–molecule	
Hydrogen bonding	E_{hb}
Dipole–dipole	dipole moment
Dipole–induced dipole	dipole moment, polarizability
Induced dipole–induced dipole	polarizability
Charge transfer	ionization poptential, electron affinity
Ion–molecule	
Ion–dipole	ionic charge, dipole moment
Ion–induced dipole	ionic charge, polarizability

$$Q_X = L\sigma_{lX} + D\sigma_{dX} + R\sigma_{eX} + M\mu_X + A\alpha_X + H_1 n_{HX} + H_2 n_{nX} + \\ + Ii_X + B_{DX} n_{DX} + B_{AX} n_{AX} + S\psi_X + B^\circ , \qquad (2)$$

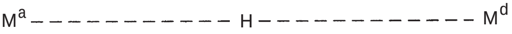
where σ_{lX} is the localized electrical effect parameter. It is identical to the σ_I and σ_F constants, σ_{dX} is the intrinsic delocalized electrical effect parameter, σ_{eX} is the electronic demand sensitivity electrical effect parameter, α is a polarizability parameter. It is defined by the equation

$$\alpha \equiv \frac{MR_X - MR_H}{100} = \frac{MR_X}{100} - 0.0103 , \qquad (3)$$

where MR_X and MR_H are the group molar refractivities of X and H, respectively, n_H and n_n are hydrogen bonding parameters, n_H is equal to the number of OH or NH bonds in X while n_n is equal to the number of lone pairs on O or N atoms in X, i is the ionic charge parameter. It takes the value 1 when the substituent is ionized and 0 when it is not. n_D and n_A are charge transfer parameters. Parameter n_D is 1 when X acts as an electron donor and 0 when it cannot, n_A is 1 when X can function as an electron donor and 0 when it cannot. ψ is an appropriate steric effect parametrization, ψ may be monoparametric, using the υ steric parameters; multiparametric, or composite. μ is the dipole moment of MeX when the substituent X is bonded to sp^3 hybridized carbon, it is the dipole moment of PhX when X is bonded to sp^{2s} hybridized carbon.

Hydrogen Bonding

In many phenomena of interest hydrogen bonding makes a major contribution to the intermolecular forces. Hydrogen bonds vary in strength from barely measurable interactions to those which are comparable in energy to some covalent bonds. They are characterized by their length and by the location of the hydrogen atom (or atoms) with respect to the atoms interacted with[10,11]. The strongest hydrogen bonds involve a hydrogen atom centered between the two other atoms it interacts with, M^d and M^a; its length is very much less than the sum of their van der Waals radii (*I*). The superscripts d and a indicate the hydrogen donor and hydrogen acceptor atoms, respectively. Hydrogen bonds of intermediate strength have the hydrogen atom lying close to or on the line joining M^d and M^a but much closer to M^d (*II*).The bond length is usually slightly less than the sum of the van der Waals radii. It must be noted that Jeffrey and Saenger[11] have questioned the use of the van der Waals radii as a method of characterizing hydrogen bonds. Hydrogen bonds can also result from the interaction of a π orbital acting as the hydrogen acceptor with a hydrogen donor (*III*).



I

II

III

Weaker hydrogen bonds result when due to structural or conformational restraints the hydrogen must lie a considerable distance from the line joining M^d and M^a. An example is the hydrogen bonded dimer of ammonia in the gas phase.

Ideally, we should know what hydrogen bonds form in the initial and final phases of a phenomenon and what their energies are. The hydrogen bonding contribution to the intermolecular forces is then given by the difference in hydrogen bond energies in the initial and final phases. Regrettably, we do not have this knowledge at the present time.

A number of attempts have been made to define hydrogen bonding parameters. The earliest was based on an observation by Hansch and Dunn[12] that $\log P_{O/W} - \log P_{C/W}$ ($\Delta\log P$) where $P_{O/W}$ and $P_{C/W}$ are the octanol/water and cyclohexane/water partition coefficients, respectively. Seiler[13] then defined a hydrogen bonding parameter I_H for some 230 compounds as $\Delta\log P - 0.16$. This is more likely a composite polarity parameter including both hydrogen bonding and contributions from dipole moments. Moriguchi and coworkers[14] defined a parameter V_H as a lipophilicity parameter which again is a polarity parameter. We proposed in a study of amino acid hydrophobicities the use of the n_H and n_n parameters described above[15]. Hydrogen bonding parameters based on equilibrium constants for hydrogen bonded complexes formed from various hydrogen donors with a reference acceptor or various hydrogen acceptors with a reference donor in an “inert” solvent such as CCl_4 were first proposed by Taft and Kamlet[16]. They have been reviewed by Abraham[17], by Raevsky[18], and by Laurence and Berthelot[19]. El Tayar and coworkers[20] have proposed a scale of polarity parameters Λ defined from the difference between $\log P$ calculated from a correlation of $\log P_{O/W}$ values for alkanes with molar volume and the experimental value of $\log P$. Again, these are composite parameters representing both dipole moment and hydrogen bonding. Finally, Rey and coworkers have proposed molecular hydrogen bonding potentials[21].

We have pointed out elsewhere that our parametrization of hydrogen bonding is at best only an approximation. A proper parametrization should take into account the energy of a particular hydrogen bond type and the number of hydrogen bonds of that type. The method we have used takes into account the probability of forming hydrogen bonds but makes the assumption that all those that form have the same energy. We have also assumed that only lone pairs on oxygen or nitrogen are generally likely to be acceptors. In order to model water–organic solvent partition coefficients and other phenomena in which at least one of the phases involved is aqueous in whole or in part we require some measure of hydrogen bonding in aqueous systems. In this work we extract such a measure from the partition coefficients.

METHOD

A basis set of 1-octanol/water log *P* values for XAk where Ak is methyl or ethyl and X is any substituent was selected (set 01, Table II). The substituents in the data set were chosen on the basis of a minimal hydrogen bond acceptor capacity. The set was correlated by multiple linear regression analysis with a form of the IMF equation

$$\log P_{\mathrm{X}} = M\mu_{\mathrm{X}} + A\alpha_{\mathrm{X}} + B_{\mathrm{C}} n_{\mathrm{C}} + B^{\circ} . \qquad (4)$$

The program used is one written in our group many years ago. The parameters used in this work are from the sources reported in[8b,8c].

The term in n_{C}, the number of C atoms in the Ak group (Ak = alkyl), is intended to account for its polarizability. No steric term was included in Eq. (*4*) for either the X or the Ak group as:

1. The interaction between the alkyl groups and the solvent should be minimal. Shielding of the alkyl groups from solvation due to a steric effect of the X group should therefore have little or no effect.

2. Values of the υ steric parameter for the methyl and ethyl groups are 0.52 and 0.56, respectively, thus the steric effect of the alkyl groups is essentially constant.

The results of the correlation are presented in Table III. Compounds with Ak equal to propyl or butyl were added to set 01 to give the extended basis set 02 (Table II). Correlation of this data set with Eq. (*4*) gave the results

Table II
Data sets[a]

01. log *P*, XAk, 1-octanol/water. X, Ak, log *P*: Me, Me, 1.81; Et, Me, 2.36; iPr, Me, 2.76; Pr, Me, 2.89; Bu, Me, 3.23; tBu, Me, 3.11; CCl_3, Me, 2.49; Cl, Me; 0.91; Br, Me; 1.19; I, Me, 1.51; CH_2Cl, Me, 1.43; CH_2Br, Me, 1.61; CH_2I, Me, 2.00; Me, Et, 2.36; Et, Et, 2.89; iPr, Et, 3.23; tBu, Et, 3.82; Cl, Et, 1.43; Br, Et, 1.61; I, Et, 2.00; CH_2Cl, Et, 2.04; CH_2Br, Et, 2.10; SiMe3, Et, 3.85; H, Me, 1.09; H, Et, 1.81

02. Set 01 + Cl, Pr, 2.04; Br, Pr, 2.10; Cl, Bu, 2.64; Me, Pr, 3.23; Me, Bu, 3.39; Et, Pr, 3.39; $ClCH_2$, Pr, 2.64

1. X_{HA} = Ac. Set 02 + Ac, Me, –0.24; CH_2Ac, 0.26; Ac, Et, 0.26; CH_2Ac, Et, 0.91; Ac, Pr, 0.91; Ac, Bu, 1.38; CH_2Ac, Pr, 1.38

2. X_{HA} = CN. Set 02 + CN, Me, –0.34; CN, Et, 0.16; CH_2CN, Me, 0.16

3. X_{HA} = Vi. Set 02 + Vi, Me, 1.77; Vi, Et, 2.40; CH_2Vi, Me, 2.40

4. X_{HA} = Ph. Set 02 + Ph, Me, 2.73; Ph, Et, 3.15; CH_2Ph, Me, 3.15; CH_2Ph, Et, 3.68; Ph, Pr, 3.68; CH_2Ph, Pr, 4.26; Ph, Bu, 4.26

5. X_{HA} = NMe_2. Set 02 + NMe_2, Me, 0.16; NMe_2, Et, 0.70; NMe_2, Bu, 1.70

6. X_{HA} = OAk. Set 02 + OMe, Me, 0.10; OEt, Et, 0.89; OEt, Bu, 2.03

7. X_{HA} = OPh. Set 02 + OPh, Me, 2.11; OPh, Et, 2.51; OPh, Pr, 3.18

8. X_{HA} = SPh. Set 02 + SPh, Me, 2.74; SPh, Et, 3.20

9. X_{HA} = CO_2Ak. Set 02 + CO_2Me, Me, 0.18; CO_2Et, Me, 0.66; CO_2Et, Et, 1.21

10. X_{HA} = SOMe. Set 02 + SOMe, Me, –1.35

11. X_{HA} = F. Set 02 + F, Me, 0.51; CH_2F, Bu, 2.33

12. X_{HA} = NO_2. Set 02 + NO_2, Me, –0.35; NO_2, Et, 0.18; NO_2, Pr, 0.87; NO_2, Bu, 1.47

13. X_{HA} = SSAk. Set 02 + SSMe, Me, 1.77

14. X_{HA} = Bz. Set 02 + Bz, Me, 1.68; Bz, Et, 2.19

15. X_{HA} = OBz. Set 02 + OBz, Me, 2.12; OBz, Et, 2.64

16. X_{HA} = SO_2Ph. Set 02 + SO_2Ph, Me, 0.47

17. X_{HA} = C_2Ak. Set 02 + C_2Me, Me, 1.46

18. X_{HA} = SAk. Set 02 + SEt, Et, 1.95

19. X_{HA} = CHO. Set 02 + CH_2CHO, Bu, 1.78

20. X_{HA} = SCN. Set 02 + SCN, Bu, 2.03

21. X_{HA} = CO_2Ph. Set 02 + CO_2Ph, Me, 1.49

22. X_{HA} = $CONMe_2$. Set 02 + $CONMe_2$, Me, –0.77

23. X_{HA} = Nh. Set 02 + 1-Nh, Me, 3.86; 2-Nh, Me, 3.87

24. X_{HA} = (*E*)-2-VnPh. Set 02 + (*E*)-2-VnPh, Me, 3.35

25. X_{HA} = NCS. Set 02 + NCS, Et, 1.47

TABLE II
(*Continued*)

---

26. X_{HA} = COSPh. Set 02 + COSPh, Me, 2.23

27. X_{HA} = SePh. Set 02 + SePh, Me, 2.87

28. X_{HA} = O_2CH. Set 02 + O_2CH, Me, 0.03; O_2CH, Pr, 0.83

51. X_{HAD} = OH. Set 02 + OH, Me, –0.77; OH, Et, –0.31; CH_2OH, Me, –0.31; CH_2OH, Et, 0.25; OH, Pr, 0.25; CH_2OH, Pr, 0.88; OH, Bu, 0.88

52. X_{HAD} = CO_2H. Set 02 + CO_2H, Me, –0.31; CO_2H, Et, 0.25; CH_2CO_2H, Me, 0.25; CH_2CO_2H, Et, 0.79; CO_2H, Pr, 0.79

53. X_{HAD} = C_2H. Set 02 + C_2H, Me, 1.46; C_2H, Et, 1.46; CH_2C_2H, Me, 1.46; CH_2C_2H, Et, 1.98; CH_2C_2H, Pr, 1.98

54. X_{HAD} = SH. Set 02 + CH_2SH, Pr, 2.28; SH, Bu, 2.28

55. X_{HAD} = $CONH_2$. Set 02 + $CONH_2$, Pr, –0.21; CH_2CONH_2, Et, –0.21

56. X_{HAD} = NH_2. Set 02 + NH_2, Me, –0.57; NH_2, Et, –0.13, CH_2NH_2, Me, –0.13; NH_2, Pr, 0.48; CH_2NH_2, Et, 0.48; NH_2, Bu, 0.88; CH_2NH_2, Pr, 0.88

57. X_{HAD} = NHMe. Set 02 + NHMe, Et, 0.15; NHMe, Pr, 0.70

58. X_{HAD} = CHF_2. Set 02 + CHF_2, Me, 0.75

59. X_{HAD} = $CHCl_2$. Set 02 + $CHCl_2$, Me, 1.79

60. X_{HAD} = CH=NOH. Set 02 + CH=NOH, Me, –0.13

61. X_{HAD} = O_2CNH_2. Set 02 + O_2CNH_2, Me, –0.66; O_2CNH_2, Et, –0.15; O_2CNH_2, Pr, 0.36

62. X_{HAD} = $NHCONH_2$. Set 02 + $NHCONH_2$, Et, –0.74

63. X_{HAD} = NHAc. Set 02 + NHAc, Me, –1.05

64. X_{HAD} = NHPh. Set 02 + NHPh, Me, 1.66; NHPh, Et, 2.16; NHPh, Pr, 2.45

65. XHA = 3-indolyl. Set 02 + 3-indolyl, Me, 2.60

66. XHA = 4(5)-imidazolyl. Set 02 + 4(5) -imidazolyl, Me, 0.23

---

[a] log $P_{O/W}$ values from Hansch C., Leo A.: *Substituent Constants for Correlation Analysis in Chemistry and Biology*. Wiley, New York 1979.

TABLE III
Results of correlations with Eqs (*4*) and (*5*)

Set	M	S_M[a]	A	S_A[a]	B_C	S_{nC}[a]
01	–0.513	0.0362	9.50	0.549	0.510	0.0666
02	–0.512	0.0313	9.39	0.534	0.573	0.0340
1	–0.512	0.0288	9.41	0.479	0.567	0.0273
2	–0.512	0.0302	9.38	0.515	0.571	0.0325
3	–0.511	0.0318	9.45	0.542	0.571	0.0341
4	–0.511	0.0290	9.38	0.483	0.56	0.0276
5	–0.511	0.0305	9.33	0.516	0.563	0.0302
6	–0.511	0.0310	9.25	0.516	0.559	0.0304
7	–0.512	0.0305	9.37	0.519	0.570	0.0318
8	–0.512	0.0308	9.38	0.527	0.570	0.0332
9	–0.511	0.0303	9.43	0.512	0.574	0.0324
10	–0.512	0.0313	9.39	0.534	0.573	0.0340
11	–0.512	0.0314	9.26	0.522	0.557	0.0308
12	–0.512	0.0299	9.43	0.507	0.580	0.0295
13	–0.512	0.0313	9.39	0.534	0.573	0.0340
14	–0.512	0.0308	9.38	0.525	0.572	0.0331
15	–0.512	0.0307	9.38	0.525	0.572	0.0331
16	–0.512	0.0313	9.39	0.534	0.573	0.0340
17	–0.512	0.0313	9.39	0.534	0.573	0.0340
18	–0.512	0.0313	9.39	0.534	0.573	0.0340
19	–0.512	0.0313	9.39	0.534	0.573	0.0340
20	–0.520	0.0273	9.33	0.517	0.566	0.0314
21	–0.512	0.0313	9.39	0.534	0.573	0.0340
22	–0.512	0.0313	9.39	0.534	0.573	0.0340
23	–0.511	0.0308	9.39	0.527	0.573	0.0335
24	–0.512	0.0313	9.39	0.534	0.573	0.0340
25	–0.512	0.0313	9.39	0.534	0.573	0.0340
26	–0.512	0.0313	9.39	0.534	0.573	0.0340
27	–0.512	0.0313	9.39	0.534	0.573	0.0340
28	–0.511	0.0318	9.32	0.542	0.559	0.0332

TABLE III
(*Continued*)

Set	M	S_{M}[a]	A	S_{A}[a]	B_{C}	S_{nC}[a]
51	–0.512	0.0286	9.40	0.478	0.570	0.0272
52	–0.511	0.0298	9.45	0.503	0.569	0.0308
53	–0.511	0.0297	9.42	0.501	0.566	0.0307
54	–0.512	0.0309	9.41	0.528	0.571	0.0334
55	–0.512	0.0309	9.41	0.528	0.571	0.0334
56	–0.511	0.0298	9.37	0.500	0.553	0.0283
57	–0.512	0.0307	9.39	0.525	0.572	0.0330
58	–0.512	0.0313	9.39	0.534	0.573	0.0340
59	–0.512	0.0313	9.39	0.534	0.573	0.0340
60	–0.512	0.0313	9.39	0.534	0.573	0.0340
61	–0.512	0.0303	9.36	0.516	0.568	0.0316
62	–0.512	0.0313	9.39	0.534	0.573	0.0340
63	–0.512	0.0313	9.39	0.534	0.573	0.0340
64	–0.511	0.0315	9.31	0.536	0.559	0.0328
65	–0.512	0.0313	9.39	0.534	0.573	0.0340
66	–0.512	0.0313	9.39	0.534	0.573	0.0340

Set	$\acute{\eta}$	$S_{\acute{\eta}}$[a]	B°	$S_{B^{\circ}}$[a]	$100R^2$[b]	A$100R^2$[c]
01	–	–	0.868	0.128	96.52	96.20
02	–	–	0.801	0.105	96.27	96.01
1	–1.10	0.0853	0.809	0.0893	97.97	97.80
2	–0.251	0.132	0.805	0.101	97.88	97.67
3	–0.213	0.103	0.798	0.106	95.93	95.54
4	–0.827	0.104	0.824	0.0900	97.44	97.22
5	–2.33	0.101	0.825	0.0969	97.21	96.94
6	–1.41	0.0994	0.840	0.0963	97.06	96.78
7	–1.15	0.131	0.808	0.0996	96.34	95.98
8	–1.05	0.168	0.807	0.103	96.39	96.03
9	–1.36	0.103	0.794	0.0996	97.32	97.06
10	–1.89	0.194	0.801	0.105	97.75	97.52

TABLE III
(*Continued*)

Set	η'	$S_{\eta'}$[a]	B°	S_{B°[a]	$100R^2$[b]	A$100R^2$[c]
11	–0.044	0.129	0.843	0.0972	96.66	96.32
12	–0.463	0.115	0.783	0.0948	97.73	97.52
13	–0.560	0.179	0.801	0.105	96.33	95.95
14	–1.37	0.158	0.804	0.102	96.34	95.98
15	–1.18	0.161	0.804	0.102	96.29	95.92
16	–1.51	0.237	0.801	0.105	96.81	96.48
17	–1.14	0.170	0.801	0.105	96.41	96.03
18	–0.852	0.172	0.801	0.105	96.30	95.92
19	–1.01	0.188	0.801	0.105	96.33	95.95
20	–0.113	0.208	0.801	0.105	96.29	95.90
21	–2.16	0.208	0.801	0.105	96.40	96.02
22	–2.14	0.197	0.801	0.197	97.46	97.20
23	–1.13	0.192	0.801	0.103	96.92	96.61
24	–0.763	0.202	0.801	0.105	96.45	96.08
25	–0.160	0.189	0.801	0.105	96.40	96.03
26	–2.06	0.224	0.801	0.105	96.27	95.89
27	–1.23	0.213	0.801	0.105	96.32	95.94
28	–1.17	0.126	0.833	0.104	97.05	96.76
51	–1.47	0.0717	0.805	0.0717	98.50	98.35
52	–1.32	0.0796	0.802	0.0969	97.95	97.76
53	–0.834	0.0749	0.810	0.0965	96.70	96.40
54	–0.730	0.131	0.803	0.103	96.22	95.85
55	–1.58	0.148	0.803	0.103	97.65	97.42
56	–1.75	0.0688	0.839	0.0927	98.23	98.07
57	–2.16	0.119	0.802	0.102	97.24	96.97
58	0.159	0.178	0.801	0.105	96.68	96.33
59	0.062	0.172	0.801	0.105	96.33	95.94
60	–1.90	0.169	0.801	0.105	97.12	96.82
61	–1.82	0.109	0.813	0.0991	97.99	97.80
62	–1.73	0.199	0.801	0.105	97.44	97.18
63	–1.83	0.191	0.801	0.105	97.60	97.36

TABLE III
(*Continued*)

Set	$\acute{\eta}$	$S_{\acute{\eta}}^{a}$	B°	$S_{B^{\circ}}^{a}$	100R$^{2\,b}$	A100$R^{2\,c}$
64	–1.70	0.146	0.834	0.103	96.04	95.66
65	–1.11	0.217	0.801	0.105	96.28	95.90
66	–1.18	0.188	0.801	0.105	96.93	96.62

Set	F^{d}	S_{est}^{a}	$S^{\circ\,e}$	n^{f}	$r_{\mu\alpha}^{g}$	$r_{\mu nC}^{g}$
01	194.2	0.165	0.204	25	0.072	0.019
02	240.9	0.163	0.206	32	0.078	0.066
1	411.1	0.150	0.152	39	0.030	0.164
2	345.7	0.158	0.157	35	0.183	0.056
3	176.8	0.166	0.218	35	0.082	0.090
4	324.0	0.152	0.171	39	0.236	0.011
5	261.4	0.159	0.180	35	0.097	0.046
6	247.8	0.162	0.185	35	0.080	0.070
7	197.3	0.159	0.207	35	0.028	0.070
8	193.7	0.161	0.206	34	0.048	0.051
9	271.9	0.159	0.177	35	0.010	0.021
10	301.4	0.163	0.163	33	0.031	0.022
11	209.7	0.164	0.198	34	0.149	0.102
12	333.8	0.157	0.162	36	0.207	0.192
13	183.7	0.163	0.208	33	0.012	0.033
14	190.9	0.161	0.207	34	0.134	0.034
15	188.1	0.161	0.209	34	0.107	0.040
16	212.5	0.163	0.194	33	0.268	0.040
17	187.8	0.163	0.206	33	0.091	0.092
18	182.1	0.163	0.209	33	0.052	0.069
19	183.6	0.163	0.208	33	0.072	0.168
20	248.1	0.161	0.207	33	0.033	0.250
21	187.3	0.163	0.206	33	0.029	0.039
22	268.6	0.163	0.173	33	0.072	0.017
23	227.9	0.161	0.160	34	0.169	0.098

TABLE III
(*Continued*)

Set	F^d	$S_{est}{}^a$	$S^{\circ}{}^e$	n^f	$r_{\mu\alpha}{}^g$	$r_{\mu nC}{}^g$
24	189.9	0.163	0.205	33	0.086	0.071
25	187.6	0.163	0.206	33	0.011	0.073
26	180.8	0.163	0.210	33	0.012	0.048
27	183.2	0.163	0.208	33	0.017	0.054
28	238.9	0.166	0.186	34	0.113	0.071
51	557.1	0.150	0.131	39	0.213	0.118
52	381.8	0.156	0.154	37	0.117	0.054
53	234.4	0.155	0.195	37	0.076	0.063
54	184.7	0.162	0.210	34	0.162	0.210
55	301.7	0.162	0.166	34	0.041	0.153
56	470.6	0.156	0.143	39	0.120	0.086
57	255.6	0.161	0.180	34	0.078	0.068
58	203.7	0.163	0.198	33	0.119	0.023
59	183.5	0.163	0.208	33	0.052	0.034
60	236.2	0.163	0.184	33	0.078	0.065
61	366.1	0.159	0.153	35	0.046	0.079
62	266.9	0.163	0.174	33	0.012	0.072
63	285.2	0.163	0.168	33	0.017	0.016
64	182.0	0.165	0.215	35	0.102	0.073
65	181.3	0.163	0.209	33	0.073	0.029
66	221.2	0.163	0.190	33	0.028	0.006

Set	$r_{\mu nX}{}^g$	$r_{\alpha nC}{}^g$	$r_{\alpha nX}{}^g$	$r_{nCnX}{}^g$
01	–	0.077	–	–
02	–	0.340	–	–
1	0.658	0.283	0.139	0.180
2	0.669	0.302	0.182	0.161
3	0.168	0.343	0.041	0.161
4	0.247	0.067	0.758	0.180
5	0.099	0.272	0.212	0.144

TABLE III
(*Continued*)

Set	$r_{\mu nX}{}^g$	$r_{\alpha nC}{}^g$	$r_{\alpha nX}{}^g$	$r_{nCnX}{}^g$
6	0.101	0.277	0.018	0.144
7	0.135	0.220	0.642	0.049
8	0.148	0.304	0.697	0.091
9	0.264	0.349	0.228	0.161
10	0.494	0.347	0.074	0.160
11	0.245	0.309	0.321	0.160
12	0.687	0.345	0.222	0.216
13	0.194	0.369	0.309	0.160
14	0.306	0.319	0.626	0.091
15	0.251	0.313	0.653	0.091
16	0.581	0.368	0.553	0.160
17	0.172	0.348	0.086	0.160
18	0.110	0.326	0.212	0.030
19	0.286	0.311	0.008	0.383
20	0.515	0.288	0.065	0.383
21	0.158	0.366	0.582	0.160
22	0.468	0.368	0.294	0.160
23	0.154	0.378	0.786	0.221
24	0.039	0.367	0.569	0.160
25	0.447	0.329	0.178	0.030
26	0.107	0.358	0.659	0.160
27	0.071	0.363	0.617	0.160
28	0.215	0.329	0.179	0.040
51	0.327	0.357	0.428	0.180
52	0.301	0.328	0.155	0.017
53	0.054	0.335	0.007	0.017
54	0.153	0.315	0.011	0.402
55	0.590	0.337	0.036	0.172
56	0.152	0.348	0.295	0.180
57	0.021	0.338	0.039	0.172
58	0.249	0.302	0.179	0.160

TABLE III
(*Continued*)

Set	$r_{\mu nX}$[g]	$r_{\alpha nC}$[g]	$r_{\alpha nX}$[g]	r_{nCnX}[g]
59	0.188	0.353	0.125	0.160
60	0.002	0.331	0.028	0.160
61	0.464	0.325	0.049	0.049
62	0.539	0.335	0.098	0.030
63	0.463	0.351	0.110	0.160
64	0.228	0.203	0.690	0.049
65	0.215	0.363	0.619	0.160
66	0.406	0.364	0.240	0.160

[a] Standard errors of the regression coefficients and of the estimate. [b] Per cent of the variance of the data accounted for by the regression equation. [c] The per cent of the variance of the data accounted for by the regression equation adjusted for the number of independent variables. [d] F statistic for the sigificance of the regression coefficients. [e] The standard error of the estimate divided by the root mean square of the data. [f] Number of points in the set. [g] Zero-th order partial correlation coefficients.

presented in Table III. A comparison of the regression coefficients for set 02 with those obtained for set 01 shows that there is no significant difference. Set 02 was therefore used as the final basis set in order to provide the largest set for this purpose. It should be noted that compounds EtX in which X is an HA or HAD group may appear in the data set twice, once as EtX and once as $MeCH_2X$ so that $\acute{\eta}$ values can be determined for both the X and CH_2X groups.

We now introduce the term $\acute{\eta}_{XHA}n_{XHA}$ into Eq. (*4*) giving the correlation equation

$$\log P_X = M\mu_X + A\alpha_X + B_C n_C + \acute{\eta}_{XHA} n_{XHA} + B^\circ , \qquad (5)$$

where n_{XHA} takes the value 1 when the hydrogen acceptor group X_{HA} is present and 0 otherwise. Correlation of a data set consisting of all the available log *P* values for $X_{HA}Ak$ plus the basis set 02 with Eq. (*5*) gives a regression equation with a value of the regression coefficient $\acute{\eta}_{XHA}$ which serves as a measure of the hydrogen acceptor activity of the X_{HA} group. A regression equation obtained in this manner is considered a reliable source of an $\acute{\eta}_{XHA}$ value only if it meets the following criteria:

1. The regression coefficients of the equation of interest are not significantly different from those of the basis set 02.

2. The goodness of fit of the equation of interest as measured by the values of $100R^2$, $\Delta 100R^2$, and S_{est} is not significantly poorer than that of the regression equation for the basis set 02.

The standard error of $\acute{\eta}_{XHA}$ serves as the error in the measured hydrogen acceptor activity of the X group. Sets 1 through 28 are the data sets used to determine values of $\acute{\eta}_{XHA}$. The data sets used are given in Table II.

We have used a similar method to determine overall hydrogen bonding activities $\acute{\eta}_{XHAD}$ for groups which can act as both hydrogen donors and hydrogen acceptors. Sets 51 through 66 are the data sets used to determine values of $\acute{\eta}_{XHAD}$. Again, the data sets used are given in Table II.

THE HYDROGEN BONDING PARAMETERS

The results of the correlations with Eq. (*5*) are given in Table III. Values of $\acute{\eta}_{XHA}$ and $\acute{\eta}_{XHAD}$, and their standard errors are given in Table IV. These quantities represent the difference in hydrogen bonding between the water saturated with octanol phase and the octanol saturated with water phase. The $\acute{\eta}_{XHAD}$ are composite parameters since they represent the overall effect of the group, including both hydrogen donor and hydrogen acceptor effects. The $\acute{\eta}_{XHA}$ parameters represent only the hydrogen acceptor effect of a group. The regression equations from which the $\acute{\eta}$ values were obtained have coefficients in excellent agreement with those of the basis set. The quality of the correlation equations is generally as good as that of the basis set correlation equation. We believe that the $\acute{\eta}$ values are reasonable measures of hydrogen bonding differences in aqueous media.

We had originally intended to obtain $\acute{\eta}_{HD}$ values which would characterize the hydrogen donor capacity of a group from the equation

$$\acute{\eta}_{XHD} = \acute{\eta}_{XHAD} - \acute{\eta}_{X'HA}\ , \qquad (6)$$

where XHAD designates a group which functions as both hydrogen acceptor and hydrogen donor, and X′HA a hydrogen acceptor group in which the hydrogen bonding H atoms of the XHAD group are replaced with alkyl groups on the assumptions that $\acute{\eta}_{XHA} \approx \acute{\eta}_{X'HA}$ and that the hydrogen bonding interactions of a group are additive. In Table V we have compared values of $\acute{\eta}_{XHAD}$ and $\acute{\eta}_{X'HA}$ by means of the quantity N_{SD}, the number of standard deviations. Generally, there is no significant difference between the

TABLE IV
ή values for XAk

X	ή	N	$S_{ή}$	Set	hb sites	Orbitals	n_a; n_d
				HA			
F	–0.044	2	0.130	11	F	n	3
OAk	–1.41	3	0.0994	6	O	n	2
OPh	–1.15	2	0.131	7	O, Ph	n, π	1, 1
O_2CPh	–1.18	2	0.161	15	O, O, Ph	n, n, π	2, 2, 1
O_2CH	–1.17	2	0.126	28	O, O	n, n	2, 2
O_2CAk^a	–1.36	3	0.103	9	O, O	n, n	2, 2
CHO	–0.776	4	0.105	19	O	n	2
COAk	–1.10	7	0.0856	1	O	n	2
COPh	–1.37	2	0.158	14	O, Ph	n, π	2, 1
CO_2Ak	–1.36	3	0.103	9	O, O	n, n	2, 2
CO_2Ph	–2.16	1	0.208	21	O, O, Ph	n, n, π	2, 1, 1
$CONMe_2$	–2.14	1	0.197	22	O, N, Ph	n, n, π	2, 2, 1
COSPh	–2.06	1	0.224	26	O, S, Ph	n, n, π	2, 2, 1
CN	–0.251	3	0.132	2	N	n	1
SAk	–0.852	1	0.172	18	S	n	2
SPh	–1.05	2	0.168	8	S, Ph	n, π	1, 1
SSAk	–0.560	1	0.179	13	S, S	n, n	1, 1
SCN	–0.113	1	0.208	20	S, N	n, n	2, 1
SOMe	–1.89	1	0.194	10	O	n	2
SO_2Ph	–1.51	1	0.237	16	O, O	n, n	2, 2
SePh	–1.23	1	0.213	27	Se, Ph	n, π	2, 1
NAk_2	–2.33	3	0.101	4	N	n	1
NCS	–0.160	1	0.189	25	N, S	n, n	1, 2
NO_2	–0.463	4	0.115	12	O, O	n, n	2, 2
C_2Ak	–1.14	1	0.170	17	C_2	π	2
Vi	–0.213	3	0.103	3	Vi	π	1
(*E*)-2-VnPh	–0.763	1	0.202	24	Vi, Ph	π, π	1, 1
Ph	–0.827	7	0.104	4	Ph	π	1
Nh	–1.13	2	0.192	23	Nh	π	1

TABLE IV
(*Continued*)

X	$\acute{\eta}$	N	$S_{\acute{\eta}}$	Set	hb sites	Orbitals	n_a; n_d
				HAD			
OH	–1.47	7	0.0717	51	O, OH	n	2: 1
CO_2H	–1.32	5	0.0796	52	O, O, OH	n, n	2, 2; 1
$OCONH_2$	–1.82	3	0.109	61	O, O, N, NH	n, n, n	2, 2, 1; 2
$CONH_2$	–1.58	2	0.148	55	O, N, NH	n, n	2; 2; 2
(*E*)-CH=NOH	–1.90	1	0.169	60	N, O, H	n, n	1, 2; 1, 3
NH_2	–1.75	7	0.0688	56	O, NH	n	1; 2
NHAk	–2.16	2	0.119	57	O, NH	n	1; 1
NHPh	–1.71	3	0.146	64	N, NH	n	1; 1
NHAc	–1.83	1	0.191	63	O, N, NH	n, n	2, 1; 1
$NHCONH_2$	–1.73	1	0.194	62	O, N, N, NH	n, n, n	2, 1, 1; 3
SH	–0.730	2	0.131	54	S, SH	n	2; 1
C_2H	–0.834	5	0.0749	53	C_2, C_2H	π	1; 1
CHF_2	–0.159	1	0.178	58	F, F, CH	n, n	3, 3; 1
$CHCl_2$	–0.0621	1	0.172	59	CH		0; 1
3-indolyl	–1.11	1	0.217	65	indolyl, NH	π	1; 1
4,(5)-imidazolyl	–1.18	1	0.188	66	N, 4(5)-imidazolyl, NH	n, π	1, 1; 1

hb sites are the substructures (atoms or groups of atoms) in a substituent which act as hydrogen donors or acceptors. Orbitals are those used by subgroups as acceptors. n_a and n_d are the number of orbitals available for each substructure and the number of MH bonds which can act as hydrogen donors respectively.

[a] This value is identical to that obtained for CO_2Ak.

$\acute{\eta}_{XHAD}$ and $\acute{\eta}_{X'HA}$ values except for the NH_2 and NHMe pair. Thus, hydrogen bonding in the 1-octanol/water system does not involve hydrogen donor activity. This conclusion is supported by the results of El Tayar, Testa, and Carrupt who have studied hydrogen bonding by the solvatochromic method[20]. There are two possible explanations of this observation. Either the amount of hydrogen bonding resulting from hydrogen donation by the substituent is negligible, which seems highly unlikely, or the extent of hydrogen bonding due to hydrogen donation is about the same in both phases and cancels out. We have also compared the $\acute{\eta}$ values for ZAk groups with ZPh groups in Table V. Of the five pairs available Δ was clearly significant in two pairs, possibly significant in another pair, and not significant in the remaining two pairs. The effect of replacing an alkyl group by a phenyl group results from three factors:

1. The phenyl group is capable of acting as a hydrogen acceptor.

2. The extent to which it does so depends on the nature of the substituents attached to it.

3. When an O or N atom is attached to a phenyl group delocalization of a lone pair occurs, resulting in a decreased probability of hydrogen acceptor activity by that atom.

The first factor should increase hydrogen acceptor activity of the group when Ph is substituted for Ak while the third should decrease it. The effect of the second varies with the electrical effect of groups attached to Ph. The number of pairs available is too small to justify a detailed analysis of the results.

TABLE V

Values of the Student *t*-test for the significance of the differences Δ between $\acute{\eta}_{XHAD}$ and $\acute{\eta}_{X'HA}$, and between $\acute{\eta}_{XHA,Ak}$ and $\acute{\eta}_{X'HA,Ph}$

X_{HAD}	$\acute{\eta}_{XHAD}$	X'_{HA}	$\acute{\eta}_{X'HA}$	$S_{\acute{\eta}}$	Δ	N_{SD}
OH	–1.47	OAk	–1.41	0.0994	–0.06	0.604
CO_2H	–1.32	CO_2Ak	–1.36	0.103	0.04	0.388
SH	–0.730	SAk	–0.852	0.172	0.122	0.709
C_2H	–0.834	C_2Ak	–1.14	0.170	0.280	1.647
NHMe	–2.16	NMe_2	–2.33	0.119	0.17	1.429
NH_2	–1.75	NHMe	–2.16	0.119	0.41	3.445
$X_{HA,Ph}$	$\acute{\eta}_{X'HA,Ph}$	$X_{HA,Ak}$	$\acute{\eta}_{XHA,Ak}$	$S_{\acute{\eta}}$	Δ	N_{SD}
OPh	–1.15	OAk	–1.41	0.131	0.26	1.985
CO_2Ph	–2.16	CO_2Ak	–1.36	0.208	–0.80	3.846
SPh	–1.05	SAk	–0.852	0.172	–0.20	1.163
COPh	–1.37	COAk	–1.10	0.158	–0.27	1.709
NHPh	–1.71	NFMe	–2.16	0.146	0.45	3.082

The Structural Dependence of the $\acute{\eta}$ Parameters

We have examined the structural dependence of two small subsets of the $\acute{\eta}_{XHA}$ values. Correlation of these values for the π type hydrogen acceptors vinyl, phenyl, styryl, and naphthyl with the number of π electrons, n_π, gave the regression equation

$$\acute{\eta}_{X_\pi HA} = -0.107(\pm 0.241)n_\pi - 0.0410(\pm 0.172) \quad (7)$$

$100r^2 = 90.69$; $F = 19.48$; $S_{est} = 0.143$; $S° = 0.432$; $n = 4$,

while correlation with the ionization potential of XH gave the regression equation

$$\acute{\eta}_{X_\pi HA} = 0.331(\pm 0.0991)\mathrm{IP}_{X_\pi HA} - 3.74(\pm 0.905) \quad (8)$$

$100r^2 = 84.83$; $F = 11.18$; $S_{est} = 0.182$; $S° = 0.551$; $n = 4$.

The $\acute{\eta}_{HA}$ values for groups of the type COZ were correlated with the equation

$$\acute{\eta}_{COZ,HA} = D\sigma_{dZ} + R\sigma_{eZ} + B° \quad (9)$$

to give as the best regression equation

$$\acute{\eta}_{COZ,HA} = 1.77(\pm 0.399)\sigma_{dZ} - 1.11(\pm 0.146) \quad (10)$$

$100r^2 = 79.70$; $F = 19.63$; $S_{est} = 0.227$; $S° = 0.533$; $n = 7$.

No term in σ_l was included in Eq. (*8*) because this parameter is approximately constant for the COZ groups. As expected, the $\acute{\eta}_{XHA}$ values are a function of the structure of X.

Though it seems clear that the original IMF model is deficient in its parametrization of hydrogen bonding it seemed of interest to determine the way in which it functioned and the extent to which it was successful. Considering the IMF equation (Eq. (*2*)) suggests that in addition to the hydrogen bonding parameters n_H and n_n which represent the probability of hydrogen donor and acceptor activity, respectively, the electrical effect parameters may also be involved. The $\acute{\eta}_{XHAD}$ parameters were therefore correlated with the equation

$$\acute{\eta}_{\mathrm{XHAD}} = H_1 n_{\mathrm{HX}} + H_2 n_{\mathrm{nX}} + L\sigma_{\mathrm{lX}} + D\sigma_{\mathrm{dX}} + R\sigma_{\mathrm{eX}} + B^\circ . \quad (11)$$

The resulting regression equation is

$$\acute{\eta}_{\mathrm{XHAD}} = -0.227(\pm 0.0552)n_{\mathrm{nX}} + 3.30(\pm 1.10)\sigma_{\mathrm{lX}} + 0.558(\pm 0.260)\sigma_{\mathrm{dX}} - - 1.75(\pm 0.302) \quad (12)$$

A100R^2 = 66.96; F = 7.198; S_{est} = 0.259; S° = 0.637; n = 12.

The $CHCl_2$ and CHF_2 groups were excluded from the correlation as their $\acute{\eta}_{\mathrm{XHAD}}$ values were not significant. The 3-indolyl and 4(5)-imidazolyl groups were excluded due to a lack of σ_l and/or σ_d constants for them.

Note that in agreement with the comparisons made in Table V there seems to be no dependence on n_{H} which represents the probability that X will act as a hydrogen donor. The $\acute{\eta}_{\mathrm{XHA}}$ values were correlated with a form of Eq. (*11*) in which the term in n_{H} was dropped, giving best results after the values for SOMe, and COSPh were excluded. Also excluded were the values for SCN, NCS, and F, which were not significant. The regression equation is

$$\acute{\eta}_{\mathrm{XHA}} = -0.394(\pm 0.0586)n_{\mathrm{nX}} + 1.92(\pm 0.568)\sigma_{\mathrm{lX}} + 0.717(\pm 0.302)\sigma_{\mathrm{dX}} + + 3.42(\pm 1.66)\sigma_{\mathrm{eX}} - 0.657(\pm 0.326) \quad (13)$$

100R^2 = 74.75; A100R^2 = 70.97; F = 14.06; S_{est} = 0.357; S° = 0.565; n = 24.

A comparison of the coefficients of Eqs (*12*) and (*13*) shows that they are not significantly different for the σ_l and σ_d parameters though they may be for n_n. We have therefore combined the two sets into one and correlated them with Eq. (*11*). After excluding the data points for Vi, SOMe, and COSPh and all nonsignificant values we obtained the regression equation

$$\acute{\eta}_{\mathrm{XHA}} = -0.325(\pm 0.0381)n_{\mathrm{n}} + 2.36(\pm 0.434)\sigma_{\mathrm{lX}} + 0.522(\pm 0.191)\sigma_{\mathrm{dX}} + + 2.44(\pm 1.19)\sigma_{\mathrm{eX}} - 1.06(\pm 0.247) \quad (14)$$

100R^2 = 78.82; A100R^2 = 76.77; F = 27.91; S_{est} = 0.304; S° = 0.497; n = 35.

Clearly, the electrical effect parameters are accounting for the contribution of the X group to the hydrogen bond energy. Unfortunately, they do not do it well enough.

The Application of the ή Parameters

Four test data sets have been studied in order to determine the applicability of the ή parameters. The first of these sets consists of the log *P* values for $Ph(CH_2)_mX$. n_{CG} is the number of C atoms in the group to which Ph and X are bonded. They were correlated with the IMF equation in the form

$$\begin{aligned}\log P_{XG} &= L\sigma_{lX} + D\sigma_{dX} + R\sigma_{eX} + M\mu_X + A\alpha_X + H\acute{\eta}_X + \\ &+ Ii_X + S_1\upsilon_{1X} + S_2\upsilon_{2X} + S_3\upsilon_{3X} + B_C n_{CG} + B^\circ .\end{aligned} \tag{15}$$

The υ_i are the segmental steric parameters for X [22]. The data set is reported in Table VI (set T21). The best regression equation results on the exclusion of the data point for $PhCH_2SiMe_3$, it is

$$\begin{aligned}\log P_{XG} &= 0.353\,(\pm 0.126)\sigma_{lX} - 0.488\,(\pm 0.0204)\mu_X + 9.49\,(\pm 0.244)\alpha_X + \\ &+ 0.886\,(\pm 0.0235)\acute{\eta}_X + 0.428\,(\pm 0.0200)n_{CG} + 2.39\,(\pm 0.0592) .\end{aligned} \tag{16}$$

$100R^2 = 98.46$; $A100R^2 = 98.37$; $F = 832.3$; $S_{est} = 0.132$; $S^\circ = 0.130$; $n = 71$.

The goodness of fit is excellent. All of the coefficients other than *L* are significant at the 99.9% confidence level, and that is significant at the 95.0% level. The only serious colinearity is between σ_l and μ. If the dependence on σ_l is real it probably represents the effect of X on hydrogen bonding to the phenyl group.

The second test set (set T22, Table VI), consists of log *P* values in 1-octanol/water for the compounds $X^1(CH_2)_nX^2$ with *n* ranging from 2 to 9. We have considered two different models for this data set. In the first model the substituents were parametrized separately. The correlation equation was

$$\begin{aligned}\log P_{X^1X^2} &= L_1\sigma_{lX^1} + M_1\mu_{X^1} + A_1\alpha_{X^1} + H_1\acute{\eta}_{X^1} + S_1\upsilon_{1X^1} + \\ &+ L_2\sigma_{lX^2} + M_2\mu_{X^2} + A_2\alpha_{X^2} + H_2\acute{\eta}_{X^2} + S_2\upsilon_{1X^2} + B_{CG}n_{CG} + B^\circ .\end{aligned} \tag{17}$$

The best regression equation was obtained on the exclusion of the values for NH_2, Me, 6; NH_2, Me, 9; OH, Me, 9; Ac, Me, 7; and F, OH, 2; it is

$$\begin{aligned}\log P_{X^1X^2} &= -0.342\,(\pm 0.0340)\mu_{X^1} + 6.56\,(\pm 0.830)\alpha_{X^1} + 0.991\,(\pm 0.0543)\acute{\eta}_{X^1} - \\ &- 0.395\,(\pm 0.0590)\mu_{X^2} + 9.28\,(\pm 0.980)\alpha_{X^2} + 0.614\,(\pm 0.0871)\acute{\eta}_{X^2} + \\ &+ 0.414\,(\pm 0.0318)n_{CG} + 1.11\,(\pm 0.134) .\end{aligned} \tag{18}$$

$100R^2 = 96.53$; $A100R^2 = 96.07$; $F = 178.6$; $S_{est} = 0.250$; $S° = 0.202$; $n = 53$.

In the second model the effects of the substituents were assumed to be additive. The correlation equation used was

$$\log P_{X^1X^2} = L\Sigma\sigma_{lX^1X^2} + M\Sigma\mu_{X^1X^2} + A\Sigma\alpha_{X^1X^2} + H\Sigma\acute{\eta}_{X^1X^2} + S\Sigma\upsilon_{X^1X^2} + + B_{CG}n_{CG} + B^\circ . \qquad (19)$$

The best regression equation was obtained on the exclusion of the values for NH_2, Me, 6; NH_2, Me, 9; OH, Me, 9; Ac, Me, 7; and $OCONH_2$, Me, 4; it was

$$\log P_{X^1X^2} = -0.365\ (\pm 0.0340)\Sigma\mu_{X^1X^2} + 8.87\ (\pm 0.615)\Sigma\alpha_{X^1X^2} + + 0.813\ (\pm 0.0395)\Sigma\acute{\eta}_{X^1X^2} + 0.362\ (\pm 0.0346)n_C + 1.09\ (\pm 0.162) \qquad (20)$$

$100R^2 = 94.72$; $A100R^2 = 94.40$; $F = 215.4$; $S_{est} = 0.305$; $S° = 0.241$; $n = 53$.

The exclusion of the data points for NH_2, Me, 6; NH_2, Me, 9; OH, Me, 9; and Ac, Me, 7; may be due to difficulty in measuring log *P* values for these compounds.

It is of interest to compare coefficients of μ, α, $\acute{\eta}$, and n_C for Eqs (*2*) (Table III), (*16*), and (*20*). They are in generally good agreement with each other.

The third test set we have studied consists of log *P* values for amino acids. The value for Car was excluded from the correlation as steric parameters were unavailable for it. The data set (set aah4, Table VI) was correlated with the IMF equation in the form

$$\log P_{Xaa} = L\sigma_{lX} + M\mu_X + A\alpha_X + H\acute{\eta}_X + S_1\upsilon_{1X} + S_2\upsilon_{2X} + S_3\upsilon_{3X} + B^\circ . \qquad (21)$$

The best regression equations are

$$\log P_X = 2.34\ (\pm 1.31)\sigma_{lX} + 6.69\ (\pm 0.497)\alpha_X + 0.894\ (\pm 0.0827)\acute{\eta}_X - 2.97\ (\pm 0.129) \qquad (22)$$

$100R^2 = 93.65$; $A100R^2 = 93.10$; $F = 108.1$; $S_{est} = 0.257$; $S° = 0.274$; $n = 26$

and

TABLE VI
Test data sets

1. log *P*, 1-octanol/water, PhX[a] X, log *P*: H, 2.13; Me, 2.69; Et, 3.15; iPr, 3.66; tBu, 4.11; $CH=CH_2$, 2.95; Ph, 4.04; $CH_2CH=CH_2$, 3.23; CH_2Ph, 4.14; cPr, 3.27; F, 2.27; Cl, 2.84; Br, 2.99; I, 3.25; CF_3, 2.79; CCl_3, 2.92[x]; SMe, 2.74; SPh, 4.45; $SiMe_3$, 4.72; SCF_3, 3.57; $PhCH_2CH_2$, 4.82; CH_2Br, 2.92; CH_2Cl, 2.30; SF_5, 3.36; OMe, 2.11; OEt, 2.51; OPh, 4.21; CHO, 1.48; Ac, 1.58; Bz, 3.18; CO_2Me, 2.12; CO_2Et, 2.64; CO_2iPr, 3.18; CN, 1.56; NMe_2, 2.31; SCN, 2.54; CH_2CN, 1.56; CH_2OAc, 1.96; CH_2Ac, 1.44; CH_2CO_2Et, 2.28; NO_2, 1.85; SOMe, 0.55; NH_2, 0.90; NHMe, 1.66; NHAc, 1.16; $CONH_2$, 0.64; OH, 1.46; CH_2OH, 1.10; CO_2H, 1.87; CH_2CO_2H, 1.41; CH_2NH_2, 1.09; $CH_2CH_2NH_2$, 1.41; $CH_2CH_2CO_2H$, 1.84; $CH=CHCO_2H$, 2.13; $NHCONH_2$, 0.83; COEt, 2.19; OBz, 3.59; $CH=CHNO_2$, 2.24; Pr, 3.57; Bu, 4.26; NHEt, 2.26; NHPh, 3.50; $CH=C(CN)_2$, 2.12; CH=CHPh, 4.81; CH_2SiMe_3, 4.13

4. log *P*, 1-octanol/water, Xaa[b] Xaa, log *P*: Ada, –0.08; Ala, –2.69; Asn, –3.41; Asp, –3.38; Bug, –1.76, Car, 0.86; Cit, –3.19; Cys, –2.49; Gln, –3.15; Glu, –2.94; Gly, –3.25; His, –2.84; Ile, –1.72; Leu, –1.61; Lys, –3.31; Lys(Me), –2.77; Met, –1.84; Nle, –1.53; Nva, –1.86; Pen, –1.78; Phe, –1.63; Ser, –3.30; Thr, –2.91; Trp, –1.06; Tyr, –2.42; Tyr(Me), –1.89; Val, –2.08

21. log *P*, 1-octanol/water, $X(CH_2)_mPh$[a] X, *n*, log *P*, ... : H, 1, 2.73, 2, 3.15, 3, 3.68, 4, 4.26; Me, 1, 3.15, 2, 3.68, 3, 4.26; Et, 1, 3.68; Pr, 1, 4.26; F, 3, 2.95; Cl, 1, 2.30, 2, 2.95, 3, 3.55; Br, 1, 2.92, 2, 3.09, 3, 3.72; I, 3, 3.90; OH, 1, 1.10, 2, 1.36; OMe, 1, 1.92, 3, 2.70; OAc, 1, 1.96, 2, 2.30, 3, 2.77; NH_2, 1, 1.09, 2, 1.41, 3, 1.83; NMe_2, 1, 1.98, 3, 2.73; NO_2, 1, 1.75, 2, 2.08; CO_2H, 1, 1.41, 2, 1.84, 3, 2.42; CO_2Me, 1, 1.83, 2, 2.32, 3, 2.77; CO_2Et, 1, 2.28, 2, 2.73; CHO, 1, 1.78; Ac, 1, 1.44, 3, 2.42; $CONH_2$, 1, 0.45, 2, 0.91, 3, 1.41; CN, 1, 1.56, 2, 1.72, 3, 2.21; $CH=CH_2$, 1, 3.23; $SiMe_3$[x], 1, 4.13; OPh, 1, 3.79; Ph, 1, 4.14, 2, 4.82; OBz, 1, 3.97; CH_2Cl, 1, 2.95, 2, 3.55; CH_2Br, 1, 3.09, 2, 3.72; CH_2I, 2, 3.90; CH_2OH, 1, 1.36; CH_2OAc, 1, 2.30, 2, 2.77; CH_2NH_2, 1, 1.41, 2, 1.83; CH_2CO_2H, 1, 1.84, 2, 2.42; CH_2CN, 1, 1.72, 2, 2.21; CH_2Ph, 1, 4.82; CH_2NMe_2, 2, 2.73; CH_2NO_2, 1, 2.08; CH_2Ac, 2, 2.42; CH_2CO_2Et, 1, 2.73.

22. log *P*, $X^1(CH_2)_mX^2$[a] X^1, X^2, *n*, log *P*: Cl, Cl, 2, 1.48; F, F, 2, 0.75; Cl, OH, 2, 0.03; F, OH, 2, –0.92; NO_2, OH, 2, –0.42; H, H, 2, 1.81; OH, OH, 2, –1.93; NH_2, OH, 2, –1.31; Br, Me, 2, 2.10; Cl, Me, 2, 2.04; NO_2, Me, 2, 0.87; OH, Me, 2, 0.25; Cl, Cl, 3, 2.00; NH_2, Me, 2, 0.15; OMe, OH, 2, –0.37; CO_2H, CO_2H, 2, –0.59; CO_2H, Me, 2, 0.79; Cl, Me, 3, 2.64; $CONH_2$, Me, 2, –0.21; NO_2, Me, 3, 1.47; Me, Me, 2, 2.89; OH, Me, 3, 0.88; OEt, OH, 2, –0.54; NH_2, Me, 3, 0.88; SCN, Me, 3, 2.03; CO_2Me, Ac, 2, –0.13; F, Me, 4, 2.33; NO_2, Me, 4, 2.03; Me, Me, 3, 3.39; $NHCONH_2$, Me, 3, 0.41; OH, Me, 4, 1.56; PrO, OH, 2, 0.05; NH_2, Me, 4, 1.49; Ac, Vi, 2, 1.02; CO_2H, CO_2H, 4, 0.08; CHO, Me, 4, 1.78; CO_2H, Me, 4, 1.88; $OCONH_2$, Me, 4, 1.35; OH, Me, 5, 2.03; NH_2, Me, 5, 1.98; NMe_2, NMe_2, 2, 0.30; $SiMe_3$, Me, 2, 3.84; $OCONH_2$, Me; 5, 1.85; NH_2, Me, 6, 2.57; Ac, CO_2Me, 4, 0.55; $OCONH_2$, Me, 6, 2.36; NCS, Ph, 2, 3.47;, CO_2H, CO_2H, 7, 1.57; Ac, Me, 7, 1.97; OH, Me, 9, 1.70; NH_2, Me, 9, 1.92; CO_2H, SH, 2, 0.43; C_2H, Me, 2, 1.98; MeO, MeO, 2, –0.21; EtO, EtO, 2, 0.33; OH, OH, 3, –1.04; Me, OAc, 2, 1.24; Me, OAc, 3, 1.82

---

[a] log $P_{O/W}$ values from Hansch C., Leo A.: *Substituent Constants for Correlation Analysis in Chemistry and Biology*. Wiley, New York 1979. [b] Pliska V., Schmidt M., Fauchere J. L.: *J. Chromatogr.* **1981**, *216*, 79. [x] Excluded from the best regression equation.

$$\log P_{\mathrm{X}} = 6.33\,(\pm 0.477)\alpha_{\mathrm{X}} + 0.819\,(\pm 0.0745)\acute{\eta}_{\mathrm{X}} - 2.88\,(\pm 0.125) \qquad (23)$$

$100R^2 = 92.72$; $\mathrm{A}100R^2 = 92.42$; $F = 146.5$; $S_{\mathrm{est}} = 0.269$; $S^{\circ} = 0.287$; $n = 26$.

For the purpose of comparison the set was also correlated with an earlier version of the IMF equation

$$\log P_{\mathrm{X}} = L\sigma_{\mathrm{lX}} + M\mu_{\mathrm{X}} + A\alpha_{\mathrm{X}} + H_1\acute{\eta}_{\mathrm{X}} + H_2\acute{\eta}_{\mathrm{X}} + S_1\upsilon_{1\mathrm{X}} + S_2\upsilon_{2\mathrm{X}} + S_3\upsilon_{3\mathrm{X}} + B^{\circ}. \qquad (24)$$

The best regression equation obtained is

$$\begin{aligned}\log P_{\mathrm{X}} = {} & -4.51\,(\pm 1.25)\sigma_{\mathrm{lX}} + 3.85\,(\pm 0.861)\alpha_{\mathrm{X}} - 0.565\,(\pm 0.0672)\acute{\eta}_{\mathrm{HX}} + \\ & + 0.854\,(\pm 0.276)\upsilon_{2\mathrm{X}} - 3.07\,(\pm 0.142)\end{aligned} \qquad (25)$$

$100R^2 = 92.80$; $\mathrm{A}100R^2 = 91.81$; $F = 67.62$; $S_{\mathrm{est}} = 0.280$; $S^{\circ} = 0.299$; $n = 26$.

Equation (*23*) is the best of these models.

log *P* values for those PhX for which $\acute{\eta}$ values are available (set 1, Table VI) were correlated with the IMF equation in the form

$$\log P_{\mathrm{X}} = L\sigma_{\mathrm{lX}} + D\sigma_{\mathrm{dX}} + R\sigma_{\mathrm{eX}} + M\mu_{\mathrm{X}} + A\alpha_{\mathrm{X}} + H\acute{\eta}_{\mathrm{X}} + S_1\upsilon_{1\mathrm{X}} + S_2\upsilon_{2\mathrm{X}} + B^{\circ}. \qquad (26)$$

Whenever possible $\acute{\eta}$ values for groups bonded to sp^2 hybridized carbon were used, thus for X equal to OMe the value in Table III for OPh was used. Best results were obtained on exclusion of the value for CCl_3. The regression equation is

$$\begin{aligned}\log P_{\mathrm{X}} = {} & 1.69\,(\pm 0.280)\sigma_{\mathrm{lX}} - 3.10\,(\pm 0.817)\sigma_{\mathrm{eX}} - 0.482\,(\pm 0.0412)\mu_{\mathrm{X}} + \\ & + 7.90\,(\pm 0.454)\alpha_{\mathrm{X}} + 0.779\,(\pm 0.0632)\acute{\eta}_{\mathrm{X}} + 0.495\,(\pm 0.173)\upsilon_{1\mathrm{X}} + \\ & + 2.15\,(\pm 0.131)\end{aligned} \qquad (27)$$

$100R^2 = 93.91$; $\mathrm{A}100R^2 = 93.39$; $F = 146.5$; $S_{\mathrm{est}} = 0.284$; $S^{\circ} = 0.261$; $n = 64$.

The results are somewhat better than those obtained with the earlier form of the IMF equation but leave room for improvement. We suspect that the use of $\acute{\eta}$ values determined for groups attached to sp^3 hybridized carbon is only an approximation when these same groups are bonded to sp^2 hybridized carbon.

THE IMF MODEL

Results for the Test Sets

The test sets in which the substituent is bonded to sp^3 hybridized carbon validate the IMF model. Equations (*16*), (*20*), and (*23*) are all a function of the dipole moment, the hydrogen bonding parameter, and the polarizability of the substituent X. Both Eqs (*16*) and (*20*) are also as expected a function of the polarizability of the methylene groups in G. There is a small dependence on the σ_l parameter in the case of Eq. (*16*) which if real is probably due to the effect of X on the hydrogen bonding capacity of the phenyl group. The number of parameters used in these equations is small and their relationship to intermolecular forces is clear cut.

$\acute{\eta}$ *Values and the Parametrization of Hydrogen Bonding*

The $\acute{\eta}$ values are certainly valid in 1-octanol/water. In view of the fact that hydrogen donor capability does not seem to be of importance in this medium it is very likely that they are not generally applicable. We had originally hoped to be able to develop a set of additive substructural parameters which would generally represent the hydrogen bonding contribution of any substituent in any medium. Such substructures would be for hydrogen donors OH, NH, and CH bonds; while for hydrogen acceptors they would be n orbitals on O, N, and S, or π orbitals in vinyl, ethynyl, or aryl groups. It now seems likely that the best we can do is define aqueous phase hydrogen bonding parameters for substituents that will represent their overall contribution. We can of course consider a segmental model of a substituent and attempt to parametrize the hydrogen bonding contribution of each segment. Thus if the substituent X can be written as the series of segments $Z_1Z_2Z_3 \ldots Z_n$, we may write the expression

$$\acute{\eta}_X = \sum_{i=1}^{n} \acute{\eta}_{Z_i} . \qquad (28)$$

As the hydrogen bonding capacity of the segment Z_i should be dependent on the electrical and steric effects of the adjacent segments, the electronegativity of the Z atom bearing lone pairs and/or bonds to hydrogen atoms, and the number of these hydrogen acceptor and donor sites the parametrization of these segmental contributions will be too cumbersome to be practical. It is also very inconvenient to require several sets of overall hydrogen bonding parameters for use in different media. We hope to develop a way out of this dilemma. This work represents a beginning in our attempt to develop a set of overall hydrogen bonding parameters for a specific medium and to interpret their structural dependence.

ABBREVIATIONS AND SYMBOLS

hb	hydrogen bonding
dd	dipole–dipole
di	dipole–induced dipole
ii	induced dipole–induced dipole
ct	charge transfer
Id	ion–dipole
Ii	ion–induced dipole
vdW	van der Waals
IMF	intermolecular force
Me	methyl
Et	ethyl
Pr	propyl
Bu	butyl
Pe	pentyl
Hx	hexyl
Hp	heptyl
Ak	alkyl
Ac	acetyl
Oc	octyl
No	nonyl
Dc	decyl
Und	undecyl
Dod	dodecyl
Vi	vinyl
Ph	phenyl
C_2	ethynylene
Bz	benzyl

REFERENCES

1. Charton M. in: *The Chemistry of Arsenic, Antimony and Bismuth* (S. Patai, Ed.), pp. 367–439. Wiley, New York 1994.
2. Charton M., Charton B. I.: *J. Org. Chem.* **1979**, *44*, 2284.
3. Charton M. in: *QSAR in Design of Bioactive Compounds* (M. Kuchar, Ed.), pp. 41–51. J. R. Prous, Barcelona 1985.
4. Charton M. in: *Rational Approaches to the Synthesis of Pesticides* (P. S. Magee, J. J. Menn and G. K. Koan, Eds), pp. 247–278. American Chemical Society, Washington, D.C. 1984.
5. Charton M., Ziffer H.: *J. Org. Chem.* **1987**, *52*, 2400.
6. Kwei T. K., Pearce E. M., Pennacchia J. R., Charton M.: *Macromolecules* **1987**, *20*, 1174.

7. Charton M.: *J. Polym. Sci., Part A: Polym. Chem.* **1988**, *26*, 1265.
8. a) Charton M. in: *Trends in Medicinal Chemistry '88* (H. van der Goot, G. Domany, L. Pallos and H. Timmerman, Eds), pp. 89–108. Elsevier, Amsterdam 1989; b) Charton M. in: *Classical and 3-D QSAR in Agrochemistry and Toxicology* (C. Hansch and T. Fujita, Eds), pp. 75–95. American Chemical Society, Washington, D.C. 1995; c) Charton M.: *Adv. Quant. Struct. Prop. Relationships* **2002**, *3*, 137.
9. Charton M.: *Prog. Phys. Org. Chem.* **1990**, *18*, 163.
10. Vinogradov S. M., Linnell R. H.: *Hydrogen Bonding*. Van Nostrand Reinhold, New York 1971.
11. Jeffrey G. J., Saenger W.: *Hydrogen Bonding in Biological Structures*. Springer-Verlag, Berlin 1991.
12. Hansch C., Dunn W. J.: *J. Pharm. Sci.* **1972**, *61*, 1.
13. Seiler P.: *Eur. J. Med. Chem.* **1974**, *9*, 473.
14. Moriguchi I., Kanada Y., Katsuichiro K.: *Chem. Pharm. Bull.* **1976**, *24*, 1799.
15. Charton M., Charton B. I.: *J. Theor. Biol.* **1982**, *99*, 629.
16. Kamlet M. J., Abboud J. L. M., Taft R. W.: *Prog. Phys. Org. Chem.* **1981**, *13*, 485.
17. Abraham M. H.: *Chem. Soc. Rev.* **1993**, *22*, 73.
18. Raevsky O. A.: *Russ. Chem. Rev.* **1999**, *68*, 505.
19. Laurence C., Berthelot M.: *Perspect. Drug Discovery Des.* **2000**, *18*, 39.
20. El Tayar N., Testa B., Carrupt P. A.: *J. Phys. Chem.* **1992**, *96*, 1455.
21. Rey S., Caron G., Ermondi G., Gaillard P., Pagliara A., Carrupt P. A., Testa B.: *J. Mol. Graphics Mod.* **2001**, *19*, 521.
22. Charton M.: *Stud. Org. Chem.* **1992**, *42*, 629.