Model Solvent Systems for QSAR.† Part 3.‡ An LSER Analysis of the 'Critical Quartet.' New Light on Hydrogen Bond Strength and Directionality

David E. Leahy, Jeffrey J. Morris, Peter J. Taylor* and Alan R. Wait
ICI Pharmaceuticals, Alderley Park, Macclesfield, Cheshire SK10 4TG, UK

An LSER analysis of log *P* for the 'critical quartet' of solvent systems has been carried out using, as initial variables, V_I for volume, μ^2 for dipolarity, and proton donor $\Sigma\alpha$ and proton acceptor β_f scales based on log K_α and log K_β respectively. A common data matrix and an unprecedented range of functionalities have been employed. By making the analysis stepwise, starting with the simplest solutes and adding more in order of increasing complexity, we have been able to identify hitherto unrecognised variables and 'fine-tune' established ones in such a way as to derive self-consistent proton donor and acceptor values applicable to the whole range of solvent systems. By this 'LSER in reverse' we have established, *inter alia*, the following new facts: (*a*) β_f possesses a constant effective zero whereas that for $\Sigma\alpha$ is solvent-sensitive; (*b*) a new term $n\beta_f$ is required for acceptor solutes with two or more available lone pairs; (*c*) when neither lone pair is available, the acceptor strength of carbonyl is sharply reduced; (*d*) a second term specific to NH_2 is required for $\Sigma\alpha$ in alkane and chloroform; (*e*) there is mutual shielding of XH and one lone pair in structures such as CO_2H and $CONH_2$; (*f*) ureas and other structures with parallel NH functions are proton donors of exceptional strength; (*g*) the acceptor ability of dipolar bases (P=O and S=O) varies with the solvent system.

Cooperativity in solute–solvent bonding exists but takes complex forms, and does not appear strong enough to account *e.g.* for the hydrogen bonding properties of bulk water and the alcohols, for which mass action appears a likelier explanation. We present evidence (see Appendix) that mass action will most probably explain certain well known anomalies in the apparent proton acceptor ability of water as revealed by partitioning studies.

The present results throw new light on previously anomalous octanol–water log *P* values and can predict *f*-values for other solvent systems. Most importantly, they provide new information not only on the strength of hydrogen bonding for more than 60 functional groups, but also on its directionality: we are able to predict, with reasonable certainty, which XH groups and lone pairs are actually available for bonding. This information is applicable to water, other solvents, and by implication to the biophase, so should find direct and immediate use in rationalising and quantifying drug–receptor interactions.

Over the past decade, the linear solvation energy relationship (LSER) methodology of Taft and Kamlet *et al.*[1] has become the technique of choice for disentangling solute–solvent interactions.[2,3] Originally derived from UV spectroscopy (the 'solvatochromic comparison method'[1]), its key parameters have since found application throughout spectroscopy,[1] in chemical kinetic and equilibrium processes,[1] and more recently in solute transfer phenomena such as solubility,[3–5] partitioning,[6–9] and by a logical extension, the whole field of mechanistic biology categorised as QSAR.[3,10] Abraham *et al.* have been particularly successful in adapting the methodology to gas–liquid transfer phenomena which range from GLC[11] to irritation of the upper respiratory tract.[12] There seems no doubt that many more such applications await discovery.

Nevertheless, there are disquiets. LSER, a multivariate regression analysis (MRA) technique,[1] has been attacked by the proponents of principal components analysis (PCA) as claiming too much for itself. It is said[13] to display not 'fundamental laws of chemistry' but only 'local empirical rules,' of which PCA can give a better account. To the medicinal chemist, most of this criticism misses the point: as we have emphasised elsewhere,[14] the purpose of a QSAR equation is to predict, and the local rules which are the self-confessed limits to PCA[13] cannot be transposed to a different set of data. To do this requires that the derived components be given some physical meaning, so that effectively we are back to MRA again.[15] On the rare occasions this can be done PCA has proved itself most valuable,[16] but its application to solution phenomena has produced results of little relevance that have verged occasionally on the ludicrous.[14,17] Hence LSER in some suitable form still seems the way forward.

Despite the success of the solvatochromic parameters that the pioneers of LSER unearthed,[1] legitimate doubt attaches to their universal applicability. The parameters π^* for dipolarity/polarisibility, and β and α for proton acceptor and proton donor ability, respectively, are solvent quantities[18–20] (we shall designate the latter as β_{solv} and α_{solv} in this paper). Their use unchanged as solute parameters however, as required *e.g.* for log *P*, makes certain rather gross assumptions which we have discussed elsewhere.[21,22] Attempts have been made to overcome some of these problems by deriving the monomer quantities α_m and β_m from various hydrogen bonding equilibria and then scaling the results,[4,6,8,23,24] but still only for alcohols, so that there has been till recently no general attack on this problem (*vide infra*). It has to be emphasised that QSARs, and

† Acronyms and abbreviations used in this paper: Alk = alkyl; CMR = molar refraction calculation facility; cons = constant; FR = fake residual; LSER = linear solvation energy relationship; MO = molecular orbital; MR = molar refraction; MRA = multivariate regression analysis; p = primary; PCA = principal components analysis; PGDP = propylene glycol dipelargonate; QSAR = quantitative structure–activity relationship; res = residual; s = secondary; sd = standard deviation; t = tertiary; TCE = 1,1,1-trichloroethane.

‡ Part 2, preceding paper.

quantities such as log P employed therein, are unequivocally Gibbs energies, so that parameters related to ΔG are required for their correlation. That is not the case for all the phenomena that LSER has been used to investigate, especially in spectroscopy; it is arguable[21] that the widespread nature of its success is largely due to its—almost impenetrable—blend of ΔH with ΔG.

Of these parameters, π^* is the most obscure. It possesses no simple relation with dipolarity: while cyclohexane which defines $\pi^* = 0.00$ has zero dipole moment, and dimethyl sulfoxide which defines $\pi^* = 1.00$ has a very large one, benzene with $\mu = 0$ possesses $\pi^* = 0.59$.[25] The excellent relation between π^* and μ for simple, highly polar solvent molecules[1] does nothing to resolve this dilemma. It is presumed[1,20b] to possess a polarisibility content which matters more in some contexts than in others. Taft *et al.*[25,26] have attempted a partial solution by inventing a new independent variable δ which modifies π^* for certain rather arbitrary classes of compound to an extent that varies with context. In eqn. (1) for example, their most recent

$$\log P_{oct} = 0.32(4) + 5.35(5)\,V_I/100 - 1.04(4)\,\pi^* + 0.35(3)\,\delta - 3.84(5)\,\beta_m + 0.10(4)\,\alpha_m \quad (1)$$
$$(n = 245 \quad r^2 = 0.992 \quad s = 0.131)$$

LSER for log P (octanol),[8] it reduces the impact of π^* for certain compounds by 30–100%. Abraham *et al.*[11] have attempted direct estimation of the polarisibility component in the π^* mix through a quantity R_2 based on the difference in molar refraction (MR) between the solute itself and a hypothetical alkane of equal volume. This seems a less arbitrary proceeding and we examine it further below.

Finally we consider the most serious charge levelled by the champions of PCA:[13] that all MRA treatments assume knowledge in advance of the relevant variables. If they have missed something, it is only too easy to blend this away in the statistics. This worry, like Hamlet's ghost, has kept coming back to haunt us. For all these reasons, we determined on a new approach.

Results

Compound Inventory.—This has been described.[27] One compound, *p*-nitroanisole (**103**), is present solely for the derivation of *f*-values[27] and does not belong to the LSER set (see Table 1, ref. 27 for all log P values). In addition, all S=O and P=O compounds, all primary and secondary aliphatic amines, and four other data points, are omitted from the LSER analysis for reasons discussed below. Also omitted are six outliers of which only one is not readily accountable.[27] This leaves 92 compounds of which 46, 78, 33 and 83 respectively appear in the LSER analyses for 'alkane', octanol, chloroform and PGDP (see ref. 27 for the definition of 'alkane'). They are categorised in Table 2 of ref. 27. These 102 compounds encompass 71 distinct hydrogen bonding functionalities (or 64 out of 92), more than twice the number to appear in the most comprehensive previous study.[8]

Candidate Parameters.—Volume. A volume term * is required if the LSER involves solute transfer;[4–9] its origin probably lies in cavity formation (endoergic) balanced by solute–solvent dispersion interactions (exoergic) as suggested by Abraham and Fuchs.[28] Originally, Kamlet and Taft *et al.*[4,6] used molar volume $\bar{V}$, changing[5,8] to intrinsic or van der Waals volume V_I after Leahy[7] demonstrated its superiority for log P. We use V_I in units of 10^{-2} dm^3 mol^{-1} to allow direct comparison with their equations.

Dipolarity/polarisibility. As a solvent term, π^* is inapplicable to solids, such as nearly all those in this data set, or indeed of interest to medicinal chemists. Kamlet *et al.*[4,8,29] have produced extensive parameter rules for calculating π^* for solids. We regard this effort as misplaced. Solutes require solute terms, the appropriate term for dipolarity, when interaction is with a continuous dielectric, being[11,30] dipole moment as μ^2. In preliminary trial runs, and at a number of intermediate stages in the analysis, we have produced parallel sets of equations featuring μ and μ^2 as alternatives; in every case, μ^2 emerged as clearly superior. It was not possible to analyse the amphiprotics using μ at all.

Nevertheless polarisibility cannot be ignored, and its effects must be hiding somewhere in these data. We have made two attempts to flush it out. As stated above, Abraham's R_2 parameter[11] looks promising, but as an added term it has proved statistically insignificant. Our other attempt made use of the MR calculation facility (CMR) inside Leo's CLOGP[31] to derive a quantity, ΔCMR, of similar significance to R_2. We were delighted to find no cross-correlation between ΔCMR and CMR itself, V_I, or any other candidate parameter; less delighted when its addition to the equations proved insignificant and did not even alter the coefficients of the other terms. We have to conclude that polarisibility is somehow 'lost' between V_I and the β-term, which is not altogether surprising since both volume and electron density (as refractive index) appear in the equation that defines MR.[11] Abraham and Fuchs[28] have reached similar conclusions. The use of δ as a modifier to π^*[25,26] was an attempt to cope with this problem, since the larger the value of δ, the less polarisibility matters, and δ is particularly large for partitioning.[8] Nonetheless anomalies remain which we discuss later.

We need also to consider whether μ^2 (or μ) should be summed on scalar or vector assumptions when two well separated polar groups are present. There are ten such cases. Two of these (**77** and **78**) possess the same two groups in different alignments; their identical log P values, where an appreciable difference would otherwise be expected, helps to suggest the scalar assumption as the correct one. The same assumption is implicit in CLOGP.[31] Hence we have used $\Sigma\mu^2$ in these cases. We list the μ values used in Table 1; no obvious anomalies are present. We have been able to use published values for 57 of 102 compounds and good model values for all but two of the remainder, where calculation was by MOPAC.[32] Any error in these last is likely to show as some balancing error in the derived β-term, but there is no indication (*vide infra*) that in practice this is serious.

The α and β terms. The recent creation of 'reasonably general' proton donor[33] and acceptor[34] scales for solvent tetrachloromethane, along with our own[21] log K_α and log K_β scales for solvent 1,1,1-trichloroethane (TCE), permits the final abandonment of solvent-based in favour of genuine solute scales for use in this context. We have preferred to base our scales on the latter as specifically tuned, through use of a much more polar solvent, to biological systems.[21]

We also make one other innovation. The scaling of α_{solv} and β_{solv} between nominal limits of zero and unity was reasonable and indeed inevitable in its original context.[1] It has been followed by Abraham *et al.*,[33,34] who have scaled log K_A^H and log K_B^H in a similar manner to give the quantities α_2^H and β_2^H which permit direct comparison with the corresponding solvent scales.[1,24,25] We regard this exercise as redundant in the present context. Partitioning is an equilibrium process, and the use of log K for hydrogen bonding puts this on the same scale as log P, so that coefficients become directly comparable.† It has one

* Some authors prefer cavity surface area to volume. As we[14] have pointed out, both concepts are riddled with ambiguity and there is no way of distinguishing between them at the present time.

† Note that μ^2 also, unlike π^*, is closely related to ΔG.[11]

Table 1 Actual and model dipole moment values[a,b]

Compound	μ	Model	μ
1 PhH	0.03[c]		
2 PhMe	0.36[d]		
3 PhEt	0.37[c]		
4 $PhCH{=}CH_2$	0.43[d]		
5 $PhCH_2CH{=}CH_2$	0.5[d]		
6 $PhCF_3$	2.61[d]		
7 PhF	1.43[d]		
8 PhCl	1.60[d]		
9 PhBr	1.55[d]		
10 PhI	1.36[d]		
11 PhCN	4.08		
12 $PhNO_2$	4.4		
13 $PhNH_2$	1.77		
14 PhNHMe	1.74		
15 $PhNMe_2$	1.66		
16 PhOH	1.76		
17 PhOMe	1.30[d]		
18 PhOCOMe	1.69[d]		
19 PhCHO	3.04		
20 PhCOMe	2.89		
21 PhCOPh	3.24		
22 $PhCO_2H$	1.76[e]		
23 $PhCONH_2$	3.76		
24 $PhCSNH_2$		$PhCSNMe_2$	4.58[d]
25 $PhCONHNH_2$	3.13		
26 PhCONHOH	3.67		
27 PhCONHMe		PhCONHEt	3.60[d]
28 PhCONHEt	3.60[d]		
29 PhNHCOMe	3.88		
30 PhNHCSMe	4.64		
31 PhCONHPh	3.94		
32 $PhCONMe_2$	3.92[d]		
33 PhN(Me)COMe	3.57		
34 $PhNHCONH_2$	4.31		
35 $PhNHCSNH_2$	5.16		
36 $PhN(Me)CONH_2$		Me_2NCONH_2	4.66
37 PhNHCONHMe		MeNHCONHMe	4.60
38 PhNHCSNHMe		EtNHCSNHEt	5.20[d]
39 $PhNHCONMe_2$		$PhNHCONEt_2$	3.20
40 PhNHCONHPh	3.94		
41 $PhNHCO_2Me$	4.11		
42 $PhN{=}C(NH_2)$	1.81[f]		
43 PhSOMe	3.98[d]		
44 $PhSO_2Me$	4.78		
45 $PhSO_2NH_2$	5.13		
46 $PhSO_2NHMe$	4.75[d]		
47 $PhSO_2NMe_2$	5.12[d]		
48 $PhNHSO_2Me$	4.60[d]		
49 $PhNHSO_2NH_2$		$NH_2SO_2NH_2$	3.90
50 Ph_3PO	4.55		
51 NpH[g]	0.00[d]		
52 Np-2-$O(CH_2)_3SOMe$[g]		Np-2-OMe	1.29[d]
		$Me_2S{=}O$	4.00
53 Np-2-$O(CH_2)_3SO_2Me$[g]		Np-2-OMe	1.29[d]
		Et_2SO_2	4.48
54 $PhCH_2OH$	1.80		
55 $PhCH_2OMe$	1.38[d]		
56 $PhCH_2NH_2$	1.38[d]		
57 $PhCH_2NHMe$		Et_2NH	1.26
58 $PhCH_2COMe$		$Me_2C{=}O$	2.83
59 $PhCH_2CO_2Et$		$MeCO_2Et$	1.84[d]
60 $PhCH_2CO_2H$	1.86[e]		
61 $PhCH_2CONH_2$		$MeCONH_2$	3.87
62 $PhCH_2NHCONH_2$		$MeNHCONH_2$	4.34
63 $PhCH_2NHCSNH_2$		$BuNHCSNH_2$	5.70
64 $PhCH_2NHCSNHMe$		EtNHCSNHEt	5.20
65 $PhCH_2OCONH_2$		$EtOCONH_2$	2.59
66 $Ph(CH_2)_2CN$		EtCN	3.60
67 $Ph(CH_2)_2OH$		EtOH	1.73[d]
68 $Ph(CH_2)_2OMe$		Et_2O	1.27[d]
69 $Ph(CH_2)_2NH_2$		$PrNH_2$	1.35[d]
70 $Ph(CH_2)_2NHMe$		Et_2NH	1.26
71 $Ph(CH_2)_2NHEt$		Et_2NH	1.26
72 $Ph(CH_2)_2NMe_2$		Et_3N	0.69
73 $Ph(CH_2)_2COMe$		EtCOMe	2.79[d]

Table 1 (*continued*)

Compound	μ	Model	μ
74 $Ph(CH_2)_2OCOMe$	1.86[d]		
75 $Ph(CH_2)_2NHCOMe$		MeCONHEt	3.90
76 $Ph(CH_2)_2NHCSNH_2$		$BuNHCSNH_2$	5.70
77 *o*-$ClPh(CH_2)_2CONEt_2$		PhCl	1.60[d]
		$MeCONEt_2$	3.70[d]
78 *p*-$ClPh(CH_2)_2CONEt_2$		PhCl	1.60[d]
		$MeCONEt_2$	3.70[d]
79 *p*-$ClPh(CH_2)_2NHCONHMe$		PhCl	1.60[d]
		MeNHCONHMe	4.60
80 $Ph(CH_2)_2CO_2Me$		$PrCO_2Me$	1.81[d]
81 $Ph(CH_2)_3CONHSO_2Et$		$MeCONHSO_2Ph$	7.71
82 $Ph(CH_2)_3CN$		EtCN	3.60
83 $Ph(CH_2)_3OH$		BuOH	1.78
84 $Ph(CH_2)_3OMe$		Et_2O	1.27[d]
85 $Ph(CH_2)_3NH_2$		$BuNH_2$	1.44[d]
86 $Ph(CH_2)_3NMe_2$		Et_3N	0.69
87 $PhCO_2Me$	1.97		
88 $PhCO_2Et$	1.85[d]		
89 $PhCO_2Pr^i$	1.82[d]		
90 $PhCO_2(CH_2)_4CN$		$PhCO_2Et$	1.85[d]
		EtCN	3.60
91 $PhCO_2(CH_2)_4CONH_2$		$PhCO_2Et$	1.85[d]
		$PrCONH_2$	3.85
92 *p*-$NO_2PhO(CH_2)_3SMe$		*p*-$NO_2C_6H_4OMe$	4.89[d]
		Me_2S	1.45[d]
93 *p*-$NO_2PhO(CH_2)_3SOMe$		*p*-$NO_2C_6H_4OMe$	4.89[d]
		$Et_2S{=}O$	4.02
94 *p*-$NO_2PhO(CH_2)_3SO_2Me$		*p*-$NO_2C_6H_4OMe$	4.89[d]
		Et_2SO_2	4.48
95 *p*-$NO_2PhO(CH_2)_3SO_2NH_2$		*p*-$NO_2C_6H_4OMe$	4.89[d]
		$MeSO_2NH_2$	4.60
96 $PhCH(Me)CH_2OH$	1.73[d]		
97 $PhC(CF_3)_2OH$	1.71[c]		
98 PrNHC(=NCN)NHMe		$(NH_2)_2C{=}NCN$	8.22
99 $C_6H_{13}NHCSNHMe$		EtNHCSNHEt	5.20[d]
100 $C_3F_7CH_2NHCSNHMe$	5.74[f]		
101 EtOEt	1.27[d]		
102 CH_3CO_2Et	1.84[d]		

[a] A. L. McLellan, *Tables of Experimental Dipole Moments*, vol. 1, W. H. Freeman, San Francisco, 1963; vol. 2, Rahara Enterprises, El Cerrito, CA 94530, 1974. [b] Permittivity in debyes, at 25 °C or as near to this as possible, and in dioxane unless otherwise stated (1 D = 3.336×10^{-30} C m). [c] As liquid. [d] In benzene. [e] Explicitly stated as value for monomer. [f] Value calculated by MOPAC.[32] [g] Np = naphthyl.

other advantage. Unlike V_1 or μ, log K possesses no definable minimum, yet there must be some point at which hydrogen bonding *per se* fades into a generalised weakly dipolar interaction,[21] already covered by the μ^2 term. For log K_A^H and log K_B^H, this point has been fixed[33,34] with reasonable precision at log $K \approx -1.1$, that value being incorporated into the α_2^H and β_2^H scales. In more polar solvents this value should be higher, and for log K_α and log K_β we suspect[21] a value close to -0.6. In water, one would expect a higher value still (*vide infra*). With the α and β terminology it is easy to forget that this minimum must exist, and indeed previous LSER studies[6–9] of log P have entirely neglected that factor. Its key consequence, which we demonstrate below, is that α_{solv} and β_{solv} are incorrectly zeroed for use as solute terms, quite apart from the other problems we discuss above.

Nevertheless the α/β terminology is useful, and we use both here as shorthand for some form of log K. We use β_f to stand for the functional group log K value corrected for the scale zero noted above; β_{ar} for the contribution to β of the aryl ring, where present; and $\Sigma\beta$ for the sum of these. The equivalent term α_f for proton donors is in practice replaced by $\Sigma\alpha$ since its zero is calculated differently (*vide infra*). These and all other independent variables used for the final regression equations are collected in Table 2.

Table 2 Final parameter list for regression equations[a,b]

Compound	V_1	μ^2	β_{Ar}	β_f[c]	$\Sigma\beta$	$n\beta_f$	$\Sigma\alpha$[d]
1 PhH	0.495	0.00	0.3		0.3		
2 PhMe	0.598	0.13	0.3		0.3		
3 PhEt	0.674	0.14	0.3		0.3		
4 $PhCH{=}CH_2$	0.635	0.18	0.3		0.3		
5 $PhCH_2CH{=}CH_2$	0.733	0.25	0.3		0.3		
6 $PhCF_3$	0.670	6.81	0.0		0.0		
7 PhF	0.523	2.04	0.3		0.3		
8 PhCl	0.586	2.56	0.0		0.0		
9 PhBr	0.628	2.40	0.0		0.0		
10 PhI	0.670	1.85	0.0		0.0		
11 PhCN	0.604	16.65	0.0	1.3	1.3		
12 $PhNO_2$	0.615	19.36	0.0	0.8	0.8		
13 $PhNH_2$	0.568	3.13	0.4	1.6	2.0		0.1
14 PhNHMe	0.647	3.03	0.4	1.2	1.6		0.0
15 $PhNMe_2$	0.742	2.76	0.4	0.9	1.3		
16 PhOH	0.540	3.10	0.4	0.4	0.8		1.9
17 PhOMe	0.633	1.69	0.4	0.6	1.0		
18 PhOCOMe	0.743	2.86	0.2	1.8	2.0	1.8	
19 PhCHO	0.616	9.24	0.0	1.5	1.5		
20 PhCOMe	0.702	8.35	0.0	1.9	1.9		
21 PhCOPh	1.025	10.50	0.0	1.7	1.7		
22 $PhCO_2H$	0.656	3.10	0.0	1.1	1.1		2.2
23 $PhCONH_2$	0.662	14.14	0.0	2.7	2.7		1.0
24 $PhCSNH_2$	0.717	20.98	0.0	1.4	1.4	1.4	0.8
25 $PhCONHNH_2$	0.737	9.80	0.0	3.8	3.8		0.9
26 PhCONHOH	0.702	13.47	0.0	3.3	3.3		1.6
27 PhCONHMe	0.765	12.96	0.0	3.1	3.1		1.0
28 PhCONHEt	0.860	12.96	0.0	3.1	3.1		1.0
29 PhNHCOMe	0.766	15.05	0.3	2.1	2.4		1.6
30 PhNHCSMe	0.814	21.53	0.3	1.1	1.4	1.1	1.6
31 PhCONHPh	1.095	15.52	0.3	2.0	2.3		1.4
32 $PhCONMe_2$	0.865	15.37	0.0	3.3	3.3	3.3	
33 PhN(Me)COMe	0.863	12.75	0.3	2.7	3.0	2.7	
34 $PhNHCONH_2$	0.736	18.58	0.3	2.3	2.6		2.2
35 $PhNHCSNH_2$	0.783	26.63	0.3	1.8	2.1	1.8	1.8
36 $PhN(Me)CONH_2$	0.842	21.72	0.3	3.4	3.7		1.4
37 PhNHCONHMe	0.833	21.16	0.3	2.4	2.7		2.1
38 PhNHCSNHMe	0.892	27.04	0.3	1.8	2.1	1.8	1.8
39 $PhNHCONMe_2$	0.930	10.24	0.3	2.8	3.1	2.8	1.1
40 PhNHCONHPh	1.171	15.52	0.6	1.8	2.4		1.8
41 $PhNHCO_2Me$	0.809	16.89	0.3	1.6	1.9		0.4
42 $PhN{=}C(NH_2)$	0.752	3.28[e]	0.3	3.1	3.4		2.4
43 PhSOMe	0.732	15.84	0.0	*3.2*	3.2	*f*	
44 $PhSO_2Me$	0.784	22.85	0.0	2.0	2.0	6.0	
45 $PhSO_2NH_2$	0.750	26.32	0.0	1.8	1.8	5.4	1.0
46 $PhSO_2NHMe$	0.855	22.56	0.0	1.8	1.8	5.4	0.8
47 $PhSO_2NMe_2$	0.954	26.21	0.0	1.9	1.9	5.7	
48 $PhNHSO_2Me$	0.849	21.16	0.3	1.6	1.9	4.8	0.8
49 $PhNHSO_2NH_2$	0.829	15.21	0.3	2.0	2.3	6.0	1.1
50 Ph_3PO	1.516	20.70	0.0	*4.2*	4.2	*f*	
51 NpH[g]	0.766	0.00	0.4		0.4		
52 $NpO(CH_2)_3SOMe$[g]	1.365	17.66	1.1[h]	*3.3*	4.4	*f*	
53 $NpO(CH_2)_3SO_2Me$[g]	1.405	21.73	1.1[h]	2.1	3.2	6.3	
54 $PhCH_2OH$	0.614	3.24	0.3	1.7	2.0		1.0
55 $PhCH_2OMe$	0.725	1.90	0.3	1.5	1.8		
56 $PhCH_2NH_2$	0.644	1.90	0.3	*2.1*	2.4		*0.5*
57 $PhCH_2NHMe$	0.742	1.59	0.2	*2.2*	2.5		*0.5*
58 $PhCH_2COMe$	0.802	8.01	0.2	1.9	2.1	1.9	
59 $PhCH_2CO_2Et$	0.942	3.39	0.2	1.8	2.0	1.8	
60 $PhCH_2CO_2H$	0.737	3.46	0.2	1.8	2.0		2.1
61 $PhCH_2CONH_2$	0.762	14.98	0.2	3.3	3.5		1.0
62 $PhCH_2NHCONH_2$	0.827	18.84	0.3	2.9	3.2		2.6
63 $PhCH_2NHCSNH_2$	0.882	32.49	0.3	1.9	2.2	1.9	2.0
64 $PhCH_2NHCSNHMe$	0.987	27.04	0.3	2.0	2.3	2.0	2.0
65 $PhCH_2OCONH_2$	0.812	6.71	0.3	2.6	2.9		1.2
66 $Ph(CH_2)_2CN$	0.811	12.96	0.3	1.9	2.2		
67 $Ph(CH_2)_2OH$	0.738	2.99	0.3	2.0	2.3		1.1
68 $Ph(CH_2)_2OMe$	0.840	1.61	0.3	1.9	2.2		
69 $Ph(CH_2)_2NH_2$	0.766	1.82	0.3	*2.5*	2.8		*0.7*
70 $Ph(CH_2)_2NHMe$	0.863	1.59	0.3	*2.7*	3.0		*0.7*
71 $Ph(CH_2)_2NHEt$	0.964	1.59	0.3	*2.7*	3.0		*0.7*
72 $Ph(CH_2)_2NMe_2$	0.967	0.48	0.3	2.9	3.2		
73 $Ph(CH_2)_2COMe$	0.920	7.78	0.3	1.9	2.2	1.9	
74 $Ph(CH_2)_2OCOMe$	0.967	3.46	0.3	1.8	2.1	1.8	
75 $Ph(CH_2)_2NHCOMe$	0.980	15.21	0.3	3.3	3.6	3.3	1.0

Table 2 (*continued*)

Compound	V_I	μ^2	β_{Ar}	β_f[c]	$\Sigma\beta$	$n\beta_f$	$\Sigma\alpha$[d]
76 $Ph(CH_2)_2NHCSNH_2$	1.012	32.49	0.3	2.0	2.3	2.0	2.0
77 *o*-$ClPh(CH_2)_2CONEt_2$	1.383	16.25[i]	0.0	3.6	3.6	3.6	
78 *p*-$ClPh(CH_2)_2CONEt_2$	1.380	16.25[i]	0.0	3.6	3.6	3.6	
79 *p*-$ClPh(CH_2)_2NHCONHMe$	1.152	23.72[i]	0.0	3.1	3.1	3.1	2.0
80 $Ph(CH_2)_2CO_2Me$	0.966	3.28	0.3	1.8	2.1	1.8	
81 $Ph(CH_2)_3CONHSO_2Et$	1.405	59.44	0.3	3.8[j]	4.1	*j*	1.0
82 $Ph(CH_2)_3CN$	0.903	12.96	0.3	1.9	2.2		
83 $Ph(CH_2)_3OH$	0.843	3.17	0.3	2.0	2.3		1.1
84 $Ph(CH_2)_3OMe$	0.955	1.61	0.3	1.9	2.2		
85 $Ph(CH_2)_3NH_2$	0.884	2.07	0.3	*2.5*	2.8		*0.7*
86 $Ph(CH_2)_3NMe_2$	1.086	0.48	0.3	2.9	3.2		
87 $PhCO_2Me$	0.743	3.88	0.0	1.6	1.6		
88 $PhCO_2Et$	0.839	3.42	0.0	1.6	1.6		
89 $PhCO_2Pr^i$	0.942	3.31	0.0	1.6	1.6		
90 $PhCO_2(CH_2)_4CN$	1.151	16.38[i]	0.0	3.5[i]	3.5		
91 $PhCO_2(CH_2)_4CONH_2$	1.237	18.25[i]	0.0	4.9[i]	4.9		1.0
92 *p*-$NO_2PhO(CH_2)_3SMe$	1.157	26.01[i]	1.2[k]	0.7	1.9		
93 *p*-$NO_2PhO(CH_2)_3SOMe$	1.217	40.07[i]	1.2[k]	*3.3*	4.5	*f*	
94 *p*-$NO_2PhO(CH_2)_3SO_2Me$	1.260	43.98[i]	1.2[k]	2.1	3.3	6.3[l]	
95 *p*-$NO_2PhO(CH_2)_3SO_2NH_2$	1.238	45.07[i]	1.2[k]	1.9	3.1	5.7[m]	1.1
96 $PhCH(Me)CH_2OH$	0.820	2.99	0.3	2.0	2.3		1.1
97 $PhC(CF_3)_2OH$	0.967	2.92	0.0	1.2	1.2		0.9
98 $PrNHC(=NCN)NHMe$	0.843	43.56[e]		2.8	2.8		2.8
99 $C_6N_{13}NHCSNHMe$	1.073	27.04		2.0	2.0	2.0	2.0
100 $C_3F_7CH_2NHCSNHMe$	1.056	32.95[e]		1.5	1.5	1.5	2.2
101 EtOEt	0.510	1.61		1.9	1.9		
102 CH_3CO_2Et	0.526	3.39		1.8	1.8	1.8	

[a] As used in eqns. (17)–(20). [b] V_I is intrinsic or van der Waals volume as 10^{-2} dm^3 mol^{-1} (ref. 7); μ is permittivity in debyes; β_{Ar} and β_f are the aryl ring and functional group contributions, as apparent log K_β to $\Sigma\beta$; $(n - 1)$ is the number of available lone pairs; and $\Sigma\alpha$ is apparent log K_α. Estimates of α and β values are to the nearest 0.1. Italicised values have been obtained by back-calculation (see Tables 11 and 12) and are not used in the regressions. [c] Incorporates zero correction of 0.3. [d] Solvent-variable zero corrections required as follows: 'alkane' and chloroform, add 0.4; octanol, add 0.9; PGDP, subtract 0.1. No value less than zero is permitted. [e] Calculated using MOPAC[32] (see Table 1). [f] Solvent-variable quantity (see the text and Table 12). [g] Np = 2-naphthyl. [h] For naphthoxy moiety. [i] Summation. [j] No estimate possible concerning relative contributions of β_f and $n\beta_f$. [k] For *p*-nitrophenoxy moiety. [l] Inapplicable to 'alkane' and chloroform (see the text and Table 12). [m] Inapplicable to chloroform (see the text and Table 12).

Statistical Methodology.[35]—The normal procedure in MRA is to regress all data against all relevant variables. Because of our suspicion that all of these might not be known, we decided instead on a stepwise approach, starting with the simplest compounds in functional group terms, and working progressively towards the most complex.

In broad terms, these 102 compounds break down into three categories.[27] Category A comprises the 21 compounds that possess either no hydrogen bonding functionality as commonly understood, or proton acceptors with no more than one lone pair: nitriles, ethers and tertiary amines. We have reliable log K_β values (see Table 3) for all of these. (It may cause surprise to see ethers placed in this class, but Hine[36] has unequivocal evidence for ethers in solution which we[21] have confirmed, and there is furthermore evidence[37] that alcohols behave similarly). Category B encompasses the remaining proton acceptors, and Category C the amphiprotics. All our preliminary work was carried out on Category A, using therefore the variables V_I, μ^2 and log K_β alone.

Our key tool in this analysis was a form of back-calculation in which the V_I and μ^2 values are assumed to be accurately known but the β-term is not. If the residual (res) of eqn. (2) is treated as

$$\log P = \text{cons} + aV_I + s\mu^2 + b\Sigma\beta + \text{res} \qquad (2)$$

part of the β-term, then once this equation has been set up for the four solvent systems, revised values of $\Sigma\beta$ may be back-calculated as in eqn. (3) and averaged across the solvent set. This

$$\text{FR}(\beta)/b = (b\Sigma\beta + \text{res})/b = \text{new estimate for } \Sigma\beta \qquad (3)$$

averaging process is carried out for all members of a class where more than one exists, *e.g.* **72** and **86** in Table 4, which shows the position for nitriles, ethers and tertiary amines after recycling the data nine times. These revised compromise $\Sigma\beta$ values (to the nearest 0.1) are then used to generate four new eqns. (2), and the iterative process is repeated as set out previously[22] until successive cycles yield constant values of $\Sigma\beta$. The result of using these new $\Sigma\beta_{app}$ values in the tenth regression cycle is shown in Fig. 1 [FR(β)[35] in the Figures is defined by eqn. (3)]. It should be emphasised that Table 4 and the associated Fig. 1 represent snapshots at a moment in time [*cf.* eqns. (5)–(8) in Table 5], and not all parameter values will be quite the final ones.

While any variable may in principle be examined by the FR procedure, V_I is a context-independent quantity and μ is nearly so; the latter rarely varies by more than about 10% across the range of solvents. Hence in practice the 'method of fake residuals' was confined to α and β. There is a justification for this. It is now known[16,21,33,34] that no universal scale of hydrogen bonding ability can exist. Compounds vary not only in strength but in ranking order as a function, especially, of solvent: there are differences in this respect between tetrachloromethane[33,34] and TCE,[21] so it may reasonably be expected that water-based partitioning systems will be different again. This is specially true for β since the behaviour of proton acceptors in all partitioning systems is dominated by the exceptional donor properties of water.[1,6,8,38] Hence we regarded even the log K_α and log K_β scales[21] as merely starting-points on the way to a comprehensive picture of hydrogen bonding in partitioning systems.

We next have to interpret these $\Sigma\beta$ values, which as noted above, are composite of β_f, β_{Ar}, and an unknown zero correction. Our procedure is exemplified as follows. Suppose first that this zero correction is nil. Then, using log K_β for β_f

(Table 3), this places β_{Ar} for PhCN (**11**: $\Sigma\beta = 1.3$) and PhOMe (**17**: $\Sigma\beta = 1.0$) at *ca.* 0.25 and 0.7 respectively, so that, for PhOMe, the ring is a much better proton acceptor than oxygen, which seems unreasonable. If we place the scale zero at log K_β −0.4, however, the β_{Ar} values for PhCN and PhOMe become −0.15 and 0.3; the first is inadmissible (no β can be negative) but the second now looks acceptable—Abraham[39] also, by a route based on HPLC, finds bonding to the ring not much less than to the functional group in this compound. Given scale zeros of −1.1 for tetrachloromethane[33,34] and *ca.* −0.6 for TCE,[21] a higher value is expected for the more polar solvent water, whose donor properties, as noted above, are expected to dominate the β-term in any solvent–water partitioning system. By detailed cross-comparison it was possible to narrow the

Table 3 Comparison of β_f and $\Sigma\alpha$ with log K_β and log K_α

Functional group[a]	$(\beta_f - 0.3)$[b]	log K_β[c,d]	$\Sigma\alpha$	log K_α[c,d]
$ArNO_2$	0.5	*0.74*		
ArCN	1.0	1.06		
AlkCN	1.6	1.23[e]		
ArOH	0.1		1.9	2.14
ArOAlk	0.3	0.30		
$ArCH_2OH$	1.4		1.0	0.90
$ArCH_2OAlk$	1.2			
AlkOH	1.7	1.41[f]	1.1	1.11[g]
AlkOAlk	1.6	1.46[h]		
AlkSAlk	0.4	*0.40*[i]		
$ArNH_2$	1.3	*0.96*	0.1	*0.81*
ArNHAlk	0.9		0.0	*0.44*
ArN(Alk)Alk	0.6	*0.80*		
$ArCH_2NH_2$	(1.8)	2.36	(0.5)	*j*
$ArCH_2NHAlk$	(1.9)	*2.55*	(0.5)	*j*
$AlkNH_2$	(2.2)	2.84[k]	(0.7)	*j*
AlkNHAlk	(2.4)	*2.92*[l]	(0.7)	*j*
AlkN(Alk)Alk	2.6	*2.68*[m]		
ArCHO	1.2	*1.18*		
ArC=OAr	1.4	*1.44*		
ArC=OAlk	1.6	1.46(*1.76*)		
AlkC=OAlk	1.6	1.61[n]		
$ArCO_2H$	0.8		2.2	2.07
ArCOOAlk	1.3	1.23[o]		
$ArCH_2CO_2H$	1.5		2.1	2.04[q]
$ArCH_2COOAlk$	1.5	1.43		
AlkCOOAr	1.5	*1.08*[r]		
$ArCONH_2$	2.4		1.0	
ArCONHAlk	2.8		1.0	
ArCON(Alk)Alk	3.0	2.82		
$ArCH_2CONH_2$	3.0		1.0	
AlkCONHAlk	3.0	*2.99*[s]	1.0	0.64[t]
AlkCON(Alk)Alk	3.3	*3.08*[u]		
ArCONHAr	1.7		1.4	
ArCONHOH	3.0		1.6	
$ArCONHNH_2$	3.5		0.9	
AlkCONHAr	1.8		1.6	1.34
AlkCON(Alk)Ar	2.4	*2.55*[v]		
$AlkCONHSO_2Alk$	3.5	0.99[w]	1.0	1.0
$ArCH_2OCONH_2$	2.3	*2.42*[x,y]	1.2	
AlkOCONHAr	1.3		0.4	0.6
$ArNHCONH_2$	2.0		2.2	
ArNHCONHAlk	2.1		2.1	
ArNHCON(Alk)Alk	2.5		1.1	
$ArCH_2NHCONH_2$	2.6		2.6	
AlkNHCONHAlk	2.8	3.19[y,z]	2.0	
ArNHCONHAr	1.5		1.8	
$ArN(Alk)CONH_2$	3.1		1.4	
$ArCSNH_2$	1.1		0.8	
AlkCSNHAr	0.8		1.6	1.52
$ArNHCSNH_2$	1.5		1.8	
ArNHCSNHAlk	1.5		1.8	
$ArCH_2NHCSNH_2$	1.6		2.0	
$ArCH_2NHCSNHAlk$	1.7		2.0	
$AlkNHCSNH_2$	1.7		2.0	
AlkNHCSNHAlk	1.7	1.96[y,aa]	2.0	2.1
ArS=OAlk	2.9	*2.91*		

Table 3 (*continued*)

Functional group[a]	$(\beta_f - 0.3)$[b]	log K_β[c,d]	$\Sigma\alpha$	log K_α[c,d]
$ArSO_2Alk$	1.7	*1.77*[bb]		
$ArSO_2NH_2$	1.5		1.0	1.15[cc]
ArSONHAlk	1.5		0.8	0.90[dd]
$ArSO_2N(Alk)Alk$	1.6	1.36[ee]		
AlkS=OAlk	3.0	3.06[ff]		
$AlkSO_2Alk$	1.8	*1.83*[gg]		
$AlkSO_2NH_2$	1.6	*1.74*[hh]	1.1	
$AlkSO_2NHAr$	1.3		0.8	
$ArNHSO_2NH_2$	1.7		1.1	
Ar(Ar)(Ar)P=O	3.9	3.85		

[a] This listing corresponds to that in Table 4 of Part 2,[27] with the order slightly changed to emphasise inter-relationships. [b] Scaled by removal of the intercept term to permit direct comparison with log K_β. [c] Ref. 21. Italicised values are scaled from log K_B^H (ref. 34) or log K_A^H (ref. 33) as previously[21] detailed. [d] Models are indicated by footnotes, otherwise the compound itself was employed. [e] MeCN. [f] EtOH. [g] PrOH. [h] MeOBu^t. [i] Et_2S. [j] Immeasurably low. [k] Pr^iNH_2. [l] Et_2NH. [m] $PrNMe_2$. [n] $Me_2C{=}O$. [o] Scaled from pK_{HB} (R. W. Taft, D. Gurka, L. Joris, P. von R. Schleyer and J. W. Rakshys, *J. Am. Chem. Soc.*, 1969, **91**, 4801). [q] $MeCO_2H$. [r] Vinyl acetate; probably a poor model (see the text). [s] MeCONHMe. [t] $C_6H_{13}CONHC_6H_{13}$. [u] $MeCONEt_2$. [v] Ph_2NCOMe. [w] *N*-Methyl derivative of saccharin **105**; probably a bad model (see the text). [x] $EtOCONEt_2$. [y] Tertiary compound for which higher K_β is expected. [z] $Me_2NCONMe_2$. [aa] $Me_2NCSNMe_2$. [bb] Ph_2SO_2. [cc] *p*-Tolyl-SO_2NH_2. [dd] *p*-TolylSO_2NHMe. [ee] $PhSO_2N(Me)CH_2Ph$. [ff] $Me_2S{=}O$. [gg] Tetramethylenesulfone. [hh] $MeSO_2NHMe$.

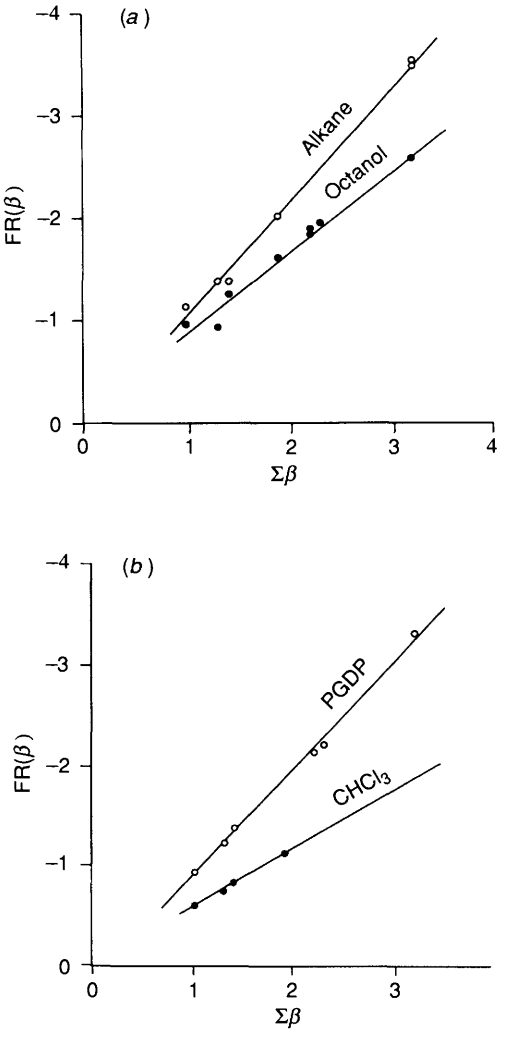


Fig. 1 Plot of FR(β) *vs.* $\Sigma\beta$ after the tenth regression cycle: (*a*) for 'alkane' and octanol; (*b*) for chloroform and PGDP (© Elsevier 1991, *i.e.* ref. 35, and reproduced with permission)

Table 4 Derivation of Σβ for simple acceptors[a]

Compound		'Alkane'	Octanol	Chloroform	PGDP	Mean
	b	−1.07	−0.83	−0.57	−0.96	
11 PhCN	FR(β)	−1.35	−0.91	−0.73	−1.21	
	$\Sigma\beta_{app}$	1.3	1.1	1.3	1.3	1.3
15 $PhNMe_2$	FR(β)	−1.35	−1.23	−0.81	−1.36	
	$\Sigma\beta_{app}$	1.3	1.5	1.4	1.4	1.4
17 PhOMe	FR(β)	−1.11	−0.94	−0.60	−0.93	
	$\Sigma\beta_{app}$	1.0	1.1	1.0	1.0	1.0
66 $Ph(CH_2)_2CN$	FR(β)		−1.85		−2.12	
	$\Sigma\beta_{app}$		2.2		2.2	2.2
82 $Ph(CH_2)_3CN$	FR(β)		−1.80			
	$\Sigma\beta_{app}$		2.2			
72 $Ph(CH_2)_2NMe_2$	FR(β)	−3.43			−3.28	
	$\Sigma\beta_{app}$	3.2			3.4	3.2
86 $Ph(CH_2)_3NMe_2$	FR(β)	−3.43	−2.53			
	$\Sigma\beta_{app}$	3.2	3.1			
68 $Ph(CH_2)_2OMe$	FR(β)				−2.19	
	$\Sigma\beta_{app}$				2.3	2.3
84 $Ph(CH_2)_3OMe$	FR(β)		−1.90			
	$\Sigma\beta_{app}$		2.3			
101 Et_2O	FR(β)	−1.98	−1.57	−1.11		
	$\Sigma\beta_{app}$	1.9	1.9	1.9		1.9

[a] Here *b* is the slope of the Σβ term for each of the four regression equations at the end of the ninth regression cycle; FR(β) is the residual for each point if the *b*Σβ term in each equation is omitted; and $\Sigma\beta_{app}$ (apparent Σβ) is the quantity obtained, to one place of decimals, by dividing FR(β) by *b*. The mean $\Sigma\beta_{app}$ obtained for each substituent, shown in the last column, was then used as its Σβ value in the following (tenth) regression cycle. Plots of FR(β) *vs.* Σβ for all four solvent systems at the end of this tenth cycle are shown as Fig. 1.

Table 5 Coefficients for LSER correlation equations[a]

	Cons	V_1[b]	μ^2	Σβ	*n*β	α_1[c,d]	I_1[f]	α_2[d,e]	I_2[g]	*n*	r^2	*s*	*F*
Alkane													
5	0.29	4.71	−0.045	−1.07						13	0.993	0.08	414
9	0.23	5.05	−0.050	−1.14	−0.22					24	0.984	0.14	284
13	0.18	4.98	−0.055	−1.11	−0.24	−1.08	−0.41	−1.52	−0.51	46	0.996	0.12	1268
17	0.20	4.95	−0.055	−1.10	−0.24	−1.07		−1.44		46	0.996	0.117	1725
	(0.08)	(0.13)	(0.003)	(0.02)	(0.02)	(0.02)		(0.06)					
Octanol													
6	0.06	4.80	−0.030	−0.83						19	0.984	0.09	312
10	0.25	4.47	−0.025	−0.81	−0.16					40	0.988	0.09	725
14	0.20	4.44	−0.023	−0.77	−0.19	−0.11	−0.08	−0.08	−0.10	78	0.990	0.10	853
18	0.21	4.42	−0.023	−0.77	−0.19	−0.10				78	0.990	0.095	1406
	(0.05)	(0.08)	(0.001)	(0.01)	(0.01)	(0.01)							
Chloroform													
7	0.00	5.89	−0.007	−0.57						9	0.990	0.07	170
11	0.59	4.96	−0.012	−0.58	−0.20					15	0.983	0.10	144
15	0.41	5.10	−0.0006	−0.60	−0.23	−0.98	−0.39	−1.56	−0.53	33	0.993	0.11	413
19	0.43	5.07	−0.0006	−0.60	−0.23	−0.98		−1.49		33	0.993	0.108	610
	(0.11)	(0.20)	(0.0027)	(0.03)	(0.01)	(0.02)		(0.06)					
PGDP													
8	0.10	5.17	−0.021	−0.96						17	0.983	0.09	255
12	0.16	5.15	−0.019	−1.02	−0.22					39	0.989	0.10	798
16	0.02	5.39	−0.021	−1.08	−0.20	−0.59	0.05	−0.65	0.07	83	0.995	0.10	1830
20	0.03	5.42	−0.021	−1.09	−0.20	−0.61				83	0.995	0.097	2961
	(0.05)	(0.08)	(0.001)	(0.01)	(0.01)	(0.02)							

[a] For definition of parameters and list of values see Table 2. [b] In units of 10^{-2} dm^3 mol^{-1}. [c] For NH and OH in 'alkane' and chloroform; common value for all proton donors in octanol and PGDP. [d] Equal to Σα in eqns. (13)–(16); intercept term (see notes to Table 2) incorporated in eqns. (17)–(20). [e] For NH_2 in 'alkane' and chloroform. [f] Intercept term for α_1. [g] Intercept term for α_2.

permissible limits to between −0.2 and −0.4. Eventually the scale zero settled at −0.3, where it has remained; occasional checks using values of −0.2 or −0.4 invariably gave worse results.

The next stage was to feed in those compounds from Category B for which we possessed good log K_β values. To obtain Σβ, we added in the zero correction and a provisional estimate for β_{Ar}. We then assumed the previous—tenth—cycle of regression equations to apply, and used these to calculate the corresponding FR(β) values. For PGDP, the most comprehensive data set, the results are shown as Fig. 2(*a*). There is clearly some scatter, but not more than would normally be considered acceptable. However, if the regression line for PGDP from Fig. 1(*b*) is added, it becomes clear that two types of

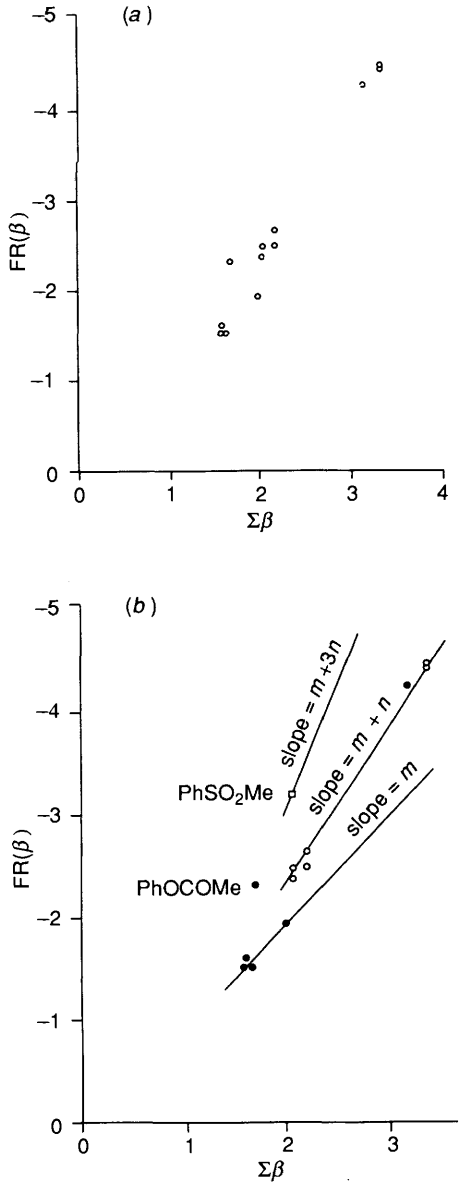


Fig. 2 (*a*) Plot of FR(β) *vs.* $\Sigma\beta$ for carbonyl acceptors in PGDP on the assumption that the equation employed for PGDP in Fig. 1(*b*) can also be used to calculate their expected FR(β) values (© Elsevier 1991 and reproduced with permission); (*b*) The same data as for Fig 2(*a*), showing different relations between FR(β) and $\Sigma\beta$ for carbonyls with one available lone pair (slope *m*) or two (slope $m + n$) [○, aliphatic carbonyl compounds and ●, aromatics—the single point for $PhSO_2Me$ (slope $m + 3n$) is added; that for PhOCOMe is an outlier (see the text)] (© Elsevier 1991, *i.e.* ref. 35, and reproduced with permission)

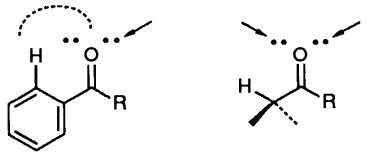


Fig. 3 Shielding by ring C–H

carbonyl are involved. All those compounds that fit this line [Fig. 2(*b*)] are aromatic: shielding by ring CH (Fig. 3) could limit these effectively to one lone pair. All but two of the remainder are aliphatic, to which this restriction cannot apply. The two exceptions are revealing. Rekker[40] noted the anomalously negative (octanol) *f*-value of aromatic tertiary amide (Fig. 4); on the UV evidence that phenyl is twisted out of the amide plane, he attributed this to decoupling of resonance. In

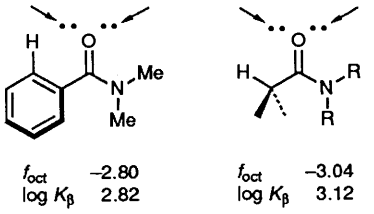


Fig. 4 *f*-Values of amides

fact (Table 2) β_f is only slightly elevated, and most of the effect comes from release of the second lone pair. Similarly, phenyl acetate (**18**) is highly twisted,[41] which accounts for the long-standing puzzle that (Alk)COO(Ar)* is much more hydrophilic than (Ar)COO(Alk) despite the expected cross-conjugation—*f*-values of −1.18 and −0.58 on the octanol scale [(Alk)COO(Alk) = −1.49].[31] Hence the major determinant of *f* is again the number of available lone pairs, not β_f. In fact, the point for PhOCOMe on Fig. 2(*b*) is a mis-plot; the model compound here (see Table 3) was vinyl acetate, which compound class we now know to be planar.[41] Every guess we have made as to compound planarity, based on these and similar considerations, has since been confirmed from the crystal structure evidence[41] where available.

We are now in a position to define a new and hitherto unsuspected variable: $n\beta_f$, where *n* is the number of lone pairs after the first. Here *n* is a multiplier for β_f alone, not for $\Sigma\beta$. (Values are listed in Table 2). In addition to lines of slope *m* for one available lone pair and $(m + n)$ for two, Fig. 2(*b*) shows a line of slope $(m + 3n)$ drawn through the solitary point for $PhSO_2Me$ (**44**) which should possess four lone pairs. All three lines converge roughly to a single point. Hence a simple indicator variable would not handle this phenomenon (as we have confirmed): $n\beta_f$ is strictly proportional to β_f, as expected† on chemical grounds.

It will be noticed that these three lines do not meet at the origin. In fact, for the four solvent systems, two initially had positive and two had negative intercepts. This is typical of the problems one encounters with the FR methodology at the start of any phase of the investigation; repeated re-cycling eventually eliminates it.

From this point it was quite easy to incorporate the rest of the proton acceptors. Certain problems remained, such as those attaching to the double-functionality compounds **92** and **94**. For the sulfide **92** we assumed a log K_β value for (Alk)S(Alk) scaled from log K_B^H, and obtained that for the *p*-nitrophenoxide moiety (classified in Table 2 as β_{ar}) from $\Sigma\beta$ by difference. This value fed into the sulfone **94** then yielded β_f for (Alk)-SO_2(Alk), slightly higher than for (Ar)SO_2(Alk) as expected (Table 3). Exactly the same value was found to fit the sulfone **53**, whose β_{Ar} value was calculated *de novo* as 0.1 higher than for PhOMe (**17**) from the difference between benzene (**1**) and naphthalene (**51**). Such results slowly built up our confidence in the methodology.

Finally we investigated the amphiprotics, to which two special problems attach. In the first place, where α is known, few of the corresponding β values are known, alcohols being among the rare exceptions (Table 3). Secondly, we had to estimate both together. The task was horrendous and would probably have proved impossible, but for one fortunate circumstance. It is known[6] that the octanol–water system has so little sensitivity to

* Alk = Alkyl; see Table 3, footnote *a*.

† It should be noted that not all lone pairs on a single atom are equivalent. They are for ketones, but the *E*-lone pair of esters forms the weaker bond,[42] presumably because of σ-resonance,[21] and the same would be expected for amides. Effects such as these, however, are too subtle for our present treatment.

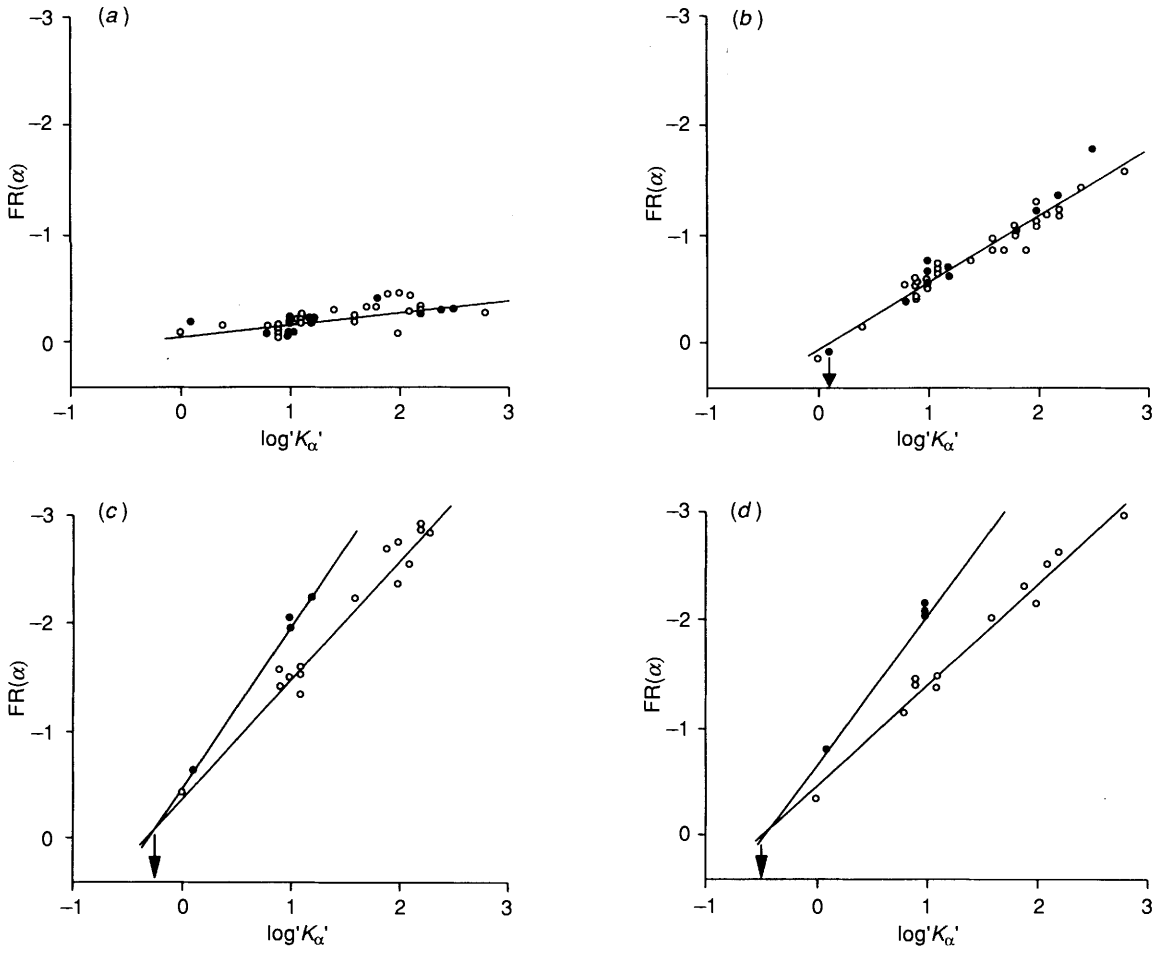


Fig. 5 Plots of FR(α) *vs.* Σα (log 'K_α') after the 45th regression cycle: (*a*) for octanol; (*b*) for PGDP; (*c*) for 'alkane'; and (*d*) for chloroform [○, NH and OH; ●, NH_2].

proton donors that even the sign of its coefficient is in doubt.[8,9] Suppose we set this coefficient to zero. Then the whole of FR for octanol becomes apparent $b\Sigma\beta$, which subtracted from FR for the other solvent systems, allows some preliminary estimate of FR(α). Unsurprisingly, this 'octanol assumption' did not last very long, but at least it did enable a start to be made.

Some 30 recyclings were required and detail would be tedious, but one point needs emphasis. We have insisted on imposing chemical criteria as well as statistical ones. Most but perhaps not quite all of these will be obvious. Examples include the following: (a) NH must never be a stronger donor than the corresponding NH_2; (b) C=O must be a stronger acceptor—usually much stronger—than the corresponding C=S; (c) the expected trend alkyl > benzyl > aryl in acceptor strength must always be present; (d) any sequence primary, secondary, tertiary (p,s,t) may reasonably lie, for any property, in the order p > s > t or p < s < t but the orders p > s < t or p < s > t are not allowed.* Application of these criteria has enabled us to demonstrate, *inter alia*, that while all primary amides have (at most) a single lone pair available, and all tertiary amides have two, aromatic CONH comes into the first category but aliphatic CONH into the second. These and other results are discussed below. It would have been possible to improve even on the statistics we have obtained by ignoring these criteria, but chemistry and not statistics has been our prime concern.

One statistical elaboration did, however, yield an unexpected dividend. As noted above, water is so dominant as a proton donor as entirely to dictate the behaviour of acceptor solutes. But since there is no dominant proton acceptor solvent, the same may not hold for donors. One consequence could be that the scale zero is a function of the system. As a precaution against this possibility, we used the two-term eqn. (4), where the second

$$\mathrm{FR}(\alpha) \equiv a\,\text{'log } K_\alpha\text{'} + zI \qquad (4)$$

term is intended to define the intercept. In addition to this, acting on certain indications that NH_2 might behave differently from NH and OH, we employed separate pairs of terms for these two categories. The use of eight independent variables at one stage of the investigation was a particular embarrassment for chloroform with only 33 data points, but we are vindicated by the final results.

These appear on Fig. 5. The slope for octanol [Fig. 5(*a*)] is very shallow, so that the regression line has little meaning,† and its main use is to demonstrate that subtraction of $b\Sigma\beta$ has left no glaring discrepancies. Originally we tried to 'force' a positive slope (α lipophilic) in line with Kamlet *et al.*[8] [eqn. (1)], which resulted in a peculiar curvature such that weak and strong donors came out as lipophilic but moderate donors as hydrophilic, so we abandoned the attempt. A trace of this curvature

* It may be objected that the order p < s > t is found, *e.g.*, for the pK_a values of aliphatic amines. However, protonation is not a unitary process; it is the complex resultant of electronic, solvational and dispersion forces.[43] These mixed orders are also to be found in a number of *f*-value sequences which are similarly composite.[27]

† The scale zero of −0.9, while statistically required, similarly has little precision and we discount it as chemically meaningless.

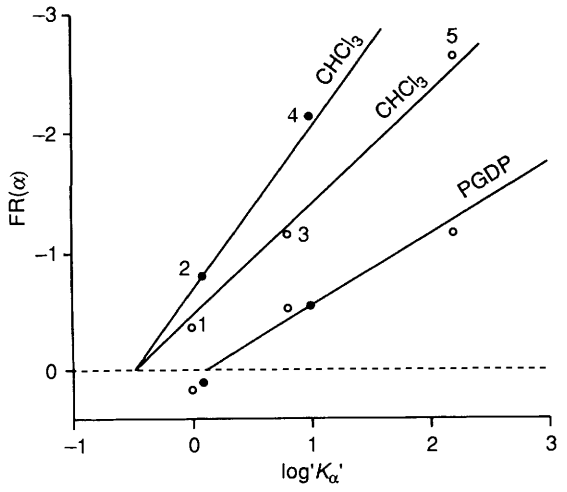


Fig. 6 Selected data points shown relative to the regression lines of Fig. 5 for chloroform and PGDP (for discussion see text) (© Elsevier 1991, *i.e.* ref. 35, and reproduced with permission): **1**, PhNHMe; 2, $PhNH_2$; 3, $PhSO_2NHMe$; 4, $PhSO_2NH_2$; 5, $PhCO_2H$

still remains on Fig. 5(*a*). El-Tayar *et al.*[10] find a slight negative slope. For PGDP [Fig. 5(*b*)] the slope is much steeper, as expected, with its zero at 'log K_α' + 0.1. For both solvents, NH_2 lies on the same line as NH and OH.

The results are spectacularly different for 'alkane' and chloroform [Figs. 5(*c*) and 5(*d*)]. Here the slopes are much steeper again, as expected,[8] but there is now a separate and still steeper slope for NH_2. This can have nothing to do with the aqueous phase, since confined to these solvents and (and never before reported). We believe this result to stem from dipole–dipole repulsion between NH or OH and the CH of the solvent. Of course the CH dipole is very small in hydrocarbons, but with C[H] > 100 mol dm^{-3} mass action effects may become important (see the Appendix). The CH of chloroform is present at much lower concentration, *ca.* 12.5 mol dm^{-3}, but by contrast is highly polarised. Any such effect should be greater for NH_2 since this contains two protons per functional group; in fact, the slope is about 50% greater in each solvent. One wonders how far even the extra slope of the line for NH and OH *vis-à-vis* PGDP may be dictated by this phenomenon. Association of NH and OH with the solvent's functional group should prevent this effect from showing itself in proton acceptor solvents. Probably the lower scale zero, at *ca.* −0.4, is another consequence of this repulsion. This scale zero must vary so much partly because, as noted above, there is no proton acceptor phase which approaches the dominance of water as proton donor.

This phenomenon points to a sidelight on 'hydrophobicity'[14] which has so far gone unremarked. The fact that alkanes do not form hydrogen bonds conceals an important asymmetry. Proton acceptors are not attracted by alkane CH, but with their surface of electrons, they are not repelled either. Such repulsion however is perfectly possible for proton donors. One wonders what role these repulsive forces may occasionally play in drug–receptor interactions.

Fig. 6, which shows the regression lines for chloroform and PGDP and how certain chosen compounds fit onto them, illustrates some consequences. The low scale zero for chloroform greatly exaggerates the effect of weak proton donors such as aniline (**13**), which has no donor ability at all on the PGDP scale. The modest donor $PhSO_2NH_2$ (**45**) has nearly the strength on the chloroform scale of the strong donor benzoic acid (**22**), but *N*-methylation restores its modesty. It is difficult to see how any sense could be made of these apparently arbitrary experimental data without the conceptual framework here provided. Such large qualitative shifts in ranking order may be among the ways in which biological membranes can discriminate; they are, of course, reflected by the fragment values of Part 2.[27]

The final stage in the statistical treatment was to use eqn. (4) to combine the slope and intercept terms into a single α term by incorporating the scale zero unique to each set (see Table 2). Inevitably therefore r^2 and F improve, and the test of uniqueness is that the regression coefficients remain substantially unchanged (Table 5).

Statistical Overview.—Table 5 summarises the regression equations, not only for each solvent system but for all at each stage of the analysis. Table 6 presents the correlation matrix and Table 7 the residuals. We have chosen this rather than the conventional tabulation of observed results *vs.* calculated since, when the former cover six decades, it is easy that way to give a spuriously favourable impression of goodness-of-fit.

Given an expected sd for good data of *ca.* ±0.07,[14,44] our standard errors s, at 0.095–0.117, approach the 'level of exhaustive fit.'[13] In fact, of 240 data points, the sd for only 27 exceed twice this value. Nevertheless it must be pointed out that our methodology is such as to exaggerate goodness-of-fit, in that the continual readjustment of β and α means that these are no longer truly independent of log P. Hence, as discussed,[27] these equations are unsuitable for its *de novo* calculation, and there are further caveats which we note below.

Our procedure in recycling our data was to persist until we had succeeded in roughly cancelling the residuals for any one compound across all solvents taken together. In a few cases,* we have had to be content with striking a balance between relatively large errors; there is a persistent tendency, which we cannot explain, for the larger sd to attach to certain compounds. We can find no common thread: for example, the bifunctional ester/nitrile **90** is one such deviant, whereas the ester/amide **91** is notably well behaved. In view of our earlier comments one might expect exclusively aliphatic compounds to behave badly, which is true for ethyl acetate **102** and the thiourea **99**, but diethyl ether (**101**) and the thiourea (**100**) are exemplary. Hence we conclude that these are artefacts, not systematic deviations.

Eqns. (5)–(8) are for Category A, (9)–(12) add in all proton acceptors, and (13)–(16) incorporate the amphiprotics, while (17)–(20) are the final regression equations. These are extremely robust: once introduced, there is scarcely any change in the coefficient of any polar term. This is specially important for $\Sigma\beta$ and $n\beta_f$, where the lack of any influence of the second on the first is clear evidence that the effect of multiple lone pairs has been cleanly separated. (Note that $\Sigma\beta$ and $n\beta$ are very poorly correlated: Table 6). We emphasise that this new variable $n\beta$ has never been detected before.†

The exceptions to this stability are the regression constants and the coefficients of V_I which, especially for chloroform, 'see-saw' in a mutually compensating manner. We are unsure why this happens, but it may be connected with our failure to disentangle the polarisibility factor. Ideally, the regression constant should be zero, since a compound of zero volume and having no other properties should possess log $P = 0$; of the final correlation equations, (19) is particularly offensive in this respect. This instability in the V_I term has the unfortunate consequence of invalidating the coefficient of V_I as a predictor for $f(CH_2)$. Table 8 compares these values[27] with those deduced from

* Those not discussed here comprise **16**, **33** and **64**.

† In a parallel development concerning *de novo* calculation of log P, Richards[45] has found that the fit is much improved if the energy of interaction with polar moieties is treated as a discontinuous function of their surface area.

Table 6 Correlation matrix for parameters of Table 5

	V_1	μ^2	$\Sigma\beta$	$n\beta$	α_1	I_1	α_2
μ^2	0.540						
$\Sigma\beta$	0.595	0.383					
$n\beta$	0.417	0.485	0.204				
α_1	0.109	0.286	0.109	−0.067			
I_1	0.036	0.137	0.174	−0.080	0.885		
α_2	0.005	0.279	0.265	0.060	−0.251	−0.292	
I_2	−0.069	0.175	0.291	0.063	−0.301	−0.331	0.890

$\Delta V_1(CH_2)$. Even given some imprecision in the latter, as noted above, agreement is poor. Others[6,8,9] have fared no better; the very different coefficients of V_1 according to whether μ^2 or π^* is used for dipolarity suggests different ways of blending away the polarisibility term. With μ^2 in use this is probably split between $\Sigma\beta$ and (mostly) V_1; note the 35% correlation between these variables (Table 6). Hence these equations may poorly predict the homologues that have dominated most previous series[6–9] and are unlikely to be suitable for compound sets, *e.g.* the polychlorobiphenyls, which possess little polarity but where polarisibility may be important. Significantly perhaps, the non-hydrogen bonders **1–10** and **51** contain more than their share of high residuals (Table 7). However, interpolation as we have used it for deriving approximate *f*-values[27] should still have some limited validity, and extension, using the FR methodology, to compounds that differ only in their functional group may also be permissible. With Kamlet *et al.*[6] we emphasise that the prediction of log *P* is not our primary intention. As will become clear below, LSER in our hands is a way of disentangling the chemistry.

Discussion

The final regression equations show intriguing regularities. Given that alkanes do not form hydrogen bonds, it is intuitively pleasing that the coefficients of $\Sigma\alpha$ and $\Sigma\beta$ for 'alkane' should be substantially identical, at −1.07 and −1.10 respectively. It is equally pleasing that both coefficients should be close to unity: that is, the strength of hydrogen bonding is reflected by log *P* in almost linear fashion. The first helps to substantiate our previous suggestion[21] that, fortuitously, log K_α and log K_β on which these scales are based do indeed carry roughly equal weight. Both are examples of a serendipity not at all to be found in previous studies.[6–9]

There are other symmetries. Chloroform, a pure proton donor, rejects donor solutes equally with 'alkane' while rejecting acceptors, relative to water, only half as well (coefficients of $\Sigma\alpha$ and $\Sigma\beta$ are −0.98 and −0.60 respectively). PGDP behaves in the precise mirror image of this (−0.61 and −1.09 correspondingly). Hence both solvents were well chosen for their purpose. In contrast is the established[6–9] lack of symmetry for octanol, with almost the same affinity for donors as water, but with a lack of affinity for acceptors, relative to water, half-way between chloroform and 'alkane' (coefficient −0.77). The fact that these coefficients are power relations, *i.e.* they imply a constant ratio not a constant difference between solvent log *P* values, is one reason why Fujita's treatment[46] of the difference in behaviour between chloroform and octanol cannot be valid; we have seen[27] that it does not work in practice.

Not all differences are so straightforward. It is unsurprising that the coefficent of μ^2 should be greatest for 'alkane', but very surprising indeed that it should virtually vanish for chloroform. Both features appear (using π^*) in previous treatments,[8,9] but have received no comment. On the face of it, this implies some close-range similarity between chloroform and water which conventional measures of dipolarity do not reflect. Dipolarity is

Table 7 Residuals for the final regression equations[a,b,c]

Compound	Alkane	Octanol	$CHCl_3$	PGDP
1	−0.08	−0.04	0.04	−0.02
2	0.07	0.11	0.13	−0.04
3	0.10	0.19	0.01	0.02
4		0.17		−0.10
5		0.01		−0.01
6		−0.01		−0.25
7	0.11	0.02	−0.05	0.01
8	−0.03	0.10	0.06	−0.06
9	−0.08	0.06	0.00	−0.11
10	−0.09	0.12		−0.13
11	0.12	0.05	0.00	0.13
12	0.14	−0.03	−0.13	0.08
13	0.05	−0.11	0.05	0.09
14	0.00	−0.02	0.03	0.15
15	−0.01	−0.12	0.12	−0.05
16	−0.22	−0.16	−0.08	0.25
17	−0.08	−0.09	0.08	0.08
18	0.05	−0.07		0.11
19	−0.02	−0.09		0.04
20	−0.02	−0.09	−0.06	0.05
21	0.20	−0.02		−0.11
22	−0.12	−0.01	−0.10	0.11
23	0.00	0.09	0.02	−0.19
24		0.10		0.02
25		0.05		−0.02
26		0.04		−0.02
27	−0.11	0.14	−0.09	−0.03
28	0.08		−0.03	
29	−0.07	0.01	−0.07	0.06
30		−0.07		0.11
31		−0.07		0.04
32		0.09	−0.08	−0.14
33	−0.21	0.19		−0.23
34		0.02		−0.07
35		−0.14		0.02
36		0.06		−0.10
37		0.09		0.06
38				−0.04
39		0.00	0.10	−0.02
40		−0.05		0.02
41		−0.05		0.04
42		0.02		−0.03
43	*d*	*d*	*d*	*d*
44	−0.09	0.01	0.03	0.04
45		−0.02	−0.07	0.01
46		0.02	0.03	−0.09
47		0.05	−0.14	0.03
48		0.00		0.03
49		−0.03		0.01
50	*d*	*d*	*d*	*d*
51	−0.16	0.01		−0.01
52				*d*
53				0.13
54	−0.02	−0.02		0.05
55		−0.03		0.05
56	*d*	*d*	*d*	*d*
57		*d*		*d*
58	0.03	−0.17		0.05
59	0.14			0.02
60	0.16	−0.14	−0.08	0.04
61		0.09		−0.09
62		0.10		−0.14
63				0.07
64	−0.14	−0.17	0.20	0.09
65		−0.01		−0.02
66		−0.09		0.15
67	0.10	−0.07	−0.02	−0.11
68				0.07
69	*d*	*d*	*d*	*d*
70				*d*
71				*d*
72	−0.04			0.01
73				−0.01
74		−0.16		0.02
75				0.12

Table 7 (*continued*)

Compound	Alkane	Octanol	$CHCl_3$	PGDP
76				−0.05
77				−0.08
78				−0.06
79				0.04
80	0.04	−0.14		
81		−0.06		−0.02
82		−0.01		
83	0.05	−0.01		
84		−0.01		
85		*d*		
86	0.01	0.18		
87	−0.10	−0.06	−0.23	
88	−0.02	0.03	*−0.84*	0.08
89	*−0.35*	0.11	*−1.66*	−0.02
90	−0.11	−0.19	*−1.07*	0.33
91	−0.02	0.08	0.03	−0.06
92	−0.15	−0.04	*−2.19*	0.02
93	*d*	*d*	*d*	*d*
94	*−0.81*[d]	0.04	*−0.92*[d]	−0.03
95	0.00	−0.03	*−0.65*[d]	−0.02
96				−0.08
97		0.09		−0.06
98		0.00	0.09	−0.03
99	0.30	0.21	*0.50*	−0.14
100	0.01	0.02	*−0.82*[d]	0.06
101	0.08	−0.08	0.00	
102	0.14	−0.01	0.22	

[a] Residual = log P(obs) − log P(calc). [b] According to eqns. (17), (18), (19) and (20), respectively. [c] Italicised values are for compounds omitted from the regressions. [d] Statistical analysis inapplicable: see text.

Table 8 Experimental and calculated $f(CH_2)$ values

		$f(CH_2)$ derived from		
Solvent	Obs.[a]	ΔV_1[b]	ΔV_1[c]	ΔV_1[d]
Cyclohexane	0.64		(0.84)[e]	
Heptane	0.62			0.77
'Alkane'	0.62	0.56		
Chloroform	0.62	0.57	0.69	0.68
Diethyl ether	0.56			0.63
Octanol	0.53	0.50	0.58	0.66
PGDP	0.51	0.61		

[a] Ref. 27. [b] This work. [c] Ref. 8. [d] Ref. 9. [e] From $\Delta\bar{V}$ (ref. 6).

the least important term in any equation (except for octanol), but even so, this somewhat tarnishes our suggestion[14,47] that chloroform may prove a good model for potentially donor membranes such as polysaccharides. Chloroform is also unique in showing, as 0.38, by far the highest coefficient ratio of $n\beta_f$ to $\Sigma\beta$; the rest lie at 0.18–0.25. Possibly chloroform is too bulky for two molecules to bond with ease *e.g.* to the two lone pairs of carbonyl. If so, this must also limit its use as a model.

It is of interest to explore the relation between our hydrogen bonding parameters and those of Kamlet *et al.*[8] On the limited comparison which is possible that between $\Sigma\beta$ and β_m is only moderate, $r^2 = 0.945$ for $n = 34$, but most significantly, there is an intercept term: $\Sigma\beta = 0$ at $\beta_m = 0.08 \pm 0.02$. Some imprecision may be due to the muddle caused by $n\beta_f$, which is presumably 'lost' in their treatment, but the intercept is probably a medium effect. Just such an intercept is to be found for log K_β, based on TCE, *vs.* log K_B^H, based on tetrachloromethane,[21] and is due essentially to their different scale zeros. Since β_m is scaled to β_{solv} which in turn contains a contribution from log K in tetrachloromethane,[1] this suggestion is plausible. Similarly, $\Sigma\alpha$ *vs.* α_m gives $r^2 = 0.92$ with $\Sigma\alpha = 0$ at $\alpha_m = 0.18 \pm 0.03$ ($n = 8$). We conclude that neither β_m nor α_m is appropriate for partitioning studies; both will tend to exaggerate the strength of weak hydrogen bonds. Unfortunately, MRA can lose effects such as these in the blend,[13] and great vigilance is required to detect them.

However, the chief importance of this study is as LSER in reverse: the deduction, from solute–solvent interactions, of the solute's contribution. We explore in the following sections what we have discovered.

Carbonyl and Other C=Z *Species.*—Ketones and esters have been discussed. For both, difference in lone pair availability accounts for most of the difference in f-value between the aliphatic and aromatic series, the intrinsic β_f value changing very little (A and B in Table 9). Aldehydes if unhydrated should behave similarly, but we have no evidence. Amides are more complex. While in both series the expected β_f order $t > s > p$ is found,* only secondary amides (D) precisely parallel the above. Whereas all tertiary amides (C) have both lone pairs available, one is shielded by NH in all structures that incorporate $CONH_2$. Two lines of evidence support this hypothesis. Firstly, all primary and secondary amidic part-structures of the same type possess the identical $\Sigma\alpha$ value, indicating that Z-NH and the Z lone pair are mutually shielded. Secondly, the abnormal drop in β_f for aromatic *vs.* aliphatic $CONH_2$ (E) can be rationalised if both lone pairs are shielded, so that the only hydrogen bond now possible is along the C=O axis.† The same abnormality is found for carboxylic acids (F) and primary ureas (G), with a drop in β_f of 0.6–0.7 in these three cases as against a mean of *ca.* 0.2 elsewhere ($n = 5$). [The most crowded of the aromatic series, the primary ureas, still possesses a mean torsion angle of only 18 ± 15° ($n = 45$) in the solid state.[41] However the corresponding figure for thioureas is 62 ± 12° ($n = 21$):[41] see below]. We have noted[21] that, while carboxylic acids remain strong donors, they are *ca.* $10^{2.5}$ weaker than would be expected *e.g.* on pK_a arguments. Since all three compound types form dimers in the solid state,[41] no crystal structure evidence can be adduced for this phenomenon. Nevertheless this is how they must bind to other molecules, *e.g.* at the biological receptor. So far as we are aware, this point has not previously been considered.

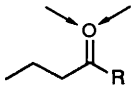
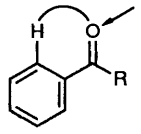
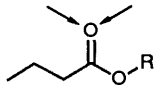
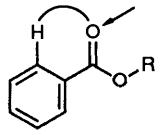
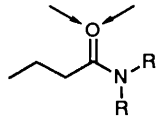
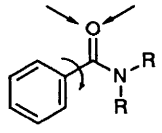
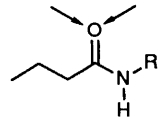
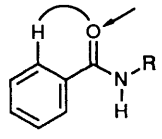
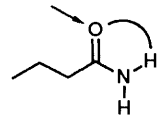
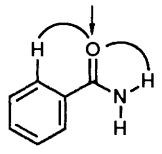
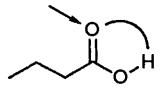
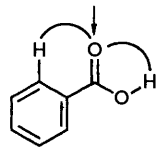
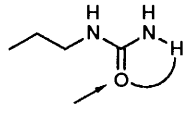
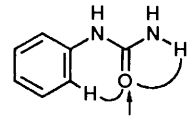
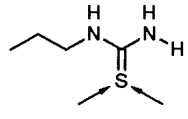
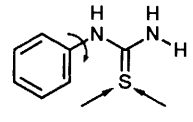
Important evidence relevant to the above propositions has been reported by Symons *et al.*[49–52] Acetone,[49] and all three classes of amide,[51,52] bond at both lone pairs in water. However in methanol, with very little 'free' OH,[53] secondary and tertiary amides form both mono- and di-solvates (of C=O), whereas primary amides form only a monosolvate.[52] Acetone in methanol,[49] and methyl acetate in both methanol and water,[50] form a mixture of mono- and di-solvates.

In interpreting these results, it is necessary to distinguish between binding saturation and bond strength. On the extreme assumption of 55 mol dm^{-3} excess OH in water, 90% saturation of the second lone pair would result in $K_2 \approx 0.2$. This value represents a rather weak bond and is likely to contribute little to overall free energy. Of course, beyond that minimum value we do not know what K_2 actually is; that cannot be deduced from Symons' work. Nevertheless in methanol, which should behave

* It was forcing this order on chemical grounds (*cf.* comments above) that revealed the phenomena here described. Note that **27** possesses[41] a torsion angle of only 13°.

† In an important solution study, Laurence *et al.*[48] have shown the presence of both types of bonding, with the linear much more favoured. However, their work was in a non-polar solvent, used much bulkier probes than water, and did not allow of multiple contacts, all of which should favour linearity. There is no necessary clash between their work and ours.

Table 9 Strength and directionality of bonding to carbonyl[a]

Case		$\beta_f/\Sigma\alpha$	f_{oct}[b]		$\beta_f/\Sigma\alpha$	f_{oct}[b]
(A)		1.9	−1.90		1.9	−1.09
(B)		1.8	−1.49		1.6	−0.56
(C)		3.6	−3.04		3.3	−2.80
(D)		3.3/1.0	−2.71		3.1/1.0	−1.81
(E)		3.3/1.0	−2.18		2.7/1.0	−1.26
(F)		1.8/2.1	−1.11		1.1/2.2	−0.03
(G)		2.9/2.6	−2.18		2.3/2.2	−1.07
(H)		2.0/2.0	−1.29		1.8/1.8	−1.17

[a] Straight arrows indicate hydrogen bond direction, curved arrows indicate twisting (non-planarity). A curved line indicates where mutual shielding of H and lone pair is believed to occur. [b] Ref. 31.

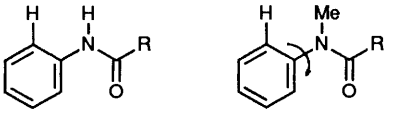
in a qualitatively similar manner to water, it is clear that the K_1/K_2 ratio is very much greater for primary than for secondary or tertiary amides. Hence, for primary amides, it is probably fair to discount the second lone pair as making much contribution to binding strength as matters in any competitive process, such as partitioning or drug–receptor binding, itself a form of partitioning. (Fujita's approach[46] to hydrogen bonding as a factor in log *P*, which we have shown does not work in practice,[27] is vitiated by just this confusion between strength and saturation.) Symons' picture is entirely compatible with ours, given that ours reveals only those bonds that contribute appreciably to ΔG for the overall binding process. It should be noted that Symons' results are all for aliphatic species, so throw no light on possible shielding by *peri*-CH.

Thiones (H) are exceptional. Here it seems that the greater bond length of C=S *vs.* C=O lifts the sulfur lone pairs clear of both types of shielding (their lesser directionality[54] may also help, as also the severe twisting noted above). Hence we can explain the anomalous f_{oct} order, (Ar)NHCSNH$_2$ < (Ar)NHCONH$_2$ (Table 9), as the difference between two lone pairs and none, without having to invoke anomalous intrinsic acceptor ability. This difference may, again, be relevant to receptor binding. The expected f_{oct} order is restored for the aliphatic pair. The only sets where lone pair availability is the same for both series, (C) and (H), also show much the smallest inter-series f_{oct} differences.

The above-mentioned contrast in behaviour between (E)–(G) and the remainder amounts to an extra drop of 0.4–0.5 in β_f when hydrogen bonding along the line of the lone pair is blocked out. It follows that, typically, almost two-thirds of the

strength of hydrogen bonding to carbonyl is due to the directional or charge-transfer component. This conclusion is in total contrast with most MO analyses[55] which tend to treat the strength of hydrogen bonding as largely electrostatic in origin, charge transfer being responsible only for its directionality. There is no real conflict: MO calculation effectively refers to the gas phase at $\varepsilon = 1$, while our results pertain to water-based systems where dipole–dipole interaction must be greatly attenuated. In fact it is well established that increasing solvent polarity favours the charge transfer *vs.* the electrostatic component in hydrogen bonding.[56] Nevertheless there has been some tendency to treat MO calculation as directly applicable to the biological receptor, whereas this new evidence makes clear a major limitation. Another relevant factor is that MO calculation on the isolated molecule is a measure of ΔH not ΔG, *cf.* our previous discussion[21] on the relation between β_{sm} and log K_β.

Electronic effects are generally as expected: aromatics are weaker acceptors than aliphatics, diaryls *e.g.* PhCOPh (**21**) are weaker again, C=S is much weaker than C=O, and urethanes are weaker than the corresponding ureas (Table 3). Some apparent anomalies in β_f are explicable as due to planarity or the lack of it, as in the contrasts of Scheme 1; here *N*-alkylation increases

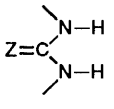


Scheme 1

the torsion angle from 21 ± 13° ($n = 24$) to 83 ± 9° ($n = 13$).[41] One that must be genuine, and so far as we know unsuspected, is the consistently weaker proton acceptor ability of ureas *vs.* the corresponding carboxamide. This effect is appreciable (mean $\Delta\beta_f$ −0.3) and unexpected since ureas are (slightly) stronger proton transfer bases than carboxamides. We presume that σ-resonance, which operates in esters[42] and amides to make the *E* lone pair less available (Scheme 2), in ureas and *e.g.* urethanes now operates on both. Experimental data are available only for tetrasubstituted ureas (Table 3) whose geometry may be abnormal.

Scheme 2

The value of Σα *ca.* 1.0 for amide NH is slightly greater than expected (Table 3) and may indicate a small degree of co-operativity. It is enhanced to *ca.* 1.6 on *N*-aryl substitution for electronic reasons. No such reason will explain its remarkable enhancement for ureas (G) and thioureas (H). We have IR evidence[21] for this phenomenon in one compound but it now appears to be general wherever the part-structure **104** is present.

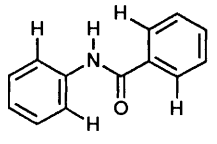


104

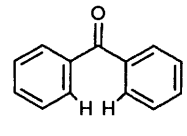
We believe it to originate in dipole–dipole repulsion between the two NHs which either leads to specially strong bifurcated bonds, or simply makes bonding more favourable (these alternatives are not mutually exclusive). Its considerable variability (Σα 1.8–2.6) is much more than can be explained on electronic grounds and may point to small but significant differences in H–H separation. That donors of type **104** should be as strong as phenols or carboxylic acids has, again, not been suspected so far as we are aware, and may have important biological implications. Presumably urea itself possesses this property. One such compound, the cyanoguanidine **98** with Σα = 2.8, is the strongest proton donor in this set.

In confirmation of this phenomenon's origin, *N*-methylation of either NH as in **36** and **39** destroys it. Equally, it is absent in the sulfamide $PhNHSO_2NH_2$ **49** relative to the sulfonamide $PhSO_2NH_2$ **45** since, in the virtual absence of resonance,[57] the constraints which force planarity in **104** do not apply to **49**.

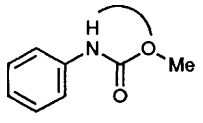
Of lesser phenomena, we single out the following. (a) Based on Scheme 1 (see above), 1,5 and (inevitably) 1,6 H–methyl interactions are likely to release one carbonyl lone pair whereas the corresponding H–H clashes apparently do not except when two are present, as in **31**, which behaves as if it possesses one



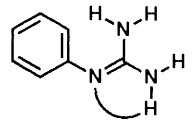
31



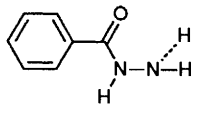
21



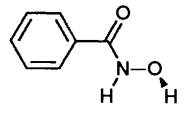
41



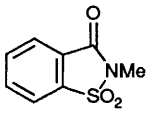
42



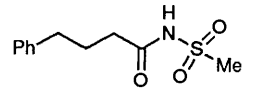
25



26



105



81

lone pair. In the solid state,[41] this structural unit shows torsion angles of 31 ± 8° and 66 ± 8° for phenyl adjacent to CO and NH respectively ($n = 11$). Benzophenones show twisting between the ring planes and CO of 30 ± 7° in the crystal state ($n = 18$)[41] and, in solution, this 1,7 H–H interaction also appears to release one effective lone pair in **21** itself. It will be noticed, here and previously, how imprecise is the guidance provided by torsion angle to solution lone pair availability. (b) The abnormally low Σα = 0.4 for the urethane **41** is probably due to lone pair shielding as shown; its more usual value in $PhCH_2OCONH_2$ (**65**) may be caused by twisting of the bulky substituent. (c) Phenylguanidine (**42**), while a strong acceptor, is *ca.* $\Delta\beta_f = 1.2$ weaker than it should be from its pK_a by comparison with other imines.[21] On the analogy of carboxamide, this is probably due to lone pair shielding as shown. By contrast, the cyanoguanidine (**98**) is *ca.* $\Delta\beta_f = 1.8$ stronger. This adds to the evidence[21] that nitrile is its principal hydrogen bond acceptor site. (d) Values for Σα of 0.9 and 1.6 for benzoylhydrazine (**25**) and benzenehydroxamic acid (**26**) respectively suggest NH as the principal proton donor of each, given their known[58] conformations as shown, in which NH_2 for the former and OH in the latter do not bond to carbonyl. The surprisingly high β_f 3.8 for **25** may indicate an appreciable residual acceptor ability for its amino-group; **26** also shows some elevation.

Table 10 β_{Ar} Values

β_{Ar}	Species	β_{ring} [a]
0.4	Naphthalene	
	$PhNR_2$ (R = H or alkyl)	0.17
	PhOR (R = H or alkyl)	0.13
0.3	Benzene, styrene, vinylbenzene	
	PhR (R = H or alkyl)	0.14
	$PhCH_2R$ (R = OR′ or NR_2′)	0.13
	PhNHR (R = COR′)	
	PhF	0.10
0.2	$PhCH_2$ (R = COR′)	
	PhOR (R = COR′)	
0.0	PhCl, PhBr, PhI, PhCOR	0.09
	$PhCF_3$, PhSOR, $PhSO_2R$	
	PhCN	0.06
	$PhNO_2$	0.04

[a] Ref. 39.

(e) The most spectacular predictive failure of Table 3 concerns saccharin's *N*-methyl derivative **105** for the acylsulfonamide **81**. While C=O and SO_2 in **105** are held well apart, the preferred *trans*-conformation of the amide group could bring them close together in **81** as shown, resulting in a sort of 'α-effect'.[59]

Phenyl as Proton Acceptor.—The benzenoid π-cloud is a potential proton acceptor for which values have been measured[25,34] and used in the LSER analysis of partitioning.[8] Abraham[39] has attempted to dissect the β_{Ar} and β_f contributions to some simple aromatics and some of his results for β_{ring} (scaled differently) are shown alongside ours in Table 10. Their general correspondence is as good as can be expected. Two points are specially significant. Firstly, electron donors increase β_{Ar} by far less than electron acceptors decrease it; both scales agree on this. In terms of σ this is understandable if σ_I contributes more to the blend than σ_R, as we[14,38] have argued should be the case for hydrogen bonding. (One slight surprise is that PhF alone among the halobenzenes should possess acceptor ability; we do not know whether this is due to substituent or ring). Secondly, the β_{Ar} scale clearly has a much higher cut-off point than β_{ring}; by eye, zero for the first corresponds roughly to 0.1 for the latter. This is consistent with $\Sigma\beta = 0$ at Taft's[8] $\beta_m = 0.08$ as noted above, and again probably stems from the use of very different solvents for generating the two scales; the cut-off point for β_{Ar} is essentially that relevant to water.

Table 10 throws some unexpected light on alkyl–aryl *f*-value differences. For carbonyl compounds, as seen above, most of the difference is due to lone pair shielding. The other main effect is the extinction of β_{Ar}, and this applies to electronegative groups in general, *e.g.* nitrile. We can also show that the need for Leo's special category of benzyl *f*-value[31] has two distinct origins. Relative to the alkyl value, *f* for electron donors becomes more positive through a drop in β_f. That for electron acceptors rises through a drop in β_{Ar}. It is encouraging that these quite small effects should be so accurately reproduced.

Cooperative Effects in OH *and* NH.—It has been seen above that cooperative effects are quite small when, as *e.g.* in carboxamides, donor and acceptor involve distinct heteroatoms. Where these are the same, as in OH and NH, there is clearly more scope for cooperativity, and indeed the very high α_{solv} and β_{solv} value for alcohols has been attributed to this.[1] We indeed find cooperativity, but its pattern is peculiar.

Relative to ether, alkyl OH shows less than a twofold enhancement in acceptor and none in donor ability (Table 3). This is vastly less than required to account for the bulk solvent properties of alcohols, and points to mass action as a major factor (see Appendix for further discussion). There is no sign of enhancement in either for phenol. Aromatic NMe_2 gives about the expected β_f but this is enhanced for NHMe and much enhanced for NH_2; Kamlet *et al.*[8] list a similar trend in β_m but do not comment. At the same time, $\Sigma\alpha$ shows a spectacular fall; aniline is not a donor in the PGDP–water system (Fig. 4). This too is echoed by Kamlet *et al.*,[8] again without comment. Some of this peculiarity may stem from the amine inversion process, which leads to an abnormal relation between acceptor ability and pK_a[21] and may in some manner be reflected asymmetrically by the two simultaneous hydrogen bonding processes.

Primary and secondary alkylamines were omitted from the regression analysis since these are known to be vanishingly poor proton donors[21,33,60] so no prior estimate of $\Sigma\alpha$ was possible. We have attempted to back-calculate their β_f and $\Sigma\alpha$ values using the 'octanol assumption'. The results are not very satisfactory, owing chiefly to solvent inconsistencies such that PGDP tends to estimate $\Sigma\alpha$ higher than the remainder; the most coherent results that we can manage are given in Table 11. Nevertheless the sequence in β_f is reasonable ($t > s > p$) and the results show clearly that $\Sigma\alpha$ increases with amine basicity, the alkyl > benzyl > aryl sequence of 0.7 > 0.5 > 0.1 being almost quantitatively that of pK_a (at a Brønsted α of *ca.* 0.1). We are forced to rationalise this totally unexpected result by postulating that hydrogen bonding to the amine lone pair, much greater of course for the stronger bases, polarises N–H progressively towards $N–H^+$ as it proceeds, so that $\Sigma\alpha$ follows β_f instead of opposing it. This goes clean contrary to the usual electronic arguments and, so far as we are aware, has no precedent. Of course all these donors are still very weak, which adds greatly to the difficulty of determining them.

P=O *and* S=O *Bases.*—These were omitted for an entirely different reason: no consistent β_f has proved possible. That is, acceptor ability varies with the solvent system. It is known that X=O bases where X is a second-row element are better represented in the dipolar form $X^+–O^-$ by MO calculation,[61] in sharp contrast with C=O and N=O. Hence their proton acceptor ability is likely to show abnormal sensitivity to solvent, as indeed we have demonstrated.[21] If falling solvent polarity leads to a smooth transition from X=O to $X^+–O^-$ then effectively an extra lone pair comes into being. We may test this hypothesis by back-calculation. If a constant β_f, on the basis of some suitable model (Table 3), is assumed for each species, we may calculate $FR(n\beta_f)$ by difference to yield an approximate estimate of the number of available lone pairs. The result of this calculation appears in Table 12. A single lone pair appears for PhSOMe (**43**) in octanol and PGDP, so that, allowing for shielding of one lone pair as in acetophenone, sulfoxide here is represented as S=O. In 'alkane' and chloroform it is close to $S^+–O^-$. By contrast, $PhSO_2Me$ (**44**) behaves regularly. Aliphatic sulfoxide seems to be more polarisible, appearing as S=O in octanol, $S^+–O^-$ in 'alkane', half-way between in PGDP, and (impossibly) in excess of either in chloroform. (In considering errors, note that a single lone pair electron here accounts for $\Delta\log P = 0.6–1.0$ according to β_f and the slope of $n\beta_f$). Aliphatic sulfone is also more polarisible, with octanol and PGDP still behaving regularly but 'alkane' and chloroform apparently giving rise to highly dipolar species. Sulfonamide is less polarisible than sulfone, with only chloroform in **95** 'abnormal', while P=O is possibly the most polarisible species of all.

The order of apparent solvent polarity revealed by these data is octanol > PGDP > 'alkane' > chloroform. In view of its high permittivity, the position of chloroform comes as a

Table 11 Best-fit residuals for primary and secondary aliphatic amines[a,b]

Compound	β_f	$\Sigma\alpha$	'Alkane'	Octanol	Chloroform	PGDP
56 $PhCH_2NH_2$	2.1	0.5	0.46	0.06	0.26	−0.44
57 $PhCH_2NHMe$	2.2	0.5		0.13		−0.15
69 $Ph(CH_2)_2NH_2$	2.5	0.7	0.23	0.17	0.35	−0.22
85 $Ph(CH_2)_3NH_2$	2.5	0.7		0.07		
70 $Ph(CH_2)_2NHMe$	2.7	0.7				−0.10
71 $Ph(CH_2)_2NHEt$	2.7	0.7				−0.14

[a] By back-calculation (see text). [b] See Table 2 for definition of β_f and $\Sigma\alpha$.

Table 12 Lone pair involvement in S=O and P=O bases[a]

Compound	β_f	'Alkane'	Octanol	Chloroform	PGDP
43 PhSOMe	3.2	2.2	1.1	2.1	0.9
93 p-$NO_2PhO(CH_2)_3SOMe$	3.3	3.0	1.5	3.5	2.6
52 $NpO(CH_2)_3SOMe$	3.3				2.7
44 $PhSO_2Me$	2.0	4.2	4.0	3.9	3.9
94 p-$NO_2PhO(CH_2)_3SO_2Me$	2.1	5.6	3.9	5.9	4.1
53 $NpO(CH_2)_3SO_2Me$	2.1				3.7
45 $PhSO_2NH_2$	1.8		4.1	4.2	4.0
95 p-$NO_2PhO(CH_2)_3SO_2NH_2$	1.9	4.0	4.1	5.3	4.1
50 $Ph_3P{=}O$	4.2	2.7	1.5	3.8	3.0

[a] By back-calculation (see the text). Numbers refer to total number of lone pair electrons required to fit data at constant functional group β (β_f).

surprise. There are two possible explanations. Perhaps there is some special repulsive force due to the C–Cl dipole, but if so, it is difficult to see why dipolar bases should be singled out. The other lies in the nature of hydrogen bonding to CH. It is known[16] that chloroform's CH forms virtually a pure electrostatic hydrogen bond; if so, this solvent even more than 'alkane' may induce charge separation. Regardless of its explanation, however, this is perhaps the most extreme variability so far reported in the behaviour of X=O bases, and suggests a number of ways in which these may be used as probes in the context of membrane binding and penetration.

It should be emphasised that the above is only one possible way of analysing these data, and indeed its central presumption, of a solvent-invariable β_f, cannot be wholly correct [*cf.* **50** and **93** in chloroform]. Nevertheless, it appears remarkably successful.

Other Species.—Four deserve comment. The heavily fluorinated thiourea **100** is well behaved elsewhere but lower by $\Delta\log P = 0.82$ than expected in chloroform; we attribute this to repulsion between the C–F and C–Cl dipoles. Repulsion of the incoming proton acceptor by C–F will also account for the weaker donor properties of $PhC(CF_3)_2OH$ (**97**) relative to benzyl alcohol (**54**). Its close analogue $(CF_3)_2CHOH$ is a strong donor,[21,33] but **97** is much more crowded. We also draw attention to the very poor acceptor ability of aromatic NO_2. Despite its formal similarity to SO_2, it is inconceivable therefore that all four lone pairs are available in a similar way. There is indeed crystal structure evidence[62] that NO_2 forms a single bifurcated bond involving both oxygens. Finally we note the evidence of Symons *et al.*,[63] that aliphatic nitriles form more than one hydrogen bond to water, as a possible explanation for this group's anomalously high β_f value (Table 3).

Conclusions

The present study has provided not only an unprecedented selection of quantitative hydrogen bond strengths, tuned to the needs of the medicinal chemist, but information on hydrogen bond directionality that might have been obtained in no other way. It needs to be emphasised that crystal structure, in this context, is an equivocal guide. Not only is the information it provides on planarity apt to be misleading when translated to solution chemistry (particularly because compounds tend to pack in the most compact manner possible); the degree of non-planarity revealed in the solid state, as noted above, appears to have very imprecise consequences for lone pair availability. Probably a torsion angle of 15° still implies eclipsing by CH, and one of 40° implies release, but between these limits, little can be said. Furthermore, the solid state gives no information concerning the mutual eclipsing of lone pairs by OH or NH, since in practice this is avoided by dimerisation. It requires a solution technique to reveal the subtleties of interactions in solution.

A project that started as a study in LSER has finished as LSER in reverse. Information of the type presented *e.g.* in Table 9, ripe for the medicinal chemist's immediate use, has no precedent that we know of. All previous treatments have had to assume some type of extrapolation, as that of MO theory from gas phase to solution. Our evidence is direct. It derives from water-based solvent systems of known relevance to biology;[64] it returns the compliment by quantifying the properties of the solute. These should be equally applicable to water, other solvents, and the biophase. We look forward to their application to the biological receptor.

Appendix

The Water Paradox

In terms of solvent–water partitioning, there are two anomalies in the behaviour of water as proton acceptor that have never been satisfactorily explained, nor even an analysis attempted. We attempt that analysis here.

The original solvent listing of Kamlet *et al.*[1,25] gives water as $\beta_{solv} = 0.18$, far less than for bulk alcohols (typically 0.6–0.9). This would imply a large positive coefficient of $\Sigma\alpha$ for octanol, yet Kamlet *et al.*[8] find a slight positive slope while El-Tayar *et al.*[9] and ourselves find slight negative slopes. That is the first anomaly.

Nevertheless, we[38] have established quite clearly that, in heterocycles containing amphiprotic substituents, increase in proton donor strength is just as effective as decrease in proton

Table 13 Some partitioning solvent parameters

Solvent	a[a]	log K_β[b]	β_{solv}[c]	$[S_\beta]$[d]
Octanol	−0.10	*ca.* 1.4	0.88[e]	7.86[f]
PGDP	−0.61	*ca.* 1.4	0.45[g]	5.07
Chloroform	−0.98		0.0	
'Alkane'	−1.07		0.0	
Water		*ca.* 1.2[h]	0.18	55.5

[a] Coefficient of $\Sigma\alpha$ term in final correlation equations (Table 5). [b] Ref. 21. [c] Ref. 25. [d] Molar concentration in solvent of proton acceptor groups. [e] For BuOH. [f] Includes equilibrium concentration of water. [g] For EtOAc. [h] Scaled from log K_B^H (ref. 34).

Table 14 Δf Values for ether and hydroxy[a]

	'Alkane'	Octanol	Chloroform	PGDP
AlkOAlk	−2.28	−1.56	−1.30	−1.67
ArOAlk	−0.80	−0.55	−0.30	−0.46
Δf	1.48	1.01	1.00	1.21
AlkOH	−3.73	−1.67	−2.52	−2.48
ArOH	−2.90	−0.50	−2.23	−1.03
Δf	0.83	1.17	0.29	1.45
$\Delta\Delta f$	−0.65	0.16	−0.71	0.24

[a] f-Values from Table 3.

acceptor strength for raising log P_{oct}. This result can be generalised to multisubstituted benzenes[65] and other heterocycles.[66] This second anomaly is clearly in conflict with the first and helps to re-establish the original[1,25] position.

Part of the problem must lie in the nature of the solvatochromic process. Despite bulk water's exceptional proton donor properties (α_{solv} = 1.17),[25] only about 3.5 of the possible 4 hydrogen bonds are formed at ambient temperature,[67] which given a concentration of 55 mol dm^{-3}, points to an extreme reluctance of the second water lone pair to form a hydrogen bond. This is consistent with Hine's evidence[36] for ethers quoted above. Hence a solute proton donor in competition with water's excess protons has largely to make do with water's second lone pair, giving a quite weak hydrogen bond which is reflected by β_{solv} = 0.18. Alcohols contain no excess protons so this situation does not arise.

Some solvent parameters relevant to acceptor ability are assembled in Table 13. Eqn. (21) demonstrates an excellent

$$a = 1.04(8)\beta_{solv} - 1.04(4) \quad (21)$$

$$(n = 4 \quad r^2 = 0.989 \quad s = 0.06 \quad F = 178)$$

relation between the slope of the $\Sigma\alpha$ term and β_{solv} for the organic phase. No other relation is apparent, and no second term is significant. Its simplicity is no doubt helped by the nearly equal log K_β values. There would be no problem except that eqn. (21) predicts β_{solv} = 1.00 for water (*i.e.* the point at which a = 0). We believe the source of this paradox to lie in number density: the enormous discrepancy between $[S_\beta]$ for water and for every other solvent. That is: it depends on mass action. In view of our failure to detect any very large degree of cooperativity involving alcohols in water (see above), we suspect that mass action, and not as supposed[1] cooperativity, may be the dominant factor in the high α_{solv} and β_{solv} values of the alcohols.

The substitution of X for H in RH to give RX is a displacement process in which the substituent X carries with it its complete baggage of enthalpic and entropic terms. Strictly this shows as a substituent π-value,[64] but π and f are related very simply by eqn. (22) and little error is introduced if we use f-values instead. By contrast, the process described by eqn. (23) is

$$f_X = \pi_X + f_H \quad (22)$$

$$\Delta\pi_X = \pi_X(\text{heterocycle}) - \pi_X(\text{benzene}) \quad (23)$$

essentially isodesmic, as Scheme 3 shows: here it is known[38,66] that $\Delta\pi$ is positive whenever X in the heterocycle is a stronger proton donor than X in benzene. Such an isodesmic process is largely isoentropic—there is no change in volume, and none or very little in rigidity or conformation—so that the enthalpic component of log P will tend to dominate $\Delta\pi$. Hence the mass

Scheme 3

action or number density effect, which has nothing to do with intrinsic affinity, largely disappears, and we are back with water's β_{solv} value which reflects the latter. We have previously argued[21] that solvatochromic β-values are largely enthalpic in nature.

If this argument is correct, it should be reflected in our data. We may regard the substitution of X by another X′ identical except in electronic properties as a parallel process to that of Scheme 3. Two such substitutions are those of aromatic for aliphatic ether and OH; these have the desirable feature that they parallel the electronic changes of Scheme 3. Table 14 lists Δf values for all four solvent systems, and also $\Delta\Delta f$, the extent to which Δf for OH exceeds or falls short of Δf for ether. As expected, all Δf values are positive: the aromatic species are less hydrophilic with respect to all solvents. However, for octanol and PGDP, $\Delta\Delta f$ is also positive: increase in proton donor ability aids partitioning into the solvent. The opposite effect for 'alkane' and chloroform is much more marked, but inevitable since these possess no proton acceptor properties.

To summarise: mass action dominates the first paradox but is much more muted in the second. Both appear to be resolved.

Acknowledgements

We thank Dr. D. J. Gilman for help and advice concerning crystal structure.

References

1 M. J. Kamlet, J.-L. M. Abboud and R. W. Taft, *Prog. Phys. Org. Chem.*, 1981, **13**, 485.
2 M. H. Abraham, R. M. Doherty, M. J. Kamlet and R. W. Taft, *Chem. Br.*, 1986, **22**, 551.
3 M. J. Kamlet, R. M. Doherty, J.-L. M. Abboud, M. H. Abraham and R. W. Taft, CHEMTECH, 1986, 566.
4 M. J. Kamlet, R. M. Doherty, J.-L. M. Abboud, M. H. Abraham and R. W. Taft, *J. Pharm. Sci.*, 1986, **75**, 338.
5 M. J. Kamlet, R. M. Doherty, M. H. Abraham, P. W. Carr and R. W. Taft, *J. Phys. Chem.*, 1987, **91**, 1996.
6 (*a*) M. J. Kamlet, M. H. Abraham, R. M. Doherty and R. W. Taft, *J. Am. Chem. Soc.*, 1984, **106**, 464; (*b*) R. W. Taft, M. H. Abraham, G. R. Famini, R. M. Doherty, J.-L. M. Abboud and M. J. Kamlet, *J. Pharm. Sci.*, 1985, **74**, 807.
7 D. E. Leahy, *J. Pharm. Sci.*, 1986, **75**, 629.
8 M. J. Kamlet, R. M. Doherty, M. H. Abraham, Y. Marcus and R. W. Taft, *J. Phys. Chem.*, 1988, **92**, 5244.
9 N. El-Tayar, R. S. Tsai, B. Testa, P. A. Carrupt and A. Leo, *J. Pharm. Sci.*, 1991, **80**, 590.
10 See, *e.g.*, M. J. Kamlet, R. M. Doherty, M. H. Abraham and R. W. Taft, *Nature (London)*, 1985, **313**, 384; M. J. Kamlet, D. J. Abraham, R. M. Doherty, R. W. Taft and M. H. Abraham, *J. Pharm. Sci.*, 1986, **75**, 350.
11 M. H. Abraham, G. S. Whiting, R. M. Doherty and W. J. Shuely, *J. Chem. Soc., Perkin Trans. 2*, 1990, 1451.

12 M. H. Abraham, G. S. Whiting, Y. Alarie, J. J. Morris, P. J. Taylor, R. M. Doherty, R. W. Taft and G. O. Nielsen, *Quant. Struct.-Act. Relat.*, 1990, **9**, 6.
13 M. Sjöström and S. Wold, *Acta Chem. Scand., Ser. B*, 1981, **35**, 537; S. Wold and M. Sjöström, *Acta Chem. Scand., Ser. B*, 1986, **40**, 270.
14 P. J. Taylor, *Hydrophobic Properties of Drugs*, in *Comprehensive Medicinal Chemistry*, ed. C. Hansch, Pergamon, Oxford, 1990, vol. 4 (ed. C. A. Ramsden), p. 241.
15 M. J. Kamlet and R. W. Taft, *Acta Chem. Scand., Ser. B*, 1985, **39**, 611; M. J. Kamlet, R. M. Doherty, G. R. Famini and R. W. Taft, *Acta Chem. Scand., Ser. B*, 1987, **41**, 589.
16 P.-C. Maria, J.-F. Gal, J. de Franceschi and E. Fargin, *J. Am. Chem. Soc.*, 1987, **109**, 483.
17 M. Chastrette, M. Rajzmann, M. Chanon and K. F. Purcell, *J. Am. Chem. Soc.*, 1985, **107**, 1.
18 M. J. Kamlet and R. W. Taft, *J. Am. Chem. Soc.*, 1976, **98**, 377.
19 M. J. Kamlet and R. W. Taft, *J. Am. Chem. Soc.*, 1976, **98**, 2886; R. W. Taft and M. J. Kamlet, *J. Chem. Soc., Perkin Trans. 2*, 1979, 1723.
20 M. J. Kamlet, J.-L. M. Abboud and R. W. Taft, *J. Am. Chem. Soc.*, 1977, **99**, (*a*) 6027, (*b*) 8325.
21 M. H. Abraham, P. P. Duce, D. V. Prior, D. G. Barratt, J. J. Morris and P. J. Taylor, *J. Chem. Soc., Perkin Trans. 2*, 1989, 1355.
22 M. H. Abraham, G. J. Buist, P. L. Grellier, R. A. McGill, D. V. Prior, S. Oliver, E. Turner, J. J. Morris, P. J. Taylor, P. Nicolet, P.-C. Maria, J.-F. Gal, J.-L. M. Abboud, R. M. Doherty, M. J. Kamlet and R. W. Taft, *J. Phys. Org. Chem.*, 1989, **2**, 540.
23 B. Frange, J.-L. M. Abboud, C. Benamou and L. Bellon, *J. Org. Chem.*, 1982, **47**, 4553; J.-L. M. Abboud, L. Sraidi, G. Guiheneuf, A. Negro, M. J. Kamlet and R. W. Taft, *J. Org. Chem.*, 1985, **50**, 2870.
24 R. W. Taft, J.-L. M. Abboud, M. J. Kamlet and M. H. Abraham, *J. Solution Chem.*, 1985, **14**, 153.
25 M. J. Kamlet, J.-L. M. Abboud, M. H. Abraham and R. W. Taft, *J. Org. Chem.*, 1983, **48**, 2877.
26 R. W. Taft, M. H. Abraham and M. J. Kamlet, *J. Am. Chem. Soc.*, 1981, **103**, 1080.
27 Part 2, D. E. Leahy, J. J. Morris, P. J. Taylor and A. R. Wait, *J. Chem. Soc., Perkin Trans. 2*, 1992, preceding paper.
28 M. H. Abraham and R. Fuchs, *J. Chem. Soc., Perkin Trans. 2*, 1988, 523.
29 M. J. Kamlet, personal communication.
30 J. G. Kirkwood, *J. Chem. Phys.*, 1934, **2**, 351.
31 THOR Masterfile 351; CLOGP version 3.51, Daylight Chemical Information Systems, 3951 Claremont St., Irvine, CA, 92714.
32 J. J. P. Stewart, MOPAC, *QCPE* No. 455, *QCPE Bull.*, 1983, **3**, 43.
33 M. H. Abraham, P. L. Grellier, D. V. Prior, P. P. Duce, J. J. Morris and P. J. Taylor, *J. Chem. Soc., Perkin Trans. 2*, 1989, 699.
34 M. H. Abraham, P. L. Grellier, D. V. Prior, J. J. Morris and P. J. Taylor, *J. Chem. Soc., Perkin Trans. 2*, 1990, 521.
35 Preliminary communication: P. J. Taylor, in *QSAR: Rational Approaches to the Design of Bioactive Compounds*, eds. C. Silipo and A. Vittoria, Elsevier, Amsterdam, 1991, p. 75; P. J. Taylor, paper presented at the 8th European Symposium on QSAR, Sorrento, Italy, Septermber 1990.
36 J. Hine, S. Hahn and D. E. Miles, *J. Org. Chem.*, 1986, **51**, 577.
37 F. Franks and D. J. G. Ives, *Q. Rev. Chem. Soc.*, 1966, **20**, 1.
38 S. J. Lewis, M. S. Mirrlees and P. J. Taylor, *Quant. Struct.-Act. Relat.*, 1983, **2**, 100.
39 M. H. Abraham, personal communication.
40 R. F. Rekker, *The Hydrophobic Fragmental Constant*, Elsevier, Amsterdam, 1977.
41 Cambridge Crystallographic Data Base. Cambridge, UK.
42 P. L. Huyskens, H. Marshal and Th. Zeegers-Huyskens, *J. Mol. Struct.*, 1987, **158**, 379.
43 R. W. Taft, M. Taagepera, J.-L. M. Abboud, J. F. Wolf, D. J. DeFrees, W. J. Hehre, J. E. Bartmess and R. T. McIver, Jr., *J. Am. Chem. Soc.*, 1978, **100**, 7765.
44 A. Leo, C. Hansch and D. Elkins, *Chem. Rev.*, 1971, **71**, 525.
45 N. G. J. Richards, personal communication.
46 T. Fujita, T. Nishioka and M. Nakajima, *J. Med. Chem.*, 1977, **20**, 1071.
47 D. E. Leahy, P. J. Taylor, and A. R. Wait, *Quant. Struct.-Act. Relat.*, 1989, **8**, 17.
48 C. Laurence, M. Berthelot and M. Helbert, *Spectrochim. Acta, Part A*, 1985, **41**, 883.
49 M. C. R. Symons and G. Eaton, *J. Chem. Soc., Faraday Trans. 1*, 1985, **81**, 1963.
50 K. B. Patel, G. Eaton and M. C. R. Symons, *J. Chem. Soc., Faraday Trans. 1*, 1985, **81**, 2775.
51 G. Eaton and M. C. R. Symons, *J. Chem. Soc., Faraday Trans. 1*, 1988, **84**, 3459.
52 G. Eaton, M. C. R. Symons and P. P. Rastogi, *J. Chem. Soc., Faraday Trans. 1*, 1989, **85**, 3257.
53 M. C. R. Symons, J. M. Harvey and S. E. Jackson, *J. Chem. Soc., Faraday Trans. 1*, 1980, **76**, 256; W. A. P. Luck and W. Ditter, *Adv. Mol. Relaxation Processes*, 1972, **3**, 321.
54 E. Flood and J. E. Boggs, *J. Mol. Struct.*, 1976, **34**, 147.
55 See, *e.g.*, P. Kollman, J. McKelvey, A. Johansson and S. Rothenberg, *J. Am. Chem. Soc.*, 1975, **97**, 955; S. Rothenberg and H. F. Schaeffer, III, *J. Chem. Phys.*, 1970, **53**, 3014.
56 M. J. Kamlet, C. Dickinson, T. Gramstad and R. W. Taft, *J. Org. Chem.*, 1982, **47**, 4971.
57 M. Charton, *Prog. Phys. Org. Chem.*, 1981, **13**, 119.
58 O. Exner, in *The Chemistry of Double-bonded Functional Groups, Supp. A*, ed. S. Patai, Wiley, New York, 1977, p. 1.
59 J. D. Aubort and R. F. Hudson, *Chem. Commun.*, 1970, 937; R. W. Taft, F. Anvia, M. Taagepera, J. Catalan and J. Elguero, *J. Am. Chem. Soc.*, 1986, **108**, 3237.
60 D. D. Nelson, Jr., G. T. Fraser and W. Klemperer, *Science*, 1987, **238**, 1670.
61 F. Bernardi, H. B. Schlegel, M.-H. Whangbo and S. Wolfe, *J. Am. Chem. Soc.*, 1977, **99**, 5633.
62 T. W. Panunto, Z. Urbáncyak-Liplowska, R. Johnson and M. C. Etter, *J. Am. Chem. Soc.*, 1987, **109**, 7786.
63 G. Eaton, A. S. Pena-Nuñez and M. C. R. Symons, *J. Chem. Soc., Faraday Trans. 1*, 1988, **84**, 2181.
64 C. Hansch, *Acc. Chem. Res.*, 1969, **2**, 232.
65 A. J. Leo, *J. Chem. Soc., Perkin Trans. 2*, 1983, 825.
66 J. Bradshaw and P. J. Taylor, *Quant. Struct.-Act. Relat.*, 1989, **8**, 279.
67 G. Némethy, *Angew. Chem., Int. Ed. Engl.*, 1967, **6**, 195.

Paper 1/04347A
Received 20*th August* 1991
Accepted 6*th January* 1992