Model Solvent Systems for QSAR.† Part 2.‡ Fragment Values ('*f*-Values') for the 'Critical Quartet'

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

A data matrix has been prepared of log *P* values for 103 compounds distributed across four highly contrasted solvent–water partitioning systems: the 'critical quartet' of octanol (amphiprotic), alkane (inert), chloroform (proton donor) and propylene glycol dipelargonate (PGDP; proton acceptor). Here 'alkane' is defined as the straight-chain sequence from hexane to octane and (possibly) higher; it is shown that cyclohexane is out of line. In principle, these log *P* values can now be used to construct a comparative table of fragment values (*f*-values) for all four systems. In practice, those for non-polar substituents must first be established. Here the key quantity is $f(CH_2)$. This has been re-determined, and in the process its variability rationalised, for 24 water-saturated solvent systems; here the key factors (dry solvents are different) turn out to be the molarity, in the organic phase, of water and the solvent's own functional group. There results an almost complete data matrix of 82 *f*-values for all four solvents, about 25% of which are derived from the linear solvation energy relationship (LSER) equations of Part 3.[7] It is shown that these four sets are very distinct, a fact that misleading statistical treatments can easily disguise. How the medicinal chemist might use these contrasting data sets is critically discussed, with particular reference to the rationalisation of biological selectivity.

Cell membranes protect the cell from the intrusion of unwanted substances, drugs included. Hence an important aspect of drug design is to build into the favoured biomolecule such physical characteristics as are expected to aid its absorption into the cell. Since the pioneering work of Hansch and his co-workers[1,2] it has been widely assumed that octanol–water partitioning forms a reasonably general model for this process.

Recently, we have challenged this view.[3] We have pointed to the inherent improbability that all cell membranes should possess the same physical characteristics. We have postulated[3,4] that, for animals including man, at least two limiting types need to be considered: amphiprotic, as typified by the polypeptide backbone; and proton acceptor, as typified by phospholipids. There is even evidence[5] that the second type may dominate, and for this reason we have developed[3] propylene glycol dipelargonate (PGDP) as a complementary—not rival—partitioning solvent to octanol which the medicinal chemist may find useful for exploring that possibility.

Nevertheless, a proliferation of solvent models is not in our view either necessary or desirable. We argue[3,4] that four limiting types should be sufficient. One is amphiprotic; here octanol picks itself. One is inert; here, in principle, any alkane will do (but see later). Chloroform is the classical proton donor solvent[6] and, while the LSER analysis to be reported in Part 3[7] suggests that some of its partitioning characteristics may be atypical, a large body of useful information exists and a plausible rival has yet to present itself. Finally, our development of PGDP as the putative standard proton acceptor[3] plugs this gap in a much more satisfactory way than heretofore. While no actual membrane is likely to match exactly any of these models, any real membrane should lie within the quadrilateral they define.

To the practising medicinal chemist, a random collection of log *P* values for a given model solvent is at best irritating and may well prove entirely unhelpful. What is required is a set of fragment values (*f*-values) and factors (*F*)[8] from which log *P* may be calculated with reasonable accuracy. Such estimations play an important part in that prediction of biological activity ahead of synthesis which is one chief aim of drug design.[1] So far, only the octanol system has been explored to an adequate extent,[8–10] although we have covered 80 fragment values in our study of PGDP[3] while Rekker[10,11] has more tentative information on a variety of solvent systems. The aim of this paper is to present fragment values for as many functional groups as possible across this 'critical quartet' of model solvent systems. Our methodology is general and could equally be applied elsewhere, should others choose to do so.

There are certain inherent limitations in this approach. In principle, if the complete set of non-polar *f*-values is known for a given solvent system, any polar *f*-value may be obtained *via* eqn. (1),[10] where *Z* might represent the latter

$$\log P_{XYZ} = f_X + f_Y + f_Z + \Sigma F \qquad (1)$$

and *X* and *Y* a summation of the former, and provided that correction factors (ΣF) may be discounted. In practice, things are not so simple. For octanol, Leo[8,9] has shown that bond (F_b) and branching (F_{cbr} or F_{gbr}) factors may be present even in the absence of perturbing influences of other sorts. We do not possess the information to estimate these factors elsewhere. The chief resulting problem, as we have discussed for PGDP,[3] is that the aryl–alkyl interface becomes difficult to handle: polar *f*-values derived from compounds that do, or do not, contain aryl moieties tend to differ. Typically the latter, which tend (as here) to be in the minority, are too positive by $\Delta\log P = 0.1$–0.3, fortunately a quite small margin in most cross-comparisons (*cf.* Table 4). For consistency, therefore, we here treat octanol–water in the same way as the rest, so that its *f*-values as quoted in Table 4 are for use in this context only and in no way rival, let alone supersede, those of the Pomona school.[9]

A more fundamental approach to solvent–water partitioning is *via* the LSER methodology of Taft and Kamlet *et al.*,[12,13] since this should reveal the underlying chemical factors[14] whose

† Acronyms used in this paper: LSER = linear solvation energy relationship; PGDP = propylene glycol dipelargonate; QSAR = quantitative structure–activity relationship; sd = standard deviation; TMP = 2,2,4-trimethylpentane.
‡ Part 1, ref. 3.

changing balance across a range of solvent systems must cause these *f*-values to vary as they do. This is our subject in Part 3.[7] It is mentioned here for its ability to pinpoint probable outliers among the log *P* values themselves, since we have made some retrospective use of this analysis in deciding which shall be considered authentic and which doubtful. Individual cases are discussed later; meanwhile, this is one reason why the *f*-values of Table 4 differ in some details from those we have previously published,[3,4,15] and which this listing supersedes. Nevertheless, since most polar *f*-values vary so markedly across these four solvent systems, it has been considered worth retaining as approximate several whose origin lies in log *P* values that may be quite appreciably in error (*vide infra*).

Despite the superior rigour of the LSER approach it must be pointed out, as Taft and Kamlet *et al.* indeed do,[12] that it does not possess the precision for calculating individual log *P* values to the accuracy that medicinal chemistry requires. Essentially LSER characterises the *system* rather than the solute.[13] Hence a fragment listing is a practical necessity and the present exercise is not redundant.

Results

Compound Inventory.—Our intention behind this selection of compounds was twofold: to produce as complete a data matrix as possible, ideally examining every compound in every solvent; and to encompass the widest possible selection of functional groups. The first aim has not been fulfilled: only 36 data points are held in common and for chloroform, the worst case, only 33 of a possible 92 compounds appear in the regression analyses of Part 3[7] (this becomes 38 out of 102 in terms of log *P* data deemed authentic). However, if we may define a 'distinct functional group' as one expected to possess its own *f*-value,[8–10] then these 92 compounds incorporate 64 different hydrogen bonding functionalities, a total which rises to 71 out of 102 for the whole data set. This total is incomparably greater than for any previous LSER analysis, and gives us considerable confidence in the validity of our results.[7] Even for chloroform, 26 such functionalities appear among the 33 data points analysed, or 31 out of 38 *in toto*. Table 1 sets out the log *P* values, while Table 2 analyses them by category in the above terms.

The great majority of compounds in Table 1 comprise those we have previously employed to derive *f*-values for PGDP,[3] with additions intended to strengthen the representation of functional groups for the other solvent systems. A few of these PGDP values have been slightly revised (see Table 1). Except for our own determinations or re-determinations, which are noted, the octanol list consists of Leo's favoured 'LOGPSTAR'[9] values. Apart from our own new data, which again are noted, we have used the determinations or recommendations of Fujita *et al.*[6] wherever possible for chloroform, since we regard these as much the most painstaking set available in the literature (see later).

The published alkane data presented us with a quandary. If the aim is to cover the widest possible range of functional groups, then no single alkane solvent provides remotely enough data, and some type of blending or interconversion process becomes essential. Seiler[16] has published a number of regression equations for the interconversion of log *P* values between cyclohexane and other hydrocarbon solvent systems, from which it would appear that all the latter (hexane, heptane, octane, dodecane, hexadecane) differ slightly but significantly among one another. However, Seiler's database, while large, is also polyglot, and our own examination of the data[9] for compounds contained within the present series suggests that no more than experimental scatter is involved. We believe that hexane, heptane, all forms of octane, and (probably) decane and hexadecane, give equal log *P* values to within the limits of error. The 'alkane' values of Table 1 have hence been obtained by treating all these solvent systems as equivalent. Most are based in fact on hexane, heptane, or 2,2,4-trimethylpentane ('isooctane': TMP); our own data (see footnotes) are based entirely on the latter. Most are blends in which we have taken the mean value, either of all published determinations, or of those which form a tight cluster around a central value in cases where there are obvious outliers in one direction or both. The standard deviation (sd) is indicated in all such cases. Mostly it makes little difference to the value, though some to its precision, whatever reasonable procedure is followed, as in the case of phenol, for which our chosen value of -0.87 ± 0.05 for $n = 9$ becomes -0.89 ± 0.15 when all 15 published values are employed.

It is equally clear, however, that cyclohexane (and, less certainly, decalin*) differs from the remainder. Comparative data for cyclohexane, where available, are included in Table 1. It will be seen that these are consistently greater, at least for log $P > 0$; we obtain the very tight relationship of eqn. (2). We

$$\log P \text{'alkane'} = 0.948(21)\log PC_6H_{12} - 0.102(34) \quad (2)$$
$$(n = 18\ r^2 = 0.992\ s = 0.12\ F = 1991)$$

consider this relation good enough for predictive purposes where 'alkane' data are not available, and have in fact added three useful values to Table 1 in this way (see footnotes). This relation is close to Seiler's[16] for hexane but not elsewhere. These greater log *P* values for cyclohexane, as also its perceptibly higher value for $f(CH_2)$, probably originate from its cyclic structure; see further discussion below.

Outliers.—Ten compounds were wholly excluded from the LSER analysis for reasons which became apparent as it proceeded and will be discussed in Part 3.[7] The same reasons probably exclude **94** in 'alkane' and **94** and **95** in chloroform, while another special reason, to be discussed later,[7] invalidates the LSER analysis of **100** in chloroform. None of these reasons, however, calls the actual log *P* value into question, and all in principle could therefore be used for deriving fragment values. There remain six genuine outliers to be considered. In our opinion, outliers are only permissible in the statistical analysis of what should be good chemical data if some good reason for questioning the experimental value can be put forward. (Here biological data differ, since one cannot even pretend that one understands all the factors). No outliers have been permitted for octanol or PGDP. For alkane, we exclude log *P* for isopropyl benzoate (**89**), since it is out of line with the methyl (**87**) and ethyl (**88**) esters so a chain-branching effect[8] may operate. The position for chloroform, which has five outliers, is the least satisfactory. However, special experimental problems attach to chloroform. All volatile solvents are difficult to use at the high volume ratios required for high log *P* values by the shake-flask methodology;[4] the risk is then that solvent evaporation will drive solute into the aqueous phase and result in a spuriously low log *P*. The intense UV opacity of chloroform makes the measurement of solute concentration in the organic phase commonly impossible, so that the usual mass balance checks become a problem. Difficulties begin at log $P \approx 2$, and in our experience, values $\geqslant 3$ can only be obtained even approximately by taking the most stringent experimental precautions. In the present set of compounds an extreme example is presented by the sulfide (**92**), log *P* measured as 2.96 but predicted[7] as 5.13, an entirely inaccessible value. If the measured value were correct, then aliphatic sulfide would have to be more

* Perhydronaphthalene.

Table 1 Log *P* values.[a,b]

	Compound	C_6H_{12}	'Alkane'[c]	Octanol	$CHCl_3$	PGDP[d]
1	PhH	2.47	2.24 ± 0.05	2.13	2.80	2.36
2	PhMe		2.89 ± 0.11	2.73	3.41	2.89
3	PhEt		3.30 ± 0.04	3.15	3.68	3.37
4	$PhCH{=}CH_2$			2.95		3.03
5	$PhCH_2CH{=}CH_2$			3.23		3.65
6	$PhCF_3$			3.01		3.26
7	PhF		2.46 ± 0.01	2.27	2.85	2.50
8	PhCl	3.13	2.93 ± 0.01	2.84	3.46	3.08
9	PhBr		3.10 ± 0.02	2.99	3.61	3.27
10	PhI		3.33	3.25		3.48
11	PhCN	1.11 ± 0.05	0.96 ± 0.04	1.56	2.71	1.66
12	$PhNO_2$	1.69	1.44 ± 0.03	1.85	2.93	2.16
13	$PhNH_2$	0.01 ± 0.05	−0.04 ± 0.06	0.90	1.42	0.95
14	PhNHMe	1.18 ± 0.04	1.04 ± 0.01	1.66	2.40[e]	1.87
15	$PhNMe_2$	2.47	2.28 ± 0.07	2.31	3.54[e]	2.52
16	PhOH	−0.80 ± 0.11	−0.87 ± 0.05	1.46	0.36	1.17
17	PhOMe	2.19 ± 0.11	2.06 ± 0.04	2.11	3.12	2.41
18	PhOCOMe		1.13	1.49		1.57
19	PhCHO	1.24 ± 0.11	1.07 ± 0.03	1.48		1.57
20	PhCOMe	1.27 ± 0.02	1.10 ± 0.05	1.58	2.79	1.63
21	PhCOPh	3.29	3.02[f]	3.18		3.40
22	$PhCO_2H$	−0.85 ± 0.11	−0.84 ± 0.44	1.87	0.46	1.15
23	$PhCONH_2$		−2.30	0.64	0.11	−0.36
24	$PhCSNH_2$			1.49		1.26
25	$PhCONHNH_2$			0.19		−0.83
26	PhCONHOH			0.26		−0.98
27	PhCONHMe		−1.76	0.86	1.00	−0.05
28	PhCONHEt		−1.10		1.54	
29	PhNHCOMe	−1.51	−1.70	1.16	0.85	0.40
30	PhNHCSMe			1.71		1.44[g]
31	PhCONHPh			2.62		2.38
32	$PhCONMe_2$			0.62	2.01	0.00
33	PhN(Me)COMe		−0.40	1.12		0.40
34	$PhNHCONH_2$			0.75[e]		−0.55
35	$PhNHCSNH_2$			0.70[e]		0.06
36	$PhN(Me)CONH_2$			0.42		−0.78[g]
37	PhNHCONHMe			1.12		0.00[g]
38	PhNHCSNHMe					0.58
39	$PhNHCONMe_2$			0.98	1.29	0.29
40	PhNHCONHPh			2.86		2.42
41	$PhNHCO_2Me$			1.76		1.84
42	$PhN{=}C(NH_2)_2$			0.53		−1.10
43	PhSOMe		*−1.49*	*0.55*	*1.41*[e]	*−0.41*
44	$PhSO_2Me$		−0.92	0.50	1.87[e]	0.47
45	$PhSO_2NH_2$			0.31	−0.24	−0.03
46	$PhSO_2NHMe$			0.92	1.31	0.64
47	$PhSO_2NMe_2$			1.35	2.69	1.48
48	$PhNHSO_2Me$			0.95		0.77
49	$PhNHSO_2NH_2$			0.40		−0.09
50	Ph_3PO	*0.68*	*0.18 ± 0.01*	*2.83*	*2.95*	*1.60*
51	Naphthalene (NpH)		3.39 ± 0.01	3.30		3.73
52	Np-2-$O(CH_2)_3SOMe$					*1.18*
53	Np-2-$O(CH_2)_3SO_2Me$					2.58
54	$PhCH_2OH$	−0.62	−0.67 ± 0.10	1.10		0.61
55	$PhCH_2OMe$			1.96[h]		1.99
56	$PhCH_2NH_2$		*−0.21*	*1.09*	*1.18*	*0.18*
57	$PhCH_2NHMe$			*1.52*		*0.90*
58	$PhCH_2COMe$		0.98	1.44		1.59
59	$PhCH_2CO_2Et$	2.40	2.17[f]			2.54
60	$PhCH_2CO_2H$	−1.23	−1.07	1.41	0.45	0.59
61	$PhCH_2CONH_2$			0.45		−0.61
62	$PhCH_2NHCONH_2$			0.73		−1.03
63	$PhCH_2NHCSNH_2$					0.27[g]
64	$PhCH_2NHCSNHMe$		−2.13	1.35	1.45	0.84
65	$PhCH_2OCONH_2$			1.20		0.44
66	$Ph(CH_2)_2CN$			1.72		1.90
67	$Ph(CH_2)_2OH$		−0.36 ± 0.04	1.36	1.31[e]	0.74
68	$Ph(CH_2)_2OMe$					2.22
69	$Ph(CH_2)_2NH_2$		*−0.56*	*1.41*	*1.36*	*0.50*[g]
70	$Ph(CH_2)_2NHMe$					*0.94*[g]
71	$Ph(CH_2)_2NHEt$					*1.44*
72	$Ph(CH_2)_2NMe_2$		1.39			1.78[g]
73	$Ph(CH_2)_2COMe$					2.06
74	$Ph(CH_2)_2OCOMe$			2.30		2.57
75	$Ph(CH_2)_2NHCOMe$					0.02

Table 1 (*continued*)

	Compound	C_6H_{12}	'Alkane'[c]	Octanol	$CHCl_3$	PGDP[d]
76	$Ph(CH_2)_2NHCSNH_2$					0.60
77	o-$ClPh(CH_2)_2CONEt_2$					2.47
78	p-$ClPh(CH_2)_2CONEt_2$					2.47
79	p-$ClPh(CH_2)_2NHCONHMe$					0.67
80	$Ph(CH_2)_2CO_2Me$	2.31	2.09[f]	2.32		
81	$Ph(CH_2)_3CONHSO_2Et$			1.67[e]		1.37
82	$Ph(CH_2)_3CN$			2.21		
83	$Ph(CH_2)_3OH$		0.10 ± 0.06	1.88		
84	$Ph(CH_2)_3OMe$			2.70		
85	$Ph(CH_2)_3NH_2$			*1.83*		
86	$Ph(CH_2)_3NMe_2$		2.03	2.73		
87	$PhCO_2Me$	2.08	1.80 ± 0.02	2.12	3.01[e]	
88	$PhCO_2Et$		2.38 ± 0.04	2.64	*2.89*	2.84
89	$PhCO_2Pr^i$		*2.57*	3.18[e]	*2.59*	3.30
90	$PhCO_2(CH_2)_4CN$		1.03	2.05[e]	*3.10*	2.44
91	$PhCO_2(CH_2)_4CONH_2$		−2.13	1.39[e]	1.72[e]	0.41
92	p-$NO_2PhO(CH_2)_3SMe$		2.26[e]	3.24[e]	*2.96*	3.70
93	p-$NO_2PhO(CH_2)_3SOMe$		*−2.53*	*0.93*[e]	*2.00*	*−0.13*
94	p-$NO_2PhO(CH_2)_3SO_2Me$		*−1.95*[e]	1.10[e]	*2.48*[e]	1.06
95	p-$NO_2PhO(CH_2)_3SO_2NH_2$		−3.11	0.97[e]	*0.75*[e]	0.66
96	$PhCH(Me)CH_2OH$					1.21
97	$PhC(CF_3)_2OH$			3.41		3.35
98	$PrNHC(=NCN)NHMe$			0.42[e]	−0.03[e]	−1.03
99	$C_6H_{13}NHCSNHMe$		−0.94	2.34[e]	*2.36*	1.41
100	$C_3F_7CH_2NHCSNHMe$		−1.17	2.39[e]	*1.17*[e]	1.91
101	EtOEt		0.62 ± 0.04[j]	0.89	1.88[e]	
102	CH_3CO_2Et	0.34	0.34 ± 0.05	0.73	1.84[e]	
103	p-$NO_2C_6H_4OMe$[k]		1.32	2.03	3.18	2.40

[a] Ref. 9 unless otherwise stated. [b] Italicised values are omitted from the regression equations of Part 3.[7] [c] For definition see text. [d] Ref. 3. [e] This work. [f] Scaled from log P (cyclohexane) using eqn. (2). [g] Revised value. [h] Dr. J. Bradshaw, personal communication. [j] Mean of two results for hexadecane. [k] Not part of LSER data set; needed for the calculation of f-values using compounds **92–95**.

Table 2 Distribution of compounds between categories[a]

	Functional group category				
Solvent	Neutral[b]	Acceptor[c]	Amphiprotic[c]	Total[d]	Total[e]
'Alkane'	8	16 + 5	16 + 1	32	46
Octanol	11	21 + 7	37 + 2	58	78
Chloroform	9	11 + 0	15 + 1	26	33
PGDP	11	22 + 6	42 + 2	64	83
Maximum[f]	11	22	42	64	92

[a] Not including 10 compounds excluded from statistical analysis (see the text). [b] Containing no proton donor or acceptor functionality. [c] First figure is of distinct functional groups, second is of duplicates. [d] Of distinct functionalities. [e] Included in regression analyses. [f] Except in last column, maximum score without duplication.

hydrophilic, for chloroform alone, than aromatic sulfoxide, which is inconceivable; elsewhere, it is about on a level with aromatic ether (see Table 4). Similarly, compound **90** is reported as log P = 3.10 where 4.17 is predicted.[7] In one case, we have more definite evidence. For methyl (**87**), ethyl (**88**) and isopropyl (**89**) benzoates, log P is reported as 2.17, 2.89 and 2.59, respectively; for the first we now find log P = 3.01, a much more reasonable value that automatically invalidates the other two. The only result we cannot reasonably explain is that for the thiourea **99**, which breaks the pattern in being, by Δlog P = 0.50, more hydrophobic than predicted. One can only say that this is one of several compounds which, for no obvious reason, have given inconsistent results across the solvent set in our LSER analysis.[7]

Fragment Values.—General. The use of eqn. (1) for determining polar fragment values assumes that those for the non-polar fragments are already known. Since most organic molecules of any reasonable size are dominated by their carbon skeleton, these latter moieties will tend numerically to preponderate (*cf.* Table 1), and any error in their values will be cumulative. Hence alkyl and simple aryl (phenyl) fragment values must be established first.

Methylene.[17] Methylene is uniquely important as the repeating unit of homologous series. Hence $f(CH_2)$ is the most fundamental of all fragment values and has to be determined with great care, since every other fragment value depends on it and an accurate additive log P system is impossible otherwise.

Two important compilations of $f(CH_2)$ exist, due to Davis *et al.*[18] and to Rekker[10] (Table 3). Unfortunately, each was compiled before the existence of bond and branching factors was known,[8] and at least the latter is not entirely based on homologues, as is essential if these perturbing factors, F_b and F_{cbr},[8] are to be avoided. Hence we have undertaken a fundamental re-examination of this question, incorporating more recent data[9] and basing our results on homologues alone. Our results are contained in the column entitled 'Re-run.' We regard these twelve results as accurately validated. In the light of these, 12 more in Davis's compilation appear reasonable (footnote *a* indicates two that are not). Table 3 also contains, re-expressed as $f(CH_2)$, recent results by Abraham *et al.*[19,20] on the gas → solvent free energy of transfer for methylene. These are for the pure liquid and hence are obtainable for water-miscible solvents (see footnote *e*); we discuss them later.

$$f(CH_2) = 0.568(7) - 0.136(8)\log[H_2O] \quad (3)$$
$$(n = 9\ r^2 = 0.974\ s = 0.01\ F = 265)$$

For the nine alcohols of Table 3, Davis *et al.*[18] found the relation between $f(CH_2)$ and water content that is expressed by eqn. (3). In fact, our revised f-values turn this line into a curve [Fig. 1(*a*)]. It will be seen that a further, steeper curve exists for carbonyl-containing solvents. There is no reason, however, why

Table 3 Methylene fragment values and associated parameters[a]

Solvent	$f(CH_2)$ according to					$\delta_H^2/100$[f]	$(n^2 - 1/n^2 + 2)$[g]	$[H_2O]$[h]	$[S]$[j]
	Davis[b]	Rekker[c]	Re-run[d]	Consensus	Abraham[e]				
Cyclohexane	0.64	0.646 ± 0.008	0.639 ± 0.084	0.64	0.69	0.672	0.257	0.002 45[k]	
Hexane	0.62			0.62	0.67	0.528	0.229	0.004 06[l]	
Heptane	0.62			0.62		0.553	0.236	0.003 46[l]	
Octane	0.62			0.62		0.570	0.240	0.003 69[l]	
Decane					0.67	0.598	0.248	0.002 9[l]	
Hexadecane					0.67	0.641	0.261		
Benzene	0.62	0.619 ± 0.010	0.615 ± 0.073	0.62	0.67	0.838	0.295	0.035[l]	
Toluene	0.60	0.56		0.60		0.794	0.292	0.015[l]	
Chlorobenzene					0.67	0.935	0.306	0.020[l]	
Tetrachloromethane	0.62	0.66	0.650 ± 0.090	0.65	0.65	0.738	0.274	0.008 8[l]	
Chloroform	0.62	0.628 ± 0.007	0.621 ± 0.093	0.62		0.887	0.267	0.067[l]	
Dichloromethane	0.60			0.60		0.977	0.254	0.145[l]	
Oleyl alcohol	0.58	0.535 ± 0.010	0.566 ± 0.021	0.57		0.659	0.274	0.712[k]	3.12[k]
Octan-1-ol	0.52	0.527 ± 0.006	0.533 ± 0.074	0.53	0.63	1.033	0.258	1.72[k]	6.14[k]
4-Methylpentan-2-ol	0.51			0.51		1.000	0.247	2.84[b]	7.38
Pentan-1-ol	0.51	0.51	0.518 ± 0.041	0.51		1.198	0.248	3.36[b]	8.54
Pentan-2-ol	0.47			0.47		1.158	0.245	6.6[b]	8.06
Isobutanol	0.44			0.44		1.243	0.240	7.49[b]	8.96
Butan-1-ol	0.44	0.404 ± 0.012	0.450 ± 0.033	0.44	0.62	1.295	0.242	9.53[k]	9.08[k]
2-Methylbutan-2-ol	0.43			0.43		1.166	0.244	10.51[b]	7.0
Butan-2-ol	0.38	0.32	0.328 ± 0.013	0.33		1.222	0.241	19.67[b]	6.07
Diethyl ether	0.56	0.531 ± 0.007	0.559 ± 0.112	0.56		0.562	0.217	0.578[l]	9.43
Nitrobenzene	0.48			0.48	0.60	1.222	0.319	0.16[l]	9.76
Olive Oil	0.53	0.589 ± 0.010	0.587 ± 0.124	0.59			0.278	0.072 5[k]	1.04
PGDP			0.506 ± 0.033	0.51				0.665[m]	5.07
Ethyl acetate	0.45			0.45	0.63	0.612	0.240	1.60[k]	9.88
Butan-2-one	0.33			0.33	0.58	0.860	0.231	4.40[b]	10.0

[a] A very few values from ref. 18 have been rejected as inherently improbable. These include, *e.g.*, 0.40 for isopentyl acetate (7 carbon atoms) allegedly below 0.45 for ethyl acetate (4 carbon atoms), and 0.33 for pentan-3-one (5 carbon atoms) identically with butan-2-one (4 carbon atoms) [b] Ref. 18. [c] Ref. 10. [d] This work, based on log *P* for strict homologues only. [e] Ref. 19: from gas–liquid partioning to pure solvents. Results for water-miscible solvents: acetone, 0.59; ethanol, 0.58; dimethylformamide, 0.57; *N*-methylpyrrolidinone, 0.54; methanol, 0.54; trimethylene carbonate, 0.48; dimethyl sulfoxide, 0.48; ethylene glycol, 0.39. [f] δ is the Hildebrand solvent parameter; data from ref. 19. [g] n is refractive index at 25 °C; data from ref. 19, in which this function is given the mnemonic $f(n^2)$. [h] Equilibrium concentration of water in organic solvent. [j] Equilibrium concentration of solvent functional group in water-saturated solvent. [k] R. N. Smith, C. Hansch and M. M. Ames, *J. Pharm. Sci.*, 1975, **64**, 599. [l] J. A. Riddick and W. B. Bunger, *Organic Solvents* (*Techniques of Chemistry*, vol. 2), ed. A. Weissberger, Academic Press, New York, 3rd. edn., 1970. [m] Ref. 3.

a logarithmic relation in $[H_2O]$ is required; many free-energy-related quantities show linear relations with mole fraction x_i rather than its logarithm. An example is the linear relation with chemical reaction rate in water as log k that is commonly shown by the concentration of added salts or co-solvents.[21] The result of re-expressing $f(CH_2)$ in these terms is shown in Fig. 1(*b*) and, for the alcohols, as eqn. (4). This equation, which as it stands

$$f(CH_2) = 0.552(8) - 0.0118(9)\,[H_2O] \quad (4)$$
$$(n = 9\ r^2 = 0.963\ s = 0.014\ F = 184)$$

$$f(CH_2) = 0.593(12) - 0.0114(5)\,[H_2O] - 0.0061(16)\,[S] \quad (5)$$
$$(n = 9\ r^2 = 0.989\ s = 0.0085\ F = 273)$$

leaves oleyl alcohol* appreciably off-line, is notably improved (note the *F*-statistic) by a second term $[S]$ for the functional group (as proton acceptor) concentration of the water-saturated solvent in itself (Table 3). The significance of eqn. (5) appears to be that all polar functionalities contribute to the task of making the organic phase more water-like and, therefore, less hospitable to non-polar moieties such as methylene. In fact, for $[S] = 0$, eqn. (5) extrapolates to a value of $[H_2O] = 52$ mol dm^{-3} at $f(CH_2) = 0$, whereas water is 55 mol dm^{-3}. In view of the necessity that all fragment values become zero for partitioning between water and itself, this is a very satisfying result. It contrasts with eqn. (3), for which $f(CH_2)$ becomes zero at the improbable concentration of $[H_2O] = 1.5 \times 10^4$ mol dm^{-3}.

The four carbonyl acceptors of Fig. 1(*b*) continue to lie on a curve which can be linearised, unconvincingly, as eqn. (6). This again is much improved [eqn. (7)] by a term in $[S]$. The use of

$$f(CH_2) = 0.563(22) - 0.055(9)\,[H_2O] \quad (6)$$
$$(n = 4\ r^2 = 0.945\ s = 0.03\ F = 34)$$

$$f(CH_2) = 0.595(17) - 0.039(8)\,[H_2O] - 0.009(4)\,[S] \quad (7)$$
$$(n = 4\ r^2 = 0.992\ s = 0.017\ F = 65)$$

two parameters to describe four points is scarcely robust statistics, but nevertheless, the higher coefficients of $[H_2O]$ and $[S]$ than for eqn. (5) do correctly reproduce what is visible to the eye, namely that ester and ketonic solvents are much more sensitive to both than are alcohols. This accounts for the initially surprising fact that the less polar solvent PGDP should possess a lower value of $f(CH_2)$ than the more polar solvent octanol.[3] Its explanation most probably lies in the much greater disruption of solvent structure expected for a pure proton acceptor when water enters it than is likely for an alcohol, whose amphiprotic structure should be much less affected. Possibly the proton acceptor is actually cross-linked by added water. It is particularly interesting, therefore, that neither diethyl ether nor nitrobenzene should fit eqn. (7). The former is much less sensitive, lying close to the alcohol line, while the second is much more so [Fig. 1(*a*)]. In view of the strong evidence,[22] which we have confirmed,[23] that ethers (and presumably alcohols) possess for bonding purposes in solution only one effective lone

* $CH_3(CH_2)_7CH{=}CH(CH_2)_8OH$.

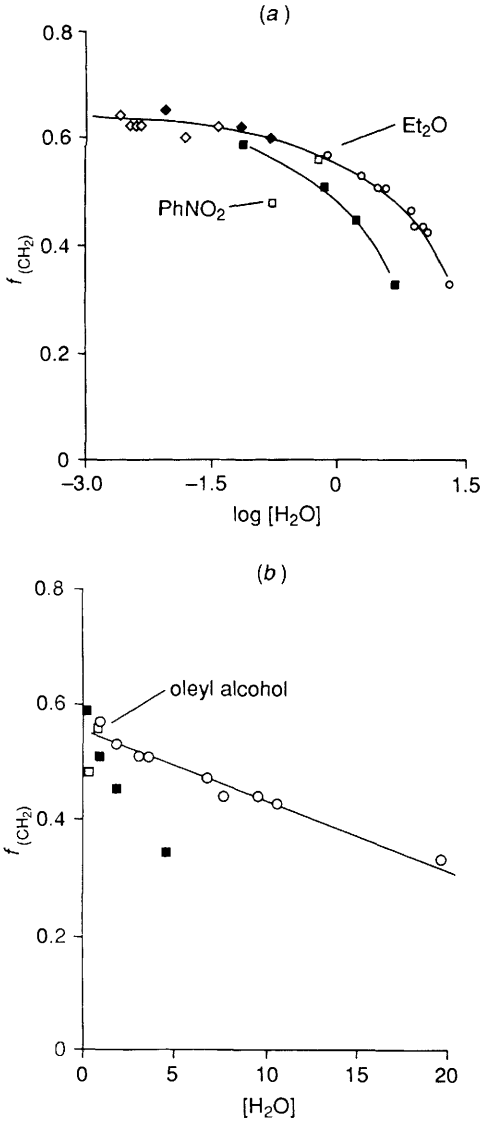

Fig. 1 $f(CH_2)$ as a function: (*a*) of log $[H_2O]$; (*b*) of $[H_2O]$ [Hydrocarbons and chlorocarbons have been omitted from (*b*) to avoid crowding the origin (© Elsevier 1991, *i.e.* ref. 17, and reproduced with permission)]: ◇, hydrocarbons; ◆, chlorocarbons; ○, alcohols; ■, carbonyl acceptors; □, other acceptors

pair, these sensitivities may be a simple function of the number of available lone pairs; the nitro-group nominally has four, though it is likely that not all are available for bonding.[24]

It is also interesting that eqns. (5) and (7) possess the same intercept to within the limits of error. This may be interpreted as $f(CH_2)$ for some notional solvent of high polarity but possessing neither water content nor polar functionality (of those in Table 3, CH_2Cl_2 fits the bill quite well). We may then interpret the effect of water content in the organic phase as resulting from an increase in solvent drag; this, to the extent it is present, offsets that of the aqueous phase and so reduces the difference in entropy of solvation between the phases.

For dry solvents, Abraham *et al.*[20] have demonstrated that the major factor in $f(CH_2)$ is cohesive energy density as represented by Hildebrand's δ_H^2; if alcohols are excluded, the equation can be further improved by incorporation of $f(n^2)$, a refractive index term. (Abraham's *f*-values, and these terms, are listed in Table 3). These terms are reasonably interpreted[20] as indices of the energy required for cavity formation in the solvent (endergonic) more than offset by that due to dispersion interactions with the solute (exergonic); possibly the failure of alcohols to fit well is connected with their special status as associated solvents. The difference between Abraham's and our 'consensus' values for $f(CH_2)$ in highly non-polar solvents is a little surprising in view of their very low water content, but leaving this aside, it is clear from comparison of the two lists that cyclohexane stands apart from the other alkanes (hexane to hexadecane), which are otherwise essentially equivalent. We have noted above the exceptional behaviour of cyclohexane with respect to log *P*, with which this enhancement in $f(CH_2)$, from 0.62 to 0.64, is consistent. Cyclohexane is also out of line in that Abraham's equation[20] would predict $f(CH_2)$ as slightly lower than for the (dry) linear alkanes, not higher as is found (0.66 *vs.* 0.69); the difference lies in a too-high δ_H^2 term. Leahy,[25] in his comparison of molar volume $\bar{V}$ with intrinsic volume V_I for various classes of hydrocarbon, noted that cycloalkanes are consistently denser than expected from the relation for linear alkanes. Possibly all these effects share a common origin. As rigid structures, cycloalkanes should pack better, hence higher δ_H and density, but at the same time may inflict less solvent drag on solutes, which therefore benefit from a less negative than usual ΔS of solvation that δ_H^2, an enthalpic term (it relates to ΔU),[26] does not reveal.

Some other points deserve brief mention. The order in $f(CH_2)$ of $CCl_4 > CHCl_3 > CH_2Cl_2$ is that predicted by Abraham,[20] but values are a little high, especially for CCl_4, so may indicate some slight favourable interaction between the C–H and C–Cl dipoles. It is interesting that $f(CH_2)$ for CCl_4 possesses the highest known value for a 'wet' solvent. In this regard, we may speculate concerning the perfluorocarbons, about which nothing is known. These are notable for their very low cohesive energy density, presumably the result of intermolecular dipole–dipole repulsion; Abraham's equation in δ_H^2 alone[20] predicts a limiting value of 0.706 for $f(CH_2)$ at $\delta_H^2 = 0$, which these solvents may then possibly approach.

For all 'wet' solvents taken together, an equation in $[H_2O]$ and $[S]$ fits $f(CH_2)$ better than one in δ_H^2 and $f(n^2)$, though neither is good.[27] We do not interpret this, however, as indicating any superiority in the present approach. Rather, we suggest that the former, molecular variables represent major contributing factors that help to determine the magnitude of the latter. Hence there is really no contradiction.

Other Non-polar Moieties. Since the only molecular property relevant to the partitioning of alkanes consists in their volume, surface area, or both, it seems probable that $f(CH_3)$, $f(CH_2)$ and $f(CH)$ should remain in constant ratio for all solvent–water partitioning systems. Since this ratio is accurately known for octanol,[8,9] all that is required for any other solvent system is that $f(CH_2)$ be accurately determined. We have used this principle already for PGDP;[3] the start of Table 4 incorporates the values we believe to apply to chloroform and 'alkane'. The latter are substantially identical with the values of 0.833, 0.615 and 0.397 once suggested by Rekker[11] for $f(CH_3)$, $f(CH_2)$ and $f(CH)$, respectively. These values are rounded to the nearest 0.01 since, at the present time, we do not believe greater precision for these or any fragment values to be justified, except inside Leo's CLOGP.[9] It is particularly pleasing that $f(C)$ remains substantially a constant (0.19–0.20) for all these solvent systems, since alkyl carbon is at all times shielded from the solvent.

The alternative, of course, would be to calculate the ratio of $f(CH_3)$ to $f(CH_2)$ to $f(CH)$ directly from volume or surface area. As we have detailed elsewhere,[4] we believe both quantities to be hung around with complications and ambiguities that render the necessary precision impossible. Even in the present study and using V_I, a quantity calculated with some care,[25] we find the apparent volume increment for methylene to be surprisingly inconstant.[7] Far better to trust an established[8,9] empirical ratio.

Table 4 Fragment values for the 'critical quartet' of solvent systems[a,b]

Fragment	'Alkane'	Octanol	Chloroform	PGDP
Context-independent values				
CH_3	0.83	0.70	0.83	0.67
CH_2	0.62	0.53	0.62	0.51
CH	0.41	0.36	0.41	0.35
H	0.21	0.17	0.21	0.16
C	0.20	0.19	0.20	0.19
Ph	2.03	1.96	2.59	2.20
Context-dependent values[c]				
$ArCH{=}CH_2$	*0.97*	0.99	*0.88*	0.83
$ArCH_2CH{=}CH_2$	*0.84*	0.74	*0.76*	0.94
$ArCF_3$	*1.12*	1.05	*1.23*	1.06
ArF	0.43	0.31	0.26	0.30
ArCl	0.90	0.88	0.87	0.88
ArBr	1.07	1.03	1.02	1.07
ArI	1.30	1.29	*1.24*	1.28
$ArNO_2$	−0.59	−0.11	0.34	−0.04
ArCN	−1.07	−0.40	0.12	−0.54
AlkCN	*−2.28*	−1.32	*−0.68*	−1.32
ArOH	−2.90	−0.50	−2.23	−1.03
$ArCH_2OH$	−3.32	−1.39	*−2.23*	−2.10
AlkOH	−3.73	−1.67	−2.52	−2.48
ArOAlk	−0.80	−0.55	−0.30	−0.46
$ArCH_2OAlk$	*−1.78*	−1.23	*−1.01*	−1.39
AlkOAlk	−2.28	−1.56	*−1.30*[d]	−1.67
AlkSAlk	−0.92	−0.38	*0.11*	−0.23
$ArNH_2$	−2.07	−1.06	−1.17	−1.25
$ArCH_2NH_2$	−2.86	−1.40	*e*	−2.53
$AlkNH_2$	−3.83	−1.67	−2.47	−2.72
ArNHAlk	−1.82	−1.00	−1.02	−1.00
$ArCH_2NHAlk$	*e*	−1.67	*e*	−2.48
AlkNHAlk	*e*	*−2.08*[f]	*e*	−2.96
ArN(Alk)Alk	−1.41	−1.05	−0.71	−1.02
AlkN(Alk)Alk	−3.53	−2.22	*−2.07*	−2.78
ArCHO	−0.96	−0.48	*0.07*	−0.63
ArC=OAr	−1.04	−0.74	*−0.57*	−1.00
ArC=OAlk	−1.76	−1.08	−0.63	−1.24
AlkC=OAlk[g]	−2.50	−1.75	*−1.27*	−1.81
$ArCO_2H$	−2.87	−0.09	−2.13	−1.05
$AlkCO_2H$[g]	−3.72	−1.08	−2.76	−2.12
ArCOOAlk	−1.08	−0.54	−0.41	−0.57
AlkCOOAr	−1.73	−1.17	*−0.82*	−1.30
AlkCOOAlk	−1.96	−1.41	*−1.01*[d]	−1.34
$ArCONH_2$	−4.33	−1.32	−2.48	−2.56
$AlkCONH_2$[g]	*−5.39*	−2.04	*−3.05*	−3.32
ArCONHAr	*−3.76*	−1.30	*−2.34*	−2.02
ArCONHAlk	−4.60	−1.80	−2.46	−2.92
AlkCONHAr	−4.56	−1.50	−2.57	−2.47
AlkCONHAlk	*−6.16*	*−3.10*[h]	*−3.53*	−3.87
ArCON(Alk)Alk	*−4.49*	−2.74	−2.24	−3.54
AlkCON(Alk)Ar	−4.09	−2.24	*−1.85*	−3.14
AlkCON(Alk)Alk	*−5.56*	*−3.69*[j]	*−2.92*	−3.83
ArCONHOH	*−4.88*	−1.70	*−2.53*	−3.18
$ArCONHNH_2$	*−4.31*[k]	−1.77	*−1.96*[k]	−3.03
$AlkCONHSO_2Alk$	*−7.47*[l]	−3.11	*−2.19*[l]	−3.54
$ArCH_2OCONH_2$	*−4.31*	−1.29	*−2.78*	−2.27
AlkOCONHAr	*−2.54*	−0.90	*−0.81*	−1.03
$ArNHCONH_2$	*−5.83*[k]	−1.21	*−3.86*[k]	−2.75
$ArCH_2NHCONH_2$	*−7.25*[k]	−1.76	*−4.97*[k]	−3.74
$ArN(Alk)CONH_2$	*−6.37*	−2.24	*−3.22*	−3.65
ArNHCONHAr	*−3.92*	−1.06	*−2.40*	−1.98
ArNHCONHAlk	*−5.36*	−1.54	*−2.83*	−2.87
AlkNHCONHAlk	*−6.95*	*−2.93*[m]	*−3.96*	−3.94
ArNHCON(Alk)Alk	*−5.16*	−2.38	−2.96	−3.25
$ArCSNH_2$	*−3.05*	−0.47	*−1.48*	−0.96
AlkCSNHAr	*−3.77*	−0.95	*−1.92*	−1.43
$ArNHCSNH_2$	*−5.34*[k]	−1.26	*−3.14*[k]	−2.14
$ArCH_2NHCSNH_2$	*−6.22*[k]	*−1.43*	*−3.64*[k]	−2.44
$AlkNHCSNH_2$	*−6.62*[k]	*−1.52*[n]	*−3.99*[k]	−2.62
ArNHCSNHAlk	*−4.84*	*−1.35*[o]	*−2.30*	−2.29
$ArCH_2NHCSNHAlk$	−5.61	*−1.43*	−2.59	−2.54
AlkNHCSNHAlk	−5.70[p]	−1.71[p]	−2.40[p]	−2.48[p]
$ArN{=}C(NH_2)_2$	*−6.09*[k]	−1.43	*−4.55*[k]	−3.30
AlkNHC(=NCN)NHAlk	*−7.44*[p]	−2.04[p]	−2.93[p]	−3.39[p]

Table 4 (*continued*)

Fragment	'Alkane'	Octanol	Chloroform	PGDP
ArS=OAlk	−4.35	−2.11	−2.01	−3.28
AlkS=OAlk	−5.71	−2.69	−3.04	−4.10
$ArSO_2Alk$	−3.78	−2.16	−1.55	−2.40
$AlkSO_2Alk$	−5.13	−2.52	−2.56	−2.80
$ArSO_2NH_2$	*−4.88*[q]	−1.65	−2.83	−2.23
$AlkSO_2NH_2$	−5.46	−1.95	−3.46	−2.60
$ArSO_2NHAlk$	*−4.25*	−1.74	−2.11	−2.23
$AlkSO_2NHAr$	*−4.16*	−1.71	*−2.10*	−2.10
$ArSO_2N(Alk)Alk$	*−3.68*	−2.01	−1.56	−2.06
$ArNHSO_2NH_2$	*−4.72*[k]	−1.56	*−2.93*[k]	−2.29
Ar(Ar)(Ar)P=O	−5.91	−3.05	−4.82	−5.00

[a] For derivation see text. [b] Italicised values are calculated from the final regression equations of Part 3 (following paper). [c] Attachment points indicated as follows: Alk = alkyl, Ar = aryl (phenyl). For substituents other than OR, SR and NR_2, benzyl and alkyl values are essentially identical (see text). Aryl and (probably) benzyl values are inapplicable to heterocycles. [d] Calculated value preferred to experimental (see the text). [e] Incapable of calculation. [f] Assumes no proton donor component (see the text); Leo[9] gives −2.15. [g] Derived in part or in whole from benzylic compounds (see note *c*). [h] Leo[9] gives −2.71. [j] Leo[9] gives −3.04. [k] May be too negative (see the text). [l] Of doubtful validity (see the text). [m] Leo[9] gives −2.18. [n] Leo[9] gives −1.29. [o] Leo[9] gives −1.79. [p] Doubtful: derived from wholly aliphatic molecules (see the text). [q] Valid if 'well-behaved': see the text.

The other essential non-polar quantity is f(Ph), readily obtained from the general formula $f(C_6H_{6-n}) = \log P(\text{benzene}) - nf(H)$. For 'alkane', this log P is a statistically derived quantity as for much else. For octanol we use Leo's LOGPSTAR[9] value for log P, for chloroform Fujita's,[6] and for PGDP our own.[3] Values for any other aromatic, or any other system, can of course be obtained in a similar manner.

One topic that lies beyond the scope of this paper concerns the possible inconstancy of f(H). Testa *et al.*[28] have suggested this on the reasonable grounds that hydrogen in H_2 or a benzene ring is more exposed than in an alkyl chain. Values of *ca.* 0.23 (Leo[8,9]) for the former and 0.17 (Rekker[10]) for the latter would go a long way to reconciling the problems that are associated with the aryl–alkyl interface and which Leo's bond and branching factors were intended to handle.[27] But we do not possess the knowledge to pursue that point here.

Polar moieties. With the non-polar values established these can now be calculated *via* eqn. (1), for compounds from which interaction terms (ΣF) can reasonably be assumed to be absent. Conditions for this include the presence, in benzene, only of alkyl and halogen if multi-substituted;[8,29] at least three alkyl carbons between polar groups; and the absence of chain branching. All are satisfied for all compounds in the present set which have actually been used for this purpose (a few are in for special reasons[7] and have not). These derived f-values are set out in Table 4. For 18 these are the means of two or more derivations; it is reassuring that the maximum sd in these cases is ±0.07, with a mean of 0.027 ± 0.018.

It is also possible, if somewhat dangerous, to derive f-values from log P data calculated by means of LSER. There are so many gaps for 'alkane' and chloroform in Table 4 that this has been considered a worthwhile aim. It would certainly be impermissible for poor LSER equations or ones based on a narrow range of functional groups, but with good equations[7] derived from four very diverse solvent systems, and an unprecedented 71 distinct hydrogen bonding functionalities, the attempt has been felt to be justified. For any f-value its precision must depend on the breadth of the predictive database, and there is some evidence in the footnotes to Table 4 that this can be poor, with Δf as great as 0.7, where prediction depends on one solvent alone. The reason for this probably lies in the failure, which we discuss elsewhere,[7] of the volume term in

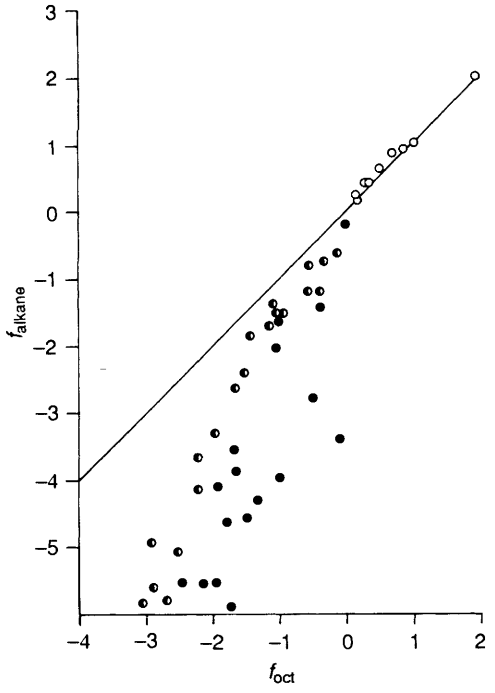

Fig. 2 Alkane fragment values *vs.* those for octanol: ●, amphiprotic groups; ◐, proton acceptor groups; ○, functional groups with no hydrogen bonding ability—the line is isolipophilic [© Pergamon Press 1990, *i.e.* ref. 4, and reproduced with permission]

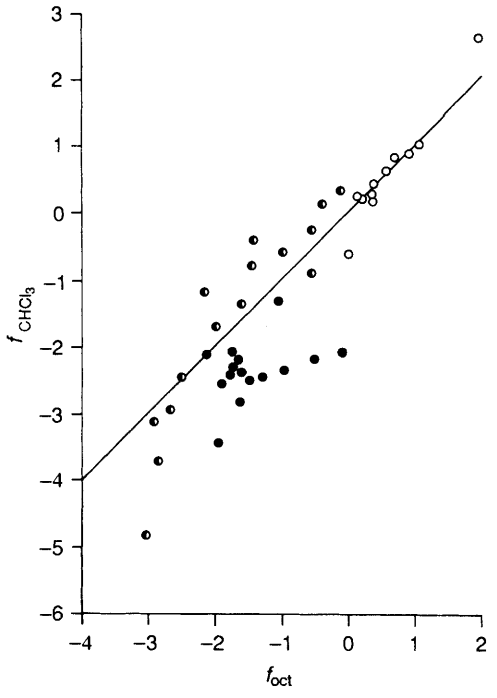

Fig. 3 Chloroform fragment values *vs.* those for octanol; key and © as for Fig. 2

the LSER equations to predict $f(CH_2)$ with sufficient accuracy. Prediction based on more than one solvent is likely to be more accurate since these errors are of both signs and tend to cancel. When however these *f*-values are used for generating part-structures—we wish to re-emphasise that they are not intended for extended log *P* calculation—then illuminating cross-comparisons can be drawn.

Discussion

Table 4 consists in a data matrix of 82 fragment values for four solvent–water systems that represent[3,4] a quadrilateral of properties outside which, we believe, no actual biological membrane is likely to lie. About one-quarter are calculated, but allowing for this, it is complete except for a handful of points that our LSER treatment cannot handle.[7] A few comments are in order. Attachment points are defined as alkyl, benzyl, or aryl; the first two give distinct values when attached to a hetero-atom[8,9] though, if the group is electronegative *e.g.* carbonyl, the difference is a matter of very fine tuning (typically $\Delta f < 0.1$) and we have chosen to ignore it. The calculated 'alkane' value for $ArSO_2NH_2$ assumes this substituent to be 'well-behaved'[7] in that context (we think it will be). We cannot calculate *f*-values for primary and secondary amines since their proton donor strength is problematical,[7] hence the gaps; the exception concerns octanol, but only because the contribution of this property to log *P* is very small, so that an estimate for Alk_2NH * may be hazarded. We also doubt our own predictions for primary ureas and the like in 'alkane' and chloroform (see footnote *k*). These turn out to be exceptional proton donors, probably through the parallel alignment of NH groups,[2,3] and in addition, NH_2 of any sort is anomalously hydrophilic in those two solvent systems,[7] but for these two effects to operate together may be piling Pelion on Ossa, and we have no actual

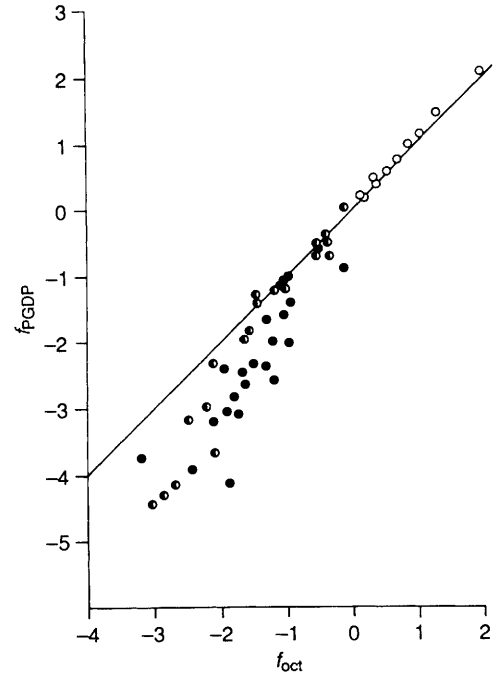

Fig. 4 PGDP fragment values *vs.* those for octanol; key and © as for Fig. 2

evidence that they do. The most generally doubtful category consists of those *f*-values that have had to be derived from purely aliphatic compounds. It was mentioned above that our unavoidable neglect of bond and branching factors[8,9] may put these out of line with the remainder. For PGDP we estimate[3] that *f*-values derived this way may be up to 0.3 log units too positive. For 'alkane' and chloroform we have no estimate, but for this reason have preferred calculated to observed *f*-values for two cases (see footnotes) in the interests of a level cross-comparison. We must emphasise, once again, that these four lists of *f*-values are derived on simplistic assumptions and are intended for cross-comparison alone; the task of constructing

* Secondary alkyl amine; see Table 4, footnote *c*

an analogue to CLOGP[9] for anything other than octanol lies in the future.

Nevertheless, these fragment values will serve their purpose if they enable the medicinal chemist to compare one solvent system with another, however tentatively. Most discussion is deferred to Part 3[7] when we shall use LSER to dissect the fine structure that makes these *f*-values as they are, but an impressionistic picture may be attempted here. On Figs. 2–4 we plot 'alkane,' chloroform and PGDP fragment values (experimental only!) against those for octanol; the line on each is isolipophilic, *i.e.* the deviation from this line for any group shows how much more lipophilic (positive) or hydrophilic (negative) it is with respect to the reference solvent. Unsurprisingly, most deviations are negative. 'Alkane' (Fig. 2) is the easiest to interpret. All polar groups lie below the isolipophilic line, and all amphiprotics are more negative than the pure proton acceptors, since any form of hydrogen bonding enhances hydrophilicity. PGDP (Fig. 4) is rather more complex, with less extreme deviations—it is much more polar—but considerable overlap between these classes, so that amphiprotics are sometimes more and sometimes less relatively hydrophilic than the nearest comparable proton acceptor. PGDP *f*-values tend to resemble those in octanol for weak acceptors and in chloroform for strong amphiprotics, while for strong acceptors they are more negative than in either. The most intriguing case is that of chloroform (Fig. 3). Fujita *et al.*[6] have suggested that proton acceptors should be better extracted from water by chloroform than by octanol for simple mass-action reasons; *i.e.*, the former solvent possesses twice the proton donor concentration. For a number of aromatic proton acceptors *vs.* non-hydrogen bonders, they demonstrate the expected[6] twofold rise in relative log *P*. As a general picture, this dissolves in the light of Fig. 3, which shows not only a lack of constancy, but a tendency for this effect to reverse for the stronger proton acceptors. There are subtle reasons behind this which we shall explore.[7] It is clear that, to put it mildly, each set needs to be considered on its merits, and that nothing remotely resembling a Collander relation[30] holds.

This leads to a final point of some importance to medicinal chemistry. If Figs. 2–4 are treated as plots of simple linear regressions, then fits of 87%, 81% and 89%, respectively are obtained, despite the obviously disparate nature of the data concerned. Fits of this order are frequently used to claim equivalence between octanol–water log *P* values and, *e.g.*, some chromatographic capacity factor or theoretical construct.[4] Such spurious demonstrations are even easier, of course, for data sets—most, indeed—less heterogeneous than the present one. Here they are given away by their slopes, which at *ca.* 1.7, 1.2 and 1.3 for the three cases above are far from isolipophilic. Hence any attempt at log *P* prediction through such misleading statistics risks being disastrously wrong, to an extent that increases with the distance of the key substituent from the isolipophilic line and which may totally vitiate the biological correlation for which it is intended.

On the positive side, the characterisation of biological membranes by the use of log *P* for discriminating solutes is an important future task for medicinal chemistry. Solutes may penetrate membranes preferentially in a way that is independent of lipophilicity as such. As we have previously argued,[3,4] there is a potential pay-off here for drug design. If two membranes contain differing receptors such that one is responsible for wanted and the other for some form of unwanted biological activity, the potential exists for biological selectivity based on different partitioning behaviour.[3,4,17] Two reports of selectivity based on differential log *P* have already appeared.[17,31] We hope that the fragment values of Table 4 will help the medicinal chemist to explore this possibility.

Experimental

Measurement of log *P* was carried out in duplicate or more by standard methods,[3,4,32] with special precautions to avoid evaporation in the case of TMP and chloroform. Solute concentration was measured by GLC for **101** and **102** and by UV methods elsewhere.

Acknowledgements

It is a pleasure to acknowledge frequent and stimulating discussions with Dr. M. H. Abraham, and the original inspiration of Dr. R. H. Davies.

References

1 C. Hansch, *Acc. Chem. Res.*, 1969, **2**, 232.
2 R. N. Smith, C. Hansch and M. M. Ames, *J. Pharm. Sci.*, 1975, **64**, 599.
3 D. E. Leahy, P. J. Taylor and A. R. Wait, *Quant. Struct.–Act. Relat.*, 1989, **8**, 17.
4 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.
5 R. H. Davies, R. D. Bagnall and W. G. M. Jones, *Int. J. Quantum. Chem., Quantum Biol. Symp.*, 1974, **1**, 201.
6 T. Fujita, T. Nishioka and M. Nakajima, *J. Med. Chem.*, 1977, **20**, 1071.
7 Part 3. D. E. Leahy, J. J. Morris, P. J. Taylor and A. R. Wait, *J. Chem. Soc., Perkin Trans. 2*, 1992, following paper.
8 C. Hansch and A. J. Leo, *Substituent Constants for Correlation Analysis in Chemistry and Biology*, Wiley, New York, 1979.
9 THOR Masterfile 351, CLOGP version 3.51, Daylight Chemical Information Systems, 3951 Claremont St., Irvine, CA 92714.
10 R. F. Rekker, *The Hydrophobic Fragmental Constant*, Elsevier, Amsterdam, 1977.
11 R. F. Rekker, personal communication to A. L. J. de Meere.
12 M. J. Kamlet, M. H. Abraham, R. M. Doherty and R. W. Taft, *J. Am. Chem. Soc.*, 1984, **106**, 464; 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.
13 M. J. Kamlet, R. M. Doherty, J.-L. M. Abboud, M. H. Abraham and R. W. Taft, CHEMTECH., 1986, 566.
14 J.-L. M. Abboud, M. J. Kamlet and R. W. Taft, *Prog. Phys. Org. Chem.*, 1981, **13**, 485.
15 A. L. J. de Meere, P. J. Taylor, H. M. Grünbauer and E. Tomlinson, *Int. J. Pharm.*, 1986, **31**, 209.
16 P. Seiler, *Eur. J. Med. Chem.–Chim. Ther.*, 1974, **9**, 473.
17 Preliminary communication: D. E. Leahy, J. J. Morris, P. J. Taylor and A. R. Wait, in *QSAR: Rational Approaches to the Design of Bioactive Compounds*, eds. C. Silipo and A. Vittoria, Elsevier, Amsterdam, 1991, pp. 75, 123; P. J. Taylor, poster presented at the 8th European Symposium on QSAR, Sorrento, Italy, September, 1990.
18 S. S. Davis, T. Higuchi and J. H. Rytting, *Adv. Pharm. Sci.*, 1974, **4**, 73.
19 M. H. Abraham, *J. Am. Chem. Soc.*, 1982, **104**, 2085.
20 M. H. Abraham, G. S. Whiting, R. M. Doherty and W. J. Shuely, *J. Chem. Soc., Perkin Trans. 2*, 1990, 1451.
21 F. A. Long and W. F. McDevit, *Chem. Rev.*, 1952, **51**, 119.
22 J. Hine, S. Hahn and D. E. Miles, *J. Org. Chem.*, 1986, **51**, 577.
23 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.
24 T. W. Panunto, Z. Urbáncyak-Lipkowska, R. Johnson and M. C. Etter, *J. Am. Chem. Soc.*, 1987, **109**, 7786.
25 D. E. Leahy, *J. Pharm. Sci.*, 1986, **75**, 629.
26 J. H. Hildebrand, *Chem. Rev.*, 1949, **44**, 37; A. F. M. Barton, *Chem. Rev.*, 1975, **75**, 731; M. H. Abraham, *Prog. Phys. Org. Chem.*, 1974, **11**, 1.
27 P. J. Taylor, unpublished analysis.
28 J. M. Mayer, H. van de Waterbeemd and B. Testa, *Eur. J. Med. Chem.–Chim. Ther.*, 1982, **17**, 17.
29 A. J. Leo, *J. Chem. Soc., Perkin Trans. 2*, 1983, 825.
30 R. Collander, *Acta Chem. Scand.*, 1951, **5**, 774.
31 R. C. Young, R. C. Mitchell, T. H. Brown, C. R. Ganellin, R. Griffiths, M. Jones, K. K. Rana, D. Saunders, I. R. Smith, N. E. Sore and T. J. Wilks, *J. Med. Chem.*, 1988, **31**, 656.
32 A. Leo, C. Hansch and D. Elkins, *Chem. Rev.*, 1971, **71**, 525.

Paper 1/04345E
Received 20*th August* 1991
Accepted 6*th January* 1992