JOURNAL OF PHYSICAL ORGANIC CHEMISTRY, VOL. 7, 743–750 (1994)

MODEL SOLVENT SYSTEMS FOR QSAR. PART IV. THE HYDROGEN BOND ACCEPTOR BEHAVIOUR OF HETEROCYCLES

DAVID E. LEAHY, JEFFREY J. MORRIS, PETER J. TAYLOR* AND ALAN R. WAIT
Zeneca Pharmaceuticals, Alderley Park, Macclesfield, Cheshire SK10 4TG, UK

An LSER analysis based on the partitioning of 15 proton acceptor heterocycles has succeeded in extracting $\Sigma\beta$ values, but only at the cost of demonstrating solvent dependence for some of them. As noted by Abraham, the division lies between protic and aprotic organic phases. His observation that pyridine and quinoline are less effective acceptors when surrounded by solvent than in 1 : 1 association was confirmed, and possible reasons for this are discussed. Two other such cases are *N*-methylimidazole and pyridazine, both of which give lower $\Sigma\beta$ values in octanol than in PGDP. For both, $\Sigma\beta$ in PGDP is what would be expected on the basis of log K_β. The value for pyridazine in octanol suggests that, here, the 'α-effect' is no longer operative; it is possible that this result can be generalized to other such heterocycles. Elsewhere, the most remarkable finding is that, where there are two proton acceptor sites in one heterocyclic ring, $\Sigma\beta$ is the simple unattenuated sum of the separate β_f values. If this result is general, it leads to a very simple way of estimating $\Sigma\beta$ for heterocycles by calculation where data are unavailable. Evidence was also found, in certain cases, for hydrogen bonding to the π-donor heteroatom or the aromatic ring. The QSAR implications of these results are discussed.

INTRODUCTION

The creation of hydrogen bonding scales, which allow for the first time quantitative comparison of all or nearly all proton donors or acceptors with one another, is a surprisingly recent achievement.[1–4] The most comprehensive such scales are due to Abraham and coworkers,[1,2] who for solvent tetrachloromethane have derived a general equation:[5]

$$\log K = 7{\cdot}354\,\alpha_2^H\beta_2^H - 1{\cdot}094 \quad (1)$$

where α_2^H and β_2^H represent scaled proton donor and acceptor values that relate to the defining equilibrium constants as in the equations

$$\alpha_2^H = (\log K_A^H + 1{\cdot}1)/4{\cdot}636 \quad (2)$$

$$\beta_2^H = (\log K_B^H + 1{\cdot}1)/4{\cdot}636 \quad (3)$$

Equation (1) is based on 1312 data points and allows the accurate calculation (s.d. = 0·09) of thousands more. Raevsky *et al.*[4] have independently derived an equation of similar form to equation (1).

A major driving force behind this work has been the need for solute proton donor and acceptor scales for use in drug design.[3,6–9] For this purpose, it is not self-evident that the very non-polar tetrachloromethane should be a good model solvent; the type of biological process in which we are interested[6,7] probably takes place in regions of considerably higher polarity.[3] In deriving our log K_α and log K_β scales, therefore, we employed as solvent 1,1,1-trichloroethane, about the most polar solvent devoid of proton donor or acceptor properties that it is possible to obtain. We find, in fact, certain significant changes in ranking order, e.g. for sulphoxides, as between these two solvents.[3] In addition, the very low solubility in tetrachloromethane of many polar molecules of biological interest limits the utility of that solvent in the present context.

For a molecule with multiple hydrogen bonding sites, the individual equilibrium constants K_{ij} relate to K_{obs} as in the equation

$$K_{obs} = K_1 + K_2 + \ldots \quad (4)$$

Provided that $K_1 \gg K_2$, then log $K_{obs} \approx \log K_1$ and there is no ambiguity; in the special case that $K_1 = K_2$, statistical correction is in principle required (but may not always be appropriate[3]). Equation (4) is for 1 : 1 association. If, however, the solute is immersed in a great excess of solvent, it is possible that each of these sites will be fully utilized. In that case, equation (4) ceases to apply and the expected relation is multiplicative; that is, the (logarithmic) α- and β-values of equations (1)–(3) become additive. For multiply substituted drug molecules in which the functional groups are

* Author for correspondence.

CCC 0894–3230/94/120743–08

Received 16 May 1994
Revised 30 June 1994

well separated, this additivity assumption has successfully been used in quantitative structure–activity relationships (QSARs).[10] The same assumption underpins *additivity rules* for log P.[11]

It is also possible for a single functional group, such as carbonyl, to contain multiple binding sites; this contrasts, e.g., with amines which do not. Recently, therefore, there have been attempts to redetermine solute α and β values under the 'real-life'[8] conditions of excess solvent, and to compare the resulting $\Sigma\alpha$ and $\Sigma\beta$ values with their originals.[7–9,12] The methodology involves back-calculation based on liquid–liquid or GLC partitioning. It turns out that the differences are much more significant for $\Sigma\beta$ than $\Sigma\alpha$, not only through multiplicity of lone pairs, but also because certain aromatic moieties allow bonding to the π-cloud.[9,12] Abraham,[12] basing his results on calculations for 16 partitioning systems, has published $\Sigma\alpha_2^H$ and $\Sigma\beta_2^H$ values for a large number of solutes; these have been used successfully in a range of biological and other correlation equations.[13]

Our parallel investigations[7–9] have concerned not only the strength of hydrogen bonding but also its directionality, in the sense that we wished to know, e.g. for carbonyl, whether the second lone pair is active, and if so, how its strength should compare with the first. We found,[9] *inter alia*, that while both lone pairs are in general available for uncrowded aliphatic carbonyl groups, *peri*-interactions in aromatic compounds can effectively shield one lone pair, and the same is true for eclipsing OH or NH, e.g. in CO_2H and $CONH_2$ (but not in CONHR with its *trans* alignment). We found that the presence or absence of a second (or subsequent) lone pair could be handled by a new quantity, $n\beta_f$, where n is the number of available lone pairs after the first, and β_f is the functional group β-value, a quantity in general closely related to log K_β. For a given solvent system the coefficient of $n\beta_f$ is variable but, for those examined, lies in the range 18–38% of β_f itself, a chemically reasonable result in that charge transfer from the first lone pair is likely to reduce charge density on the second. We also found evidence that, for certain types of carbonyl, bond formation is probably constrained along the bond axis, with appreciable loss in strength. All these considerations, and others,[9] are highly relevant in the context of receptor binding, which was and remains our chief preoccupation.[6,7]

Heterocycles, a subject of great importance to the medicinal chemist, were conspicuous by their absence from this analysis.[9] These may be seen as a special type of large and rather rigid functional group, containing an aromatic core which, in many cases, links one highly directional binding site with another. As such, unusual complications might well be anticipated. Abraham[12] has already discovered that pyridine and the alkypyridines behave abnormally; suitably warned, we have kept away from this subject until now. We now demonstrate that heterocycles as acceptors immersed in a solvent do indeed show several types of unexpected behaviour, but that some kind of rationale is possible nevertheless.

RESULTS

Our previous investigation[9] was aimed at covering the largest possible number of functional groups with the smallest possible number of compounds, since we judged this approach to be that of most value to the medicinal chemist. We have employed the same approach here, in that only parent heterocycles have been considered (where *N*-methyl derivatives, e.g. of imidazole, count as parents). We have been further constrained by the discovery[9a] that calculated permittivities are in general unreliable and lead to highly inconsistent results, so that we have considered only heterocycles for which μ has been measured (on the evidence[9a] that μ is scarcely affected by *N*-methylation, we have stretched a point in that respect). We have confined our main analysis to proton acceptors so as to avoid having to estimate α and β simultaneously. Later we present a tentative analysis of NH-containing heterocycles.

This leaves only 15 heterocycles, which, however, cover most important classes. For this reason, we felt it imperative to expand the database by including certain compounds from our previous analysis:[9] benzene, naphthalene, the halobenzenes, and certain other compounds PhR, where R is some symmetrical and conformationally rigid substituent. There are ten of these, bringing the total number of compounds to 25. It seemed to us possible that symmetrical and very rigid compounds, as are all heteroaromatics, might form a sub-set with significantly different coefficients to some or all of the parameters used in the analysis; we have indeed found evidence for this (see below).

$$\log P = c + mV_I + s\mu^2 + b\Sigma\beta + dn\beta_f \qquad (5)$$

Leaving aside terms concerned with proton donors, our previous analysis[9a] employed equation (5). Our back-calculation procedure assumed that V_I and μ are certainly known, and then optimized $\Sigma\beta$ and $n\beta_f$ iteratively across four solvent systems in such a way as to constrain the summed residuals for any one compound as closely as possible to zero. These solvent systems were the 'critical quartet'[8,9] of amphiprotic, inert, donor and acceptor which we consider sufficiently to delineate the range of biological possibilities. We have employed the same procedure here, with two important exceptions. Since so few log P values exist for parent heterocycles in chloroform and 'alkane' correlation analysis is confined perforce to octanol and PGDP. Second, the use of $n\beta_f$ seems inappropriate. This term was introduced to handle multiple lone pairs on the same heteroatom, not simply in the same molecule; where more than one potential site for hydrogen

bonding exists, we have in this case to leave their relative importance to the results of the analysis.

For the first 10 compounds in Table 1, all parameters were fixed at their former values.[9a] As a first attempt at $\Sigma\beta$ for those heterocycles containing a single nitrogen atom we employed $\log K_\beta + 0{\cdot}3$, this allowing for the apparent scale zero of $\log K = -0{\cdot}3$ in water,[9a] and then back-calculated $\Sigma\beta$ for the remainder from the resulting regression equation. It quickly became apparent that PGDP was much easier to fit than octanol; while the octanol equations satisfied most statistical criteria, the intercept term c was unacceptably large. We eventually discovered that this problem could be overcome if four compounds were omitted: pyridine, quinoline, pyridazine and *N*-methylimidazole. As we formerly found[9a] for S=O and P=O, $\Sigma\beta$ in these cases appears to depend on the solvent system, being lower for octanol than for PGDP. Abraham[12] made similar observations for all the above except pyridazine (not examined).

Our final list of parameter values appears in Table 1; $\Sigma\beta(1)$ and $\Sigma\beta(2)$ refer to the optimum values for octanol and PGDP, respectively. As previously,[9a] we did not consider it justifiable to attempt a precision of better than 0·1. Table 2 sets out the correlation equations. Both sets of $\Sigma\beta$ values have been used to analyse both sets of data, and in addition we include for comparison those derived from the first 10 compounds, the non-heterocycles. We have also extended this analysis to such data as exist for chloroform and 'alkane;'[14] it will be noted that the PGDP $\Sigma\beta$ values fit both very much better than those for octanol. This again is consistent with Abraham's observations,[12] although the possibility remains open that some other set would fit even better. Discussion below considers only the optimum equations: set (1) for octanol and set (2) elsewhere.

We recognize that the restriction of this analysis to two solvent systems, in place of our original four,[9] is a considerable drawback that poses a potential threat to its validity. In particular, as pointed out by a referee, it would require no great error in log *P* to invalidate some of our $\Sigma\beta$ values. We have attempted to counter this threat by using only well attested data. In particular, all

Table 1. Log *P* and parameter values[a]

Compound	Log *P*				V_I	μ^2	$\Sigma\beta$[b]	
	Oct	PGDP	$CHCl_3$	Alkane			1	2
Benzene	2·13	2·36	2·80	2·24	0·495	0·00	0·3	0·3
Naphthalene	3·30	3·73		3·39	0·766	0·00	0·4	0·4
PhMe	2·73	2·89	3·41	2·89	0·593	0·13	0·3	0·3
$PhCF_3$	3·01	3·26			0·670	6·81	0·0	0·0
PhF	2·27	2·50	2·85	2·46	0·523	2·04	0·3	0·3
PhCl	2·84	3·08	3·46	2·93	0·586	2·56	0·0	0·0
PhBr	2·99	3·27	3·61	3·10	0·628	2·40	0·0	0·0
PhI	3·25	3·48		3·33	0·670	1·85	0·0	0·0
PhCN	1·56	1·66	2·71	0·96	0·604	16·65	1·3	1·3
$PhNO_2$	1·85	2·16	2·93	1·44	0·615	19·36	0·8	0·8
Pyridine	0·65	0·08	1·32	−0·31[c]	0·470	4·93	2·0	2·5
Quinoline	2·03	1·62		1·26[c]	0·737	5·52	2·0	2·5
Pyrazine	−0·23	−0·61			0·447	0·44	3·1	3·1
Pyrimidine	−0·40	−0·80			0·447	5·43	3·3	3·3
Pyridazine	−0·65	−1·60			0·447	18·23	3·3	3·9
Isoxazole	0·08	0·17			0·362	9·06	1·8	1·8
Oxazole	0·12	−0·20			0·363	2·25	2·3	2·3
Thiazole	0·44	0·24			0·420	2·96	2·2	2·2
N-Methylpyrazole	0·23[d]	−0·16			0·488	5·48	2·8	2·8
N-Methylimidazole	−0·03	−1·40		−2·16[c,e]	0·488	14·44	2·8	4·0
1-Methyl-*s*-triazole	−1·08	−1·83			0·462	12·25	4·2	4·2
Furan	1·34	1·56			0·389	0·52	0·6	0·6
Thiophene	1·81	2·09		1·89[c,f]	0·442	0·29	0·3	0·3
N-Methylpyrrole	1·21	1·39			0·512	4·12	1·5	1·5
N-Methylindole	2·72	3·01			0·777	5·66[g]	1·2	1·2

[a] Log *P* values from Ref. 9b or 15 unless stated otherwise; V_I in units of 10^{-2} dm^3 mol^{-1}; μ is permittivity in debye.[26]
[b] First column, 'octanol set;' second column, 'PGDP set' (see text).
[c] Ref. 14.
[d] A. J. Leo, personal communication.
[e] In cyclohexane.
[f] In hexadecane.
[g] Assumed as for indole.

Table 2. Regression equations for the four solvent systems[a]

Solvent	*c*	*m*	*s*	*b*	*n*	r^2	s.e.	*F*
Oct P	0·20	4·52	−0·025	−0·77	10	0·992	0·07	251
	(0·18)	(0·29)	(0·005)	(0·08)				
(1)	−0·09	4·95	−0·025	−0·71	25	0·998	0·07	2812
	(0·09)	(0·14)	(0·003)	(0·01)				
(2)	−0·34	5·27	−0·019	−0·63	25	0·982	0·20	375
	(0·23)	(0·38)	(0·008)	(0·04)				
PGDP P	0·08	5·14	−0·025	−0·82	10	0·995	0·06	379
	(0·16)	(0·25)	(0·004)	(0·07)				
(1)	0·21	5·09	−0·035	−0·99	25	0·982	0·25	384
	(0·30)	(0·49)	(0·010)	(0·05)				
(2)	0·04	5·29	−0·021	−0·93	25	0·998	0·08	3718
	(0·09)	(0·15)	(0·003)	(0·02)				
$CHCl_3$ P	−0·06	5·99	−0·019	−0·42	7	0·988	0·06	83
	(0·34)	(0·59)	(0·006)	(0·10)				
(1)	0·02	5·88	−0·005	−0·69	8	0·988	0·10	113
	(0·60)	(1·06)	(0·009)	(0·09)				
(2)	0·02	5·86	−0·012	−0·55	8	0·995	0·06	289
	(0·37)	(0·66)	(0·005)	(0·05)				
'Alk' P	0·53	4·31	−0·047	−1·06	9	0·994	0·08	300
	(0·22)	(0·36)	(0·006)	(0·11)				
(1)	−0·22	5·65	−0·040	−1·34	13	0·988	0·21	237
	(0·37)	(0·60)	(0·011)	(0·08)				
(2)	−0·14	5·37	−0·050	−0·97	13	0·996	0·12	673
	(0·22)	(0·36)	(0·006)	(0·03)				
Regression equations for functional groups:[b]								
Oct	0·21	4·42	−0·023	−0·77	78	0·990	0·10	1406
	(0·05)	(0·08)	(0·001)	(0·01)				
PGDP	0·03	5·42	−0·021	−1·09	83	0·995	0·10	2961
	(0·05)	(0·08)	(0·001)	(0·01)				
$CHCl_3$	0·43	5·07	−0·0006	−0·60	33	0·993	0·11	610
	(0·11)	(0·20)	(0·0027)	(0·03)				
'Alk'	0·20	4·95	−0·055	−1·10	46	0·996	0·12	1725
	(0·08)	(0·13)	(0·003)	(0·02)				

[a] P = Preliminary regression, on the first 10 compounds in Table 1; (1) using $\Sigma\beta(1)$ in Table 1; (2) using $\Sigma\beta(2)$ in Table 1.
[b] From Ref. 9a, omitting the term in $n\beta_f$ and all terms involving α.

log *P* values for PGDP and some for octanol come from a single laboratory (our own),[15] with replicates agreeing to within 0·02, while the remaining octanol values derive from Leo's 'Starlist,'[16] to which similar standards apply. The same cannot be claimed for the chloroform or 'alkane' values, hence our much more tentative use of them.

DISCUSSION

By comparison with our original regression equations[9a] as also set out in Table 2, it will be seen that, while the coefficients of μ^2 are little changed, and (as previously) changes in the intercept term tend to cancel against those of V_I, the coefficients of $\Sigma\beta$ are significantly lower, by 10–15%. The origin of this may lie in an entropic constraint, precisely the result of the very rigid structure that all these compounds have in common; most of those examined previously were flexible molecules in which the aromatic moiety, if present, was incidental. If so, it may help to explain why a preliminary analysis in which we used the original correlation equations to obtain estimates for $\Sigma\beta$ was wildly unsuccessful. It seems likely, therefore, that the use of correlation equations to extract chemically meaningful information as is our present intention must be based on compounds of similar structural type.

There is a rough relationship between $\Sigma\beta$ (for PGDP) and Abraham's $\Sigma\beta_2^H$ as given by the equation

$$\Sigma\beta_2^H = 0{\cdot}176(0{\cdot}005)\Sigma\beta + 0{\cdot}085(0{\cdot}010) \quad (6)$$
$$(n = 23,\ r^2 = 0{\cdot}978,\ s = 0{\cdot}036,\ F = 928)$$

Its most significant feature is its intercept of 0·085; that is, the effective zero for $\Sigma\beta$ is higher. As noted previously,[9a] this means that most benzenes substituted with electron-withdrawing non-bonding functionalities possess $\Sigma\beta = 0$, whereas their proton acceptor ability is

still perceptible on the $\Sigma\beta_2^H$ scale. While $\Sigma\beta$ is tuned to water, it may be that $\Sigma\beta_2^H$ from its mode of derivation is tuned to the organic phase.

Heterocycles of constant $\Sigma\beta$

These may contain one nitrogen acceptor or two. The former class consists of oxazole, isoxazole, thiazole and *N*-methylpyrazole. Only thiazole possesses exactly the β_f value expected (Table 3); the rest are too positive by $\Delta\beta = 0{\cdot}2$–$0{\cdot}4$. This extra contribution may be due to the π-donor heteroatom or be caused by hydrogen bonding to the π-cloud. The former possibility is suggested by the zero $\Delta\beta$ value for thiazole. The sulphur atom even of thiophene probably possesses no proton acceptor ability; we find a $\Sigma\beta$ value for thiophene identical with that for benzene (hardly a surprise). Hence no perceptible bonding to the sulphur atom of thiazole is expected. Oxygen is a better π-donor than sulphur,[17] so the larger $\Sigma\beta$ value for furan than for thiophene is as expected; possibly the π-cloud of oxazole accounts for $\Delta\beta = 0{\cdot}2$ in that case. On the other hand, oxygen is *mutatis mutandis* a better proton acceptor than sulphur,[2,3] so fractional bonding to oxygen is also plausible. For isoxazole, a still weaker base, the appreciable value of $\Delta\beta = 0{\cdot}4$ may argue for some contribution from the 'α-effect'[18] in addition to possibly, some binding to the π-cloud. The former is also suggested by a comparison of the $\Sigma\beta$ values for furan, thiophene and *N*-methylpyrrole (Table 2). These show the expected π-donor order $N \gg O > S$, as we have found previously[19] in our analysis of the log *P* values for substituted π-donor heterocycles. In view of this, the lesser value of $\Delta\beta = 0{\cdot}3$ for *N*-methylpyrazole, almost certainly due to bonding to the π-cloud, suggests some other source for at least part of the extra bonding experienced by isoxazole.

Regular heterocycles with two nitrogen acceptors comprise pyrazine, pyrimidine and 1-methyl-1,2,4-triazole. It is intriguing that, in all three cases, $\Sigma\beta$ is a simple summation of β_f, without any of the attenuation present, e.g. in carbonyl, and which might have been expected here also. Perhaps this comes about since the degree of proton transfer in hydrogen bonding is in fact small[20] and the aromatic core of these heterocycles acts as a buffer of a type not available in simpler cases. For the triazole, we used Kenny's recent calculated values[21] of log K_β with appropriate zero correction (Table 3) and obtained a $\Sigma\beta$ value identical with that found. This contrasts with the case of 1 : 1 association [equation (4)], where the less basic site (the 2-position) contributes almost nothing to the observed log K_β. Such examples of the contrast between partitioning Gibbs energy on the one hand, and the very different binding energies potentially available for different topological arrangements at the receptor site, are of special relevance to the medicinal chemist.

Table 3. Comparison between β_f and $\Sigma\beta$ for nitrogen π-acceptor heterocycles

	$\Sigma\beta$[a]		β_f[b]		
Compound	1	2	Obs.[c]	Calc.[d]	Notes
Pyridine	2·0	2·5	2·82	2·81	See text
Quinoline	2·0	2·5	2·80		See text
Pyrazine	3·1	3·1	1·46[e]	1·60[e]	$\Sigma\beta \approx 2\beta_f$
Pyrimidine	3·3	3·3	1·67[e]	1·99[e]	$\Sigma\beta \approx 2\beta_f$
Pyridazine	3·3	3·9	2·53[e]	2·58[e]	See text
Isoxazole	1·8	1·8	1·36	1·40	$\Delta\beta \approx 0{\cdot}4$
Oxazole	2·3	2·3	1·97	2·21	$\Delta\beta \approx 0{\cdot}2$
Thiazole	2·2	2·2	2·20	2·10	$\Delta\beta \approx 0{\cdot}0$
N-Methylpyrazole	2·8	2·8	2·52	2·57	$\Delta\beta \approx 0{\cdot}3$
N-Methylimidazole	2·8	4·0	3·98	3·99	See text
1-Methyl-*s*-triazole	4·2	4·2	2·68	2·93(4) 1·31(2)	$\Sigma\beta \approx \beta_{f(4)} + \beta_{f(2)}$

[a] See Table 1 for the significance of $\Sigma\beta(1)$ and $\Sigma\beta(2)$.
[b] $\beta_f = \log K_\beta + 0{\cdot}3$.
[c] Ref. 3.
[d] Ref. 21.
[e] Statistically corrected.

Heterocycles of variable $\Sigma\beta$

Abraham[12] has previously reported that pyridine and the alkylpyridines show considerably lower $\Sigma\beta_2^H$ values than would have been expected on the basis of β_2^H itself, and has tentatively explained this as due to a kind of 'solvent sorting' by which pyridine in the octanol phase is preferentially solvated by its aqueous component. The same effect is found for two other organic phases, isobutanol and *n*-butyl acetate, which also possess a high water content; these give an average $\Sigma\beta_2^H = 0{\cdot}45$, contrasting with an average $\Sigma\beta_2^H = 0{\cdot}52$ elsewhere.[12] This is echoed by Table 3, where octanol and PGDP give apparent $\Sigma\beta$ values of 2·0 and 2·5, respectively.

However, even the latter value falls short of the expected $\beta_f = 2{\cdot}82$, and indeed Abraham *et al.*[2] had previously reported $\beta_2^H = 0{\cdot}62$ for pyridine. So, while Abraham's explanation may contribute, it cannot be the whole story. Our suggested supplementary explanation starts from the fact that benzene itself possesses a β_{ar} (= aromatic β)[9a] value, but one only just above the threshold level. An aromatic ring with β_{ar} just below this threshold would contribute nothing to $\Sigma\beta$, yet might still be capable of back-polarization by bulk water with some loss of electron density on nitrogen. This effect would still operate, e.g. in alkane–water and PGDP–water systems, so accounting for the failure of $\Sigma\beta$ to attain the expected value even there. Further electronegative substitution might then reduce π-density to the point that back-polarization can no longer occur, so accounting for the regular behaviour of deactivated pyridines,[12] and of the three diazaheterocycles above. It is possible that back-polarization also operates in some electronegatively substituted benzenes; we have

previously noted[9a] that, in terms of $\Sigma\beta$, $PhNO_2$ is apoorer proton acceptor than might have been expected.

N-Methylimidazole also appears in Abraham's list of deviant compounds, though he does not comment on it. We noted[3] that this compound behaves in 1,1,1-trichloroethane (TCE) as if considerably polarized in the sense of $RN^+{=}CH{-}N^-$, resulting in an unexpectedly high $\log K_\beta$ value; it is one of the few compounds for which our correlation equation for $\log K_\beta$ vs $\log K_B^H$ gives seriously inaccurate results. PGDP is a solvent of similar polarity to TCE and it might be expected that the PGDP–water solvent system would therefore match β_f fairly exactly, as it does. The consequence for S=O and P=O of polarization in the above sense is to generate a repulsion with aprotic solvents (and $CHCl_3$) which increases as the dipolarity of the solvent falls, since $X^+{-}O^-$ is favoured in media of low polarity.[9a] Since the opposite trend is expected for $RN^+{=}CH{-}N^-$, the origin of this superficially similar phenomenon is not understood. We shall only note that it resembles that for pyridine, although much exaggerated.

Pyridazine is our last case, and the only example in this set of a genuine 'α-effect heterocycle' (though we have log K_β values for others[3]). Our β_f value of 2·53 (Table 3) is that obtained by statistical correction, although it is not clear what such an operation means in the present case, since it is uncertain whether bonding involves a single lone pair. (Hydrogen bonding by the isolated molecule appears to take place as if along the line of one lone pair, somewhat distorted towards the other[22]). Unsurprisingly, both $\Sigma\beta$ values are much greater than this, and by much more than the statistical factor. In water, and perhaps hydroxylic solvents generally, the 'α-effect' is presumably at a minimum; it is known from kinetic studies that pyridazine addition does not show the 'α-effect' in water,[23] though it can elsewhere.[24] It may be significant, therefore, that $\Sigma\beta(1) = 3{\cdot}3$ is very much the same value as for the other diazines, which may indicate an intrinsic acceptor ability per nitrogen atom of much the same order. In that case, $\Sigma\beta(2) = 3{\cdot}9$ presumably indicates a specific repulsion by the proton acceptor solvent PGDP, which could be greater or less in another aprotic solvent. It is possible, therefore, that $\Sigma\beta$ for heterocycles that contain adjacent sp^2-hybridized nitrogen atoms will show an unusual degree of solvent sensitivity. It is also possible, for different reasons, that the same will be true for *N*-methylimidazole and any analogous very polarizable heterocycles that may exist. One such class might be 'push–pull' conjugated systems such as the aminoenones, e.g. 4-pyridone, which we have noted previously[3] to be exceptional proton acceptors. Table 4 sets out octanol and PGDP log *P* values for a number of proton acceptor heterocycles, and lists the difference $\Delta\log P$. For the first set, which are known or expected to show 'regular' behaviour, $\Delta\log P$ is never more negative than −0·4. The second set is for potential

Table 4. $\Delta\log P$ for some proton acceptor heterocycles[a,b]

Compound	Log *P*(oct)	Log *P*(PGDP)	$\Delta\log P$
Pyrazine	−0·23	−0·61	−0·38
Quinoxaline	1·32	1·20	−0·12
Pyrimidine	−0·40	−0·80	−0·40
Quinazoline	1·01	0·70	−0·31
Oxazole	0·12	−0·20	−0·32
Thiazole	0·44	0·24	−0·20
Pyridazine	−0·65	−1·60	−0·95
Cinnoline	0·93	0·29	−0·64
Phthalazine	0·57	−0·46	−1·03
1-Methyl-1,2,3-triazole	−1·20	−1·58	−0·38
1-Methylbenzotriazole	1·13	0·86	−0·27
Pyridine	0·65	0·08	−0·57
1-Methylimidazole	−0·03	−1·40	−1·37
1-Methylquinol-2-one	1·45	0·94	−0·51
1-Methylquinol-4-one	0·44	−1·92	−2·36
1-Methylquinoxalin-2-one	0·79	0·68	−0·11
3-Methylquinazolin-4-one	0·69	0·35	−0·34
3-Methylbenzotriazin-4-one	0·84	1·13	+0·27

[a] $\Delta\log P = \log P(\text{oct}) - \log P(\text{PGDP})$.
[b] Data from Ref. 15.

'α-effect' heterocycles, and these are very variable, with three (including pyridazine) showing large negative values. The third is a miscellany that includes 1-methylimidazole, $\Delta\log P = -1{\cdot}37$, and also the most extreme example of this type, 1-methylquinol-4-one, $\Delta\log P = -2{\cdot}36$, that we have yet encountered.[15] In the absence of more specific criteria, it is possible that this difference in log *P* value provides the best guide we have at present as to the likelihood of 'regular' or 'irregular' behaviour. Unfortunately, data for solvents other than octanol and PGDP still scarcely exist for oxoheterocycles, as indeed is the case for heterocycles generally.

Extension to heterocycles containing NH

Data exist for five of these (Table 5). If we are to attempt to extract $\Sigma\alpha$ values, some estimate for $\Sigma\beta$ is

Table 5. Calculations for NH-containing heterocycles[a]

	Log P^b					$\Delta\log P^c$	
Compound	PGDP	Alkane	V_I	μ^2	$\Sigma\beta^d$	PGDP	Alkane
Pyrrole	0·72		0·411	4·37	1·3	−0·19	
Indole	2·38	0·79[e]	0·675	5·66	1·0	−0·19	−1·44
Pyrazole	−0·69	−2·91[e]	0·385	5·43	2·6	−0·23	−2·04
Imidazole	−2·17	−3·70[e]	0·385	16·16	3·8	−0·37	−1·13
		−4·46					−1·89
1,2,4-Triazole	−2·5[f]		0·350	10·82	4·0	−0·44	

[a] For parameter value units, see Table 1.
[b] Ref. 15.
[c] Residual from use of equation (5).
[d] $\Sigma\beta(2)$ in Table 1 with 0·2 subtracted (see text).
[e] In cyclohexane (Ref. 14).
[f] Approximate value.

first necessary. We noted previously[9a] an average rise in $\Sigma\beta$ of *ca* 0·2 on methylation of an amine or amide, so as a first approximation the $\Sigma\beta$ values in Table 5 were obtained by subtracting 0·2 from the appropriate values in Table 1. Use of the correlation equations from Table 2 then leads to a set of residuals (the last two columns of Table 5) from which we may attempt to calculate $\Sigma\alpha$. Octanol with its very low sensitivity to $\Sigma\alpha$[9a,14] is useless for this purpose, and there is only one available datum for chloroform, so this analysis is confined in practice to 'alkane' and PGDP.

The two sets of residuals possess very different magnitudes. One reason for this lies in the different slopes of the α-term, which we found in our previous analysis[9a] to be $-1{\cdot}07$ and $-0{\cdot}61$ for 'alkane' and PGDP, respectively. Another, very significant reason lies in the different scale zeroes which we have found[9a] to apply to $\Sigma\alpha$: this lies at $\log K_\alpha = -0{\cdot}4$ for 'alkane' but $+0{\cdot}1$ for PGDP. Taken together, these factors would lead, in the case of indole, for example, to apparent $\log K_\alpha$ values of 0·4 or 0·9 for PGDP and 'alkane,' respectively. The measured value is 1·15. For imidazole, the PGDP-derived value is 0·7 and that from 'alkane' is 0·7 or 1·4, where the measured value is $\log K_\alpha$ 1·20. A lower value of the coefficient to $\Sigma\alpha$, as found for $\Sigma\beta$, would of course lead to higher values for $\Sigma\alpha$ itself, but the difference between the two sets of calculated values would then increase.

We are unable at present to resolve this problem. We note that in our previous treatment[9a] we found aliphatic amines to be anomalous, with PGDP leading to higher estimates for $\Sigma\alpha$ than other solvent systems: the opposite trend to that found here. We also note that Abraham[12] found $\Sigma\beta_2^H$ for anilines to be as anomalous as those for pyridines, but in the opposite sense, while indole behaves like an aniline. There is no mention of anomalous $\Sigma\alpha_2^H$ values, however. It is possible that cooperative effects are present, with a tendency for α- and β-values to move in compensating directions, in such a way that the overall effect, e.g. on $\log P$, is more stable than would be predicted on the basis of either term separately. We have previously noted[3] that amine K_α and K_β values display a variety of anomalies, relative to those of other compounds, which may be connected with the constraints imposed by the amine inversion process, and which show themselves, for different classes of amine, in different sorts of compensating behaviour.[9a] It is equally possible, however, that these apparent phenomena are simply unreal, the result of trying to partition an overall hydrogen bonding effect between $\Sigma\alpha$ and $\Sigma\beta$ on the basis of insufficient data. More experimental evidence is badly needed.

CONCLUSION

The results of this work, while necessarily tentative, suggest a number of generalizations concerning heterocycles that should if valid be of special value to the medicinal chemist:

(a) Hydrogen bonding to nitrogen acceptor sites is additive in terms of β_f: there is no attenuation. (This seems to echo Leo's principle[16,25] in $\log P$ calculation that 'nitrogen never ages:' aza-nitrogen exerts its maximum electronic 'pull' regardless of the number of donors.) If relative acceptor strength can be calculated, therefore, it should be possible to estimate $\Sigma\beta$ for fairly complex heterocycles, and use the results in interpreting either $\log P$ or receptor binding. Kenny's paper[21] contains several further calculations of his sort. Their extension, e.g. to oxazole, will help to quantify the possible role of π-donor heteroatoms as minor but potentially useful binding sites.

(b) We have noted previously[3] that 'α-effect' heterocycles may be of particular value in contexts where one requires the maximum degree of proton acceptor ability for the minimum degree of full proton transfer (i.e. basicity). We now have direct evidence that the extent of this enhancement varies with context, being effectively zero in water and (probably) other hydroxylic solvents, but very considerable in (at least some) aprotic surroundings. For lack of evidence, we have to leave open the question as to whether this enhancement is due to intramolecular destabilization alone, or involves actual repulsion of other proton acceptors. In the first, $\Sigma\beta$ towards 'alkane' will be similar to that for PGDP; if the second, it should be less. This again could be a relevant consideration in the context of receptor binding.

(c) It is intriguing to discover that some aromatic rings are effective proton acceptors; for example, pyrrole apparently is little weaker than isoxazole. However, in terms of short-range forces, this does not compare like with like. Whereas isoxazole in water can probably form no more than two hydrogen bonds (the nature of the second being not fully defined), pyrrole may form several of varying strength, both sides of the π-cloud being available. This last does not reflect the situation likely to obtain at the biological receptor, where more than one bond is probably unrealistic. However, that bond will not be subject to strong directional constraints, so there may be contexts in which it is important. The order $N \gg O > S$ is clearly indicated as the effect of the donor atom on the likely strength of binding to the π-cloud.

(d) We have had to leave open the question of how important the NH donor is in heterocycles that contain it, while adding to the impression[1,3,9,12] that its importance is generally not great. Tetrazole may be an exception,[3] but the required data do not yet exist.

(e) We also have evidence, the origin of which is not well understood, for anomalous behaviour by 'push–pull' systems that certainly include *N*-methylimidazole and may extend to the class of oxoheterocycle and related compounds[3] for which resonance of the type

$X-CH=CH-C=Y \leftrightarrow X^+=CH-CH=C-Y^-$ can be written. All such compounds are exceptionally strong proton acceptors and the evidence is suggestive that their $\Sigma\beta$ values are likely to show the same type of dependence on solvent system as attaches to the 'α-effect' heterocycles. Both classes may be as discriminating biologically for their inability to bind in very non-polar surroundings as for the strong hydrogen bonding of which they should be capable with a suitable proton donor.

ACKNOWLEDGEMENT

This paper is dedicated to Professor R. W. Taft, pioneer extraordinary, on the occasion of his 70th birthday.

REFERENCES

1. 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* 699–711 (1989).
2. M. H. Abraham, P. L. Grellier, D. V. Prior, J. J. Morris and P. J. Taylor, *J. Chem. Soc., Perkin Trans. 2* 521–529 (1990).
3. 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* 1355–1375 (1989).
4. O. A. Raevsky, V. Yu. Grigor'ev, D. B. Kireev and N. S. Zefirov, *Quant. Struct.–Act. Relat.* **11**, 49–63 (1992).
5. M. H. Abraham, P. L. Gellier, D. V. Prior, R. W. Taft, J. J. Morris, P. J. Taylor, C. Laurence, M. Berthelot, R. M. Doherty, M. J. Kamlet, J.-L. M. Abboud, K. Sraidi and G. Guiheneuf, *J. Am. Chem. Soc.* **110**, 8534–8536 (1988).
6. J. J. Morris, L. R. Hughes, A. T. Glen and P. J. Taylor, *J. Med. Chem.* **34**, 477–456 (1991).
7. D. E. Leahy, J. J. Morris, P. J. Taylor and A. R. Wait, in *QSAR: Rational Approaches to the Design of Bioactive Compounds*, edited by C. Silipo and A. Vittoria, pp. 75–82. Elsevier, Amsterdam (1991).
8. D. E. Leahy, J. J. Morris, P. J. Taylor and A. R. Wait, *J. Chim. Phys.* **89**, 1597–1602 (1992).
9. D. E. Leahy, J. J. Morris, P. J. Taylor and A. R. Wait, *J. Chem. Soc., Perkin Trans. 2* (a) 705–722; (b) 723–731 (1992).
10. M. J. Kamlet, R. M. Doherty, J.-L. M. Abboud, M. H. Abraham and R. W. Taft, *Chem. Tech. US* **16**, 566–576 (1986); M. H. Abraham, H. S. Chadha and R. C. Mitchell, *J. Pharm. Sci.*, **83**, 1257–1268 (1994).
11. See, e.g., R. F. Rekker, *The Hydrophobic Fragmental Constant*. Elsevier, Amsterdam (1977); C. Hansch and A. J. Leo, *Substituent Constants for Correlation Analysis in Chemistry and Biology*. Wiley, New York (1979).
12. M. H. Abraham, *J. Phys. Org. Chem.* **6**, 660–684 (1993).
13. M. H. Abraham, *Pure Appl. Chem.* **65**, 2503–2512 (1993); *Chem. Soc. Rev.* 73–83 (1993).
14. M. H. Abraham, H. S. Chadha, G. S. Whiting and R. C. Mitchell, *J. Pharm. Sci.*, **83**, 1085–1100 (1994).
15. D. E. Leahy, P. J. Taylor and A. R. Wait, *Quant. Struct.–Act. Relat.* **8**, 17–31 (1989).
16. *THOR Masterfile 351, CLOGP Version 3.51*. Daylight Chemical Information Systems, Irvine, CA.
17. M. Charton, *Prog. Phys. Org. Chem.* **13**, 119–251 (1981).
18. J. D. Aubort and R. F. Hudson, *Chem. Commun.* 937–938 (1970); R. W. Taft, F. Anvia, M. Taagepera, J. Catalan and J. Elguero, *J. Am. Chem. Soc.* **108**, 3237–3239 (1986).
19. J. Bradshaw and P. J. Taylor, *Quant. Struct.–Act. Relat.* **8**, 279–287 (1989).
20. R. W. Taft, D. Gurka, L. Joris, P. von R. Schleyer and J. W. Rakshys, *J. Am. Chem. Soc.* **91**, 4801–4808 (1969).
21. P. W. Kenny, *J. Chem. Soc., Perkin Trans. 2* 199–202 (1994).
22. P. W. Kenny, personal communication.
23. P. M. Bond, E. A. Castro and R. B. Moodie, *J. Chem. Soc., Perkin Trans. 2* 68–72 (1976).
24. J. A. Zoltewicz and L. W. Deady, *J. Am. Chem. Soc.* **94**, 2765–2769 (1972).
25. A. J. Leo, *J. Chem. Soc., Perkin Trans. 2* 825–838 (1983).
26. A. L. McLellan, *Tables of Experimental Dipole Moments*. Vol. 1, W. H. Freeman, San Francisco (1963); Vol. 2, Rahara Enterprises, El Cerrito, CA (1974).