Calculation of Abraham descriptors from solvent–water partition coefficients in four different systems; evaluation of different methods of calculation

Andreas M. Zissimos,*[a] Michael H. Abraham,[a] Matthew C. Barker,[a] Karl J. Box[b] and Kin Y. Tam †[b]

[a] *Department of Chemistry, University College London, 20 Gordon Street, London, UK WC1H OAJ. E-mail: a.zissimos@ucl.ac.uk; Fax: +44-020-7679-7463*

[b] *Sirius Analytical Instruments Ltd, Riverside, Forest Row Business Park, Forest Row, East Sussex, UK RH18 5DW*

Received (in Cambridge, UK) 7th November 2001, Accepted 19th December 2001
First published as an Advance Article on the web 25th January 2002

Partition coefficients for a set of drug compounds have been measured in four solvent–water systems, with octanol, chloroform, cyclohexane and toluene as solvents. The data have been used to test four different methods for the calculation of the three Abraham descriptors, dipolarity/polarizability **S**, hydrogen bond acidity **A**, and hydrogen bond basicity **B**. The methods involved (i) the use of Microsoft 'Solver', (ii) use of a series of regression equations developed from compounds with known descriptors, and use of two further methods that have been developed, (iii) a program similar to Solver that we denote as 'Descfit', and (iv) a program that uses a set of three simultaneous equations, and which we denote as 'TripleX'. We show that partition coefficients for a given drug in only four solvent–water systems can be used to calculate the three Abraham descriptors reliably, and we test all four methods of calculation for reproducibility and ease of use. We finally test the applicability of descriptors calculated for the set of drug compounds, to predict properties of biological importance such as human intestinal absorption and blood/brain distribution.

Introduction

The use of properties that are easy to measure in order to calculate or estimate properties that are difficult to measure is a well-known method in most fields of chemistry and biochemistry. With the advance of modern techniques such as combinatorial chemistry, high throughput screening has become very important and therefore estimates of physical properties from structure by calculations that can be performed rapidly are of primary importance. A large number of transport-related processes involve either the equilibrium transfer (K or P) or the rate of transfer (k) of a solute from one phase to another. Since log K (or log P) and log k are free-energy related, Abraham and co-workers formulated a number of solute properties or descriptors that are also free-energy related and could be used for the correlation of log K and log k values.

The original work of Kamlet and Taft and co-workers[1,2] showed that it was indeed possible to define a rather small number of descriptors that could be combined in a linear way for the correlation of solute properties. After considerable preliminary work,[3,4] Abraham and co-workers, succeeded in constructing a new and more rigorous set of five solute descriptors,[5–10] specified as follows. $\boldsymbol{E}$ is an excess molar refraction that is obtained from refractive index for solutes that are liquid at 20 °C. For solids, the refractive index of the hypothetical liquid at 20 °C can be calculated, or $\boldsymbol{E}$ can be obtained by the summation of fragments or substructures. $\boldsymbol{S}$ is the dipolarity/polarizability that can be obtained from gas liquid chromatographic measurements on polar stationary phases, or more generally from water–solvent partition coefficients. $\boldsymbol{A}$ and $\boldsymbol{B}$ are the overall or effective hydrogen bond acidity and basicity that are most easily obtained from water–solvent partitions, and $\boldsymbol{V}$ is the McGowan characteristic volume[11] that can easily be calculated from bond and atom contributions.[8] The range of solutes for which descriptors are currently available is now quite large, and encompasses compounds as far as helium, hydrogen, nitrogen, *etc.* on one hand and drugs, environmental pollutants and pesticides on the other.

Table 1 Table of the available solute descriptors

Descriptor	Maximum value	Minimum value	Total
$\boldsymbol{E}$	4.62	−1.37	4290
$\boldsymbol{S}$	5.60	−0.54	3760
$\boldsymbol{A}$	4.33	0.00	4490
$\boldsymbol{B}$	4.52	0.00	3440
$\boldsymbol{V}$	8.56	0.07	4380

These solute descriptors can be combined in a linear free energy relationship [eqn. (1)]. The dependent variable, log SP, is

$$\log SP = c + e\boldsymbol{E} + s\boldsymbol{S} + a\boldsymbol{A} + b\boldsymbol{B} + v\boldsymbol{V} \qquad (1)$$

a solute property in a given system. For example, it might be log P, for a set of solutes in a given water–solvent partition system. The coefficients in the equations are found by the method of multiple linear regression.

Descriptors for a large number of solutes have been obtained from experimental data; the maximum and minimum ranges of these descriptors in our database are shown in Table 1. The solute descriptors represent the solute influence on various solute–solvent phase interactions. Hence, the regression coefficients c, e, s, a, b and v correspond to the complementary effect of the phases on these interactions. The coefficients can then be regarded as system constants which characterize the phase and contain chemical information about the phase in question.

† *Present address*: AstraZeneca Pharmaceuticals, Mereside, Alderley Park, Macclesfield, Cheshire, UK SK10 4TG.

DOI: 10.1039/b110143a

An example to illustrate the chemical information contained in the system constants is partition of solutes between two phases; the system constants will reflect differences in properties of the two phases, and hence can take positive or negative values. The important water–octanol[12] system is characterized by eqn. (2).

$$\log P_{\text{oct}} = 0.088 + 0.562\,\boldsymbol{E} - 1.054\,\boldsymbol{S} + 0.034\,\boldsymbol{A} - 3.460\,\boldsymbol{B} + 3.814\,\boldsymbol{V} \quad (2)$$

$$(n = 613,\ r = 0.9974,\ \text{SD} = 0.116,\ F = 23161.6)$$

Thus, octanol (actually, wet octanol) is revealed to be able to interact with π- and n-electron pairs more than is water (positive *e*-coefficient), but is less dipolar/polarizable than water (hence the negative *s*-coefficient). Octanol is as strong a hydrogen-bond base as is water (almost zero *a*-coefficient), but is a weaker hydrogen-bond acid (negative *b*-coefficient). The large *v*-coefficient means that octanol is able to interact with solutes by dispersion forces and/or that the energy required to create a given sized cavity in octanol is relatively low.

Any application of the general solvation equation [eqn. (1)] depends on the availability of the solute descriptors, and the need to calculate descriptors for new compounds will always be of primary importance. As explained earlier, the descriptor ***V*** can be calculated quite simply for any structure from the molecular formula and the number of rings in the molecule, using the algorithm of Abraham for the number of bonds in the molecule.[8] The ***E*** descriptor can be calculated from the refractive index at 20 °C, using either the observed refractive index for a liquid, or a calculated refractive index for the liquid. This descriptor can also be estimated by the addition of fragment values (substructures). The remaining three descriptors ***S***, ***A*** and ***B*** have to be obtained by analogy to other compounds (within a homologous series for example), by fragment addition[13,14] and by experimental measurements of physicochemical properties such as log *P* values in a number of water–solvent systems.

In order to obtain reliable descriptors from log *P* values it is necessary to have at least three systems as different as possible. The difference in the physical properties of solvent systems is reflected in the coefficients obtained for each solvation equation as described earlier. Practical considerations are also of great importance. Such considerations include toxicity, availability, viscosity and volatility of each solvent system. Apart from the physical considerations however, a method for categorising the equations is needed. Ishihama *et al.*[15] proposed a vector (v) methodology that is defined for a particular solvation equation, [eqn. (1)], as follows. Let $v_i = (e_i, s_i, a_i, b_i, v_i)$. Then the analogy between any two or more given solvation equations SP_i and SP_j is expressed as $\cos\theta_{ij}$ between v_i and v_j as follows:

$$\cos\theta_{ij} = \frac{\varpi_i \varpi_j}{|\varpi_i|\,|\varpi_j|} = \frac{e_i e_j + s_i s_j + a_i a_j + b_i b_j + v_i v_j}{\sqrt{e_i^2 + s_i^2 + a_i^2 + b_i^2 + v_i^2}\sqrt{e_j^2 + s_j^2 + a_j^2 + b_j^2 + v_j^2}}$$

As the linear correlation between SP_i and SP_j becomes better, the value of cos θ becomes closer to 1. An Excel macro program was written to automatically calculate cos θ and θ and display the results in a matrix. The larger the θ value, the less correlation there is between SP_i and SP_j. Identical equations would give a cos θ value of 1 (Table 2). The four systems chosen for use in the measurement of partition coefficients were octanol, chloroform, cyclohexane and toluene, which fulfil the criteria set at the beginning of this work, and which have large θ values between pairs of solvents.

Table 2 The cos θ and θ matrix for the four solvent equations

cos θ	Octanol	Chloroform	Cyclohexane	Toluene
Octanol	1.000	0.848	0.874	0.902
Chloroform	0.848	1.000	0.979	0.986
Cyclohexane	0.874	0.979	1.000	0.990
Toluene	0.902	0.986	0.990	1.000
θ	Octanol	Chloroform	Cyclohexane	Toluene
Octanol	0.00	31.98	29.02	25.55
Chloroform	31.98	0.00	11.85	9.49
Cyclohexane	29.02	11.85	0.00	8.21
Toluene	25.55	9.49	8.21	0.00

Methods for the calculation of descriptors

It is relatively easy to set out various mathematical procedures for the calculation of three descriptors from four values of log *P*. It is however not so easy to compare the results of the mathematical procedures with each other in order to ascertain which method leads to the most accurate descriptors because we have no independent knowledge of what are the ‘true’ descriptors. We therefore took a set of 47 compounds for which descriptors had already been calculated from a variety of equations; these included equations for chromatographic data, as well as for a very large number of log *P* values. The descriptors obtained in this way, we refer to as the ‘database’ descriptors. We then applied our four mathematical methods to the particular four sets of log *P* values, so that we could compare the results with the ‘database’ values.

In a second procedure, we chose a test set of 13 drug compounds for which the four log *P* values were known,[16] or for which we had determined the log *P* values experimentally. Once again, we calculated four sets of descriptors by our mathematical methods. However, in order to ascertain which method, or methods, gave the most reasonable set of descriptors, we applied the calculated descriptors to the estimation of other log *P* values and biological processes that were available, and which had not been used in the calculation. The mathematical methods are set out below.

Solver. Solver is a tool in Microsoft Excel which can be used to determine the maximum or minimum value of one cell by changing other cells. Solver minimises the sum of squares on the required equations to fit the targeted cells ***S***, ***A*** and ***B*** and the values are accepted when the overall sums of squares are at a minimum. Solver uses the generalised reduced gradient (GRG2) nonlinear optimisation code developed by Leon Lasdon, University of Texas at Austin, and Allan Waren, Cleveland State University.

TripleX. If three equations are available, then three simultaneous equations can be constructed and solved for the three unknowns ***S***, ***A*** and ***B***. The TripleX program takes all combinations of the three equations from a series of solvent–water systems to calculate ***S***, ***A*** and ***B*** for each combination. The program then statistically obtains a more accurate result of ***S***, ***A*** and ***B*** than for any one combination. The five-parameter equation, eqn. (1), is reduced to a three-parameter equation by re-arranging terms to give

$$\log SP - \boldsymbol{E}e - \boldsymbol{V}v = \boldsymbol{S}s + \boldsymbol{A}a + \boldsymbol{B}b \qquad (3)$$

This is equivalent to

$$X_n = \boldsymbol{S}_n s_n + \boldsymbol{A}_n a_n + \boldsymbol{B}_n b_n \qquad (4)$$

The program has been modified to work with up to seven Abraham equations according to the needs of the user. In the present work we use a version that calculates results for ***A***, ***B*** and ***S*** based on four log *P* values and the four combinations that arise from these equations. The program utilizes matrices and a Gauss–Jordan routine for the solution of simultaneous equations.[17] TripleX was developed in Visual Basic for Applications.

Descfit SIMPLEX minimization method. Descfit has been developed to determine the three unknown descriptors, namely, ***A***, ***B*** and ***S*** for a particular solute by using three or more experimentally measured solvation properties (log *SP*) in conjunction with the solvation equations of various solvent systems derived by the Abraham group. Descfit assumes ***E*** and ***V*** are known parameters. The program uses a well known procedure namely the SIMPLEX[18] method, and treats the unknown descriptors as adjustable parameters and minimizes the root-mean-square-difference (RMSD) between log SP_{exp} and log SP_{cal} as defined below:

$$\text{RMSD} = \sqrt{\frac{\sum_{i=1}^{neqs}(\log SP_{\text{calc}}(i) - \log SP_{\text{exp}}(i))^2}{neqs}}$$

where n_{eqs} represents the number of log SP_{exp} values (*i.e.* number of solvation equations). Note that n_{eqs} must be greater than or equal to the number of adjustable parameters. To increase the reliability of the calculation, it is preferable to maintain an 'over-determined' condition by using a larger number of log SP_{exp} than the number of adjustable parameters. An added feature in Descfit is that it allows the user to fix any one or two of the three adjustable parameters in the optimization calculation. This may be useful if any of the descriptor(s) are readily available or can be obtained independently.

Regressions for obtaining descriptors. The fourth method of obtaining the three descriptors ***A***, ***B*** and ***S*** uses regression equations obtained from the 47 compound 'database' training set on the lines of eqn. (5). The training set chosen for this purpose is shown in Table 3 and includes compounds with a satisfactory range of descriptors.

$$\text{Descriptor} = w\log P_{\text{oct}} + k\log P_{\text{chl}} + q\log P_{\text{cycl}} + x\log P_{\text{tol}} + e\boldsymbol{E} + v\boldsymbol{V} \quad (5)$$

Using the method of multiple linear regression three equations were obtained of the same form as eqn. (5):

$$\boldsymbol{S} = 0.049 - 0.092\log P_{\text{oct}} + 0.229\log P_{\text{chl}} - 0.713\log P_{\text{cyc}} + 0.625\log P_{\text{tol}} + 0.355\boldsymbol{E} - 0.188\boldsymbol{V} \quad (6)$$

$$(n = 47, r^2 = 0.916, \text{SE} = 0.152, F = 73.054)$$

$$\boldsymbol{A} = 0.108 + 0.261\log P_{\text{oct}} - 0.155\log P_{\text{chl}} - 0.248\log P_{\text{cyc}} + 0.171\log P_{\text{tol}} - 0.049\boldsymbol{E} - 0.097\boldsymbol{V} \quad (7)$$

$$(n = 47, r^2 = 0.964, \text{SE} = 0.058, F = 177.194)$$

$$\boldsymbol{B} = -0.089 - 0.033\log P_{\text{oct}} + 0.338\log P_{\text{chl}} + 0.178\log P_{\text{cyc}} - 0.587\log P_{\text{tol}} + 0.137\boldsymbol{E} + 0.595\boldsymbol{V} \quad (8)$$

$$(n = 47, r^2 = 0.881, \text{SE} = 0.137, F = 49.187)$$

Here, and elsewhere, *n* is the number of data points, *r* is the correlation coefficient, SE is the standard error in the dependent variable and *F* is the Fisher *F*-statistic.

The construction of a suitable training set of compounds for use in the prediction of descriptors of drug compounds dictates consideration of the 'descriptor space' the training set covers. Because of the nature of drug compounds which are usually large molecules with extensive hydrogen bonding properties and therefore large ***A*** and ***B*** as well as large ***S*** values, the training set must cover a large 'descriptor space'. It is only within or slightly outside this space that the above equations can be considered to be valid and can be used to predict values of ***S***, ***A*** and ***B***. A good visual way of looking at the range of descriptors the training set covers is by plotting histograms of the values of the descriptors in the desired range. Three such histograms are in Fig. 1, showing the frequency of descriptors in the training set chosen.

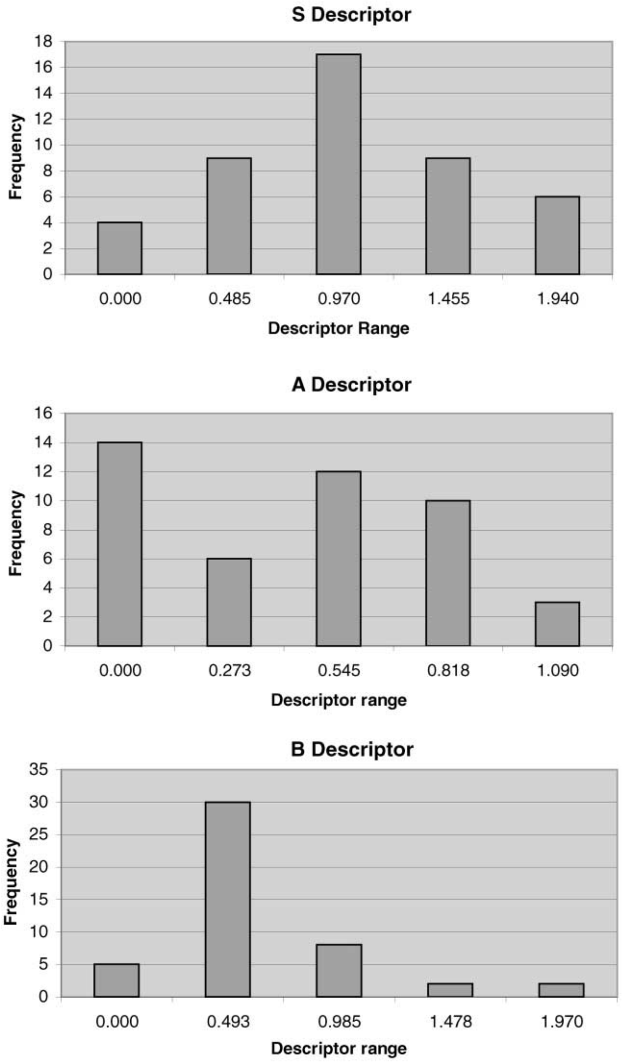

Fig. 1 Histograms of the descriptors in the 'database' training set showing the range covered by each descriptor.

By comparison to the ranges of the descriptors quoted in Table 1 the ranges shown in the histograms of Fig. 1 look rather small. However, apart from ***A*** where the range is indeed small (0–1.09) and the distribution is poor with a disproportionate number of compounds with a zero value, the range of ***S*** and ***B*** is perfectly adequate for this work. The range and distribution of ***S*** (0–1.94) is good. Finally for ***B*** the range is good (0–1.97) but the distribution is not ideal; more compounds with large ***B*** values are needed.

Experimental

Measurements of log *P* values for the test set consisting of drug compounds were measured using a GlpKa instrument utilising potentiometric methods developed by Sirius Analytical.[19] This instrument is designed specifically to determine ionisation constants (pK_a) and partition coefficients (log *P*) of weak proton acids and bases in various water–solvent systems. The

Table 3 The training set of 47 compounds used for obtaining the regression equations

	Name	log P_{oct}	log P_{chl}	log P_{cycl}	log P_{tol}	***S***	***A***	***B***	***E***	***V***
1	Krypton	0.89	1.22	1.24	1.10	0.00	0.00	0.00	0.000	0.2460
2	Xenon	1.28	1.50	1.64	1.50	0.00	0.00	0.00	0.000	0.3290
3	Hydrogen	0.45	0.54	0.69	0.58	0.00	0.00	0.00	0.000	0.1086
4	Nitrogen	0.67	0.93	1.04	0.89	0.00	0.00	0.00	0.000	0.2222
5	Propanone	−0.24	0.50	−0.96	−0.21	0.70	0.04	0.49	0.393	0.7918
6	Butanone	0.29	1.15	−0.25	0.40	0.70	0.00	0.51	0.166	0.6879
7	Methanol	−0.77	−1.33	−2.49	−1.92	0.44	0.43	0.47	0.278	0.3082
8	Ethanol	−0.31	−0.87	−1.89	−1.46	0.42	0.37	0.48	0.246	0.4491
9	Propanol	0.25	−0.30	−1.49	−0.82	0.42	0.37	0.48	0.236	0.5900
10	Butan-1-ol	0.84	0.42	−0.87	−0.30	0.42	0.37	0.48	0.224	0.7309
11	Pentanol	1.56	1.05	−0.26	0.51	0.42	0.37	0.48	0.219	0.8718
12	Benzoic acid	1.87	0.60	−0.85	0.36	0.90	0.59	0.40	0.730	0.9317
13	Phenol	1.47	0.32	−0.93	0.22	0.89	0.60	0.30	0.805	0.7751
14	*m*-Chlorophenol	2.50	1.02	−0.12	1.05	1.06	0.69	0.15	0.909	0.8975
15	*p*-Chlorophenol	2.39	1.07	−0.35	1.08	1.08	0.67	0.20	0.915	0.8975
16	Triethylamine	1.44	1.86	1.10	1.09	0.15	0.00	0.79	0.101	1.0538
17	Propylamine	0.47	0.25	−0.98	−0.64	0.35	0.16	0.61	0.225	0.6311
18	Pentanoic acid	1.39	0.32	−1.10	−0.20	0.60	0.60	0.45	0.205	0.8875
19	*p*-Toluidine	1.39	1.95	0.56	1.35	0.95	0.23	0.45	0.923	0.9571
20	Ethyl acetate	0.73	1.82	0.34	0.96	0.62	0.00	0.45	0.106	0.7466
21	Aniline	0.90	1.35	0.05	0.89	0.96	0.26	0.41	0.955	0.8162
22	Resorcinol	0.80	−1.34	−3.79	−2.17	1.11	1.09	0.52	0.980	0.8338
23	*o*-Nitroaniline	1.85	1.83	0.36	1.65	1.37	0.30	0.36	1.180	0.9904
24	*m*-Nitroaniline	1.37	1.60	−0.42	1.19	1.71	0.40	0.35	1.200	0.9904
25	*p*-Nitroaniline	1.39	1.26	−1.00	0.78	1.83	0.45	0.38	1.220	0.9904
26	Methyl acetate	0.18	1.16	−0.19	0.36	0.64	0.00	0.45	0.142	0.6057
27	Pyridine	0.65	1.29	−0.31	0.29	0.84	0.00	0.52	0.631	0.6753
28	*o*-Nitrophenol	1.85	2.53	1.45	2.26	1.05	0.05	0.37	1.015	0.9493
29	*m*-Nitrophenol	2.00	0.50	−1.51	0.34	1.57	0.79	0.23	1.050	0.9493
30	*p*-Nitrophenol	1.91	0.20	−2.05	−0.19	1.72	0.82	0.26	1.070	0.9493
31	*o*-Methoxyphenol	1.32	1.70	0.47	1.26	0.91	0.22	0.52	0.837	0.9747
32	Benzene	2.13	2.76	2.35	2.64	0.52	0.00	0.14	0.610	0.7164
33	Toluene	2.73	3.41	2.99	3.14	0.52	0.00	0.14	0.601	0.8573
34	Hexanol	2.03	1.69	0.45	1.29	0.42	0.37	0.48	0.210	1.0127
35	Diethyl ether	0.89	1.88	0.93	1.3	0.25	0.00	0.45	0.041	0.7309
36	2-Naphthol	2.70	1.74	0.29	1.68	1.08	0.61	0.40	1.520	1.1441
37	Salicylic acid	2.26	0.64	−0.50	0.39	0.84	0.71	0.38	0.890	0.99
38	Phenylacetic acid	1	0.57	−1.23	0.09	0.97	0.60	0.61	0.730	1.07
39	Atropine	1.83	2.44	−1.02	0.77[a]	1.94	0.36	1.64	1.188	2.282
40	Aspirin	0.9	0.63	−2.00[a]	−0.49	0.80	0.49	1.00	0.930	1.2879
41	Nicotine	1.17	1.89	0.36	0.86	0.75	0.00	1.14	0.865	1.371
42	1-Naphthylamine	2.25	2.6	1.26	2.43	1.26	0.2	0.57	1.670	1.185
43	1-Naphthol	2.84	1.50	0.58	1.80	1.05	0.6	0.37	1.520	1.1441
44	Ephedrine	1.13	1.10	−0.44	0.40	0.76	0.21	0.21	0.919	1.4385
45	Cyclohexane	3.44	4.16	4.15	3.96	0.1	0.00	0.00	0.305	0.8454
46	*o*-Chlorophenol	2.15	1.36	0.87	1.37	0.88	0.32	0.31	0.853	0.8975
47	Quinine	3.47	2.29[a]	0.04[a]	1.01[a]	1.23	0.37	1.97	2.469	2.5512

[a] Measured in this work.

determination of log *P* by potentiometry has been covered extensively in the literature[20,21] although the technique itself is still undergoing active development. The potentiometric measurements have all been carried out at Sirius Analytical Ltd.

The excess molar refraction descriptor, ***E***, for all the compounds in the test set was estimated by addition of fragment values and the McGowan volume, ***V***, was calculated from structure as explained earlier. The three remaining descriptors were calculated using Solver, TripleX, Descfit and the regression equations by using measured log *P* values obtained by the potentiometric method.

Results

As a first step in the assessment of the reliability of the descriptor calculations, descriptors from the four water–solvent systems (Table 3) chosen at the beginning of this work (octanol, chloroform, cyclohexane and toluene), were obtained (Table 4). The agreement between the descriptors calculated using the four different methods was quite reasonable. Table 5 tabulates the standard deviations between values of ***S*** obtained from as many sources of data as possible, that is the ‘data base set’, and the four calculated values; SD values are also given for the corresponding ***A*** and ***B*** values. The results show that there is good agreement between the four sets of calculation, and that they yield ***S***, ***A*** and ***B*** values in reasonable agreement with the data base set. This kind of comparison gives a good indication of what the maximum accuracy of these methods could be. It also provides strong evidence of self-consistency within our database training set. The results show that Regressions, Descfit, TripleX and Solver is the order of accuracy by comparison to the ‘data base set’.

A more detailed analysis of the descriptor calculations on the 13 drugs in the test set has been carried out in order to establish how successful is each of the calculation methods, and, more importantly, to show that reasonable descriptors which can be applied in predicting physicochemical and biological properties can be obtained from a set of only four partition coefficients measured in different systems (Table 6).

We first consider the two methods Solver and Descfit. These two programs are categorised together because of their similarities as far as the calculation of the descriptors. In both programs there is a minimisation process on a certain function. Usually this function is the one calculating the standard error for the descriptors to fit the particular Abraham solvation equations used. The closeness of the results (Table 7) obtained

Table 4 Calculation of the descriptors of the 47-compound training set using the four methods[a]

	Name	***S***					***A***					***B***				
		DB	Solv.	TX	Descfit	Reg.	DB	Solv.	TX	Descfit	Reg.	DB	Solv.	TX	Descfit	Reg.
1	Krypton	0.00	−0.13	−0.13	−0.13	0.00	0.00	−0.03	−0.02	−0.03	0.01	0.00	0.07	0.07	0.07	0.02
2	Xenon	0.00	−0.17	−0.18	−0.17	−0.02	0.00	0.00	0.02	0.00	0.03	0.00	0.06	0.05	0.06	−0.02
3	Hydrogen	0.00	−0.19	−0.22	−0.19	−0.02	0.00	0.00	0.04	0.00	0.06	0.00	0.06	0.03	0.06	−0.07
4	Nitrogen	0.00	−0.20	−0.22	−0.20	−0.03	0.00	−0.04	0.00	−0.04	0.01	0.00	0.12	0.10	0.12	0.00
5	Propanone	0.70	0.68	0.63	0.68	0.73	0.04	0.00	0.08	0.00	0.07	0.49	0.80	0.75	0.80	0.57
6	Butanone	0.70	0.70	0.67	0.67	0.64	0.00	−0.01	0.02	0.00	0.06	0.51	0.65	0.51	0.52	0.44
7	Methanol	0.44	0.34	0.32	0.34	0.43	0.43	0.31	0.34	0.31	0.36	0.47	0.52	0.50	0.52	0.39
8	Ethanol	0.42	0.19	0.17	0.19	0.32	0.37	0.30	0.33	0.3	0.33	0.48	0.59	0.56	0.59	0.45
9	Propanol	0.42	0.41	0.39	0.41	0.48	0.37	0.36	0.39	0.36	0.38	0.48	0.51	0.49	0.51	0.40
10	Butan-1-ol	0.42	0.41	0.42	0.41	0.44	0.37	0.36	0.34	0.36	0.34	0.48	0.51	0.52	0.51	0.51
11	Pentanol	0.42	0.03	0.57	0.57	0.56	0.37	0.39	0.42	0.43	0.41	0.48	0.62	0.42	0.41	0.42
12	Benzoic acid	0.9	0.94	0.94	0.94	0.93	0.59	0.68	0.68	0.68	0.65	0.40	0.35	0.35	0.35	0.34
13	Phenol	0.89	0.91	0.90	0.91	0.93	0.60	0.60	0.62	0.6	0.60	0.30	0.30	0.29	0.30	0.25
14	*m*-Chlorophenol	1.06	0.96	0.96	0.96	0.95	0.69	0.73	0.72	0.73	0.68	0.15	0.16	0.16	0.16	0.20
15	*p*-Chlorophenol	1.08	1.23	1.24	1.23	1.15	0.67	0.73	0.72	0.73	0.70	0.20	0.11	0.11	0.11	0.16
16	Triethylamine	0.15	−0.08	−0.10	−0.08	0.08	0.00	0.03	0.05	0.03	0.00	0.79	0.81	0.80	0.81	0.69
17	Propylamine	0.35	0.28	0.30	0.28	0.32	0.16	0.27	0.23	0.27	0.25	0.61	0.55	0.58	0.55	0.59
18	Pentanoic acid	0.60	0.53	0.53	0.53	0.56	0.6	0.60	0.60	0.60	0.56	0.45	0.48	0.48	0.48	0.45
19	*p*-Toluidine	0.95	1.00	1.01	1.00	0.96	0.23	0.11	0.10	0.11	0.12	0.45	0.53	0.54	0.53	0.53
20	Ethyl acetate	0.62	0.74	0.76	0.74	0.65	0.00	−0.02	−0.05	−0.02	0.02	0.45	0.44	0.45	0.44	0.46
21	Aniline	0.96	1.00	0.98	1.00	0.98	0.26	0.11	0.13	0.11	0.15	0.41	0.51	0.50	0.51	0.44
22	Resorcinol	1.11	1.32	1.36	1.32	1.21	1.09	1.02	0.95	1.02	0.96	0.52	0.50	0.55	0.50	0.66
23	*o*-Nitroaniline	1.37	1.38	1.37	1.38	1.31	0.30	0.33	0.35	0.33	0.35	0.36	0.35	0.34	0.35	0.32
24	*m*-Nitroaniline	1.71	1.75	1.74	1.75	1.57	0.40	0.33	0.33	0.33	0.37	0.35	0.38	0.38	0.38	0.39
25	*p*-Nitroaniline	1.83	1.87	1.88	1.87	1.66	0.45	0.47	0.46	0.47	0.50	0.38	0.35	0.36	0.35	0.41
26	Methyl acetate	0.64	0.63	0.64	0.63	0.60	0.00	−0.04	−0.04	−0.04	0.02	0.45	0.47	0.47	0.47	0.43
27	Pyridine	0.84	0.75	0.93	0.88	0.78	0.00	0.12	0.01	0.10	0.11	0.52	0.33	0.49	0.44	0.59
28	2-Nitrophenol	1.05	1.05	1.04	1.05	1.02	0.05	0.06	0.07	0.06	0.08	0.37	0.38	0.37	0.38	0.34
29	3-Nitrophenol	1.57	1.62	1.62	1.62	1.46	0.79	0.86	0.84	0.86	0.84	0.23	0.19	0.19	0.19	0.26
30	4-Nitrophenol	1.72	1.65	1.68	1.65	1.46	0.82	0.94	0.89	0.94	0.91	0.26	0.22	0.25	0.22	0.37
31	2-Methoxyphenol	0.91	0.87	0.85	0.87	0.88	0.22	0.13	0.17	0.13	0.15	0.52	0.58	0.56	0.58	0.48
32	Benzene	0.52	0.51	0.52	0.51	0.54	0.00	0.00	−0.01	0.00	0.00	0.14	0.15	0.15	0.15	0.15
33	Toluene	0.52	0.45	0.48	0.45	0.46	0.00	0.00	−0.05	0.00	−0.03	0.14	0.16	0.19	0.16	0.26
34	Hexanol	0.42	0.62	0.62	0.62	0.62	0.37	0.40	0.41	0.40	0.38	0.48	0.40	0.40	0.40	0.37
35	Diethyl ether	0.25	0.66	0.39	0.40	0.42	0.00	−0.02	−0.05	−0.08	−0.03	0.45	0.37	0.44	0.45	0.36
36	2-Naphthol	1.08	1.42	1.43	1.42	1.37	0.61	0.61	0.61	0.61	0.57	0.4	0.33	0.33	0.33	0.37
37	Salicylic acid	0.84	0.65	0.66	0.65	0.72	0.71	0.72	0.71	0.72	0.65	0.38	0.42	0.43	0.42	0.45
38	Phenylacetic acid	0.97	1.07	1.05	1.07	1.04	0.6	0.58	0.61	0.58	0.57	0.61	0.59	0.58	0.59	0.52
39	Atropine	1.94	1.78	1.77	1.78	1.64	0.36	0.37	0.39	0.37	0.31	1.64	1.66	1.65	1.66	1.56
40	Aspirin	0.80	1.44	1.45	1.44	1.32	0.49	0.50	0.48	0.50	0.49	1.00	0.91	0.92	0.91	0.92
41	Nicotine	0.75	0.62	0.60	0.62	0.70	0.00	0.03	0.05	0.03	0.00	1.14	1.14	1.13	1.14	1.01
42	1-Naphthylamine	1.26	1.48	1.46	1.48	1.43	0.20	0.19	0.21	0.19	0.20	0.57	0.50	0.48	0.50	0.45
43	1-Naphthol	1.05	1.10	1.06	1.10	1.17	0.60	−0.08	0.69	0.64	0.59	0.37	0.94	0.33	0.37	0.26
44	Ephedrine	0.76	0.67	0.61	0.67	0.82	0.21	0.03	0.34	0.24	0.22	0.21	1.14	1.14	1.2	0.92
45	Cyclohexane	0.1	0.11	0.16	0.11	0.15	0.00	−0.03	−0.11	−0.03	−0.09	0.00	0.00	0.05	0.00	0.16
46	*o*-Chlorophenol	0.88	0.38	0.36	0.38	0.53	0.32	0.38	0.42	0.38	0.35	0.31	0.41	0.38	0.41	0.30
47	Quinine	1.23	1.01	1.07	1.07	1.25	0.37	0.67	0.66	0.66	0.45	1.97	1.91	1.92	1.92	1.84

[a] DB: calculated using all available literature values. Solv.: Excel Solver. TX: TripleX program. Descfit: SIMPLEX minimization method. Reg.: Regression equation.

Table 5 Standard deviations of different methods of calculation in comparison to descriptors for the 47 compounds' 'database' set

	S	***A***	***B***
Solver	0.168	0.072	0.169
TX	0.159	0.071	0.146
Descfit	0.154	0.067	0.156
Regression	0.152	0.058	0.137

using these two methods is evidence of similarity in the calculation process. Solver has a disadvantage over Descfit, however, because it tends to get more easily stuck in local minima if the starting value is far away from the optimised value. An added feature of Descfit makes it more appealing because batches of compounds can be run, in contrast to Solver that deals with one compound at a time. In both programs the user can change the combination of solvent systems used, to obtain the best combinations of descriptors. TripleX is a simultaneous equation solver as described earlier. It can deal with many solvent systems at one time and takes batches of compounds for calculation. An added advantage of this method is that the user can distinguish between good and bad measurements by inspection. TripleX appears to be a reasonably good method for dealing with the calculation of descriptors. The regression equation method that uses equations obtained from the 'data base' training set of compounds in Table 3 is a completely different way of obtaining descriptors. Using a regression equation such as eqn. (5) provides an estimate of each descriptor independently from the rest. In some cases this can be an added advantage.

A back calculation of partition values in cyclohexane using the descriptors calculated using all methods provides a useful internal verification method for the whole process. The cyclohexane solvation equation has been chosen for this purpose because it has the largest coefficients compared to other systems and therefore the deviations in log *P* using the calculated descriptors would be at their maximum in the case where descriptor deviations occur. These values (Table 8) along with standard errors, based on the results of each method compared

Table 6 Drug compounds (test set) for which log P values have been experimentally measured

	Name	log P_{oct}	log P_{chl}	log P_{cycl}	log P_{tol}	E	V
1	Propranolol	3.480[a]	1.030[b]	−0.640[b]	2.230[b]	1.85	2.148
2	Tetracaine	3.510[a]	2.90[b]	2.045[a]	3.399[a]	1.12	2.2585
3	Papaverine	2.950[b]	4.280[b]	2.560[b]	3.000[b]	2.19	2.5914
4	Tryptamine	2.147[a]	1.526[a]	−0.599[a]	0.268[a]	1.53	1.328
5	Diclofenac	4.510[a]	2.965[a]	1.880[a]	2.960[a]	1.97	2.025
6	Chlorpromazine	5.400[a]	6.211[a]	5.240[a]	6.094[a]	2.44	2.4056
7	Ibuprofen	3.970[a]	3.025[a]	1.877[a]	2.484[a]	0.86	1.7771
8	Lidocaine	2.440[a]	4.080[a]	1.226[a]	2.120[a]	1.23	2.0589
9	Deprenyl	2.900[b]	4.292[a]	2.811[a]	3.492[a]	1.05	1.7165
10	Desipramine	4.214[a]	5.330[a]	3.379[a]	4.104[a]	1.99	2.2606
11	Fluoxetine	3.749[a]	5.483[a]	3.622[a]	4.670[a]	1.24	2.2403
12	Procaine	2.140[b]	2.130[b]	−0.130[b]	1.806[a]	1.135	1.9767
13	Miconazole	5.344[a]	5.421[a]	3.687[a]	5.710[a]	2.37	2.7229

[a] Measured in this work. [b] Obtained from Medchem 2001.[16]

Table 7 Descriptors obtained for the test set of compounds[a]

	Name	S					A					B				
		DB	Solv.	TX	Descfit	Reg.	DB	Solv.	TX	Descfit	Reg.	DB	Solv.	TX	Descfit	Reg.
1	Propranolol	1.43	1.91	1.67	1.89	2.07	0.44	0.92	1.45	1.09	1.10	1.31	1.13	0.78	1.01	0.25
2	Tetracaine	0.92	0.70	0.51	0.69	1.03	0.34	0.40	0.71	0.40	0.37	1.33	1.37	1.18	1.37	0.64
3	Papaverine	0.93	0.87	0.82	0.87	1.10	−0.10	−0.15	−0.07	−0.15	−0.27	2.04	2.09	2.04	2.09	1.80
4	Tryptamine	1.27	1.17	1.26	1.17	1.09	0.55	0.54	0.40	0.54	0.42	0.97	0.81	0.90	0.81	1.09
5	Diclofenac	1.58	0.97	0.94	0.97	1.14	0.90	0.72	0.78	0.72	0.57	0.83	0.97	0.94	0.97	0.84
6	Chlorpromazine	1.83	1.37	1.34	1.37	1.46	0.00	0.06	0.11	0.06	−0.06	0.94	1.08	1.05	1.08	0.95
7	Ibuprofen	0.97	0.45	0.46	0.44	0.56	0.60	0.57	0.56	0.57	0.42	0.70	0.85	0.86	0.85	0.85
8	Lidocaine	1.49	1.43	1.50	1.43	1.26	0.11	0.01	−0.11	0.01	−0.09	1.27	1.39	1.46	1.39	1.58
9	Deprenyl	1.00	1.01	1.00	1.01	0.99	0.00	−0.08	−0.07	−0.08	−0.12	0.94	0.94	0.94	0.94	0.88
10	Desipramine	1.64	1.38	1.43	1.38	1.32	0.10	0.07	−0.01	0.07	−0.07	0.92	1.23	1.29	1.23	1.38
11	Fluoxetine	1.33	1.36	1.33	1.36	1.31	0.08	−0.09	−0.04	−0.09	−0.14	1.06	1.18	1.15	1.18	1.05
12	Procaine	1.36	1.57	1.46	1.57	1.59	0.25	0.42	0.60	0.42	0.43	1.41	1.23	1.12	1.24	0.81
13	Miconazole	2.00	2.03	1.92	2.03	2.07	0.00	0.42	0.60	0.42	0.34	1.20	1.19	1.08	1.19	0.82

[a] DB: calculated using all available literature values. Solv.: Excel Solver. TX: TripleX program. Descfit: SIMPLEX minimisation method. Reg.: Regression equation.

Table 8 Water–cyclohexane partition coefficient, as log P_{cyc}, obtained from calculated descriptors; comparison with measured values

Name	DB	Solv.	TX	Descfit	Reg.	Measured log P_{cyc}
Propranolol	0.94	−0.78	−0.63	−0.79	2.62	−0.64
Tetracaine	2.00	1.95	2.05	1.97	5.08	2.045
Papaverine	2.50	2.54	2.57	2.54	4.03	2.56
Tryptamine	−1.53	−0.54	−0.61	−0.54	−1.35	−0.599
Diclofenac	0.86	1.87	1.84	1.87	2.79	1.88
Chlorpromazine	5.38	5.24	5.25	5.24	6.14	5.24
Ibuprofen	1.65	1.89	1.86	1.91	2.25	1.877
Lidocaine	1.38	1.26	1.24	1.26	0.99	1.226
Deprenyl	2.53	2.81	2.79	2.81	3.27	2.811
Desipramine	4.41	3.43	3.35	3.43	3.30	3.379
Fluoxetine	3.63	3.62	3.63	3.62	4.55	3.622
Procaine	−0.07	−0.17	−0.12	−0.22	1.83	−0.13
Miconazole	5.21	3.63	3.69	3.64	5.70	3.687
SD	0.72	0.06	0.02	0.06	1.38	

with the measured partitions indicate self-consistency within the model. The predictability of the calculated descriptors is best when TripleX or the Descfit method are used.

Measurements listed in the medicinal chemistry database[16] have served as an independent way to validate our descriptors. The solvent systems chosen were water–octanol, –benzene, –diethyl ether, –dibutyl ether, –ethyl acetate, –heptane and –carbon tetrachloride. Although octanol was one of the systems used in our descriptor calculations, we used the recommended log P^* values which have been measured independently. The rest of the solvent systems used are even better comparisons because these solvents were not used in the calculation of descriptors. Unfortunately only very few measurements exist on the 13 drugs in our study. The comparisons are shown in Table 9. Ionisation corrections have been applied where necessary. The predictive ability of our descriptors are reasonably good and the standard deviations observed between calculated and measured values are shown in Table 9. It has to be noted that the deviations reported incorporate the experimental error. From these results all four methods can be categorised according to their accuracy. It is evident that best predictions are obtained when descriptors are calculated from as many partition values as possible. It is however possible to obtain reasonable predictions from only four measured values. The

Table 9 Comparison of calculated log *P* values with independently measured values[16] in various water–solvent systems

Compound	Solvent	DB	Solver	TX	Desc.	Reg.	Observed
Propranolol	Octanol	3.27	3.37	4.81	3.80	6.23	2.98
	Benzene	2.38	1.45	1.56	1.49	4.86	2.49
	Diethyl ether	2.62	2.95	4.91	3.56	7.15	2.73
Tetracaine	Octanol	3.75	3.84	4.69	3.85	6.01	3.73
	Diethyl ether	3.14	3.17	4.30	3.18	6.46	3.04
	Dibutyl ether	3.24	3.32	4.35	3.33	6.67	2.76
Papaverine	Octanol	3.17	3.06	3.28	3.06	3.84	2.95
	Diethyl ether	1.72	1.54	1.84	1.54	2.77	1.85
Tryptamine	Octanol	1.30	1.96	1.56	1.96	1.07	1.35
	Benzene	−0.12	0.71	0.67	0.71	−0.18	1.07 (ic)[a]
	Ethyl acetate	2.38	2.36	2.33	2.36	2.76	1.41
Diclofenac	Octanol	4.35	4.52	4.65	4.52	4.80	4.40
Chlorpromazine	Octanol	5.45	5.45	5.59	5.45	5.80	5.35
	Benzene	6.65	6.09	6.09	6.09	6.98	5.34
	Hexane	4.59	4.50	4.52	4.50	5.37	4.80
	Heptane	4.67	4.75	4.79	4.75	5.55	4.47 (ic)[a]
	CCl_4	6.09	5.81	5.81	5.81	6.69	6.11
Ibuprofen	Octanol	3.88	3.91	3.87	3.93	3.79	3.50
Lidocaine	Octanol	2.66	2.31	2.00	2.31	1.85	2.26
	Heptane	1.06	0.94	0.87	0.94	0.74	0.99
Deprenyl	Octanol	2.92	2.91	2.92	2.91	3.13	2.90
Desipramine	Octanol	4.91	4.12	3.86	4.12	3.65	4.54
	Diethyl ether	4.96	3.69	3.35	3.69	3.01	4.13 (ic)[a]
Fluoxetine	Octanol	4.26	3.82	3.95	3.82	4.33	4.50
Procaine	Octanol	1.94	2.34	2.83	2.31	3.77	2.14
	Diethyl ether	1.04	1.69	2.34	1.64	3.77	1.80
	Dibutyl ether	0.89	1.38	1.98	1.33	3.55	0.83
Miconazole	Octanol	5.55	5.52	6.02	5.53	6.79	4.90
SD		0.41	0.46	0.76	0.49	1.13	

[a] (ic) = ion correction was applied.[20]

Table 10 Gastrointestinal absorption and blood brain barrier predictions

Name	DB	Solv.	TX	Descfit	Reg.	%Abs GI	log BB
Propranolol	89	84	79	83	100	99	—
Propranolol	0.44	−0.11	0.05	−0.10	0.29	—	0.64
Diclofenac	89	87	86	87	94	100	—
Ibuprofen	90	85	85	85	89	95	—
Desipramine	107	100	101	100	100	100	—
Desipramine	0.91	0.92	0.88	0.92	0.94	—	1.2

predicted log *P* values obtained from the database set of descriptors show a standard deviation of 0.41. Solver leads to an SD value of 0.46, and Descfit and TripleX to values of 0.49 and 0.76, respectively. Regression is the least accurate method with a standard deviation of 1.13.

It is unlikely however that one would measure four partition values for a drug compound to predict properties of a physicochemical nature. It is more likely that predictions of biological importance such as human intestinal absorbance (GI absorbance) and blood/brain distribution, would be the aim of such a project. We have predicted such values using recently reported Abraham equations for these processes.[22,23] These calculated values along with the experimental ones are shown in Table 10. By inspection of the results it can be seen that useful predictions can be made using the calculated descriptors but there were not enough data to draw any meaningful statistical comparisons of our four methods.

Conclusions

Comparisons of calculated descriptors with the 'data base' descriptors (Table 5) reveal little difference between the four calculational methods. However, the rather stringent test shown in Table 8 reveals that the Descfit and TripleX methods are much better mathematical methods than the other two. As regards the true test sets in Table 9, it is clear that the regression method performs very much worse than the other three. Overall, in terms of mathematical performance and ease of use, Descfit and TripleX methods are preferred. Both of these methods can be programmed to deal with larger numbers of systems, and can also be programmed to calculate descriptors for large numbers of compounds automatically. Although all these have been set out with water partition systems a similar procedure has been set up by Valko[24–26] and co-workers to use HPLC reverse phase chromatography to obtain the Abraham descriptors in a high throughput way. Valko and co-workers have set up their method by choosing the most orthogonal HPLC systems by non-linear mapping. A straightforward extension of the present work would be to apply the four mathematical methods we describe to the systems chosen by Valko and co-workers, and this is what we intend to do in co-operation with Valko and co-workers.

The programs used for producing Abraham descriptors (TripleX, Descfit) are available on request from the corresponding author.

Acknowledgements

We are very grateful to EPSRC for a post-doctoral fellowship to AMZ, and for a postgraduate studentship to MCB. We would also like to thank Klara Valko of GSK for useful discussions on the regression equations and also Mark England and Steve Pouros of Sirius Analytical Ltd for creating a user-friendly front phase for Descfit.

References

1 M. J. Kamlet, R. M. Doherty, J.-L. M. Abboud, M. H. Abraham and R. W. Taft, *CHEMTECH*, 1986, **16**, 566.

2 M. H. Abraham, R. M. Doherty, M. J. Kamlet and R. W. Taft, *Chem. Br.*, 1986, **22**, 551.

3 Hydrogen bonding, Part 13: M. H. Abraham, G. S. Whiting, R. M. Doherty and W. J. Shuely, *J. Chem. Soc., Perkin Trans. 2*, 1990, 1451.

4 Hydrogen bonding, Part 14: M. H. Abraham, G. S. Whiting, R. M. Doherty and W. J. Shuely, *J. Chem. Soc., Perkin Trans. 2*, 1990, 1851.

5 Hydrogen bonding, Part 16: M. H. Abraham, G. S. Whiting, R. M. Doherty and W. J. Shuely, *J. Chromatogr.*, 1991, **587**, 213.

6 Hydrogen bonding, Part 31: M. H. Abraham, *J. Phys. Org. Chem.*, 1993, **6**, 660.

7 M. H. Abraham, *Pure Appl. Chem.*, 1993, **65**, 2503.

8 M. H. Abraham, *Chem. Soc. Rev.*, 1993, **22**, 73.

9 M. H. Abraham, A. M. Zissimos and W. E. Acree, Jr, *Phys. Chem. Chem. Phys.*, 2001, **3**, 3732.

10 C. M. Du, K. Valko, C. Bevan, D. Reynolds and M. H. Abraham, *J. Liq. Chromatogr. Relat. Technol.*, 2001, **24**, 635.

11 M. H. Abraham and J. C. McGowan, *Chromatographia*, 1987, **23**, 243.

12 M. H. Abraham, H. S. Chadha, G. S. Whiting and R. C. Mitchell, *J. Pharm. Sci.*, 1994, **83**, 1085.

13 J. A. Platts, D. Butina, M. H. Abraham and A. Hersey, *J. Chem. Inf. Comput. Sci.*, 1999, **39**, 835.

14 J. A. Platts, D. Butina, M. H. Abraham and A. Hersey, *J. Chem. Inf. Comput. Sci.*, 2000, **40**, 71.

15 Y. Ishihama and N. Asakawa, *J. Pharm. Sci*, 1999, 2716.

16 A. J. Leo, MedChem 2001 database, version 4.71-2, BioByte Corp., P.O. 517, Claremont, CA 91711-0157.

17 A. F. Carley and P. H. Morgan, *Computational Methods in the Chemical Sciences*, Ellis Horwood, 1989.

18 J. A. Nelder and R. Mead, *Comput. J.*, 1965, **7**, 308.

19 Sirius Technical Application Notes (STAN)-vol. 1, 1995.

20 A. Avdeef, *Quant. Struct. Act. Relat.*, 1992, **11**, 510.

21 A. Avdeef, K. J. Box, J. E. A. Comer, C. Hibbert and K. Y. Tam, *Pharm. Res.*, 1997, **15**, 208.

22 Y. H. Zhao, J. Le, M. H. Abraham, A. Hersey, P. J. Eddershaw, C. N. Luscombe, D. Boutina, G. Beck, B. Sherborne, I. Cooper and J. A. Platts, *J. Pharm. Sci.*, 2001, **90**, 749.

23 J. A. Platts, M. H. Abraham, Y. H. Zhao, A. Hersey, L. Ijaz and D. Butina, *Eur. J. Med. Chem.*, 2001, **36**, 719.

24 K. Valko, S. Espinosa, C. M. Du, E. Bosch, M. Roses, C. Bevan and M. H. Abraham, *J. Chromatogr., A*, 2001, **933**, 73.

25 C. M. Du, K. Valko, C. Bevan, D. Reynolds and M. H. Abraham, *J. Chromatogr. Sci.*, 2000, **38**, 503.

26 K. Valko, M. Plass, C. Bevan, D. Reynolds and M. H. Abraham, *J. Chromatogr., A*, 1998, 797, 41.