Calculation of Abraham descriptors from experimental data from seven HPLC systems; evaluation of five different methods of calculation †

Andreas M. Zissimos,*[a] Michael H. Abraham,[a] Chau M. Du,[b] Klara Valko,[c] Chris Bevan,[c] Derek Reynolds,[d] John Wood[e] and Kin Y. Tam[f]

[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] *pION Inc., 5 Constitution Way, Woburn, MA 01801, USA*
[c] *Hit Generation Sciences, Analytical Technologies Unit, GlaxoSmithKline Medicines Research Centre, Stevenage, Herts., UK SG1 2NY*
[d] *REYTEK Ltd, 11C Rothsday Rd., Bedford, Bedfordshire, UK MK40 3PP*
[e] *Statistical Sciences, GlaxoSmithKline, NFSP, Third Ave, Harlow, Essex, UK CM19 5AN*
[f] *AstraZeneca Pharmaceuticals, Mereside, Alderley Park, Macclesfield, Cheshire, UK SK10 4TG*

Received (in Cambridge, UK) 17th July 2002, Accepted 8th October 2002
First published as an Advance Article on the web 4th November 2002

Solvation equations have been obtained for seven high performance liquid chromatographic (HPLC) systems, generated in the reverse phase (RP) mode with fast gradient elution. A training set of 40 compounds was used for each system. The seven equations were then used to calculate Abraham descriptors for a completely separate 40-compound test set. In this way the three descriptors dipolarity/polarizability S, hydrogen bond acidity A, and hydrogen bond basicity B were obtained. Five different procedures were used to calculate the descriptors, (i) Microsoft 'Solver', (ii) a program that uses a set of three simultaneous equations, and which we denote as 'TripleX', (iii) a program similar to Solver that we denote as 'Descfit', (iv) a series of regression equations developed from compounds with known descriptors and (v) a series of modified regression equations. We show that RP-HPLC data for a given compound in seven systems can be used to calculate the three Abraham descriptors reliably. We compare descriptors, and errors in the method, with those obtained from water–solvent partition systems.

Introduction

Most of the recent work in 'property-based design' to convert a lead or candidate molecule into a successful drug on the market, has highlighted the importance of molecular size and hydrogen bonding capacity as a way to understand 'drug-likeness'.[1] Various methods have been employed to predict biological properties that are difficult or costly to measure in the first stages of drug discovery, but in many multi-component approaches the descriptors used are poorly defined and/or are difficult to relate to specific structural elements of the properties of functional groups. The need to obtain well defined experimental molecular descriptors for drug compounds by fast and efficient processes to suit the industry remains of primary importance.

The Abraham method is one such method.[2–5] It is based on the solvation equation (1), which correlates solute properties (SP), such as partitioning,[6,7] solubility,[8] blood-brain distribution[9] and human intestinal absorption,[10] with a standard set of five molecular descriptors. Although descriptors E (Excess molar refraction) and V (McGowans volume) can be obtained from structure, the model's use is restricted due to the difficulty of measuring the molecular descriptors; S (dipolarity/polarisability), A (hydrogen bond acidity) and B (hydrogen-bond basicity). 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.

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

The experimental method often used to determine S, A and B is through the use of water/solvent partition (log P) measurements. The method has been recently reviewed,[11] and the errors involved have been evaluated. Although the results have proven to be satisfactory, the procedure itself remains relatively slow for industrial purposes due to lengthy sample preparation. In concurrent studies, Valko[12–14] and co-workers have used the fast gradient elution RP-HPLC in order to obtain Abraham descriptors in a much more rapid procedure. Valko and co-workers have set up their method by choosing the most orthogonal HPLC systems by non-linear mapping. As a straight forward extension of our earlier work[11] we apply the mathematical methods we have described to the systems chosen by Valko and co-workers.[12–14]

The descriptors for the 80 solutes used are shown in Table 1. These descriptors have been calculated from a variety of equations which included equations for chromatographic data, as well as for a very large number of log P values. We will refer to the descriptors obtained in this way as 'Table 1' descriptors. The RP-HPLC data used are in the form of CHI (chromatographic hydrophobicity index) values; these are derived from the experimental log k values using the standard procedure as described by Valko and co-workers.[12–14] In Table 2 we tabulate the CHI values for the 80 solutes on the seven systems. The

† Electronic supplementary information (ESI) available: Tables S1 to S5. See http://www.rsc.org/suppdata/p2/b2/b206927j/

DOI: 10.1039/b206927j

Table 1 Compounds and their molecular descriptors used in this study

No	Name	*E*	*S*	*A*	*B*	*V*
1	n-Heptanophenone	0.72	0.95	0	0.5	1.7184
2	n-Hexanophenone	0.719	0.95	0	0.5	1.5775
3	n-Valerophenone	0.795	0.95	0	0.5	1.4366
4	n-Butyrophenone	0.797	0.95	0	0.51	1.2957
5	n-Propiophenone	0.804	0.95	0	0.51	1.1548
6	Acetophenone	0.818	1.01	0	0.48	1.0139
7	Paracetamol	1.06	1.63	1.04	0.86	1.1724
8	Acetanilide	0.87	1.36	0.46	0.69	1.1137
9	Theophylline	1.5	1.6	0.54	1.34	1.2223
10	Caffeine	1.5	1.63	0	1.29	1.3632
11	Indazole	1.18	1.22	0.53	0.35	0.9053
12	Benzonitrile	0.742	1.11	0	0.33	0.8711
13	Chlorobenzene	0.718	0.65	0	0.07	0.8388
14	1,4-Dinitrobenzene	1.13	1.63	0	0.46	1.0648
15	Hydrocortisone	2.03	3.49	0.71	1.9	2.7976
16	Cortisone-21-acetate	1.82	3.11	0.21	2.13	3.0521
17	Progesterone	1.45	3.29	0	1.14	2.6215
18	Anisole	0.708	0.75	0	0.29	0.916
19	Benzamide	0.99	1.5	0.49	0.67	0.9728
20	Butalbarbital	1.03	1.11	0.47	1.23	1.6557
21	3,4-Dichlorophenol	1.02	1.14	0.85	0.03	1.0199
22	Phenol	0.805	0.89	0.6	0.3	0.7751
23	4-Nitrophenol	1.07	1.72	0.82	0.26	0.9493
24	2-Chlorophenol	0.853	0.88	0.32	0.31	0.8975
25	4-Iodophenol	1.38	1.22	0.68	0.2	1.0333
26	Resorcinol	0.98	1.11	1.09	0.52	0.8338
27	4-Cyanophenol	0.94	1.63	0.8	0.29	0.9298
28	4-Nitrobenzoic acid	0.99	1.07	0.62	0.54	1.1059
29	Benzoic acid	0.73	0.9	0.59	0.4	0.9317
30	3-Trifluoromethylphenol	0.425	0.87	0.72	0.09	0.9691
31	4-Hydroxybenzyl alcohol	0.998	1.15	0.88	0.85	0.9747
32	Salicylic acid	0.89	0.84	0.71	0.38	0.9904
33	Phenylacetic acid	0.73	0.97	0.6	0.61	1.0726
34	4-Nitroaniline	1.22	1.83	0.45	0.38	0.9904
35	Propranolol	1.712	1.8	0.31	1.26	2.148
36	*p*-Toluidine	0.923	0.95	0.23	0.45	0.9571
37	Aniline	0.995	0.96	0.26	0.41	0.8162
38	3-Nitroaniline	1.2	1.71	0.4	0.35	0.9904
39	Procaine	1.135	1.36	0.25	1.41	1.9767
40	Methyl 4-hydroxybenzoate	0.9	1.37	0.69	0.45	1.1313
41	n-Ethyl 4-hydroxybenzoate	0.86	1.35	0.69	0.45	1.2722
42	n-Propyl 4-hydroxybenzoate	0.86	1.35	0.69	0.45	1.4131
43	n-Butyl 4-hydroxybenzoate	0.86	1.33	0.71	0.46	1.554
44	Benzene	0.61	0.52	0	0.14	0.7164
45	Toluene	0.601	0.52	0	0.14	0.8573
46	n-Ethylbenzene	0.613	0.51	0	0.15	0.9982
47	n-Propylbenzene	0.604	0.5	0	0.15	1.1391
48	n-Butylbenzene	0.6	0.51	0	0.15	1.28
49	n-Nitroethane	0.27	0.95	0.02	0.33	0.5646
50	n-Nitropropane	0.242	0.95	0	0.31	0.7055
51	n-Nitrobutane	0.227	0.95	0	0.29	0.8464
52	Testosterone	1.54	2.59	0.32	1.19	2.3827
53	Dexamethasone	2.04	3.51	0.71	1.92	2.9132
54	Cortexalone	1.91	3.45	0.36	1.6	2.7389
55	Corticosterone	1.86	3.43	0.4	1.63	2.7389
56	Aldosterone	2.01	3.47	0.4	1.9	2.689
57	Hydroquinone	1.063	1.27	1.06	0.57	0.8338
58	3-Fluorophenol	0.667	0.98	0.68	0.17	0.7928
59	1-Naphthol	1.52	1.05	0.6	0.37	1.1441
60	Di-Et phthalate	0.729	1.4	0	0.86	1.7106
61	1,3,5-Trihydroxybenzene	1.355	1.12	1.4	0.82	0.8925
62	2-Nitrophenol	1.015	1.05	0.05	0.37	0.9493
63	Ibuprofen	0.7	0.92	0.6	0.6	1.7771
64	3-Nitrobenzoic acid	0.99	1.08	0.76	0.52	1.1059
65	Dimethylphthate	0.78	1.4	0	0.84	1.4288
66	Pentafluorophenol	0.36	0.83	0.79	0.09	0.8636
67	3-Hydroxybenzoic acid	0.91	0.88	0.86	0.58	0.9904
68	3-Hydroxybenzyl alcohol	0.998	1.12	0.88	0.81	0.9747
69	4-Fluorobenzoic acid	0.6	0.91	0.61	0.29	0.9414
70	3-Fluorobenzoic acid	0.6	0.89	0.64	0.27	0.9414
71	5-Ethylbarbituric acid	1.06	1.14	0.46	1.16	1.0921
72	3-Nitroacetanilide	1.11	2.05	0.64	0.57	1.2875
73	Indomethacin	2.24	2.85	0.4	1.08	2.5299
74	Deoxycorticosterone	1.74	3.5	0.14	1.31	2.6802
75	Cortisone	1.96	3.5	0.36	1.87	2.7546
76	3-Cyanophenol	0.93	1.55	0.84	0.25	0.9298
77	Estradiol	1.8	1.77	0.86	1.1	2.1988
78	4-Fluoroaniline	0.76	1.09	0.28	0.41	0.8339
79	2-Ethylaniline	0.962	0.85	0.23	0.45	1.098
80	Lidocaine	1.01	1.49	0.11	1.27	2.0589

The descriptors for caffeine and butabarbital are the latest calculated descriptors and they differ slightly from the ones given in ref. 12.

seven systems used in this study have been chosen to cover the widest range of the coefficients for A, B and S in the solvation equation. Most of the chosen columns are stable at relatively high pH (10.5) that is important from a practical point of view, as the compound's retention should be measured at a pH where it is not ionised. Thus, the retention times of basic compounds should be measured with high pH mobile phases. We have chosen short columns and fast generic gradients with 5 minutes cycle time. It means that gradient retention times (CHI values) for one compound can be obtained in 35 minutes under 7 different HPLC conditions. The systems themselves are listed in Table S2 (Supplementary Information †).

The five mathematical methods for the calculation of descriptors

Four of the five mathematical procedures have been discussed in detail in the evaluation of the partition coefficient method used to obtain the Abraham descriptors.[11] An outline of all five methods is given here. A set of 80 compounds that had been run on the seven RP-HPLC systems was separated into a training set and a test set each of 40 compounds. Three different training and test sets 1, 2 and 3 were chosen, based on the distribution of the three descriptors S, B and V respectively thus using the three major descriptors in the RP-HPLC equations. Descriptor A was not used to choose a training set because of its uneven distribution (too many values of zero present).

Solver

Solver is a routine in Microsoft Excel which works by minimising the sum of squares of the equations to fit the targeted cells, S, A and B. 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 simultaneous 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

$$\mathrm{SP} - E.e - V.v = S.s + A.a + B.b \qquad (2)$$

This is equivalent to

$$X_n = S_n.s_n + A_n.a_n + B_n.b_n \qquad (3)$$

The program has been modified to work with up to seven Abraham equations according to the needs of the user.

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 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[15] method, and treats the unknown descriptors as adjustable parameters and minimizes the root-mean-square-difference (RMSD) between experimental log SP and calculated log SP.

Regressions for obtaining descriptors

The fourth method of obtaining the three descriptors A, B and S uses regression equations obtained from the 40 compound 'database' training set (*i.e.* training set 1) on the lines of eqn. (4). The training set chosen for this purpose is shown in Table 4 and includes compounds with a satisfactory range of descriptors.

$$\begin{aligned}\text{Descriptor} = {} & w(\text{C18, AcN}) + q(\text{C18, MeOH}) + \\ & x(\text{C18, TFE}) + g(\text{FO, TFE}) + k(\text{PLRP, AcN}) + \\ & j(\text{DCN, MeOH}) + z(\text{DCN, AcN}) + eE + vV \end{aligned} \qquad (4)$$

Using the method of multiple linear regression three equations were obtained of the same form as eqn. (4):

$$\begin{aligned} S = {} & 0.673(0.380) - 0.013(0.016)(\text{C18, AcN}) + \\ & 0.008(0.010)(\text{C18, MeOH}) - 0.050(0.009)(\text{C18, TFE}) + \\ & 0.015(0.006)(\text{FO, TFE}) + 0.023(0.009)(\text{PLRP, AcN}) - \\ & 0.019(0.014)(\text{DCN, MeOH}) + 0.013(0.017)(\text{DCN, AcN}) + \\ & 0.273(0.161)E + 1.398(0.166)V \end{aligned} \qquad (5)$$

$N = 40$, $r^2 = 0.950$, SE = 0.212, F = 63.03

$$\begin{aligned} A = {} & 1.499(0.243) + 0.010(0.010)(\text{C18, AcN}) - \\ & 0.001(0.006)(\text{C18, MeOH}) + 0.002(0.006)(\text{C18, TFE}) - \\ & 0.011(0.004)(\text{FO, TFE}) - 0.025(0.006)(\text{PLRP, AcN}) + \\ & 0.026(0.009)(\text{DCN, MeOH}) - 0.007(0.011)(\text{DCN, AcN}) + \\ & 0.149(0.103)E - 0.369(0.107)V \end{aligned} \qquad (6)$$

$N = 40$, $r^2 = 0.896$, SE = 0.136, F = 28.58

$$\begin{aligned} B = {} & 0.103(0.193) + 0.001(0.008)(\text{C18, AcN}) - \\ & 0.007(0.005)(\text{C18, MeOH}) + 0.008(0.005)(\text{C18, TFE}) + \\ & 0.002(0.003)(\text{FO, TFE}) - 0.009(0.005)(\text{PLRP, AcN}) - \\ & 0.020(0.007)(\text{DCN, MeOH}) + 0.012(0.008)(\text{DCN, AcN}) + \\ & 0.09(0.082)E + 0.788(0.085)V \end{aligned} \qquad (7)$$

$N = 40$, $r^2 = 0.963$, SE = 0.108, F = 86.46

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 number in brackets after each coefficient is the standard error for that coefficient in the equation. The Excel statistical software was used to carry out the regressions.

Modified regressions

The main drawbacks of the regression method centre round the direct use of the descriptors E and V in the construction of the regression equations that will be used to predict S, A and B. Thus, the estimates of S, A and B, obtained from this method, will be directly susceptible to the pattern of correlation of the Abraham parameters in the training set. Also, the contribution of E and V in improving the fit of the regression equations will not necessarily be fully reflected in improved performance of the estimated Abraham parameters in predicting solute properties because of the direct linear dependence between S, A, B and E, V induced by the method of construction. Neither of the above is desirable. However, leaving E and V descriptors out of the regressions gives much poorer results and does not make sense since the totality of chemical information is assumed to be enclosed in a five dimensional space in the first place. For these reasons a modified regression method is proposed. First, training set 1 is used to construct Abraham equations in the same fashion as equation (1). The equations are shown in Table 3. $\mathrm{SP_{obs}}$ denotes the observed value of a given property.

$$\mathrm{SP_{obs}} = c + e\,E + s\,S + a\,A + b\,B + v\,V \qquad (8)$$

Table 2 Compounds and CHI values measured on the 7 HPLC systems

No	Name	C18, AcN	C18,MeOH	C18,TFE	FO,TFE	PLRP,AcN	DCN,MeOH	DCN,AcN
1	n-Heptanophenone	113.25	94.48	94.87	77.67	104.85	58.22	57.22
2	n-Hexanophenone	105.17	91.27	90.47	75.63	99.78	54.52	54.20
3	n-Valerophenone	96.25	87.33	85.92	73.30	93.79	49.67	50.06
4	n-Butyrophenone	86.68	82.31	80.92	70.82	87.01	43.13	44.23
5	n-Propiophenone	76.22	76.31	75.73	67.82	79.26	35.19	36.15
6	Acetophenone	62.40	67.98	69.54	64.42	66.98	24.86	24.78
7	Paracetamol	20.43	40.29	41.73	36.68	23.08	10.18	7.51
8	Acetanilide	41.58	57.87	56.31	52.75	41.98	17.42	15.82
9	Theophylline	19.32	32.51	45.17	46.44	21.49	6.48	1.75
10	Caffeine	25.08	49.70	52.73	51.17	27.78	13.24	8.55
11	Indazole	48.67	65.94	62.43	55.84	49.85	27.08	24.64
12	Benzonitrile	64.06	65.29	66.71	56.61	69.04	22.32	24.23
13	Chlorobenzene	89.99	85.05	81.88	74.62	93.38	40.23	43.48
14	1,4-Dinitrobenzene	70.39	68.21	58.87	56.50	75.03	29.65	37.22
15	Hydrocortisone	49.56	75.49	61.87	52.67	44.91	45.55	34.71
16	Cortisone-21-acetate	69.29	79.39	68.29	65.43	63.21	54.01	45.60
17	Progesterone	96.34	90.35	83.99	83.00	90.65	58.58	51.72
18	Anisole	75.98	77.49	71.24	60.91	81.81	27.82	31.54
19	Benzamide	29.00	48.39	52.19	47.88	29.83	7.61	4.26
20	Butalbarbital	51.08	70.37	60.96	53.50	45.08	27.50	26.92
21	3,4-Dichlorophenol	75.36	83.76	63.05	48.84	70.63	53.04	47.96
22	Phenol	47.52	59.28	48.98	39.21	48.43	13.66	14.12
23	4-Nitrophenol	56.19	67.12	47.87	41.89	54.08	34.83	31.35
24	2-Chlorophenol	61.13	70.21	57.80	46.56	60.32	33.51	34.27
25	4-Iodophenol	71.49	80.08	63.24	46.62	70.67	47.35	44.45
26	Resorcinol	25.18	40.72	40.87	22.65	27.73	11.41	10.91
27	4-Cyanophenol	45.64	59.81	47.57	43.28	46.58	29.07	27.66
28	4-Nitrobenzoic acid	56.03	71.55	52.85	47.57	51.15	36.54	34.56
29	Benzoic acid	49.70	67.29	57.98	49.15	46.46	23.92	22.72
30	3-Trifluoromethylphenol	71.90	79.10	59.33	52.32	62.71	47.16	45.30
31	4-Hydroxybenzyl alcohol	20.21	40.49	41.29	41.26	22.37	5.00	2.57
32	Salicylic acid	57.46	73.29	56.80	46.16	51.32	33.74	29.73
33	Phenylacetic acid	50.71	66.01	58.19	47.77	45.96	20.51	21.68
34	4-Nitroaniline	52.76	55.74	49.45	45.18	57.01	34.16	32.83
35	Propranolol	76.75	86.61	82.12	69.81	80.72	59.83	50.73
36	*p*-Toluidine	56.13	63.68	63.80	58.00	60.32	16.36	16.78
37	Aniline	43.22	50.92	53.05	45.72	53.08	5.65	3.27
38	3-Nitroaniline	57.67	59.51	52.50	47.83	62.17	28.14	29.99
39	Procaine	62.14	74.08	70.45	74.22	57.98	46.38	38.95
40	Methyl 4-hydroxybenzoate	51.67	66.67	54.59	49.41	49.56	35.54	29.03
41	n-Ethyl 4-hydroxybenzoate	61.45	73.42	60.64	54.60	56.22	42.30	35.56
42	n-Propyl 4-hydroxybenzoate	71.06	79.53	66.36	58.31	63.17	48.12	41.83
43	n-Butyl 4-hydroxybenzoate	80.04	84.58	71.98	61.51	69.71	52.56	46.63
44	Benzene	77.82	78.21	73.24	57.97	82.86	15.72	17.92
45	Toluene	88.65	85.76	80.64	63.61	91.53	31.29	38.36
46	n-Ethylbenzene	97.43	90.29	86.29	67.10	97.31	42.97	47.85
47	n-Propylbenzene	106.43	94.16	91.78	70.27	102.80	51.34	53.53
48	n-Butylbenzene	114.38	97.40	97.15	73.01	107.62	56.58	57.22
49	n-Nitroethane	18.31	30.05	43.94	44.63	41.39	−3.62	−8.43
50	n-Nitropropane	52.41	50.82	56.94	53.65	56.85	3.36	1.05
51	n-Nitrobutane	69.68	66.40	64.66	60.13	67.49	15.56	18.36
52	Testosterone	72.11	85.00	77.24	75.37	67.24	52.50	43.27
53	Dexamethasone	57.25	78.51	66.12	58.54	50.56	50.66	39.95
54	Cortexalone	61.44	79.79	69.78	66.12	56.43	50.05	40.32
55	Corticosterone	59.82	79.69	68.87	65.98	55.13	49.92	39.14
56	Aldosterone	44.96	70.73	64.17	64.57	42.02	41.91	32.13
57	Hydroquinone	13.47	30.25	28.98	18.64	21.49	5.29	3.71
58	3-Fluorophenol	56.05	66.53	50.21	41.69	53.32	21.07	26.88
59	1-Naphthol	73.72	78.92	64.26	51.14	72.95	49.63	45.38
60	Di-Et phthalate	81.86	78.96	79.34	61.31	76.64	44.38	44.47
61	1,3,5-Trihydroxybenzene	9.05	28.05	33.28	15.39	14.26	9.69	7.50
62	2-Nitrophenol	67.49	71.08	64.59	57.16	72.70	30.34	31.31
63	Ibuprofen	91.56	90.85	87.13	67.04	73.49	55.11	51.69
64	3-Nitrobenzoic acid	54.59	79.75	53.31	48.26	50.25	36.95	32.63
65	Dimethylphthate	65.00	68.84	70.00	65.83	64.77	34.03	33.11
66	Pentafluorophenol	72.28	82.48	56.02	50.10	61.24	43.63	41.76
67	3-Hydroxybenzoic acid	34.36	55.79	43.53	36.71	32.72	24.83	20.47
68	3-Hydroxybenzyl alcohol	24.82	45.32	42.81	35.73	25.50	7.10	4.71
69	4-Fluorobenzoic acid	54.59	71.31	57.40	50.02	48.87	29.85	28.09
70	3-Fluorobenzoic acid	55.34	71.77	57.47	49.90	49.41	31.02	28.57
71	5-Ethylbarbituric acid	15.99	36.29	41.55	16.16	18.54	−4.44	−7.74
72	3-Nitroacetanilide	52.47	65.55	53.73	52.00	50.34	38.12	32.63
73	Indomethacin	87.98	89.93	82.40	63.90	79.28	61.72	53.74
74	Deoxycorticosterone	76.65	83.73	76.14	75.54	71.06	53.78	45.24
75	Cortisone	52.58	73.09	62.81	60.01	47.07	45.61	36.19
76	3-Cyanophenol	51.55	62.05	50.21	44.69	48.87	29.53	26.99
77	Estradiol	71.34	83.87	70.72	60.53	66.32	56.92	46.26
78	4-Fluoroaniline	47.70	54.57	53.49	50.47	54.21	14.92	10.62
79	2-Ethylaniline	68.78	71.27	68.35	62.29	71.53	28.61	28.62
80	Lidocaine	85.21	85.82	89.66	84.46	72.23	46.73	44.93

Table 3 The equations of 7 HPLC columns obtained from training set 1

No	HPLC System	Coefficient						Stats			
		c	*e*	*s*	*a*	*b*	*v*	*N*	r^2	SE	F
1	C18,AcN	42.792	7.493	−20.559	−23.449	−58.645	66.108	40	0.946	6.478	120.0
		3.149	4.790	3.076	3.217	4.341	3.625				
2	C18,MeOH	45.620	2.276	−11.457	−6.661	−44.455	51.850	40	0.910	5.932	68.6
		2.883	4.386	2.816	2.945	3.975	3.319				
3	C18,TFE	51.023	6.001	−18.224	−18.897	−21.918	40.849	40	0.959	3.549	158.7
		1.725	2.624	1.685	1.762	2.378	1.986				
4	FO,TFE	45.428	−2.351	−5.646	−20.053	−17.829	29.560	40	0.839	6.473	35.3
		3.147	4.786	3.073	3.214	4.337	3.622				
5	PLRP,AcN	53.319	13.625	−15.698	−32.696	−51.760	46.804	40	0.958	5.063	156.6
		2.461	3.743	2.404	2.514	3.393	2.833				
6	DCN,MeOH	−3.334	5.623	−9.719	3.467	−43.504	53.302	40	0.922	5.419	80.8
		2.634	4.007	2.573	2.691	3.631	3.032				
7	DCN,AcN	2.259	6.568	−10.645	−1.241	−45.061	49.984	40	0.883	6.432	51.5
		3.127	4.756	3.054	3.194	4.310	3.599				

Next, the influence of E and V can be removed by considering SP_{res} (residual value) thus:

$$SP_{res} = SP_{obs} - (eE + vV) \quad (9)$$

Then the modified SP_{res} can be used as the x-term to create regression equations to predict S, A and B. The advantage over the regression equation, described previously, is that E and V are not given complete freedom to enter the regressions but can only enter as adjustment factors in a manner constrained by the Abraham equations. The modified regressions obtained using training set 1 are given below:

$$S = 0.878(0.355) - 0.014(0.016)(\text{C18, AcN})_{res} + 0.008(0.010)(\text{C18, MeOH})_{res} - 0.055(0.009)(\text{C18, TFE})_{res} + 0.018(0.006)(\text{FO, TFE})_{res} + 0.025(0.009)(\text{PLRP, AcN})_{res} - 0.023(0.014)(\text{DCN, MeOH})_{res} + 0.016(0.017)(\text{DCN, AcN})_{res} \quad (10)$$

$N = 40$, $r^2 = 0.946$, SE = 0.214, F = 79.44

$$A = 1.464(0.225) + 0.011(0.010)(\text{C18, AcN})_{res} - 0.001(0.006)(\text{C18, MeOH})_{res} + 0.004(0.005)(\text{C18, TFE})_{res} - 0.012(0.004)(\text{FO, TFE})_{res} - 0.026(0.006)(\text{PLRP, AcN})_{res} + 0.029(0.009)(\text{DCN, MeOH})_{res} - 0.008(0.011)(\text{DCN, AcN})_{res} \quad (11)$$

$N = 40$, $r^2 = 0.890$, SE = 0.135, F = 36.83

$$B = 0.085(0.175) + 0.001(0.008)(\text{C18, AcN})_{res} - 0.007(0.005)(\text{C18, MeOH})_{res} + 0.008(0.004)(\text{C18, TFE})_{res} + 0.003(0.003)(\text{FO, TFE})_{res} - 0.009(0.004)(\text{PLRP, AcN})_{res} - 0.020(0.007)(\text{DCN, MeOH})_{res} + 0.012(0.008)(\text{DCN, AcN})_{res} \quad (12)$$

$N = 40$, $r^2 = 0.963$, SE = 0.105, F = 117.35

Equations 10, 11 and 12 for predicting S, A and B are more robust. Since their correlation coefficients also closely approach those obtained by the unmodified form of the regression method (which they cannot theoretically exceed), the above results are encouraging. Incidentally, the F statistics in all three regressions improved significantly.

Experimental

All compounds listed in Table 1 are commercially available. The compounds were dissolved in 50% water:50% acetonitrile in 1 mg mL^{-1} concentration. 5 μl of the solutions was injected in the HPLC system. The HP1100 (Agilent, Amsterdam, Netherlands) equipment was used to determine the gradient retention times under each investigated chromatographic condition. The following generic fast gradient procedure was used for the seven chromatographic systems studied:

Flow rate: 2.00 ml min^{-1}

Cycle time: 5 min

Mobile phase P_1 was 0.1% (0.01 M) phosphoric acid solution (pH 2) for the neutral and acidic compounds, while for the basic compounds mobile phase P_1 was 50 mM ammonium acetate solution adjusted to pH 10.5 by concentrated ammonia solution. The columns and the organic solvent used as mobile phase P_2 are listed in Table S2†.

The gradient retention time for a calibration mixture has been measured and used to convert to Chromatographic Hydrophobicity Index (CHI) value. The CHI value approximates the volume per cent of the organic modifier in the mobile phase when the peak elutes from the column (Table S1†).[16] The calibration mixture components, their typical gradient retention times and their CHI values are summarized in Table S3†. The figure below shows a typical calibration plot. Using the slope and the intercept values, the CHI value of a compound can be obtained as CHI = slope x gradient retention time + intercept. The so-obtained CHI values for the investigated compounds are shown in Table 2.

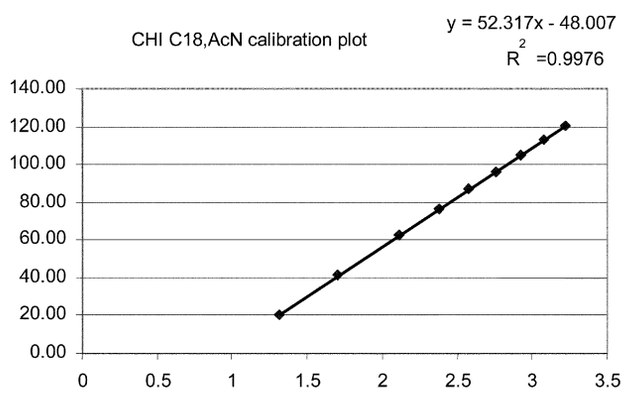

Fig. 1 Calibration graph for C18, AcN.

Results

We first applied the solvation equation (1) on the three different training sets. Results are given in Tables 6, S4 and S5†. Training set 1 shown in Table 4 is the most representative of the three training sets 1, 2 and 3, and all subsequent calculations were made using this training set and its corresponding

test set. As a first step in the assessment of the methods of calculation of descriptors we used the chosen training set. Because the five methods are applied on the training set, which was used to obtain all our equations, this exercise can only provide evidence of the likely accuracy of these methods. That is the standard deviation between calculated descriptors and those listed in Table 1 will be minimum values. The descriptors calculated for the 40 compounds in the training set are tabulated in Table 4. Standard deviation values are given (Table 5) between the calculated descriptors and those listed in Table 1. The results show that there is good agreement between the five sets of calculated descriptors and that they yield S, A, and B in reasonable agreement with those in Table 1. The Regression methods are the most accurate ones followed by Solver and Descfit exhibiting the same accuracy; TripleX is the least accurate method.

A detailed analysis of the descriptor calculations on the 40-compound test set was then carried out to establish how successful the method is for obtaining descriptors of compounds outside the training set. Our test set of compounds consists of simple mono-functional and more complicated poly-functional as well as some drug compounds, covering a large descriptor space. The results obtained from applying the five methods in the calculation of descriptors for the test set are tabulated in Table 6. Standard deviations (SD) have also been calculated comparing the descriptors obtained from each one of our four methods with those given in Table 1. The SD values are tabulated in Table 7. From inspection of the standard deviations obtained in Table 7 the overall predictive power of our methods when applied outside the training set *i.e.* the real calculation potential of the Abraham descriptors from HPLC data can be seen. The errors involved in this method are higher for S when compared to the partition method[11] set up earlier. It is however better when it comes to obtaining descriptors A and B. The regression equation methods are slightly better than Solver and Descfit and Tx is the least accurate. The results obtained by the modified regression method are slightly (probably not significant) better, in comparison to the standard regression method.

In order to have a completely different way of evaluating the HPLC method we used the descriptors obtained by all five mathematical procedures to calculate octanol/water partition coefficients using our well established octanol equation (13).[6]

$$\text{Log } P_{oct} = 0.088 + 0.562\,E - 1.054\,S + 0.034\,A - 3.460\,B + 3.814\,V \quad (13)$$

$n = 613$, $r = 0.9974$, SD = 0.116, F = 23161.6

This exercise provides a useful way to validate our results using external data, and gives a measure of the potential predictability of the descriptors calculated. Table 8 tabulates the results of calculated log P_{oct} values from our descriptors as well as experimentally measured octanol/water partition coefficients, most of which have been obtained from the Medchem database.[17] Standard deviations are also given, obtained by comparing calculated and experimental values. Not surprisingly perhaps the results reveal that the log P_{oct} values obtained from Table 1 descriptors are the best when compared to the experimental values with a SD value of 0.21. This is expected since Table 1 descriptors have been calculated from experimentally measured log P values in the first place. The regression equations obtained from our training set of 40 compounds are second in terms of accuracy with a standard deviation of 0.32 log units. The modified regression method does slightly worse with a standard deviation of 0.33 log units. Solver and Descfit are slightly less accurate. TripleX is the least accurate method with a standard deviation of 0.48 log units.

Table 5 Standard deviations of different methods of calculation in comparison to Table 1 descriptors for our 40-compound training set

	S	A	B
Solver	0.24	0.13	0.13
Tx	0.31	0.25	0.14
Descfit	0.24	0.13	0.13
Regr.	0.19	0.12	0.10
Mod-Regr.	0.19	0.12	0.10

Although the calculation of partition coefficients gives the user a good idea of where the limits of the HPLC method lie, it does not provide however, a direct comparison between the latter and the water/solvent partition method outlined in our previous work.[11] This sort of comparison can be achieved by looking at compounds studied by both methods. For this purpose eleven compounds that have been used in both exercises have been used as a measure for direct comparison of the two methods. These compounds are tabulated in Table 9 along with their calculated descriptors obtained from our methods namely Solver TripleX, Descfit and the regressions as well as the database descriptors (Table 1) for these compounds. The modified regressions have not been used in this comparison because it has not been used in the validation of the log P method. It has to be noted however that some of the common compounds used come from both the training as well as the test sets used to validate the partition coefficient method, and therefore the regression equation results are minimum values for certain compounds. Table 9 also includes calculated standard deviation values obtained from comparing the calculated descriptors with the four methods and the database descriptors. The first thing to be observed from these results is that in the case of the HPLC method the regression equation is the preferred calculational method to be used where as Solver and Descfit are the most accurate when the partition method is used. TripleX is not as good in either method.

Conclusion

Comparisons of the calculated descriptors using our five methods of calculation on HPLC CHI data reveal some differences between the four calculational methods Solver, Descfit and the two regression methods. TripleX does not perform well compared to the rest. The test shown in Table 6 reveals that the regression method is a more accurate way for obtaining the three Abraham descriptors from HPLC data. This is the result of two things. Firstly the fact that in the HPLC set-up there are seven physical properties measured from which three descriptors are calculated in comparison to the log P method where only four properties are measured. This gives the regression method an advantage. Secondly, the HPLC data are collected under the same conditions from the same user and a regression analysis deals better with it in comparison with the partition data that has been collected from various sources as well as measured. This 'inconsistency' makes the partition data a fragmented source of information with experimental errors which cannot be easily evaluated. The test shown in Table 6 gives the opportunity to compare the partition coefficient method and HPLC methods as ways for obtaining descriptors S, A and B. It shows that the two methods are quite similar with the partition coefficient method being better at yielding S and B whereas the HPLC method yields A more accurately.

Although we use HPLC column equations with $N = 40$ (Table 3) for our analysis, we suggest that in general it is better to use the equations with the largest number of data points. That is the equations with $N = 80$ (Table 10).

Table 4 Calculation of descriptors of the 40-compound training set 1 using all five methods

Compound	*S*						*A*						*B*					
	DB	Solver	Tx	Descfit	Regr.	Mod-Regr.	DB	Solver	Tx	Descfit	Regr.	Mod-Regr.	DB	Solver	Tx	Descfit	Regr.	Mod-Regr.
n-Nitroethane	0.95	1.26	0.99	1.26	0.92	1.11	0.02	−0.10	0.01	−0.10	0.01	−0.04	0.33	0.60	0.62	0.60	0.42	0.44
Benzene	0.52	0.45	0.25	0.45	0.50	0.57	0	−0.30	0.16	−0.30	−0.25	−0.23	0.14	0.29	0.14	0.29	0.11	0.16
3-Fluorophenol	0.98	1.12	0.98	1.12	1.06	1.16	0.68	0.61	0.92	0.61	0.47	0.47	0.17	0.17	0.09	0.17	0.28	0.32
Resorcinol	1.11	0.77	0.35	0.77	0.96	0.99	1.09	1.06	1.63	1.06	0.91	0.88	0.52	0.69	0.59	0.69	0.61	0.64
4-Fluoroaniline	1.09	1.28	1.32	1.28	1.05	1.14	0.28	0.11	0.12	0.11	0.24	0.22	0.41	0.47	0.45	0.47	0.36	0.41
n-Nitrobutane	0.95	0.85	0.89	0.85	0.71	0.87	0	−0.18	−0.03	−0.18	−0.09	−0.07	0.29	0.38	0.30	0.38	0.33	0.37
Pentafluorophenol	0.83	1.05	1.13	1.05	0.77	0.87	0.79	0.86	0.84	0.86	0.75	0.82	0.09	−0.16	−0.18	−0.16	−0.06	−0.04
1,3,5-Trihydroxybenzene	1.12	0.98	0.62	0.98	1.21	1.19	1.4	1.33	1.75	1.33	1.17	1.12	0.82	0.88	0.83	0.88	0.81	0.83
Indazole	1.22	0.72	1.26	0.72	0.82	0.82	0.53	0.59	−0.30	0.59	0.56	0.56	0.35	0.50	0.70	0.50	0.43	0.49
4-Cyanophenol	1.63	1.47	1.52	1.47	1.26	1.33	0.8	0.79	0.66	0.79	0.71	0.71	0.29	0.27	0.32	0.27	0.35	0.38
Benzoic acid	0.9	0.73	0.92	0.73	0.81	0.88	0.59	0.66	0.37	0.66	0.57	0.59	0.4	0.47	0.53	0.47	0.42	0.46
3-Fluorobenzoic acid	0.89	0.83	1.02	0.83	0.80	0.87	0.64	0.76	0.49	0.76	0.66	0.70	0.27	0.31	0.36	0.31	0.29	0.32
2-Nitrophenol	1.05	1.26	1.30	1.26	1.10	1.15	0.05	0.17	0.08	0.17	0.16	0.17	0.37	0.26	0.28	0.26	0.26	0.31
3-Trifluoromethylphenol	0.87	1.05	1.12	1.05	0.80	0.89	0.72	0.86	0.75	0.86	0.73	0.81	0.09	−0.06	−0.03	−0.06	0.04	0.07
4-Hydroxybenzyl alcohol	1.15	1.39	1.69	1.39	1.38	1.46	0.88	0.60	0.20	0.60	0.63	0.57	0.85	0.88	0.96	0.88	0.84	0.88
Salicylic acid	0.84	0.93	1.04	0.93	0.92	0.94	0.71	0.84	0.74	0.84	0.76	0.79	0.38	0.32	0.32	0.32	0.27	0.31
3-Nitroaniline	1.71	1.76	1.72	1.76	1.49	1.55	0.4	0.38	0.48	0.38	0.38	0.36	0.35	0.28	0.26	0.28	0.39	0.44
n-Ethylbenzene	0.51	0.35	0.20	0.35	0.43	0.47	0	0.02	0.17	0.02	−0.11	−0.03	0.15	0.12	0.11	0.12	0.07	0.12
3,4-Dichlorophenol	1.14	1.06	1.01	1.06	0.89	0.87	0.85	0.93	0.97	0.94	0.81	0.89	0.03	−0.04	−0.04	−0.04	−0.03	0.01
1,4-Dinitrobenzene	1.63	1.96	1.92	1.96	1.65	1.75	0	0.08	0.18	0.08	0.04	0.03	0.46	0.20	0.19	0.20	0.40	0.46
5-Ethylbarbituric acid	1.14	0.70	1.29	0.70	1.15	1.16	0.46	0.87	0.57	0.87	0.76	0.69	1.16	1.30	1.02	1.30	1.01	1.03
4-Nitrobenzoic acid	1.07	1.50	1.63	1.50	1.34	1.39	0.62	0.84	0.67	0.84	0.74	0.76	0.54	0.28	0.32	0.28	0.36	0.40
Acetanilide	1.36	1.19	1.46	1.19	1.21	1.29	0.46	0.42	0.00	0.42	0.40	0.38	0.69	0.77	0.86	0.77	0.71	0.76
n-Propylbenzene	0.5	0.29	0.14	0.29	0.35	0.38	0	0.04	0.22	0.04	−0.06	0.05	0.15	0.16	0.13	0.16	0.07	0.12
n-Propiophenone	0.95	1.07	1.20	1.07	0.98	1.05	0	−0.01	−0.29	−0.01	−0.04	0.00	0.51	0.42	0.50	0.42	0.39	0.44
Theophylline	1.6	1.61	2.17	1.61	1.63	1.66	0.54	0.48	−0.25	0.48	0.63	0.54	1.34	1.23	1.37	1.23	1.15	1.20
n-Butylbenzene	0.51	0.21	0.04	0.21	0.30	0.31	0	0.01	0.27	0.01	−0.07	0.06	0.15	0.24	0.18	0.24	0.11	0.16
n-Butyrophenone	0.95	1.11	1.17	1.11	1.01	1.07	0	−0.01	−0.19	−0.01	−0.06	0.00	0.51	0.40	0.46	0.40	0.38	0.44
n-Propyl 4-hydroxybenzoate	1.35	1.40	1.53	1.40	1.24	1.28	0.69	0.67	0.47	0.67	0.61	0.66	0.45	0.45	0.49	0.45	0.43	0.47
n-Valerophenone	0.95	1.11	1.11	1.11	1.01	1.07	0	−0.03	−0.07	−0.03	−0.08	0.00	0.5	0.42	0.45	0.42	0.39	0.44
n-Hexanophenone	0.95	1.12	1.02	1.12	1.01	1.07	0	−0.09	0.05	−0.09	−0.12	−0.03	0.5	0.47	0.44	0.47	0.41	0.47
Di-Et phthalate	1.4	1.22	0.85	1.22	1.28	1.35	0	0.28	0.73	0.28	0.13	0.18	0.86	0.82	0.75	0.82	0.76	0.79
Ibuprofen	0.92	0.61	0.68	0.61	0.85	0.87	0.6	0.59	0.52	0.59	0.44	0.54	0.6	0.83	0.84	0.83	0.71	0.75
Lidocaine	1.49	1.44	1.61	1.44	1.55	1.60	0.11	0.01	−0.14	0.01	0.00	0.04	1.27	1.20	1.38	1.20	1.14	1.20
Estradiol	1.77	2.55	2.52	2.55	2.40	2.35	0.86	0.62	0.63	0.62	0.56	0.57	1.1	1.06	1.06	1.06	0.99	1.05
Indomethacin	2.85	2.53	2.39	2.53	2.59	2.47	0.4	0.45	0.75	0.45	0.40	0.41	1.08	1.34	1.26	1.34	1.24	1.31
Deoxycorticosterone	3.5	3.39	3.47	3.39	3.10	3.13	0.14	0.05	−0.05	0.05	0.08	0.06	1.31	1.46	1.48	1.46	1.45	1.52
Cortexalone	3.45	3.38	3.37	3.38	3.24	3.24	0.36	0.39	0.36	0.39	0.33	0.29	1.6	1.65	1.66	1.65	1.60	1.66
Cortisone	3.5	3.56	3.49	3.56	3.41	3.42	0.36	0.52	0.59	0.52	0.45	0.39	1.87	1.73	1.72	1.73	1.71	1.77
Dexamethasone	3.51	3.57	3.33	3.57	3.48	3.47	0.71	0.60	0.92	0.60	0.48	0.44	1.92	1.81	1.74	1.81	1.75	1.80

DB: calculated using all available literature values; Solver: Excel Solver; TX: TripleX program; Descfit: Simplex minimization method; Regr.: Regression Equation; Mod-reg.: Modified Regression.

Table 6 Calculation of descriptors of the 40-compound test set 1 using all five methods

	S						*A*						*B*					
Compound	DB	Solver	Tx	Descfit	Regr.	Mod-Regr.	DB	Solver	Tx	Descfit	Regr.	Mod-Regr.	DB	Solver	Tx	Descfit	Regr.	Mod-Regr.
n-Nitropropane	0.95	0.86	0.83	0.86	0.67	0.83	0	−0.27	−0.02	−0.27	−0.07	−0.08	0.31	0.55	0.44	0.55	0.37	0.40
Phenol	0.89	1.05	0.89	1.05	1.01	1.08	0.6	0.49	0.86	0.49	0.46	0.45	0.3	0.39	0.27	0.39	0.35	0.38
Aniline	0.96	1.21	1.10	1.21	1.12	1.19	0.26	0.01	0.27	0.01	0.13	0.08	0.41	0.65	0.56	0.65	0.49	0.53
Hydroquinone	1.27	1.54	1.05	1.54	1.47	1.53	1.06	0.96	1.65	0.96	0.88	0.81	0.57	0.69	0.55	0.69	0.67	0.69
Chlorobenzene	0.65	0.66	0.99	0.66	0.53	0.59	0	−0.14	−0.71	−0.14	−0.14	−0.07	0.07	0.00	0.14	0.00	0.00	0.07
Toluene	0.52	0.41	0.24	0.41	0.50	0.57	0	−0.09	0.13	−0.09	−0.21	−0.15	0.14	0.14	0.10	0.14	0.10	0.15
Benzonitrile	1.11	0.80	0.79	0.80	0.78	0.85	0	0.06	0.06	0.06	0.06	0.07	0.33	0.40	0.41	0.40	0.33	0.38
2-Chlorophenol	0.88	1.00	0.94	1.00	0.93	0.98	0.32	0.69	0.75	0.69	0.56	0.59	0.31	0.17	0.17	0.17	0.21	0.25
Anisole	0.75	0.99	0.86	0.99	0.89	0.98	0	−0.07	0.08	−0.07	−0.12	−0.10	0.29	0.22	0.20	0.22	0.20	0.25
3-Cyanophenol	1.55	1.30	1.42	1.30	1.13	1.18	0.84	0.74	0.63	0.74	0.71	0.72	0.25	0.29	0.30	0.29	0.32	0.36
4-Fluorobenzoic acid	0.91	0.83	1.02	0.83	0.82	0.89	0.61	0.74	0.46	0.74	0.64	0.67	0.29	0.33	0.38	0.33	0.31	0.35
4-Nitrophenol	1.72	1.61	1.59	1.61	1.32	1.36	0.82	0.81	0.93	0.80	0.77	0.78	0.26	0.13	0.08	0.13	0.19	0.24
p-Toluidine	0.95	1.02	1.23	1.02	1.02	1.09	0.23	0.02	−0.23	0.02	0.07	0.06	0.45	0.60	0.63	0.60	0.51	0.56
Benzamide	1.5	0.87	1.29	0.87	1.02	1.07	0.49	0.49	−0.09	0.49	0.53	0.49	0.67	0.93	1.04	0.93	0.79	0.84
3-Hydroxybenzyl alcohol	1.12	1.17	1.22	1.17	1.26	1.32	0.88	0.72	0.72	0.72	0.70	0.65	0.81	0.85	0.83	0.85	0.77	0.80
4-Nitroaniline	1.83	1.78	1.77	1.78	1.45	1.49	0.45	0.63	0.60	0.63	0.62	0.62	0.38	0.24	0.26	0.24	0.34	0.39
3-Hydroxybenzoic acid	0.88	1.30	1.32	1.30	1.23	1.28	0.86	1.03	0.96	1.03	0.92	0.92	0.58	0.49	0.51	0.49	0.49	0.52
Acetophenone	1.01	0.98	1.25	0.98	0.93	1.00	0	0.01	−0.46	0.01	0.03	0.04	0.48	0.52	0.63	0.52	0.44	0.50
4-Iodophenol	1.22	1.10	1.02	1.10	1.05	1.00	0.68	0.85	0.93	0.85	0.72	0.77	0.2	0.11	0.10	0.11	0.10	0.16
Phenylacetic acid	0.97	0.86	0.91	0.85	1.00	1.07	0.6	0.59	0.56	0.59	0.48	0.48	0.61	0.64	0.63	0.64	0.60	0.63
2-Ethylaniline	0.85	1.26	1.41	1.26	1.14	1.20	0.23	0.03	−0.17	0.03	0.06	0.07	0.45	0.46	0.50	0.46	0.42	0.48
3-Nitrobenzoic acid	1.08	1.50	1.66	1.50	1.35	1.40	0.76	0.88	0.65	0.88	0.76	0.78	0.52	0.25	0.29	0.25	0.28	0.33
Methyl 4-hydroxybenzoate	1.37	1.41	1.54	1.41	1.22	1.27	0.69	0.75	0.51	0.75	0.71	0.73	0.45	0.42	0.47	0.42	0.39	0.43
1-Naphthol	1.05	1.39	1.44	1.39	1.24	1.18	0.6	0.73	0.65	0.73	0.68	0.72	0.37	0.17	0.19	0.17	0.17	0.24
Paracetamol	1.63	1.54	1.54	1.54	1.56	1.63	1.04	0.75	0.71	0.75	0.70	0.64	0.86	0.99	1.01	0.99	0.95	0.98
n-Ethyl 4-hydroxybenzoate	1.35	1.40	1.56	1.40	1.21	1.26	0.69	0.71	0.45	0.71	0.66	0.70	0.45	0.43	0.48	0.43	0.40	0.44
3-Nitroacetanilide	2.05	1.84	2.05	1.84	1.57	1.62	0.64	0.71	0.36	0.71	0.69	0.69	0.57	0.48	0.56	0.48	0.52	0.57
Caffeine	1.63	1.56	2.07	1.56	1.69	1.71	0	0.53	−0.26	0.53	0.55	0.48	1.29	1.21	1.38	1.21	1.10	1.16
Dimethylphthate	1.4	1.47	1.63	1.47	1.35	1.45	0	0.15	−0.19	0.15	0.12	0.14	0.84	0.72	0.82	0.72	0.71	0.75
n-Butyl 4-hydroxybenzoate	1.33	1.38	1.46	1.38	1.25	1.29	0.71	0.60	0.52	0.60	0.53	0.60	0.46	0.51	0.52	0.51	0.47	0.52
Butalbarbital	1.11	1.69	1.73	1.69	1.79	1.86	0.47	0.56	0.50	0.56	0.41	0.40	1.23	1.03	1.05	1.03	1.02	1.06
n-Heptanophenone	0.95	1.11	0.96	1.11	1.03	1.07	0	−0.15	0.15	−0.15	−0.16	−0.06	0.5	0.55	0.47	0.55	0.47	0.52
Procaine	1.36	2.34	2.82	2.34	2.08	2.16	0.25	0.28	−0.55	0.28	0.29	0.30	1.41	1.03	1.24	1.03	1.04	1.09
Propranolol	1.8	2.27	2.18	2.27	2.16	2.12	0.31	0.32	0.15	0.32	0.22	0.25	1.26	1.02	1.11	1.02	0.90	0.97
Testosterone	2.59	2.73	2.98	2.73	2.55	2.57	0.32	0.20	−0.24	0.20	0.19	0.19	1.19	1.29	1.39	1.29	1.23	1.29
Progesterone	3.29	3.32	3.28	3.32	2.89	2.96	0	−0.31	−0.15	−0.31	−0.23	−0.22	1.14	1.23	1.18	1.23	1.23	1.30
Aldosterone	3.47	3.40	3.61	3.40	3.31	3.32	0.4	0.47	0.06	0.47	0.42	0.35	1.9	1.81	1.92	1.81	1.76	1.83
Corticosterone	3.43	3.42	3.40	3.42	3.24	3.26	0.4	0.40	0.36	0.40	0.34	0.30	1.63	1.65	1.66	1.65	1.59	1.65
Hydrocortisone	3.49	3.45	3.12	3.45	3.41	3.40	0.71	0.69	1.13	0.69	0.55	0.50	1.9	1.80	1.71	1.80	1.71	1.76
Cortisone-21-acetate	3.11	4.09	3.72	4.09	3.76	3.81	0.21	0.25	0.81	0.24	0.18	0.13	2.13	1.73	1.61	1.73	1.76	1.81

DB: calculated using all available literature values; Solver.: Excel Solver; TX: TripleX program; Descfit: Simplex minimization method; Regr.: Regression Equation; Mod-reg.: Modified Regression.

Table 7 Standard deviations of different methods of calculation in comparison to database descriptors for our 40-compound test set

	S	*A*	*B*
Solver	0.29	0.15	0.15
TX	0.33	0.21	0.15
Descfit	0.29	0.15	0.15
Regressions	0.30	0.15	0.12
Mod-Regr.	0.30	0.15	0.11

Table 8 Calculation of log *P* octanol using calculated descriptors from the five mathematical methods for the 40 compound test set

Compound	log P_{DB}	log P_{solver}	log P_{Tx}	*l*og $P_{Descfit}$	log P_{Regr}	log $P_{Mod\text{-}Regr.}$	log $P_{octanol}$
n-Nitropropane	0.84	0.11	0.53	0.11	0.92	0.64	0.87
Phenol	1.54	1.06	1.64	1.06	1.24	1.04	1.47
Aniline	1.34	0.25	0.69	0.25	0.90	0.67	0.90
Hydroquinone	0.59	−0.10	0.90	−0.10	0.02	−0.12	0.59
Chlorobenzene	2.76	2.99	2.14	2.99	3.12	2.83	2.89
Toluene	2.66	2.76	3.09	2.76	2.80	2.56	2.73
Benzonitrile	1.52	1.60	1.59	1.60	1.86	1.62	1.56
2-Chlorophenol	2.00	2.35	2.44	2.35	2.30	2.11	2.15
Anisole	2.19	2.16	2.39	2.16	2.34	2.08	2.11
3-Cyanophenol	1.69	1.81	1.66	1.81	1.88	1.69	1.70
4-Fluorobenzoic acid	2.07	2.04	1.64	2.04	2.09	1.89	2.07
4-Nitrophenol	1.63	2.19	2.38	2.19	2.27	2.09	1.91
p-Toluidine	1.71	1.12	0.76	1.12	1.43	1.17	1.39
Benzamide	0.47	0.23	−0.61	0.23	0.56	0.34	0.64
3-Hydroxybenzyl alcohol	0.41	0.22	0.23	0.22	0.41	0.23	0.49
4-Nitroaniline	1.32	1.89	1.79	1.89	1.87	1.66	1.39
3-Hydroxybenzoic acid	1.47	1.34	1.26	1.34	1.44	1.27	1.50
Acetophenone	1.69	1.60	0.92	1.60	1.90	1.63	1.58
4-Iodophenol	2.85	3.30	3.43	3.30	3.37	3.24	2.91
Phenylacetic acid	1.48	1.50	1.45	1.50	1.48	1.28	1.00
2-Ethylaniline	2.37	1.89	1.61	1.89	2.15	1.90	1.74
3-Nitrobenzoic acid	1.95	2.43	2.12	2.43	2.48	2.29	1.83
Methyl 4-hydroxybenzoate	1.93	2.01	1.68	2.01	2.31	2.12	1.96
1-Naphthol	2.94	3.28	3.17	3.28	3.42	3.26	2.84
Paracetamol	0.50	0.14	0.06	0.14	0.26	0.07	0.51
n-Ethyl 4-hydroxybenzoate	2.47	2.49	2.13	2.49	2.78	2.59	2.47
3-Nitroacetanilide	1.51	2.05	1.54	2.05	2.19	1.98	1.47
Caffeine	−0.05	0.33	−0.83	0.33	0.56	0.34	−0.07
Dimethylphthate	1.59	1.94	1.42	1.94	2.11	1.84	1.56
n-Butyl 4-hydroxybenzoate	3.53	3.31	3.19	3.31	3.56	3.38	3.57
Butalbarbital	1.57	1.64	1.56	1.64	1.58	1.37	1.65
n-Heptanophenone	4.32	3.98	4.40	3.98	4.34	4.12	4.23
Procaine	1.96	2.23	0.99	2.23	2.50	2.22	2.14
Propranolol	3.00	3.33	3.10	3.33	3.85	3.66	3.48
Testosterone	3.20	2.70	2.07	2.70	3.11	2.86	3.32
Progesterone	3.49	3.12	3.35	3.12	3.57	3.28	3.87
Aldosterone	1.26	1.66	1.05	1.66	1.89	1.67	1.08
Corticosterone	2.34	2.30	2.27	2.30	2.68	2.46	1.94
Hydrocortisone	1.67	2.06	2.74	2.06	2.39	2.24	1.61
Cortisone-21-acetate	2.11	2.47	3.29	2.47	2.70	2.47	2.10
SD	**0.21**	**0.38**	**0.48**	**0.38**	**0.32**	**0.33**	

log P_{DB} = o/w partition calculated from database descriptors. log P_{solver} = o/w partition calculated using descriptors obtained from solver. log P_{Tx} = o/w partition calculated using descriptors obtained from TripleX. log $P_{Descfit}$ = o/w partition calculated using descriptors obtained from Descfit. log P_{Regr} = o/w partition calculated using descriptors obtained from Regression method. log $P_{Mod\text{-}Regr}$ = o/w partition calculated using descriptors obtained from Modified Regression method. log $P_{octanol}$ = Experimentally measured partition coefficients obtained from Medchem database.

Table 9 Comparative results of calculated descriptors from HPLC and partition coefficient methods

HPLC	*S*					*A*					*B*				
Compound	DB	Solver	Tx	Desc.	Regr.	DB	Solver	Tx	Desc.	Regr.	DB	Solver	Tx	Desc.	Regr.
Phenol	0.89	1.05	0.89	1.05	1.01	0.60	0.49	0.86	0.49	0.46	0.30	0.39	0.27	0.39	0.35
Aniline	0.96	1.21	1.10	1.21	1.12	0.26	0.01	0.27	0.01	0.13	0.41	0.65	0.56	0.65	0.49
Toluene	0.52	0.41	0.24	0.41	0.50	0.00	−0.09	0.13	−0.09	−0.21	0.14	0.14	0.10	0.14	0.10
2-Chlorophenol	0.88	1.00	0.94	1.00	0.93	0.32	0.69	0.75	0.69	0.56	0.31	0.17	0.17	0.17	0.21
4-Nitrophenol	1.72	1.61	1.59	1.61	1.32	0.82	0.81	0.93	0.80	0.77	0.26	0.13	0.08	0.13	0.19
p-Toluidine	0.95	1.02	1.23	1.02	1.02	0.23	0.02	−0.23	0.02	0.07	0.45	0.60	0.63	0.60	0.51
4-Nitroaniline	1.83	1.78	1.77	1.78	1.45	0.45	0.63	0.60	0.63	0.62	0.38	0.24	0.26	0.24	0.34
Phenylacetic acid	0.97	0.86	0.91	0.85	1.00	0.60	0.59	0.56	0.59	0.48	0.61	0.64	0.63	0.64	0.60
1-Naphthol	1.05	1.39	1.44	1.39	1.24	0.60	0.73	0.65	0.73	0.68	0.37	0.17	0.19	0.17	0.17
Procaine	1.36	2.34	2.82	2.34	2.08	0.25	0.28	−0.55	0.28	0.29	1.41	1.03	1.24	1.03	1.04
Propranolol	1.80	2.27	2.18	2.27	2.16	0.31	0.32	0.15	0.32	0.22	1.26	1.02	1.11	1.02	0.90
SD		**0.25**	**0.30**	**0.25**	**0.29**		**0.14**	**0.17**	**0.14**	**0.12**		**0.19**	**0.14**	**0.19**	**0.13**

Table 9 (*Contd.*)

Log *P* method Name	*S*					*A*					*B*				
	DB	Solver	TX	Desc.	Regr.	DB	Solver	TX	Desc.	Regr.	DB	Solver	TX	Desc.	Regr.
Phenol	0.89	0.91	0.90	0.91	0.94	0.60	0.60	0.62	0.60	0.60	0.30	0.30	0.29	0.30	0.20
Aniline	0.96	1.00	0.98	1.00	0.86	0.26	0.11	0.13	0.11	0.17	0.41	0.51	0.50	0.51	0.34
Toluene	0.52	0.45	0.48	0.45	0.46	0.00	0.00	−0.05	0.00	−0.05	0.14	0.16	0.19	0.16	0.39
2-Chlorophenol	0.88	0.38	0.36	0.38	0.49	0.32	0.38	0.42	0.38	0.38	0.31	0.41	0.38	0.41	0.17
4-Nitrophenol	1.72	1.65	1.68	1.65	1.51	0.82	0.94	0.89	0.94	0.87	0.26	0.22	0.25	0.22	0.56
p-Toluidine	0.95	1.00	1.01	1.00	0.88	0.23	0.11	0.10	0.11	0.13	0.45	0.53	0.54	0.53	0.52
4-Nitroaniline	1.83	1.87	1.88	1.87	1.62	0.45	0.47	0.46	0.47	0.49	0.38	0.35	0.36	0.35	0.48
Phenylacetic acid	0.97	1.07	1.05	1.07	1.13	0.60	0.58	0.61	0.58	0.56	0.61	0.59	0.58	0.59	0.51
1-Naphthol	1.05	1.10	1.06	1.10	1.09	0.60	0.64	0.69	0.64	0.64	0.37	0.37[a]	0.33	0.37	0.05
Procaine	1.36	1.57	1.46	1.57	1.84	0.25	0.42	0.60	0.42	0.48	1.41	1.23	1.12	1.24	0.43
Propranolol	1.80	1.91	1.67	1.89	2.39	0.31	0.92	1.45	1.09	1.22	1.26	1.13	0.78	1.01	−0.57
SD		**0.15**	**0.16**	**0.15**	**0.22**		**0.16**	**0.21**	**0.18**	**0.20**		**0.07**	**0.13**	**0.08**	**0.40**

[a] In ref. 11 the Solver calculated value for hydrogen bond basicity of 1-naphthol is erroneously given as 0.94. HPLC = this work, Log *P* method = ref. 11.

Table 10 HPLC equations obtained from all 80 compounds measured

No	HPLC System	Coefficient						Stats			
		c	*e*	*s*	*a*	*b*	*v*	*N*	r^2	SE	F
1	C18,AcN	40.891	5.803	−17.803	−22.255	−62.342	67.066	80	0.947	5.571	266.9
		2.082	2.950	1.822	2.037	2.683	2.442				
2	C18,MeOH	44.771	5.092	−12.968	−6.998	−42.918	51.677	80	0.891	5.608	121.6
		2.096	2.969	1.834	2.051	2.701	2.458				
3	C18,TFE	50.985	4.457	−14.592	−18.996	−25.324	39.643	80	0.942	3.693	242.1
		1.380	1.955	1.208	1.350	1.779	1.619				
4	FO,TFE	46.206	−2.620	−5.370	−20.939	−17.732	29.569	80	0.861	5.431	91.8
		2.030	2.0876	1.776	1.0986	2.0616	2.381				
5	PLRP,AcN	50.103	12.304	−14.237	−31.005	−55.395	49.810	80	0.947	5.571	266.9
		1.875	2.656	1.640	1.834	2.416	2.199				
6	DCN,MeOH	−5.979	9.608	−10.947	2.990	−44.016	54.145	80	0.896	5.719	127.6
		2.138	3.028	1.870	2.091	2.754	2.507				
7	DCN,AcN	−0.383	9.514	−12.028	−1.337	−44.988	51.118	80	0.866	6.119	95.7
		2.287	3.240	2.001	2.237	2.947	2.682				

Acknowledgements

We are very grateful to EPSRC for a post-doctoral fellowship to AMZ. We would also like to thank Anne Hersey of GSK for useful discussions on the regression equations.

References

1 H. van de Waterbeemd, D. A. Smith, K. Beaumont and D. K. Walker, *J. Med. Chem.*, 2001, **44**, 1313.
2 M. H. Abraham, *Pure Appl. Chem.*, 1993, **65**, 2503.
3 M. H. Abraham, Hydrogen bonding. Part 31, *J. Phys. Org. Chem.*, 1993, **6**, 660.
4 M. H. Abraham, G. S. Whiting, R. M. Doherty and W. J. Shuely, Hydrogen bonding Part16, *J. Chromatogr.*, 1991, **587**, 213.
5 M. H. Abraham, *Chem. Soc. Rev.*, 1993, **22**, 73.
6 M. H. Abraham, H. S. Chadha, G. S. Whiting and R. C. Mitchell, *J. Pharm. Sci.*, 1994, **83**, 1085.
7 M. H. Abraham, A. M. Zissimos and W. E. Acree, Jr, *Phys. Chem. Chem. Phys.*, 2001, **3**, 3732.
8 M. H. Abraham and J. Le, *J. Pharm. Sci.*, 1999, **88**, 868.
9 J. A. Platts, M. H. Abraham, Y. H. Zhao, A. Hersey, L. Ijaz and D. Butina, *Eur. J. Med. Chem.*, 2001, **36**, 719.
10 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.
11 A. M. Zissimos, M. H. Abraham, M. C. Barker, K. J. Box and K. Y. Tam, *J. Chem. Soc., Perkin Trans. 2*, 2002, 470.
12 K. Valko, S. Espinosa, C. M. Du, E. Bosch, M. Roses, C. Bevan and M. H. Abraham, *J. Chromatogr., A*, 2001, **933**, 73.
13 C. M. Du, K. Valko, C. Bevan, D. Reynolds and M. H. Abraham, *J. Chromatogr. Sci.*, 2000, **38**, 503.
14 K. Valko, M. Plass, C. Bevan, D. Reynolds and M. H. Abraham, *J. Chromatogr. A*, 1998, **41**, 797.
15 J. A. Nelder and R. Mead, *Comput. J.*, 1965, **7**, 308.
16 K. Valko, C. Bevan and D. Reynolds, *Anal. Chem.*, 1997, **69**(11), 2022–2029.
17 A. J. Leo, MedChem 2001 database, version 4.71–2, BioByte Corp., P.O. 517, Claremont, CA 91711–0157.