Eur. J. Med. Chem. 36 (2001) 719–730

PII: S0223-5234(01)01269-7/FLA

Original article

Correlation and prediction of a large blood–brain distribution data set—an LFER study

James A. Platts[a], Michael H. Abraham[b,*], Yuan H. Zhao[b], Anne Hersey[c], Luke Ijaz[b], Darko Butina[c]

[a]Department of Chemistry, Cardiff University, P.O. Box 912, Cardiff CF10 3TB, UK
[b]Department of Chemistry, University College London, 20 Gordon Street, London WC1H 0AJ, UK
[c]Mechanism and Extrapolation Technologies, GlaxoSmithKline, Park Road, Ware SG12 0DP, UK

Received 11 June 2001; accepted 7 August 2001

Abstract – We report linear free energy relation (LFER) models of the equilibrium distribution of molecules between blood and brain, as log BB values. This method relates log BB values to fundamental molecular properties, such as hydrogen bonding capability, polarity/polarisability and size. Our best model of this form covers 148 compounds, the largest set of log BB data yet used in such a model, resulting in $R^2 = 0.745$ and e.s.d. = 0.343 after inclusion of an indicator variable for carboxylic acids. This represents rather better accuracy than a number of previously reported models based on subsets of our data. The model also reveals the factors that affect log BB: molecular size and dispersion effects increase brain uptake, while polarity/polarisability and hydrogen-bond acidity and basicity decrease it. By splitting the full data set into several randomly selected training and test sets, we conclude that such a model can predict log BB values with an accuracy of less than 0.35 log units. The method is very rapid—log BB can be calculated from structure at a rate of 700 molecules per minute on a silicon graphics O^2. © 2001 Éditions scientifiques et médicales Elsevier SAS

blood–brain distribution / hydrogen-bonding / LFER / linear free energy relation / solute descriptors

1. Introduction

Brain uptake, or the ability of a molecule to enter brain tissue, has been a subject of great interest to the pharmaceutical industry for over 30 years; for recent reviews see [1, 2]. Although measures such as biological activity [3] and 'brain uptake index' [4, 5] have been proposed, the two measures that are most amenable to physicochemical analysis are: (a) brain perfusion; and (b) blood–brain distribution. The former is obtained as a rate constant from experiments over a very short time scale. Because of the variation in experimental procedure, the various sets of available data are not compatible, and so only rather restricted data sets have been analysed [6]. The latter is obtained from much longer time scale experiments, and since results from a number of workers seem self-consistent, it has been possible to build a much larger database. Blood–brain distribution, BB, is the measure we shall analyse defined through Eq. (1).

$$\text{BB} = (\text{conc. in brain})/(\text{conc. in blood}) \quad (1)$$

Young et al. [7] published the first analysis of brain uptake in terms of BB, reporting in vivo values in rats for a large number of H2 receptor histamine agonists, and modelling these values using water–octanol and water–cyclohexane partition coefficients log P(octanol) and log P(cyclohexane). Abraham et al. [8] and Chadha et al. [9] added a further 37 BB values found indirectly from solubilities in blood and brain: many attempts to model this and related data sets have been made, and are summarised below.

Lombardo et al. [10] correlated log BB against the free energy of solvation in water, obtaining a reasonable linear fit. Kelder et al. [11] used the polar surface

* Correspondence and reprints
E-mail address: m.h.abraham@ucl.ac.uk (M.H. Abraham).

area (PSA) as a descriptor of log BB, with encouraging results, but we note that the resulting model, Eq. (2),

$$\log BB = 1.33 - 0.322\ PSA \quad (2)$$

is physically unrealistic since every compound with zero PSA is predicted to have log BB = 1.33 (however, this may not be of great concern to drug designers, since no drug will have zero PSA). Clark [12] also used PSA to model log BB, this time in combination with log *P*(octanol), to yield a slightly better model. Norinder et al. [13] and Luco [14] both used large numbers of descriptors and partial least squares (PLS) methods to correlate log BB with generally excellent results. Luco collated a total of 100 BB values, but made no attempt to combine these into one data set, restricting his training model to just 58 compounds. Exactly the same data set of 100 compounds was later analysed by Feher et al. [15]; using 61 compounds as a training set, a statistically very reasonable equation was obtained with only three descriptors:

$$\log BB = 0.4275 - 0.3873\ na,\ aq + 0.1092 \log P(\text{octanol}) - 0.0017\ PSA \quad (3)$$

$$n = 61,\ R^2 = 0.730,\ R^2_{CV} = 0.688,\ \text{e.s.d.} = 0.424,\ F = 51$$

Here, and in all that follows, n is the number of points used in the regression, R^2 is the square of the overall correlation coefficient, R^2_{CV} is the cross-validated correlation coefficient, e.s.d. is the standard deviation, and F is Fischer's F-statistic. The descriptor na, aq in Eq. (3) represents the number of hydrogen-bond acceptors in an aqueous medium. Although this seems to be rather a simple descriptor, it is not obvious how it is to be determined or estimated. Eq. (3) was applied to 61 compounds out of the 100 listed. Two test sets of compounds were constructed from the remaining 39 compounds; no statistical details were given, but we have calculated that for the two test sets the e.s.d. values were 0.76 ($n = 14$) and 0.80 ($n = 25$). The most recent quantitative analysis [16] uses the free energy of solvation [13] as the sole descriptor. For a training set of 55 compounds Keserü and Molnár [16] obtained a correlation equation with $R^2 = 0.722$ and e.s.d. = 0.37 log units. A number of test sets were studied; values of e.s.d. ranged from 0.14 ($n = 5$) to 0.37 ($n = 25$). The lower e.s.d. value is clearly an artefact, because the experimental error in log BB must be around 0.30 units, but the e.s.d. value of 0.37 for 25 test compounds is impressive. It was suggested [16] that the calculation of log BB via the free energy of solvation was the fastest method to date, at >10 molecules per minute.

Summaries of the various models of blood–brain distribution that used reasonably large data sets are shown in *table I*, where details of the original model of Abraham et al. [8] are given for comparison. There are a number of models that give e.s.d. values of about 0.35 log units for a training set, and slightly higher values between observed and predicted log BB values of a test set. This appears to be close to the accuracy limit of models constructed with large datasets, and e.s.d. = 0.35 is the accuracy for which we aim here. Note that most models exclude a number of compounds from the analysis; invariably these are compounds that have much more negative observed log BB values than calculated. There are several possible reasons for such outliers. For example, if the analytical method is radiochemical detection, any biological degradation will lead to much smaller observed values than calculated. Efflux mechanisms [17], most notably by poly-glycoprotein [18], will also result in more negative observed values than calculated.

In addition to the models of blood–brain distribution, shown in *table I*, Crivori et al. [19] have recently described an analysis of brain uptake of 120 compounds using 3D VolSurf parameters. Their analysis was qualitative only, classifying compounds as +, +/−, or −, resulting in a remarkably rigorous model, considering that quite different measures of brain uptake were used; these included the equilibrium blood–brain distribution that we have dealt with, and the kinetic rate of perfusion from saline to brain. In the present work, we have been careful to use only a single measure of uptake, namely blood–brain distribution.

2. Chemistry

The general equation we use is the same as that originally employed by Abraham et al. [8], and subsequently reviewed [2, 20]. We use a simplified notation, and write the equation as follows:

$$SP = c + e.E + s.S + a.A + b.B + v.V \quad (4)$$

Here SP is a set of solute properties in a given system, e.g. a set of log BB values. The independent variables are solute descriptors: E is an excess molar refraction, S is the dipolarity/polarisability, A and B are the hydrogen-bond acidity and basicity, respectively, and V is the solute McGowan volume in units of ($cm^3\ mol^{-1}/100$). A wide range of solvation and transport processes has been modelled in this manner [20].

One difficulty over the application of Eq. (4) has been the lack of experimental descriptors, especially for drug molecules. Recently, however, we have developed [21] a general method for the prediction of the descriptors in Eq. (4) for any organic molecule. The calculated descriptors, together with equations on the lines of Eq. (4), can reproduce literature values of water–solvent partition coefficients [21b], cell permeation data [21c] and intestinal absorption in man [21d]. In this study, we will apply the same methodology to the sets of log BB data that we set out below, with the aim of establishing such calculations as a general method for the 'high-throughput' prediction of equilibrium blood–brain distribution from structure.

In the six years since the publication [8] of the log BB model, based on Eq. (4), a number of values of log BB for new compounds have appeared in the literature [11, 13, 22]. Most previous workers used the original large set of Abraham et al. [8]; values of log BB for drug molecules were obtained by essentially the same procedure set out by before [7, 23]. However, the data set of Salminen et al. [22] is quite different, and most blood–brain ratios listed are actually plasma–brain (PB) ratios. We have used these PB ratios as estimates of BB ratios, but note the particular compounds with PB ratios in *table II*. However, we have had to exclude a number of compounds that were used in the analysis of Salminen et al. [22] as follows. The PB ratio for acetylsalicylic acid can only be obtained very approximately from graphs in the original Ref. [24]. The data for valproic acid refer to experiments with 17-day-old rats [25] and so are not commensurate with the general data on mature rats; in any case equilibrium conditions were not established. The value for ibuprofen is that from a post mortem analysis [26]. In addition to the new data on drugs [11, 13, 22, 27–42], several indirect values determined in the manner of Abraham et al. [8] have also been obtained.

Table I. Summary of recent models of log BB.

Equation	Training set			Test set	
	D [a]	No	E.s.d.	No	E.s.d.
Log P(oct) [b]	1	49	0.50	0	–
Lombardo [c]	1	55	0.41	6	0.62
Kelder [d]	1	45	0.36	0	–
Clark [e]	2	55	0.35	10	0.20
Norinder et al. [f]	14	56	0.31	6	0.55
Luco [g]	18	58	0.32	37	0.37
Feher et al. [h]	3	61	0.42	14, 25	0.76, 0.80
Keserü, Molnár [i]	1	55	0.37	5, 25	0.14, 0.37
Abraham et al. [b]	5	57	0.20	0	–
This work	6	148	0.34	0	–
This work [j]	6	74	0.34	74	0.38

[a] Number of descriptors used.
[b] Calculated in Ref. [8].
[c] Ref. [10].
[d] Ref. [11].
[e] Equation DEC-I, Ref. [12].
[f] Ref. [13].
[g] Ref. [14].
[h] Ref. [15].
[i] Ref. [16].
[j] See also Table 7, where the average e.s.d. is 0.36 for five test sets each with $n = 30$.

A total of 157 values of log BB has been collated from a number of sources including directly measured and indirectly determined values: these are presented in *table II*. We do not repeat the various structures, but a complete list of SMILES strings and log BB values is available on request. Experimental LFER descriptors for many of these compounds have previously been reported [8, 20] while others had to be determined from literature log P values [43]. In total, experimental descriptors were available for 112 compounds (new compounds, mainly drugs, are reported in *table III*), and the remaining 45 had descriptors calculated using the method of Platts et al. [21a]. Coefficients e, s, a, b, and v in Eq. (3) were determined by multivariate linear regression analysis (MLRA). Statistical analyses were performed using the JMP package published by SAS Software Inc. Parameters were included in the regression analyses if a standard t-test indicated a >95% probability of significance.

3. Results

An initial regression of all 157 log BB values in *table II* against LFER descriptors showed a reasonably accurate fit, but nine molecules had to be omitted as outliers (see below for details). The remaining set of 148 values yielded the following equation:

$$\log \mathrm{BB} = 0.044 + 0.511\,E - 0.886\,S - 0.724\,A - 0.666\,B + 0.861\,V \qquad (5)$$

$$n = 148,\ R^2 = 0.710,\ R^2_{\mathrm{CV}} = 0.682,\ \text{e.s.d.} = 0.367,\ F = 71$$

Eq. (5) is a reasonable model of log BB, with many similarities to our previous [8] equation and e.s.d. close to those found in several other studies using large data sets (see *table I*). However, several large discrepancies were observed for carboxylic acid-containing molecules, such as salicylic acid and indomethacin. An improvement over Eq. (5) can be achieved by the inclusion of I_1, an indicator variable that is set to 1 for a compound containing a carboxylic acid fragment and 0 otherwise, giving Eq. (6):

$$\log \mathrm{BB} = 0.021 + 0.463\,E - 0.864\,S - 0.564\,A - 0.731\,B + 0.933\,\mathrm{Vx} - 0.567\,I_1$$

$$n = 148,\ R^2 = 0.745,\ R^2_{\mathrm{CV}} = 0.711,\ \text{e.s.d.} = 0.343,\ F = 69 \qquad (6)$$

Values of calculated and residual log BB from Eq. (6) are reported in *table II*, and the fit to observed values shown graphically in *figure 1*. The molecules omitted from Eq. (6) are **4** (calc. = 0.25), **9** (−0.40), **20** (−0.31), **21** (0.11), Y-G19 (−0.08), Y-G20 (−0.56), thioridazine (1.23), Org12692 (−0.09), and fluphenazine (0.41). With the exception of fluphenazine, all these compounds have been omitted from previous log BB models [8, 11–13, 22].

The statistics of Eq. (6) are not as impressive as those for the original model [8], see *table I*, which is to be expected given the increased chemical diversity of the data set and the number of sources from which the data were taken. Indeed, Eq. (6) is trained on the largest set of log BB data yet published; it is encouraging that data sets from different studies can be combined into one model in this way. In terms of standard deviation, Eq. (6) is close to the best models in *table I*. Considering the size and diversity of the training set used, we consider this to be the most general model of blood–brain distribution yet reported.

Eq. (6) therefore appears satisfactory, but *table IV* indicates there are several large inter-correlations of descriptors, particularly E, S, B, and V, for this set of compounds. It is possible that this non-orthogonality of descriptors affects the coefficients of Eq. (6), introducing non-physical factors and hindering interpretation. Such correlations may be removed by using principal component regression (PCR), in which log BB data are linearly regressed against orthogonal PCs formed from the original descriptors. In the present case, five PCs are found to be significant on regression against log BB, yielding $R^2 = 0.744$ and e.s.d. = 0.342. The resulting equation, expressing log BB in terms of the PCs, can be transformed back into the original descriptors to yield Eq. (7):

$$\log \mathrm{BB} = 0.062 + 0.469\,E - 0.864\,S - 0.586\,A - 0.713\,B + 0.895\,V - 0.564\,I_1 \qquad (7)$$

$$n = 148,\ R^2 = 0.744,\ R^2_{\mathrm{CV}} = 0.715,\ \text{e.s.d.} = 0.342,\ F = 83$$

Eqs. (6) and (7) are identical within the standard errors on coefficients (not shown), so we conclude that the coefficients in Eq. (6) are not greatly affected

Table II. Observed, and calculated values of log BB from Eq. (6).

Name	Observed	Calculated	Residual	References
Neon	0.20	0.04	0.16	[a]
Argon	0.03	0.15	−0.12	[a]
Krypton	−0.16	0.21	−0.37	[a]
Xenon	0.03	0.30	−0.27	[a]
Nitrogen	0.03	0.19	−0.16	[a]
Nitrous oxide	0.03	−0.12	0.15	[a]
Methane	0.04	0.22	−0.18	[a]
Pentane	0.76	0.81	−0.05	[a]
Hexane	0.80	0.96	−0.16	[a]
2-Methylpentane	0.97	0.96	0.01	[a]
3-Methylpentane	1.01	0.96	0.05	[a]
2,2-Dimethylbutane	1.04	0.96	0.08	[a]
Heptane	0.81	1.11	−0.30	[a]
3-Methylhexane	0.90	1.11	−0.21	[a]
Cyclopropane	0.00	0.31	−0.31	[a]
Cyclohexane	0.92	0.85	0.07	[a]
Methylcyclopentane	0.93	0.82	0.11	[a]
Dichloromethane	−0.11	−0.04	−0.07	[a]
Trichloromethane	0.29	0.16	0.13	[a]
1,1,1-Trichloroethane	0.40	0.43	−0.03	[a]
Trichloroethylene	0.34	0.45	−0.11	[a]
1,1,1-Trifluoro-2-chloroethane	0.08	0.07	0.01	[a]
Halothane	0.35	0.27	0.08	[a]
Teflurane	0.27	0.25	0.02	[a]
Diethylether	0.00	0.25	−0.25	[a]
Divinyl ether	0.11	0.27	−0.16	[a]
Methoxyflurane	0.25	0.14	0.11	[a]
Isoflurane	0.42	0.11	0.31	[a]
Enflurane	0.24	0.18	0.06	[a]
Fluroxene	0.13	0.36	−0.23	[a]
Propanone	−0.15	−0.38	0.23	[a]
Butanone	−0.08	−0.22	0.14	[a]
Ethanol	−0.16	−0.41	0.25	[a]
Propan-1-ol	−0.16	−0.26	0.10	[a]
Propan-2-ol	−0.15	−0.24	0.09	[a]
2-Methylpropan-1-ol	−0.17	−0.10	−0.07	[a]
SF_6	0.36	0.44	−0.08	[a]
CS_2	0.60	0.49	0.11	[a]
Benzene	0.37	0.33	0.04	[a]
Toluene	0.37	0.48	−0.11	[a]
Ethylbenzene	0.20	0.63	−0.44	[a]
p-Xylene	0.31	0.62	−0.30	[a]
m-Xylene	0.29	0.62	−0.33	[a]
o-Xylene	0.37	0.60	−0.23	[a]
1 (Cimetidine)	−1.42	−0.77	[b]	[7]
2	−0.04	−0.33	0.29	[7]
4	−1.30	0.12	[b]	[7]
5	−1.06	−0.36	−0.70	[7]
6 (Clonidine)	0.11	−0.42	0.53	[7]
7 (Mepyramine)	0.49	0.99	−0.50	[7]
8 (Imipramine)	1.06	0.69	0.37	[7]
9 (Ranitidine)	−1.23	−0.02	[b]	[7]
10 (Tiotidine)	−0.82	−0.83	0.01	[7]
13	−0.67	−0.34	−0.33	[7]
14	−0.66	−0.46	−0.20	[7]
15	−0.12	−0.13	0.01	[7]

Table II. (Continued)

Name	Observed	Calculated	Residual	References
16	−0.18	−0.22	0.04	[7]
17	−1.15	−0.75	−0.40	[7]
18	−1.57	−1.53	−0.04	[7]
19	−1.54	−1.23	−0.31	[7]
20	−1.12	−0.26	[b]	[7]
21	−0.73	0.15	[b]	[7]
22	−0.27	−0.15	−0.12	[7]
23	−0.28	−0.42	0.14	[7]
24	−0.46	−0.56	0.10	[7]
25	−0.24	0.07	−0.31	[7]
26	−0.02	0.17	−0.19	[7]
27	0.69	0.30	0.39	[7]
28	0.44	0.06	0.38	[7]
29	0.14	0.42	−0.28	[7]
30	0.22	0.25	−0.03	[7]
31	0.00	−0.43	0.43	[13]
36	0.89	1.24	−0.35	[13]
Y-G 14	−0.30	−0.24	−0.06	[9]
Y-G 15	−0.06	0.10	−0.16	[9]
Y-G 16	−0.42	−0.37	−0.05	[9]
Y-G 19	−1.30	−0.11	[b]	[9]
Y-G 20	−1.40	−0.60	[b]	[9]
SKF 89124	−0.43	−0.31	−0.12	[9]
SKF 101468	0.25	0.48	−0.23	[9]
Acetylsalicylic acid	−0.50	−0.75	0.25	[22]
Valproic acid	−0.22	−0.49	0.27	[22]
Theophylline	−0.29	−0.91	0.62	[22]
Caffeine	−0.05	−0.39	0.34	[22]
Antipyrine	−0.10	−0.24	0.14	[22]
Salicylic acid	−1.10	−0.87	−0.23	[22]
Acetaminophen	−0.31	−1.22	[b]	[22]
Ibuprofen	−0.18	−0.23	0.05	[22]
Codeine	0.55	−0.22	0.77	[22]
Pentobarbital	0.12	0.12	0.00	[22]
Alprazolam	0.04	−0.03	0.08	[22]
Indomethacin	−1.26	−0.92	−0.34	[22]
Oxazepam	0.61	0.58	0.03	[22]
Hydroxyzine	0.39	0.45	−0.06	[22]
Desipramine	1.20	0.71	0.49	[22]
Midazolam	0.36	0.49	−0.13	[22]
Promazine	1.23	0.84	0.39	[22]
Chlopromazine	1.06	0.38	0.68	[22]
Trifluoperazine	1.44	1.09	0.35	[22]
Thioridazine	0.24	1.20	[b]	[22]
32	−0.34	−0.19	−0.15	[13]
33	−0.30	−0.20	−0.10	[13]
34	−1.34	−0.99	−0.35	[13]
35	−1.82	−1.78	−0.04	[13]
Mianserin	0.99	0.69	0.30	[11]
Org4428	0.82	0.17	0.65	[11]
Org5222	1.03	0.55	0.48	[11]
Org12692	1.64	−0.09	[b]	[11]
Org13011	0.16	0.23	−0.07	[11]
Org32104	0.52	−0.14	0.66	[11]
Org30526	0.39	0.25	0.14	[11]
Mirtazapine	0.53	0.29	0.24	[11]

Table II. (Continued)

Name	Observed	Calculated	Residual	References
Tibolone	0.4	−0.19	0.59	[11]
Org34167	0.00	0.58	−0.58	[11]
Risperidone	−0.02	0.19	−0.21	[11]
9-OH Risperidone	−0.67	−0.25	−0.42	[11]
MIL-663581	−1.34	−0.83	−0.51	[11]
M2L-663581	−1.82	−1.34	−0.48	[11]
Theobromine	−0.28	−0.82	0.54	[27, 28]
Morphine	−0.16	0.08	−0.24	[29]
Propanolol	0.64	0.17	0.47	[30]
Atenolol	−1.42	−0.93	−0.49	[31]
Diazepam	0.52	0.69	−0.17	[32]
Phenytoin	−0.04	−0.55	0.51	[33]
Hexobarbital	0.10	0.10	0.00	[34]
Amobarbital	0.04	0.05	−0.01	[34]
Phenylbutazone	−0.52	−0.05	−0.47	[35]
Aminopyrine	0.00	0.05	−0.05	[36]
Desmethydesipramine	1.06	0.79	0.27	[37]
Bretazenil	−0.09	0.09	−0.18	[38]
Flumanezil	−0.29	−0.54	0.25	[38]
RO19-4603	−0.25	−0.21	−0.04	[38]
Paraxanthine	0.06	−0.75	0.81	[27, 28]
Quinidine	−0.46	0.30	−0.76	[30]
Salicyluric acid	−0.44	−1.21	0.77	[37]
Fluphenazine	1.51	0.45	[b]	[40, 41]
Haloperidol	1.34	0.80	0.54	[40, 41]
Mesoridazine	−0.36	−0.13	−0.23	[41]
Sulforidazine	0.18	0.09	0.09	[41]
Bromperidol	1.38	0.90	0.48	[41]
Northioridazine	0.75	1.10	−0.35	[41]
Nor-1-chlorpromazine	1.37	0.80	0.57	[41]
Nor-2-chlorpromazine	0.97	0.70	0.27	[41]
Desmonomethylpromazine	0.59	0.66	−0.07	[41]
Desmethyldiazepam	0.50	−0.04	0.54	[32]
1-Hydroxymidazolam	−0.07	−0.23	0.16	[32]
4-Hydroxymidazolam	−0.30	−0.21	−0.09	[32]
Triazolam	0.74	0.45	0.29	[32]
Clobazam	0.35	0.49	−0.14	[32]
Flunitrazepam	0.06	−0.09	0.15	[32]
Desmethylclobazam	0.36	0.03	0.33	[32]
Thiopental	−0.14	−0.24	0.10	[34]
Methohexital	−0.06	0.35	−0.41	[34]
Didanosine	−1.30	−1.11	−0.19	[42]
Indinavir	−0.74	−0.7	−0.04	[42]
Nevirapine	0.00	−0.38	0.38	[42]
Zidovudine	−0.72	−0.78	0.06	[42]

[a] Indirect values, [8].
[b] Omitted from final model.

by the inter-correlation of descriptors. We have used the standard deviation, e.s.d., as the main criterion for evaluating the quality of the various models of log BB. On this basis we conclude that Eq. (6) is as good as, or better than, models based on subsets of the data used here. While e.s.d. is an accurate measure of each model's ability to fit the available data, it is the predictive ability of these models that is of real interest, as the main purpose of such models is to apply them to new compounds. The R^2_{CV} value is only a measure of internal self-consistency, and to test the predictive power of any model it is necessary to split

Table III. Newly determined LFER descriptors.

Name	*E*	*S*	*A*	*B*	*V*
Cyclopropane	0.41	0.23	0.00	0.00	0.4227
Divinyl ether	0.26	0.39	0.00	0.13	0.6449
4	3.69	3.18	0.62	2.16	3.4468
5	2.94	2.98	0.61	2.43	3.1776
7	1.93	1.25	0.00	1.64	2.3870
8	1.15	1.60	0.00	1.15	2.4020
14	2.15	2.20	0.47	1.43	2.0043
23	2.58	2.60	0.46	1.98	2.5484
24	1.39	2.50	0.40	1.55	2.4317
26	1.28	1.50	0.37	1.30	2.0931
31	2.15	2.11	0.53	1.10	1.8106
36	1.87	1.30	0.00	1.10	2.3996
Acetylsalicylic acid	0.78	0.80	0.49	1.00	1.2879
Valproic acid	0.14	0.55	0.60	0.50	1.3102
Acetaminophen	1.06	1.63	1.04	0.86	1.1724
Ibuprofen	0.70	0.92	0.60	0.60	1.7771
Codeine	1.78	1.95	0.33	1.78	2.2057
Pentobarbital	1.03	1.11	0.47	1.23	1.7966
Alprazolam	2.90	2.50	0.00	1.55	2.2041
Indomethacin	2.24	2.85	0.40	1.08	2.5299
Oxazepam	2.35	1.10	0.45	1.60	1.9917
Hydroxyzine	2.00	2.21	0.10	1.89	2.9231
Desipramine	1.62	1.64	0.10	0.92	2.2606
Midazolam	2.57	2.01	0.00	1.38	2.2629
Promazine	2.05	1.70	0.00	1.01	2.2832
Chlorpromazine	2.20	1.83	0.00	1.94	2.4056
Trifluoperazine	1.99	1.80	0.00	1.50	2.8911
Thioridazine	2.71	2.05	0.00	1.32	2.9017
Phenylbutazone	1.85	2.62	0.00	1.28	2.4239
Fluphenazine	2.16	2.30	0.26	1.80	3.0907
Haloperidol	1.90	1.39	0.40	1.76	2.7980
Bromperidol	2.07	1.39	0.40	1.80	2.8506

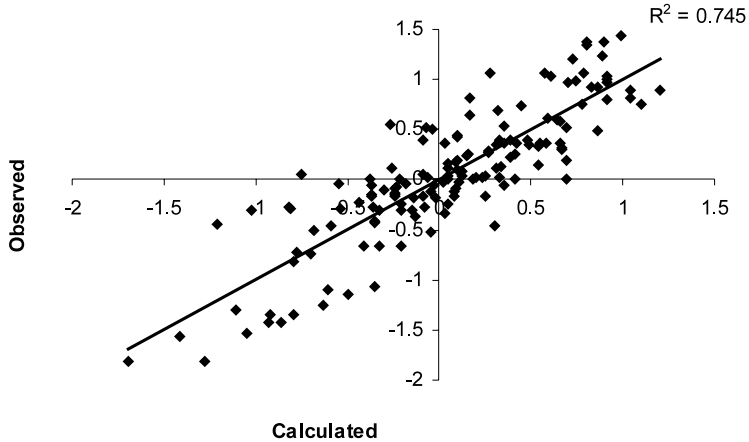

Figure 1. Observed vs. calculated values of log BB on Eq. (6).

Table IV. Inter-correlations of descriptors in Eq. (6).

Variable	E	S	A	B	V	I_1
E	1.00	–	–	–	–	–
S	0.93	1.00	–	–	–	–
A	0.34	0.39	1.00	–	–	–
B	0.85	0.88	0.46	1.00	–	–
V	0.88	0.88	0.30	0.87	1.00	–
I_1	0.09	0.05	0.25	0.09	0.04	1.00

Table V. Statistics for test sets of 30 compounds.

Set no.	R^2	E.s.d.
1	0.643	0.418
2	0.711	0.371
3	0.763	0.347
4	0.773	0.319
5	0.775	0.338
Average	0.733	0.356

the full set of data into training and test sets. A new model is developed using only those data in the training set, which is then used to predict values for the test set.

Accordingly, we have split the full set of 148 compounds in two sets of 74, using the first set to construct an equation similar to Eq. (6) and predicting log BB values for the second set, and vice versa. The two regressions yielded $R^2 = 0.767$, e.s.d. = 0.331 and $R^2 = 0.763$, e.s.d. = 0.344, with coefficients (not shown) identical within regression standard errors to Eq. (6). These regressions were then used to predict log BB values for the 74 molecules not used in training sets. We find for the 'test' sets of 74 compounds that $R^2 = 0.735$, e.s.d. = 0.376 and $R^2 = 0.701$, e.s.d. = 0.385. We have also taken five randomly selected test sets, each of 30 compounds (20% of the full set) and re-built models analogous to Eq. (6) using the remaining 118 compounds as training sets. These models, each of which is statistically identical to Eq. (6), were then used to predict log BB for the 30 test compounds. Results are summarised in *table V*. The accuracy of prediction shows considerable variation among the five test sets, with e.s.d. (defined as $\sqrt{(\Sigma(\text{obs}-\text{calc})^2/n-1)}$) ranging from 0.32 to 0.42, but the average e.s.d. is 0.36, almost identical to that found in Eq. (6), which can be taken as an estimate of the likely predictive ability of Eq. (6).

4. Discussion

Because the descriptors we use in the general Eq. (4) reflect solute/solvent interactions, we can interpret the coefficients in terms of the effect that particular interactions have on the process under consideration. The positive e- and v-coefficients in Eq. (6) indicate that increasing molecular size and (less importantly) the presence of n- and π-electron pairs tend to push compounds out of blood and into brain. In this respect, Eq. (6) is similar to many water–solvent partition (log P) equations, where also large e- and v-coefficients are found. The coefficients of the 'polar' descriptors S, A, and B are all very negative indicating that polarity and hydrogen-bond donor and acceptor ability act to keep compounds in blood and out of the brain. In turn, this indicates that the brain is less polar/polarisable, acidic and basic than is blood, but is more able to accommodate bulky solutes and interact with them via dispersive forces than blood. Again this is similar to many previously published log P equations, where the aqueous phase attracts polar, hydrogen-bonding molecules and the organic phase favours solution of bulky solutes.

The large negative coefficient of the indicator variable, I_1, in Eq. (6) and its low correlation with other descriptors (*table IV*), indicates a hitherto unknown factor affecting the transfer of solutes between blood and brain. It is apparent that the presence of a $-CO_2H$ group acts to hinder brain penetration further than would simply be due to the intrinsic hydrogen bonding and polarity properties of neutral acids. It is known [44] that acidic drugs will, in general, bind to albumin present in plasma and blood, which may account for at least some of the negative contribution of I_1. It may also be that ionisation to $-CO_2^-$ is the cause of this, as one can certainly envisage such a group increasing E and B over those for $-CO_2H$. Finally, such $-CO_2H$ groups may be removed from the brain by some efflux mechanism. It is worth noting that similar indicator variables for strong bases, such as amines, are not found to be significant in our models, or indeed in previous analyses of log BB [8, 22].

It is not possible to compare our interpretation of the chemical interactions that influence blood–brain distribution, with any similar interpretation of other models. Most of the models shown in *table I* use descriptors that are themselves combinations of solute/solvent effects. The free energy of solvation [10,

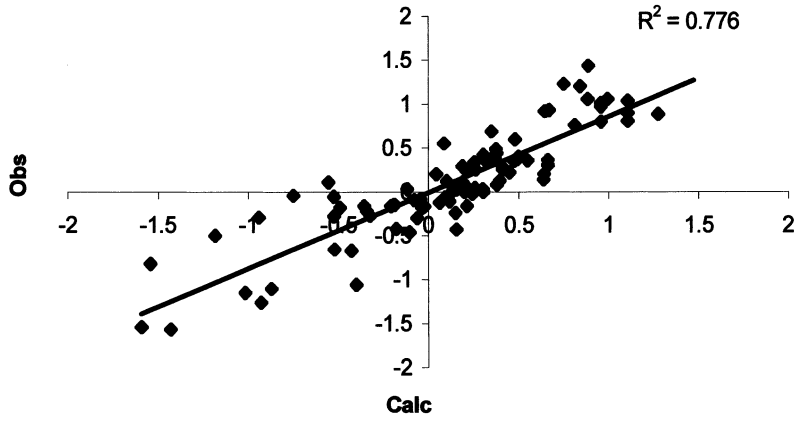

Figure 2. Observed vs. calculated values of log BB using calculated descriptors.

16], or the water–octanol partition coefficient [15] are influenced by a number of interactions, as we have shown previously [45, 46].

Keserü and Molnár [16] suggested that their calculation of log BB via a calculated free energy of solvation was the fastest method of calculation to date, at >10 molecules per minute. However, our recently developed fragment contribution method for fast, automated descriptor calculation [21a] may be employed to predict the required descriptors in Eq. (6), and thus log BB values at a rate of 700 molecules per minute, some two orders of magnitude faster than the calculations of Keserü and Molnár [16]. To test the likely accuracy of such predictions, we have taken the 105 compounds for which experimental descriptors were used in constructing Eq. (6) and have calculated their log BB values using Eq. (6) with descriptors calculated by the method of Platts et al. [21a]. The observed and calculated values agree with e.s.d. = 0.294, thus indicating that Platts et al.'s calculation method can be used in conjunction with Eq. (6) to predict further log BB values with an estimated accuracy of around 0.30 to 0.35 log units. A plot of the observed versus calculated log BB values for these 105 compounds is shown in *figure 2.*

Unlike other models for blood–brain distribution, the present one is constructed from a general equation. The various descriptors we have used are common to those we have used in the analysis of a number of processes, and so the coefficients obtained can be used to compare their correlational similarity. One elegant method of quantifying the similarity of equations that use the same descriptors is that of Ishihama and Asakawa [47]. Equations are treated as vectors in 'descriptor space', with θ defined as the angle between such vectors. If the coefficients for two LFERs constructed on the lines of Eq. (4) lead to a value of θ close to zero or 180°, then the two LFERs will be closely correlated linearly, while a value of θ near 90° indicates that the two LFERs are not linearly correlated. Given in *table VI* are values of the angle θ for several common water–solvent partitions against Eq. (5) as the standard. Note that we use Eq. (5) rather than Eq. (6) in order that the form of all the equations shall be the same. It is apparent that θ is large for all the systems shown in *table VI*, other than the standard equation, so that none of these will show a good linear relationship to log BB. It is interesting to note that two properties previously used as models of log BB for subsets of the current data do not represent good linear correlative models of log BB. $\Delta \log P$, defined [48] as $\log P$(octanol)–

Table VI. Angles between log BB model, Eq. (5) and selected partition equations.

Solvent system	Angle (θ)
Water–octanol	39.4
$\Delta \log P$[a]	43.4
Water–cyclohexane	25.9
Water–gas phase[b]	39.1
Water–chloroform	34.4
Water–ether	41.2
Water–ethyl acetate	35.3
Water–benzene	33.6
Ethylene glycol–heptane	32.1

[a] Defined as $\log P$(octanol) − $\log P$(cyclohexane).
[b] Equivalent to the free energy of solvation.

log P(cyclohexane), was found by Young et al. [7] to be superior to both water–octanol and water–cyclohexane partition in modelling log BB of 20 compounds. We find, however, that it is less correlated to log BB, using the equation, Eq. (5) that refers to many more compounds. Water–gas partition is equivalent to the free energy of solvation used [10, 16] to model log BB—here we find it only slightly better than Δ log P and worse than every other water–solvent partition considered. It is also worth noting that the octanol–water partition coefficient, as log P(octanol), widely used as a measure of lipophilicity, does not appear to be closely correlated to blood–brain distribution.

5. Conclusions

We have developed and tested a LFER model for the equilibrium distribution of solutes between blood and brain, log BB. After collating data from several sources to yield a data set of 157 log BB values, a model was constructed using MLRA. The best fit resulted from an equation in which shows that size strongly enhances brain uptake, and polarity/polarisability, hydrogen-bond acidity, basicity, and the presence of carboxylic acid groups strongly retard brain penetration. Thus, we are able to correlate log BB to six descriptors with a standard deviation of less than 0.35 log units, similar to the best reported models of log BB constructed using subsets of our current data set. Principal component regression confirms the accuracy of the linear regression equation. Splitting the data into five distinct, randomly selected test and training sets confirms this predictive ability. We have also demonstrated that our method for estimating molecular descriptors, based on functional group contributions, is a powerful predictor of log BB. Applying the equation discussed above to calculated descriptors predicts log BB with an e.s.d. of around 0.30 log units. We conclude by suggesting that the equations presented here are capable of predicting log BB values to around 0.35 log units. Additionally, the computation of log BB using Platts' method of determining descriptors is much faster than most previous methods. The calculation using the free energy of solvation is quite rapid, at >10 molecules per minute [16], but our method can generate log BB values for 700 molecules per minute on a silicon graphics O^2.

Acknowledgements

J.A.P. and Y.H.Z. are grateful to GlaxoWellcome for post-doctoral fellowships. A.H. and D.B. are members of GlaxoSmithkline's Mechanism and Extrapolation Technologies group.

References

[1] Begley M., Bradbury M.W., Kreuter J. (Eds.), The blood–Brain Barrier and Drug Delivery to the CNS, Marcel Dekker, New York, 2000.

[2] Abraham M.H., Platts J.A., in: Begley M., Bradbury M.W., Kreuter J. (Eds.), The blood–Brain Barrier and Drug Delivery to the CNS, Marcel Dekker, New York, 2000, pp. 9–32.

[3] Hansch C., Steward A.R., Anderson S.M., Bentley D.L., J. Med. Chem. 11 (1968) 1.

[4] Oldendorf W.H., Brain Res. 24 (1970) 372.

[5] Oldendorf W.H., Am. J. Physiol. 221 (1971) 1629.

[6] Gratton J.A., Abraham M.H., Bradbury M.W., Chadha H.S., J. Pharm. Pharmacol. 49 (1997) 1211.

[7] Young R.C., Mitchell R.C., Brown T.H., Ganellin C.R., Griffiths R., Jones M., Rana K.K., Saunders D., Smith I.R., Sore N.E., Wilks T.J., J. Med. Chem. 31 (1988) 656.

[8] Abraham M.H., Chadha H.S., Mitchell R.C., J. Pharm. Sci. 83 (1994) 1257.

[9] Chadha H.S., Abraham M.H., Mitchell R.C., Bioorg. Med. Chem. Lett. 4 (1994) 2511.

[10] Lombardo F., Blake J.F., Curatolo W.J., J. Med. Chem. 39 (1996) 4750.

[11] Kelder J., Grootenhuis P.D.J., Bayada D.M., Delbressine L.P.C., Ploemen J.-P., Pharm. Res. 16 (1999) 1514.

[12] Clark D.E., J. Pharm. Sci. 88 (1999) 815.

[13] Norinder U., Sjoberg P., Osterberg T., J. Pharm. Sci. 87 (1998) 952.

[14] Luco J.M., J. Chem. Inf. Comput. Sci. 39 (1999) 396.

[15] Feher M., Sourial E., Schmidt J.M., Int. J. Pharm. 201 (2000) 239.

[16] Keserü G.M., Molnár L., J. Chem. Inf. Comput. Sci. 41 (2001) 120.

[17] Begley D.J., Regina A., Roux F., Khan E.U., Rollinson C., Abbott J., in: Begley M., Bradbury M.W., Kreuter J. (Eds.), The blood–Brain Barrier and Drug Delivery to the CNS, Marcel Dekker, New York, 2000, pp. 93–108.

[18] Mayer U., in: Begley M., Bradbury M.W., Kreuter J. (Eds.), The blood–Brain Barrier and Drug Delivery to the CNS, Marcel Dekker, New York, 2000, pp. 109–120.

[19] Crivori P., Cruciani G., Carrupt P.-A., Testa B., J. Med. Chem. 43 (2000) 2204.

[20] Abraham M.H., Chadha H.S., Martins F., Mitchell R.C., Bradbury M.W., Gratton J.A., Pestic. Sci. 55 (1999) 78.

[21] (a) Platts J.A., Butina D., Abraham M.H., Hersey A., J. Chem. Inf. Comput. Sci. 39 (1999) 835;
(b) Platts J.A., Abraham M.H., Hersey A., Butina D., J. Chem. Inf. Comput. Sci. 40 (2000) 71;
(c) Platts J.A., Abraham M.H., Hersey A., Butina D., Pharm. Res. 17 (2000) 1013;

(d) Zhao Y.H., Le J., Abraham M.H., Hersey A., Eddershaw P.E., Luscombe C.N., Butina D., Beck G., Sherborne B., Cooper I., Platts J.A., J. Pharm. Sci. 90 (2001) 749.

[22] Salminen T., Pulli A., Taskinen J.J., Pharm. Biomed. Anal. 15 (1997) 469.

[23] Brown E.A., Griffiths R., Harvey C.A., Owen D.A.A., Br. J. Pharmacol. 87 (1986) 569.

[24] Niyagi M., Kondoh H., Sakurai E., Hikichi N., Niwa H., J. Pharmacobio-Dyn. 9 (1986) 704.

[25] Aly M.I., Abdel-Latif A.A., Neurochem. Res. 5 (1980) 1231.

[26] Kunsman G.W., Rohrig T.P., Am. J. Forensic. Med. Pathol. 14 (1993) 48.

[27] Kaplan G.B., Greenblatt D.J., LeDuc B.W., Thompson M.L., Shader R.I., J. Pharmacol. Exp. Therp. 248 (1989) 1078.

[28] Wilkinson J.M., Pollard I., Develop. Brain Res. 75 (1993) 193.

[29] Dambisya Y.M., Chan K., Wong C.-L., J. Pharm. Pharmacol. 44 (1992) 687.

[30] Yata N., Toyoda T., Murakami T., Nishiura A., Higashi Y., Pharm. Res. 7 (1990) 1019.

[31] De Lange E.C.M., Danhof M., de Boer A.G., Breimer D.D., Brain Res. 666 (1994) 1.

[32] Arendt R.M., Greenblatt D.J., Liebisch D.C., Luu M.D., Paul S.M., Physchopharmacology 93 (1987) 72.

[33] Chou R.C., Levy G., Pharmacol. Exp. Therp. 219 (1981) 42.

[34] Bolander H.G., Wahlström G., Norberg L., Acta Pharmacol. Toxicol. 54 (1984) 33.

[35] Bickel M.H., Gerney R., J. Pharm. Pharrmacol. 32 (1980) 669.

[36] Okuyama S., Alhara H., Jpn. J. Pharmacol. 35 (1984) 95.

[37] Argenti D., D'Mello A.P., J. Pharmacol. Exp. Therp. 270 (1994) 512.

[38] Mandema J.W., Kuck M.T., Danhof M., Br. J. Pharmacol. 105 (1992) 164.

[39] Miyagi N., Kondoh H., Sakurai E., Hikichi N., Niwa H., J. Pharmacobio-Dyn. 9 (1986) 704.

[40] Sunderland T., Cohen B.M., Psychiatry Res. 20 (1987) 299.

[41] Tsuneizumi T., Babb S.M., Cohen B.M., Biol. Psychiatry 32 (1992) 817.

[42] Glynn S.L., Yazdanian Y., J. Pharm. Sci. 87 (1998) 306.

[43] Leo A.J., MedChem 2000 database, BioByte Corp., P.O. 517, Claremont, CA 91711-0157.

[44] Hervé F., Urien S., Albengres E., Duché J.-C., Tillement J.-P., Clin. Pharmacokinet. 26 (1994) 44.

[45] Abraham M.H., Andonian-Haftvan J., Whiting G.S., Leo A., Taft R.W., J. Chem. Soc. Perkin Trans. 2 (1994) 1777.

[46] Abraham M.H., Chadha H.S., Whiting G.S., Mitchell R.C., J. Pharm. Sci. 83 (1994) 1085.

[47] Ishihama Y., Asakawa N., J. Pharm. Sci. 88 (1999) 1305.

[48] Seiler P., Eur. J. Med. Chem. 9 (1974) 473.