Rapid Calculation of Polar Molecular Surface Area and Its Application to the Prediction of Transport Phenomena. 2. Prediction of Blood–Brain Barrier Penetration

DAVID E. CLARK*

Contribution from *Rhône-Poulenc Rorer, Dagenham Research Centre, Rainham Road South, Dagenham, Essex, RM10 7XS, United Kingdom.*

Received October 7, 1998. Final revised manuscript received February 9, 1999.
Accepted for publication April 30, 1999.

Abstract □ This paper describes the derivation of a simple QSAR model for the prediction of log BB from a set of 55 diverse organic compounds. The model contains two variables: polar surface area (PSA) and calculated logP, both of which can be rapidly computed. It therefore permits the prediction of log BB for large compound sets, such as virtual combinatorial libraries. The performance of this QSAR on two test sets taken from the literature is illustrated and compared with results from other reported computational approaches to log BB prediction.

Introduction

The blood−brain barrier (BBB) is a complex cellular system whose purpose is to maintain the homeostasis of the central nervous system (CNS) by separating the brain from the systemic blood circulation.[1] In drug discovery, it is important to determine whether a candidate molecule is capable of penetrating the BBB. For drugs targeted at the CNS, BBB penetration is a necessity (unless invasive or intranasal delivery routes are being considered[2]), whereas for drugs aimed at other sites of action, passage through the BBB may lead to unwanted side-effects.

A common measure of the degree of BBB penetration is the ratio of the steady-state concentrations of the drug molecule in the brain and in the blood, usually expressed as $\log(C_{brain}/C_{blood})$ or, more simply, log BB. Experimental values of log BB published to date cover the range about −2.00 to +1.00. Within this range, compounds with log BB > 0.3 cross the BBB readily, while compounds with log BB < −1.0 are only poorly distributed to the brain.[3] The determination of log BB is difficult and time-consuming, requiring animal experiments and the synthesis (sometimes in radiolabeled form) of the compounds to be tested. Although *in vitro*[4] and artificial membrane-based methods[5] for studying BBB penetration are being developed, it would be desirable if log BB could be predicted computationally with enough accuracy to allow the early rejection of unsuitable candidates.

In this paper, we review existing computational approaches for the prediction of log BB and then describe the derivation and validation of a novel, simple QSAR (quantitative structure−activity relationship) model allowing the rapid and accurate prediction of blood-brain barrier penetration.

QSAR Models for log BB Prediction—The first purely computational approach to log BB prediction was that of Kansy and van de Waterbeemd[6] who developed the following QSAR from a set of 20 compounds taken from the work of Young et al.:[7]

$$\log BB = -0.021(\pm 0.003)PSA - 0.003(\pm 0.001)Mol\_Vol + 1.643(\pm 0.465) \quad (1)$$
$$n = 20,\ r = 0.835,\ s = 0.448,\ F = 19.5$$

where PSA is the polar surface area, Mol_Vol is the molecular volume, n is the number of compounds, r is the correlation coefficient, s is the standard error, and F is the Fisher value, a measure of the statistical significance of the equation. The standard errors of the correlation coefficients are given in parentheses.

However, subsequent application of this equation to compounds outside its training set showed it to be poorly predictive,[8] suggesting that the 20 compound training set was insufficient to derive a generally applicable QSAR for predicting log BB.[9] Thus, Abraham and co-workers[9] constructed a larger training set of 65 compounds from which (after the removal of various outliers) they derived the following two models which they denoted ACM-II and log Pplus, respectively:[10]

$$\log BB = -0.038(\pm 0.064) + 0.198(\pm 0.100)R_2 - 0.687(\pm 0.125)\pi_2^H - 0.715(\pm 0.334)\Sigma\alpha_2^H - 0.698(\pm 0.107)\Sigma\beta_2^H + 0.995(\pm 0.096)V_x \quad (2)$$
$$n = 57,\ r = 0.952,\ s = 0.197,\ F = 99.2$$

$$\log BB = +0.055 + 0.023 \log P_{oct} - 0.507\Sigma\alpha_2^H - 0.500\Sigma\beta_2^H \quad (3)$$
$$n = 49,\ r = 0.949,\ s = 0.201,\ F = 136.1$$

No standard errors of the correlation coefficients were given for eq 3.[10] In both of these equations, the various parameters (excepting the experimental quantity, log P_{oct}) are solute descriptors, specifically: R_2 is an excess molar refraction, π_2^H is a dipolarity/polarizability parameter, $\Sigma\alpha_2^H$ and $\Sigma\beta_2^H$ are the solute hydrogen-bond acidity and basicity, respectively, and V_x is the characteristic volume of McGowan.[11]

There are a number of difficulties when applying either of these equations to more than a handful of compounds. First, to estimate log BB using either of these equations, it is necessary to calculate a value for each of the descriptors for the compound in question. The descriptor values in turn are calculated by summing the contributions from the molecule's constituent fragments. While research is ongoing to automate this process, at present, manual calculations require several minutes (at least) to make an

† Corresponding author. Phone: +44 181 919 3353. Fax: +44 181 919 2029. E-mail: david-e.clark@rp-rorer.co.uk.

10.1021/js980402t CCC: $18.00 Published on Web 06/24/1999

appropriate dissection of the molecule under study and retrieve the relevant fragment values. If values for a particular fragment cannot be found, they must be calculated. Equation 3, while more compact, requires experimental values for the partition coefficient log P_{oct}, although it is conceivable that computed estimates of log *P* could be substituted for this quantity. Finally, a close examination of the descriptor values used to construct eq 2 reveals that a number of the descriptors are very highly correlated (the pairs R_2 and π_2^H, R_2 and $\Sigma\beta_2^H$, R_2 and V_x, π_2^H and $\Sigma\beta_2^H$, and π_2^H and V_x are all correlated with $r > 0.9$). Such collinearity among the descriptors may cause the regression coefficients to become unreliable.[12] For these reasons, although the above equations appear statistically impressive, other approaches have been sought.

Lombardo et al.[13] started with a set of 57 compounds drawn from the Abraham training set[9] mentioned above. Detailed conformational analyses and semiempirical calculations were used to arrive at this simple model, denoted here as LBC, which omits two outliers from the original set of 57 compounds:

$$\log BB = 0.054(\pm 0.005)\Delta G^{o}_{W} + 0.43(\pm 0.07) \quad (4)$$
$$n = 55,\ r = 0.82,\ s = 0.41,\ F = 108.3$$

where ΔG^{o}_{W} is the computed free energy of solvation of a compound in water. As described by Lombardo et al., the determination of this quantity is rather computationally expensive, making this approach unsuitable for screening large numbers of compounds. However, recent work has led to faster methods for the calculation of solvation free energies.[14]

Most recently, Norinder and co-workers[15] have developed models for log BB prediction using their computed MolSurf parameters together with statistical analysis by the Partial Least Squares to Latent Structures (PLS) method. The model from their work with the largest training set (56 compounds − model 2 in ref 15, denoted here as NSO) contained three significant PLS components with the following statistics: $n = 56$, $r = 0.913$, $s = 0.312$, $F = 86.95$. Here again, however, the computational methods employed required conformational analysis and computationally intensive semiempirical and ab initio calculations which render the approach too slow for high-throughput applications.

In what follows, we describe the derivation of a simple QSAR model for the prediction of log BB that is automatic and rapid to calculate and therefore applicable to the screening of large compound sets, such as virtual combinatorial libraries. The performance of this QSAR on two test sets taken from the literature will be illustrated and compared with the methods above where such results are available.

Computational Methods—*Training Set*—A set of 57 compounds previously studied by Lombardo et al.[13] was used as a training set. These compounds are illustrated in Figure 1 and listed in Table 1 along with experimental log BB values taken from ref 13.

Test Set 1—This set consists of the seven compounds (shown in Figure 2) used as a test set by Abraham et al.[10] The experimental log BB values listed in Table 2 for these compounds were taken from ref 10 as were the predicted log BB values from eqs 2 and 3.

Test Set 2—This test set was used by Lombardo et al.[13] and also by Norinder et al.[15] It comprises six compounds whose structures are shown in Figure 3. Table 3 shows the log BB values from the various equations and experiment.

Polar Surface Area Calculations—PSA values for the molecules under study were calculated using the methods described in the previous paper.[16]

Calculations of log *P*—Two computational methods for logP prediction were used. ClogP was calculated using the Daylight software.[17] MlogP values were computed using an implementation of the method developed by Moriguchi et al.[18] encoded in the Sybyl Programming Language and executed within the Sybyl Molecular Modeling package.[19]

Regression Analysis—All multiple linear regressions were carried out within the Tsar program.[20]

Results

Training Set—Despite the limitations found with the Kansy and van de Waterbeemd approach,[6] we decided to investigate the possibility of using PSA values to derive a generally applicable QSAR for log BB. Starting from the 57 compounds used by Lombardo et al.,[13] we first tried correlating PSA with log BB and obtained the following equation:

$$\log BB = -0.016(\pm 0.001)PSA + 0.547(\pm 0.050) \quad (5)$$
$$n = 57,\ r = 0.819,\ s = 0.455,\ F = 112.4$$

This initial result was encouraging and, incidentally, suggested a relationship between PSA and ΔG^{o}_{W}. An investigation showed the two quantities were indeed closely correlated ($r = 0.962$). However, a purely PSA-based model failed to distinguish the varying BBB-penetrating abilities of nonpolar compounds. For example, benzene and 3-methylpentane have log BB values of −0.69 and 2.01 respectively, but both have a PSA of zero. Thus, we began to search for an additional descriptor that would differentiate between the nonpolar compounds in the set. Molecular weight, molecular volume, and nonpolar surface area were tried, but none led to a significant improvement in the model.[21] Finally, calculated logP values were tried with more success generating a model we denote DEC-I:

$$\log BB = -0.0148(\pm 0.001)PSA + 0.152(\pm 0.036)ClogP + 0.139(\pm 0.073) \quad (6)$$
$$n = 55,\ r = 0.887,\ s = 0.354,\ F = 95.8$$

where ClogP is the calculated logP.[17] The two compounds omitted from the original set of 57 were N_2, for which ClogP cannot calculate an accurate value, and compound 12 which, if included in the model, shows an error of 1.5 log units in the prediction of its experimental log BB value (data not shown). Compound 12 has also been found to be an outlier by other groups.[13,15] One difficulty with using ClogP values is that there are compounds for which it cannot generate accurate values. For this reason, some workers[22] have proposed the use of the MlogP[18] approach in such circumstances. If MlogP values are substituted for ClogP values, the following model (DEC-II) is generated:

$$\log BB = -0.0145(\pm 0.001)PSA + 0.172(\pm 0.022)MlogP + 0.131(\pm 0.033) \quad (7)$$
$$n = 55,\ r = 0.876,\ s = 0.369,\ F = 86.0$$

To allow a direct comparison of these two equations with a purely PSA-based model, eq 5 was rederived based on the same 55 compounds as were used in the derivation of eqs 6 and 7 giving:

$$\log BB = -0.0156(\pm 0.001)PSA + 0.548(\pm 0.048) \quad (8)$$
$$n = 55,\ r = 0.841,\ s = 0.410,\ F = 128.4$$

When the statistics are compared, eqs 6 and 7 show superior *r* and *s* values to eq 8, and the *F* values for the three equations indicate that all are significant at the 95%

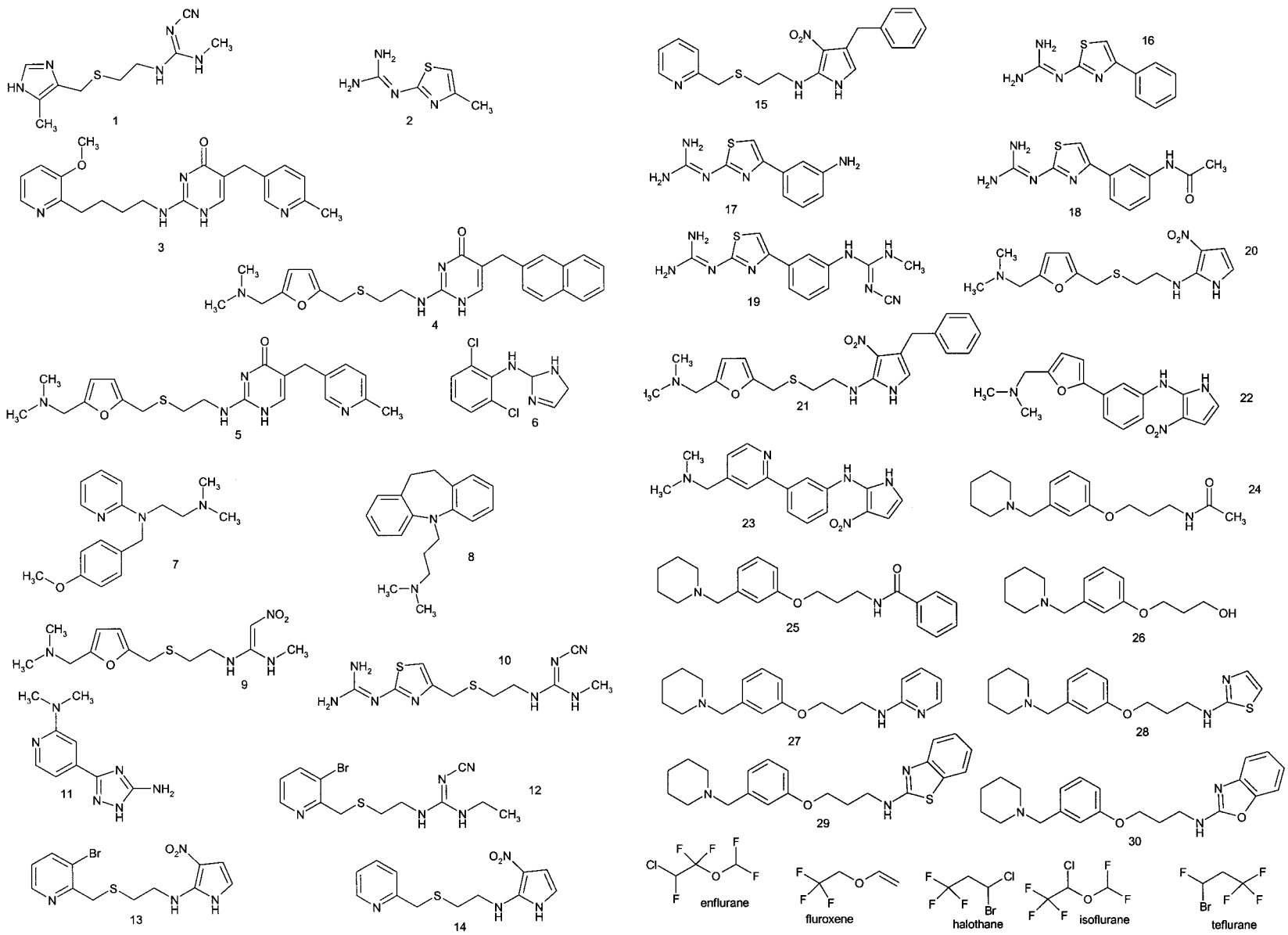

Figure 1—Compounds **1**–**30** and less familiar compounds from the training set.

Table 1—Data and Results for Training Set

compound	PSA/Å^2	ClogP	MlogP	expt log BB	DEC-I log BB	DEC-II log BB
1	92.1	0.351	0.821	–1.42	–1.17	–1.07
2	78.9	0.952	0.786	–0.04	–0.88	–0.88
3	94.0	2.297	2.300	–2.00	–0.90	–0.84
4	73.5	4.046	3.569	–1.30	–0.33	–0.32
5	87.0	1.874	2.107	–1.06	–0.86	–0.77
6	39.0	0.743	3.066	0.11	–0.32	0.09
7	26.8	2.787	2.575	0.49	0.17	0.18
8	6.0	4.413	3.876	0.83	0.72	0.71
9	84.5	1.327	0.659	–1.23	–0.91	–0.98
10	139.2	0.844	1.258	–0.82	–1.79	–1.67
11	88.8	0.911	2.592	–1.17	–1.03	–0.71
12	73.5	2.282	1.940	–2.15	n.d.[a]	n.d.[a]
13	83.9	2.747	1.642	–0.67	–0.68	–0.81
14	84.0	1.800	0.969	–0.66	–0.83	–0.92
15	78.0	3.637	2.484	–0.12	–0.46	–0.58
16	76.6	2.781	2.480	–0.18	–0.57	–0.56
17	104.4	1.784	1.960	–1.15	–1.13	–1.05
18	108.8	1.977	2.087	–1.57	–1.17	–1.09
19	135.8	1.880	2.400	–1.54	–1.58	–1.43
20	85.5	2.287	0.832	–1.12	–0.78	–0.97
21	79.5	4.124	2.253	–0.73	–0.41	–0.64
22	82.7	3.849	1.868	–0.27	–0.50	–0.75
23	85.7	3.234	1.760	–0.28	–0.64	–0.81
24	47.9	2.065	2.069	–0.46	–0.25	–0.21
25	45.2	4.004	3.228	–0.24	0.08	0.03
26	38.5	2.379	2.069	–0.02	–0.07	–0.07
27	39.1	4.259	3.272	0.69	0.21	0.12
28	40.0	4.165	2.627	0.44	0.18	0.00
29	39.2	5.759	3.902	0.14	0.43	0.23
30	54.9	5.029	4.332	0.22	0.09	0.08
butanone	22.7	0.834	0.655	–0.08	–0.07	–0.09
benzene	0.0	2.142	2.255	0.37	0.46	0.52
3-methylpentane	0.0	3.738	3.516	1.01	0.71	0.73
3-methylhexane	0.0	4.267	3.869	0.90	0.79	0.80
2-propanol	23.4	0.074	0.347	–0.15	–0.20	–0.15
2-methylpropanol	22.6	0.693	0.800	–0.17	–0.09	–0.06
2-methylpentane	0.0	3.738	3.516	0.97	0.71	0.73
2,2-dimethylbutane	0.0	3.608	3.516	1.04	0.69	0.73
1,1,1-trifluoro-2-chloroethane	0.0	1.714	2.081	0.08	0.40	0.49
1,1,1-trichloroethane	0.0	2.481	2.226	0.40	0.52	0.51
diethyl ether	11.3	0.870	0.800	0.00	0.10	0.10
enflurane	11.6	2.459	1.766	0.24	0.34	0.27
ethanol	24.4	–0.235	0.172	–0.16	–0.26	–0.19
fluroxene	10.7	1.765	1.257	0.13	0.25	0.19
halothane	0.0	2.447	2.604	0.35	0.51	0.58
heptane	0.0	4.397	3.869	0.81	0.81	0.80
hexane	0.0	3.868	3.516	0.80	0.73	0.73
isoflurane	11.0	2.999	1.766	0.42	0.43	0.27
methane	0.0	1.103	1.115	0.04	0.31	0.32
methylcyclopentane	0.0	3.314	3.124	0.93	0.64	0.67
nitrogen	54.2	n.d.[b]	–2.272	0.03	n.d.[a]	n.d.[a]
pentane	0.0	3.339	3.138	0.76	0.65	0.67
propanol	24.4	0.294	0.347	–0.16	–0.18	–0.16
propanone	22.7	0.305	0.202	–0.15	–0.15	–0.16
teflurane	0.0	2.007	2.419	0.27	0.44	0.55
toluene	0.0	2.641	2.608	0.37	0.54	0.59
trichloroethene	0.0	2.627	2.081	0.34	0.54	0.49

[a] Compounds not included in final training set for eqs 6 and 7. [b] ClogP could not calculate an accurate value for this compound. DEC-I is the predicted set of values using eq 6 and DEC-II is the predicted set of values using eq 7.

confidence level. It is apparent that the inclusion of a calculated log *P* value has improved the model for log BB prediction. Thus, eqs 6 and 7 were preferred for log BB prediction over the simple PSA-based model of eq 8.

The quantities comprising eqs 6 and 7 are rapidly calculable, and, furthermore, the descriptors involved are not significantly correlated with one another: the correlation coefficient *r* being 0.15 for PSA and ClogP, and 0.23 for PSA and MlogP. The predicted log BB values from eqs 6 and 7 are tabulated in Table 1, and plots of computed versus experimental values are shown in Figure 4. However, the real test of any QSAR equation is how well it predicts values for compounds outside its training set. Accordingly, we have applied our equations to two test sets that are available in the literature.

Test Set 1—Looking at Table 2, if all seven compounds are considered, the mean absolute errors in the log BB predictions are: ACM-II: 0.30, log Pplus: 0.54, DEC-I: 0.37 and DEC-II: 0.40. However, it can be seen that all the models overpredict log BB for compounds Y-G19 and Y-G20. This is in accord with the findings of Abraham et al.[10] who considered them as outliers. They suggested that

Table 2—Data and Results from Test Set 1[a]

compound	PSA/Å^2	ClogP	MlogP	expt log BB	ACM-II log BB	log Pplus log BB	DEC-I log BB	DEC-II log BB
Y-G14	29.3	–0.068	0.652	–0.30	–0.31	–0.47	–0.30	–0.18
Y-G15	18.8	–0.338	0.969	–0.06	–0.01	–0.34	–0.09	0.02
Y-G16	43.1	–0.603	–0.316	–0.42	–0.41	–0.50	–0.59	–0.55
Y-G19	40.6	1.495	1.656	–1.30	–0.14	–0.15	–0.24	–0.17
Y-G20	46.7	0.110	0.855	–1.40	–0.57	–0.15	–0.53	–0.40
SKF89124	67.5	2.666	2.634	–0.43	–0.44	–0.91	–0.56	–0.40
SKF101468	44.1	1.999	2.605	0.25	0.24	–0.13	–0.11	–0.06

[a] ACM-II is the predicted set of values from eq 2, log Pplus is the predicted set of values from eq 3, DEC-I and DEC-II are the predicted set of values from eq 6 and eq 7, respectively.

Table 3—Data and Results from Test Set 2[a]

compound	PSA/Å^2	ClogP	MlogP	expt log BB	LBC log BB	NSO log BB	DEC-I log BB	DEC-II log BB
31	46.7	1.980	3.139	0.00	–0.14	–0.58	–0.25	–0.01
32	62.7	0.260	2.408	–0.34	–0.28	–1.11	–0.75	–0.37
33	76.1	1.907	3.446	–0.30	–0.46	–0.75	–0.70	–0.38
34	98.5	0.380	2.698	–1.34	–0.64	–0.99	–1.26	–0.83
35	120.4	–0.932	1.959	–1.82	–0.82	–1.35	–1.77	–1.28
36	5.4	4.641	4.369	0.76–0.98	0.28	1.03	0.76	0.80

[a] LBC is the predicted set of values from eq 4, NSO is the predicted set of values from model 2 of Norinder et al.,[15] DEC-I and DEC-II are the predicted set of values from eq 6 and eq 7, respectively.

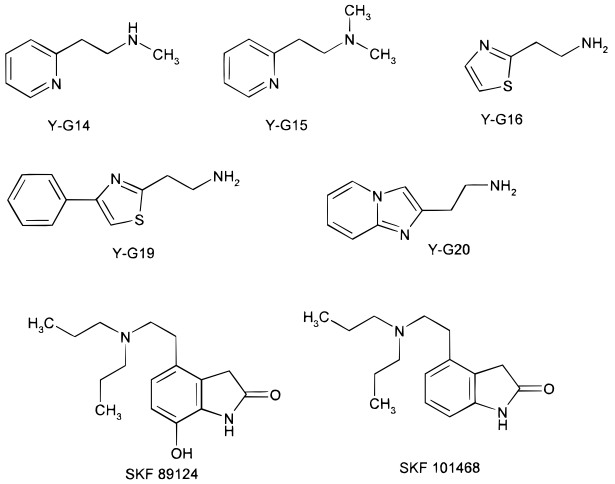

Figure 2—Structures for test set 1.

such outliers might be caused by factors such as inaccurate data due to experimental difficulties or metabolic effects not accounted for in log BB measurements.[10] The fact that our independently derived models also find large deviations for these compounds, while treating the remainder quite well, adds some weight to the argument to consider Y-G19 and Y-G20 as outliers. If this is done, for the remaining five compounds, the mean absolute errors in the log BB predictions are as follows: ACM-II: 0.02, log Pplus: 0.28, DEC-I: 0.14, and DEC-II: 0.13. Thus, whether the outliers are included or excluded, ACM-II is the best predictor for this set, and DEC-I and DEC-II perform about as well as each other and better than log Pplus.

Test Set 2—Considering the results in Table 3, it can be seen that compound **36** (amitryptyline) has a large uncertainty in its experimental log BB value; however, DEC-I and DEC-II predict log BB values within the given range. In so doing, they perform better than LBC, which underpredicts log BB, and NSO which predicts too high a value. Calculating the mean absolute errors over compounds **31**–**35**, we obtain: LBC: 0.41; NSO: 0.52; DEC-I: 0.24; DEC-II: 0.23.

In this case, the equations based on PSA and calculated log *P* perform substantially better than the more computationally expensive methods of Lombardo et al.[13] and Norinder et al.[15]

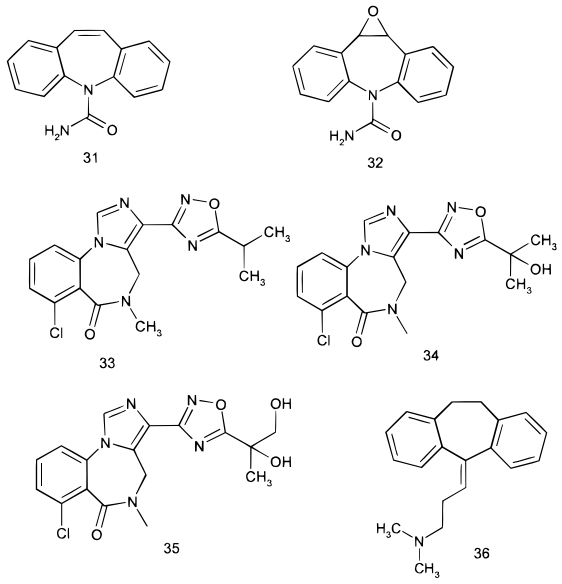

Figure 3—Structures for test set 2.

Discussion

It has been suggested by Lennernäs[23] that purely passive diffusion is universal for membranes with different physiological functions and physicochemical properties. Thus, the factors that determine a compound's intestinal absorption should also, to some extent, model its BBB permeability. Therefore, it is not surprising that PSA, which has been found to be useful in modeling intestinal absorption (see the preceding paper[16] and references therein), together with a direct estimate of lipophilicity, widely acknowledged as an important factor in transport across membranes,[24] should yield a predictive model for log BB. It would seem from our results that, with this set of compounds, quantities such as molecular weight, molecular volume, or nonpolar surface area do not correlate strongly enough with

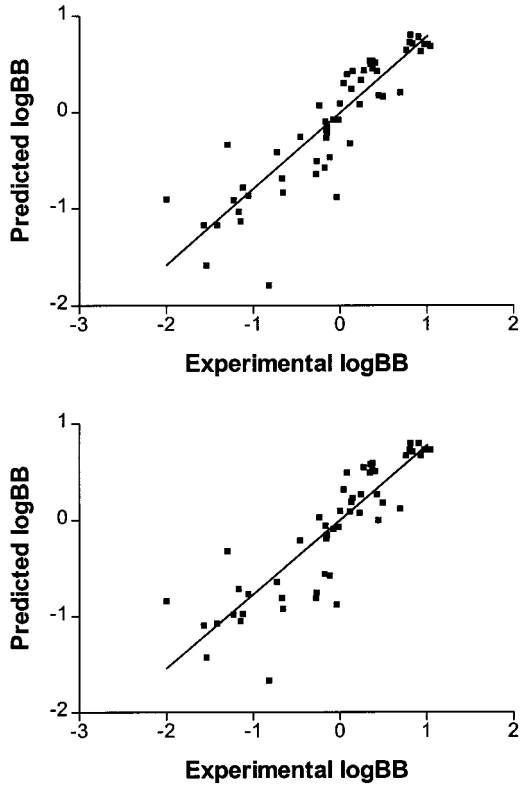

Figure 4—(a) Plot of computed (DEC-I) versus experimental log BB values. (b) Plot of computed (DEC-II) versus experimental log BB values.

lipophilicity to be useful (when combined with PSA) for modeling BBB penetration. An analysis of the 55 compounds used in the training set in this paper bears this out. The correlation coefficients, *r*, for the relationships between ClogP and molecular weight, molecular volume, and nonpolar surface area are 0.386, 0.468, and 0.599, respectively.

It is interesting to note the similarity between the models derived in this paper (DEC-I and DEC-II) and the log Pplus equation of Abraham et al.[10] In the latter (eq 3), the two parameters, $\Sigma\alpha_2^H$ and $\Sigma\beta_2^H$, denote the solute hydrogen-bond acidity and basicity, respectively. The fact that these have negative coefficients while log *P* has a positive coefficient accords with the expectation that membrane permeation will be more facile for compounds that form fewer (and weaker) hydrogen bonds with solvent and with the observation that the brain is more lipophilic than the blood, making lipophilic compounds more likely to penetrate the brain. Similar findings were reported by Norinder et al.[15] who reported that high brain/blood partitioning was favored by an absence of atoms capable of hydrogen bonding together with high lipophilicity. The same may be said for DEC-I and DEC-II, where better brain penetration is predicted for compounds with high calculated log *P* and low PSA (the polar surface area being an indication of a compound's capacity to form hydrogen bonds). In addition, van de Waterbeemd et al.[25] recently showed that, of a set of 125 drugs, all those showing CNS activity could be found within the ranges: $0 \le \text{PSA} \le 90$; $-1 \le \log D\ (\text{pH } 7.5) \le 4$ with the likelihood of CNS activity appearing to increase with decreasing PSA and increasing log *D*. Thus, the models derived in this paper would seem to be sensible from a physiological point of view and are also in agreement with insights from other models into the factors facilitating blood−brain barrier penetration.

As with the model for intestinal absorption described previously (see the preceding paper[16] and references therein), it is likely that the simple model described here is only valid for passive diffusion processes across the BBB. There are active transport systems for both influx and efflux in the brain and compounds which are affected by these are not likely to be well-predicted. Another factor influencing BBB transport is the binding of drugs by plasma proteins.[26] This too is not directly accounted for in the PSA/calculated log P-based model, although plasma binding within a given series may be correlated with log *P*.[27]

Finally, other experimental measures of blood−brain barrier permeability are now becoming available. Eddy et al.[4] have reviewed various in vitro models, and Lombardo et al.[13] showed that, for a small set of 10 compounds, it was possible to correlate permeability across a monolayer composed of endothelial cells from bovine brain microvessels with free energy of solvation. Gratton et al.[28] have reported permeability−surface area measurements (denoted log PS) using a short-duration vascular perfusion method for 18 compounds and correlated the values with Abraham's solute descriptors. A relatively new technique for measuring brain penetration in vivo is microdialysis.[29] This offers the potential for observing drug disposition between the extracellular and intracellular space in the brain, something not possible with traditional log BB measurements.[30] As more data emerge from these various novel techniques, it will be interesting to see how well computational techniques can adapt to predict new measures of blood−brain barrier penetration.

Conclusion

In summary, the equations (DEC-I and DEC-II) described in this paper for log BB prediction show a good predictive ability. Their utility is enhanced by the fact that they comprise only two simple, noncorrelated variables, values for which may be rapidly computed for almost any structure. (As described in the preceding paper,[16] PSA values can be computed in about 10−15 s on a modern workstation, and ClogP and MlogP calculations are very fast.) The procedure for log BB estimation is fully automated, allowing the prescreening of virtual libraries and other compound sets prior to synthesis or purchase. For this reason, it should be of use in drug discovery projects where blood−brain barrier penetration is an issue.

References and Notes

1. De Vries, H. E.; Kuiper, J.; De Boer, A. G.; Van Berkel, T. J. C.; Breimer, D. D. The Blood-Brain Barrier in Neuroinflammatory Diseases. *Pharmacol. Rev.* **1997**, *49*, 143−155.
2. Pardridge, W. M. CNS Drug Design Based on Principles of Blood-Brain Barrier Transport. *J. Neurochem.* **1998**, *70*, 1781−1792.
3. Abraham, M. H.; Takacs-Novak, K.; Mitchell, R. C. On the Partition of Ampholytes: Application to Blood-Brain Distribution. *J. Pharm. Sci.* **1997**, *86*, 310−315.
4. Eddy, E. P.; Maleef, B. E.; Hart, T. K.; Smith, P. L. In Vitro Models to Predict Blood-Brain Barrier Permeability. *Adv. Drug Deliv. Rev.* **1997**, *23*, 185−198.
5. Reichel, A.; Begley, D. J. Potential of Immobilized Artificial Membranes for Predicting Drug Penetration Across the Blood-Brain Barrier. *Pharm. Res.* **1998**, *15*, 1270−1274.
6. Kansy, M.; van de Waterbeemd, H. Hydrogen-Bonding Capacity and Brain Penetration. *Chimia* **1992**, *46*, 299−303.
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. Development of a New Physicochemical Model for Brain Penetration and Application to the Design of Centrally Acting H2 Receptor Histamine Antagonists. *J. Med. Chem.* **1988**, *31*, 656−671.
8. Calder, J. A.; Ganellin, C. R. Predicting the Brain-Penetrating Capability of Histaminergic Compounds. *Drug Des. Discov.* **1994**, *11*, 259−268.
9. Abraham, M. H.; Chadha, H. S.; Mitchell, R. C. Hydrogen Bonding. 33. Factors that Influence the Distribution of

Solutes Between Blood and Brain. *J. Pharm. Sci.* **1994**, *83*, 1257−1268.
10. Abraham, M. H.; Chadha, H. S.; Mitchell, R. C. Hydrogen Bonding. Part 36. Determination of Blood Brain Distribution Using Octanol−Water Partition Coefficients. *Drug Des. Discov.* **1995**, *13*, 123−131.
11. Abraham, M. H.; McGowan, J. C. The Use of Characteristic Volumes to Measure Cavity Terms in Reversed Phase Liquid Chromatography. *Chromatographia* **1987**, *23*, 243−246.
12. Livingstone, D. *Data Analysis for Chemists*. Oxford University Press: Oxford, 1995; pp 126−127.
13. Lombardo, F.; Blake, J. F.; Curatolo, W. J. Computation of Brain−Blood Partitioning of Organic Solutes via Free Energy Calculations. *J. Med. Chem.* **1996**, *39*, 4750−4755.
14. Hawkins, G. D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. OMNISOL: Fast Prediction of Free Energies of Solvation and Partition Coefficients. *J. Org. Chem.* **1998**, *63*, 4305−4313.
15. Norinder, U.; Sjöberg, P.; Österberg, T. Theoretical Calculation and Prediction of Brain−Blood Partitioning of Organic Solutes Using MolSurf Parametrization and PLS Statistics. *J. Pharm. Sci.* **1998**, *87*, 952−959.
16. Clark, D. E. Rapid Calculation of Polar Molecular Surface Area and its Application to the Prediction of Transport Phenomena. 1. Prediction of Intestinal Absorption. *J. Pharm. Sci.*, **1999**, *88*, 807−814.
17. ClogP. Daylight Chemical Information Software, version 4.51. Daylight Chemical Information Inc.: 27401 Los Altos, Suite #370, Mission Viejo, CA 92691.
18. Moriguchi, I.; Hirono, S.; Liu, Q.; Nakagome, I.; Matsushita, Y. Simple Method of Calculating Octanol/Water Partition Coefficient. *Chem. Pharm. Bull.* **1992**, *40*, 127−130.
19. Sybyl 6.4.2. Tripos, Inc., 1699 S. Hanley Rd., Suite 303, St. Louis, MO 63144-2913. The SPL script was originally developed by Dr. James Blake (Pfizer, US) and has been subsequently modified by Dr. Jon Swanson (Tripos Inc.) to accommodate changes in the Tripos Sybyl Line Notation (SLN) description of substructure queries.
20. Tsar 3.1. Oxford Molecular Ltd., The Medawar Centre, Oxford Science Park, Sandford-on-Thames, Oxford, OX4 4GA, England.
21. To allow comparison with eqs 6−8, the statistics for the equations involving these quantities are PSA and molecular weight: $n = 55$, $r = 0.845$, $s = 0.410$, $F = 64.8$; PSA and molecular volume: $n = 55$, $r = 0.849$, $s = 0.405$, $F = 66.9$; PSA and nonpolar surface area: $n = 55$, $r = 0.848$, $s = 0.406$, $F = 66.6$.
22. Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. *Adv. Drug Deliv. Rev.* **1997**, *23*, 3−25.
23. Lennernäs, H. Human Jejunal Effective Permeability and its Correlation with Preclinical Drug Absorption Models. *J. Pharm. Pharmacol.* **1997**, *49*, 627−638.
24. Conradi, R. A.; Burton, P. S.; Borchardt, R. T. Physicochemical and Biological Factors that Influence a Drug's Cellular Permeability by Passive Diffusion. In *Lipophilicity in Drug Action and Toxicology*; Pliška, V., Testa, B., van de Waterbeemd, H., Eds.; VCH: Weinheim, 1996; pp 233−252.
25. van de Waterbeemd, H.; Camenisch, G.; Folkers, G.; Chretien, J. R.; Raevsky, O. A. Estimation of Blood-Brain Barrier Crossing of Drugs Using Molecular Size and Shape, and H-bonding Descriptors. *J. Drug Target.* **1998**, *6*, 151−165.
26. Olson, R. E.; Christ, D. D. Plasma Protein Binding of Drugs. In *Annual Reports in Medicinal Chemistry*; Bristol, J. A., Ed.; Academic Press: San Diego, 1996; Vol. 33, pp 327−336.
27. Rowley, M.; Kulagowski, J. J.; Watt, A. P.; Rathbone, D.; Stevenson, G. I.; Carling, R. W.; Baker, R.; Marshall, G. R.; Kemp, J. A.; Foster, A. C.; Grimwood, S.; Hargreaves, R.; Hurley, C.; Saywell, K. L.; Tricklebank, M. D.; Leeson, P. D. Effect of Plasma Protein Binding on in Vivo Activity and Brain Penetration of Glycine/NMDA Receptor Antagonists. *J. Med. Chem.* **1997**, *40*, 4053−4068.
28. Gratton, J. A.; Abraham, M. H.; Bradbury, M. W.; Chadha, H. S. Molecular Factors Influencing Drug Transfer Across the Blood-Brain Barrier. *J. Pharm. Pharmacol.* **1997**, *49*, 1211−1216.
29. Elmquist, W. F.; Sawchuk, R. J. Application of Microdialysis in Pharmacokinetic Studies. *Pharm. Res.* **1997**, *14*, 267−288.
30. Evrard, P. A.; Ragusi, C.; Boschi, G.; Verbeeck, R. K.; Scherrmann, J.-M. Simultaneous Microdialysis in Brain and Blood of the Mouse: Extracellular and Intracellular Brain Colchicine Disposition. *Brain Res.* **1998**, *786*, 122−127.

Acknowledgments

I am very grateful to my colleagues at RPR for many helpful discussions and invaluable advice. I would especially like to thank Dr. Iain McLay for proposing the investigation of PSA values in log BB prediction and Dr. Stephen Pickett for suggesting the inclusion of ClogP as a variable in eq 6. Dr. Ulf Norinder kindly supplied a preprint of ref 15. Dr. James Blake (Pfizer, US) is thanked for making his SPL script for MlogP calculations available and the technical support of Drs. Jon Swanson and Gareth Wilden (Tripos, Inc.) in relation to this code is acknowledged.

JS980402T