European Journal of Pharmaceutics and Biopharmaceutics 70 (2008) 462–466

Contents lists available at ScienceDirect

European Journal of Pharmaceutics and Biopharmaceutics

journal homepage: www.elsevier.com/locate/ejpb

Research paper

Predicting blood–brain barrier penetration from molecular weight and number of polar atoms

Xu-Chun Fu [a], Guo-Ping Wang [b], Hai-Li Shan [c], Wen-Quan Liang [c], Jian-Qing Gao [c,*]

[a] Department of Pharmacy, Zhejiang University City College, Hangzhou, PR China
[b] Department of Chemistry, Zhejiang University, Hangzhou, PR China
[c] College of Pharmaceutical Sciences, Zhejiang University, Hangzhou, PR China

ARTICLE INFO

Article history:
Received 24 March 2008
Accepted in revised form 8 May 2008
Available online 17 May 2008

Keywords:
Blood–brain barrier
Predictive model
Molecular weight
Polar atom

ABSTRACT

A simple three-descriptor model to predict blood–brain barrier is derived from a training set of 78 compounds: $\log BB = -9.880 \times 10^{-6} M_W^2 + 7.339 \times 10^{-3} M_W - 0.2268 n_{pol} - 0.1143$ ($n = 78$, $r^2 = 0.74$), where log BB is the logarithm of the ratio of the steady-state concentration of the compound in the brain to concentration in the blood, M_W is the molecular weight, n_{pol} is the number of polar atoms (oxygen, nitrogen, and attached hydrogen), n is the number of compounds, and r is the correlation coefficient. The model is validated through use of leave-one-out procedure and an external test set (25 compounds). The model is suitable for the rapid prediction of the blood–brain barrier penetration of drug candidates because of its predictive ability and simplicity.

1. Introduction

Central nervous system (CNS) therapeutic agents must cross the blood–brain barrier (BBB) to be effective, whereas peripherally drugs must possess very limited ability to cross the BBB in order to avoid or minimize CNS side effects. A common measure of the degree of BBB penetration is the ratio of the steady-state concentration of the drug molecule in the brain to concentration in the blood, usually expressed as $\log(C_{brain/blood})$ or log BB. The measurement of BBB penetration in vivo is a time-consuming, expensive, and difficult technique, requiring animal experiments and the synthesis of the test compounds, usually in radiolabeled form [1–3]. Although in vitro models for the prediction of BBB penetration (e.g., bovine brain capillary endothelial cells) are available, they are experimentally cumbersome [4]. It would be desirable if log BB could be predicted computationally with enough accuracy to allow the early rejection of unsuitable candidates.

Young et al. showed that log BB values of 20 H_2 receptor histamine antagonists were correlated with $\Delta \log P$ (octanol–cyclohexane) [2]. van de Waterbeemd and Kansy examined the same series of 20 compounds and found a significant correlation between log BB and the cyclohexane–water partition coefficient when the molecular volume was included in the parameterization [5]. They also found that log BB was correlated with polar surface area (*PSA*, defined as the sum of the van der Waals surface areas of oxygen atoms, nitrogen atoms, and attached hydrogen atoms in a molecule), but the model showed it to be poorly predictive when tested with compounds outside its training set, suggesting that the structural diversity of the 20 H_2 receptor histamine antagonists might be insufficient to develop a generally applicable model for predicting log BB [6]. Thus, Abraham et al. constructed a larger training set of 65 compounds and derived a correlation between log BB and solvato-chromatic parameters for 57 compounds (eight compounds were excluded as outliers) [7]. With a set of 57 compounds drawn from the Abraham training set mentioned above, several researches developed the models for log BB prediction using calculated molecular structural parameters such as free energy of solvation in water, ΔG_w^{θ}, *PSA*, and calculated octanol–water partition coefficient, *C*log*P* or *M*log*P*, respectively [8–10]. More recently, a variety of models to predict BBB penetration for larger dataset have been developed using different descriptors such as the three-dimensional molecular field descriptors and electropological state indices, [11–20]. In summary, the BBB penetration of a compound is thought to be dependent on its hydrogen-bonding potential, lipophilicity and size. Weak hydrogen-bonding potential, high lipophilicity, and small size are favorable to BBB penetration.

In this paper, we derive a simple model for the prediction of log BB from a dataset of 86 compounds, in which eight compounds are removed as outliers.

* Corresponding author. College of Pharmaceutical Sciences, Zhejiang University, 388 Yuhangtang Road, Hangzhou 310058, Zhejiang, PR China. Tel./fax: +86 571 88208437.
E-mail address: gaojianqing1029@yahoo.com.cn (J.-Q. Gao).

0939-6411/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.ejpb.2008.05.005

2. Methods

The dataset of 111 compounds and their corresponding log BB values are taken from the literature [2,6–8,21–26]. Of the 111 compounds, 25 compounds are chosen as the test set used by Luco and Feher et al. [11,12]. The remaining 86 compounds are used for the training set. Molecular weight (M_W) and number of polar atoms which are oxygen atoms, nitrogen atoms, and attached hydrogen atoms (n_{pol}) are selected as the structural descriptors to develop predictive model for BBB penetration. The model to predict blood–brain barrier penetration is derived on the training set using the stepwise multiple regression analysis and then cross-validated using leave-one-out procedure in which one compound is left out from the training set and predicted from the model based on the remaining data and tested on the external prediction [27].

3. Results and discussion

3.1. Development of predictive model of blood–brain barrier penetration

The 86 compounds of training set are illustrated in Fig. 1 and listed in Table 1 along with their experimental log BB values.

Using M_W and n_{pol} as regression variables, the following regression equation is obtained from the stepwise multiple regression analysis (including quadratic terms) for the 86 compounds,

$$\log BB = -9.880 \times 10^{-6} M_W^2 + 7.339 \times 10^{-3} M_W - 0.2268 n_{pol} - 0.1143$$

$$n = 78 \quad r^2 = 0.74 \quad q^2 = 0.71 \quad s = 0.37 \quad F = 71 \qquad (1)$$

where n is the number of compounds, r is the correlation coefficient, q is the cross validation coefficient, s is the standard deviation, and F is the Fisher F-statistic. Since Compounds 2, 6, 23, 30, 60, 61, 62 and 78 have residuals greater than 2.5 standard deviations, they are removed from the above equation as outliers. Previous studies also identified Compounds 2 [11,20], 6 [11,18,20], 23 [7,8], 30 [7,8,10–12,15,16,18,20], 60 [12,15,16,20], 61 [10–12,15,18,20], 62 [10–12,15,18,20] and 78 [16,20] as outliers. This might be caused by factors such as inaccurate data due to experimental difficulties, metabolic effects or active transport systems.

Other descriptors are also tested. When n_{pol} in Eq. (1) is replaced with the number of hydrogen bond donors and hydrogen bond acceptors, or M_W is replaced with log M_W, the respective regression equations are not so significantly statistical as Eq. (1).

The calculated log BB values for the training set from Eq. (1) are presented in Table 1 and the experimental and calculated log BB values are plotted in Fig. 2.

Eq. (1) displays obvious statistical significance. The descriptors involved are not significantly correlated with one another ($r^2 = 0.16$). Because n_{pol} is a descriptor of hydrogen-bonding potential, Eq. (1) indicates that the log BB of a compound is inversely correlated with its hydrogen-bonding capacity. This is in agreement with the negative coefficients of the hydrogen bond terms such as the polar molecular surface area, the hydrogen bond acidity and basicity in other models [7,10,12,19,20]. Eq. (1) also shows the parabolic relation between log BB and molecular weight. The explicit descriptor for lipophilicity is absent from Eq. (1) and the molecular weight terms in the equation represent a combination of the impacts of molecular size and lipophilicity on BBB penetration. Increasing molecular weight decreases molecular diffusion through a lipid membrane, and therefore decreases log BB value. On the other hand, bigger molecular weight also means higher lipophilicity which facilitates BBB penetration.

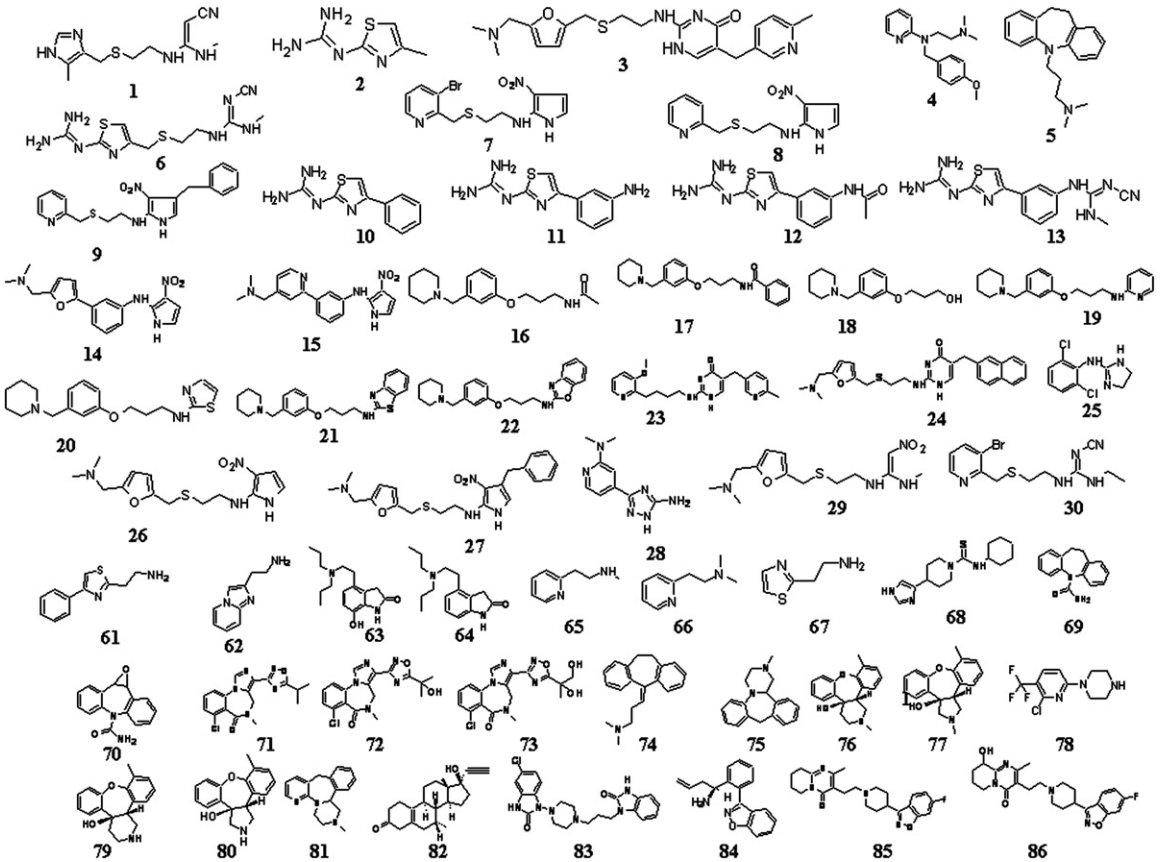

Fig. 1. Compounds 1–30 and 61–86.

Table 1
Experimental and calculated logBB values for the training set compounds and their descriptors

Compound		M_W	n_{pol}	logBB		
				Exp.	Calc.[a]	Pred.[b]
1		252.3	9	−1.42	−0.93	−0.90
2		156.2	8	−0.04	–	–
3		413.5	9	−1.06	−0.81	−0.79
4		285.4	4	0.49	0.27	0.26
5		280.4	2	0.83	0.71	0.71
6		312.4	14	−0.82	–	–
7		357.2	8	−0.67	−0.57	−0.56
8		278.3	8	−0.66	−0.65	−0.65
9		368.5	8	−0.12	−0.57	−0.58
10		218.3	8	−0.18	−0.80	−0.84
11		233.3	11	−1.15	−1.43	−1.47
12		275.3	11	−1.57	−1.34	−1.31
13		314.4	14	−1.54	−1.96	−2.04
14		326.4	9	−0.27	−0.81	−0.83
15		337.4	9	−0.28	−0.80	−0.82
16		290.4	5	−0.46	0.05	0.06
17		352.5	5	−0.24	0.11	0.13
18		249.4	4	−0.02	0.19	0.20
19		325.5	5	0.69	0.09	0.08
20		331.5	5	0.44	0.10	0.09
21		381.5	5	0.14	0.11	0.11
22		365.5	6	0.22	−0.11	−0.13
23		379.5	9	−2.00	–	–
24		448.6	8	−1.30	−0.62	−0.48
25		230.1	5	0.11	−0.08	−0.09
26		324.4	9	−1.12	−0.81	−0.80
27		414.5	9	−0.73	−0.81	−0.82
28		204.2	9	−1.17	−1.07	−1.06
29		314.4	9	−1.23	−0.82	−0.81
30		342.3	7	−2.15	–	–
31	Butanone	72.1	1	−0.08	0.14	0.15
32	Benzene	78.1	0	0.37	0.40	0.40
33	3-Methylpentane	86.2	0	1.01	0.44	0.42
34	3-Methylhexane	100.2	0	0.90	0.52	0.51
35	2-Propanol	60.1	2	−0.15	−0.16	−0.16
36	2-Methylpropanol	74.1	2	−0.17	−0.08	−0.07
37	2-Methylpentane	86.2	0	0.97	0.44	0.42
38	2,2-Dimethylbutane	86.2	0	1.04	0.44	0.42
39	1,1,1-Trifluoro-2-chloroethane	118.5	0	0.08	0.62	0.63
40	1,1,1-Trichloroethane	133.4	0	0.40	0.69	0.70
41	Diethyl ether	74.1	1	0.00	0.15	0.16
42	Enflurane	184.5	1	0.24	0.68	0.69
43	Ethanol	46.1	2	−0.16	−0.25	−0.26
44	Fluroxene	126.1	1	0.13	0.43	0.44
45	Halothane	197.4	0	0.35	0.95	0.99
46	Heptane	100.2	0	0.81	0.52	0.51
47	Hexane	86.2	0	0.80	0.44	0.43
48	Isoflurane	184.5	1	0.42	0.68	0.69
49	Methylcyclopentane	84.2	0	0.93	0.43	0.41
50	Pentane	72.2	0	0.76	0.36	0.34
51	Propanol	60.1	2	−0.16	−0.16	−0.16
52	Propanone	58.1	1	−0.15	0.05	0.07
53	Teflurane	180.9	0	0.27	0.89	0.92
54	Toluene	92.1	0	0.37	0.48	0.48
55	Trichloroethene	131.4	0	0.34	0.68	0.69
56	Acetylsalicylic acid	180.2	5	−0.50	−0.25	−0.24
57	Valproic acid	144.2	3	−0.22	0.06	0.06
58	Salicylic acid	138.1	5	−1.10	−0.42	−0.39
59	*p*-Acetamidophenol	151.2	5	−0.31	−0.36	−0.37
60	Chlorambucil	304.2	4	−1.70	–	–
61		204.3	4	−1.30	–	–
62		161.2	5	−1.40	–	–
63		276.4	6	−0.43	−0.20	−0.20
64		260.4	4	0.25	0.22	0.22
65		136.2	3	−0.30	0.02	0.03
66		150.2	2	−0.06	0.31	0.32
67		128.2	4	−0.42	−0.24	−0.24
68		292.5	6	−0.16	−0.17	−0.17
69		236.3	5	0.00	−0.07	−0.07
70		252.3	6	−0.34	−0.25	−0.25
71		357.8	7	−0.30	−0.34	−0.34
72		373.8	9	−1.34	−0.79	−0.77
73		389.8	11	−1.82	−1.25	−1.20
74		277.4	1	0.89	0.93	0.94
75		264.4	2	0.99	0.68	0.66
76		295.4	4	0.82	0.28	0.27
77		281.4	4	1.03	0.26	0.24
78		265.7	4	1.64	–	–
79		281.4	5	0.52	0.03	0.02
80		267.3	5	0.39	0.01	−0.00
81		265.4	3	0.53	0.46	0.45
82		312.5	3	0.40	0.53	0.54
83		426.9	10	−0.78	−1.05	−1.09
84		264.3	5	0.00	0.00	0.00
85		410.5	6	−0.02	−0.13	−0.14
86		426.5	8	−0.67	−0.60	−0.59

[a] Calculated from Eq. (1).
[b] Predicted using the leave-one-out cross validation procedure.

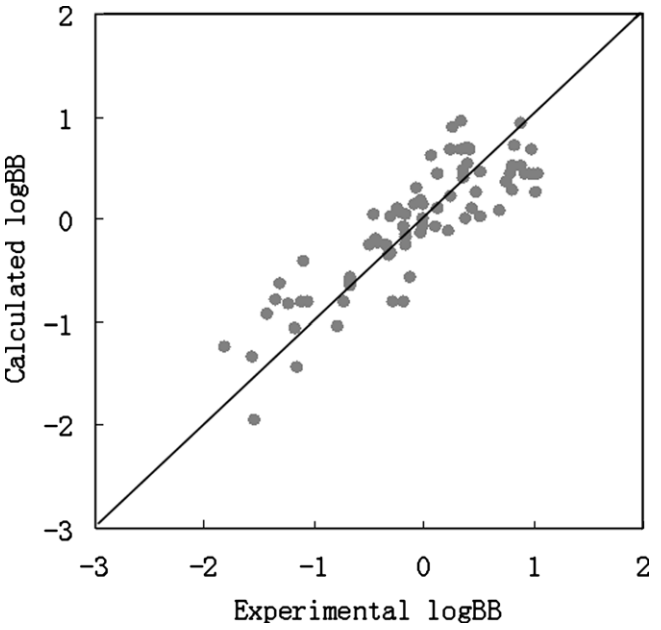

Fig. 2. Relationship between experimental and calculated logBB values for the training set.

3.2. Model validation using the leave-one-out procedure

The predictive model (Eq. (1)) is validated using leave-one-out procedure. Its cross validation coefficient (q^2 = 0.71) is almost same as its correlation coefficient (r^2 = 0.74). The predicted values using the leave-one-out cross validation procedure (shown in Table 1) are also very close to the respective calculated values from Eq. (1).

3.3. Model validation using test set outside the training set

In order to assess the predictive power of Eq. (1) further, a test set of logBB values are predicted. The experimental and predicted logBB values are listed in Table 2 and plotted in Fig. 3.

As may be seen from Table 2 and Fig. 3, the predicted logBB values from Eq. (1) are in good agreement with the respective experimental ones and only 3 compounds (94, 95, and 105) are predicted above or near two standard deviations. The RMSE value (root mean square error) calculated on the 25 validation compounds is 0.53. Considering the experimental difficulties and the varied experimental conditions under which the logBB values have been obtained, the predictive model for BBB penetration containing only

Table 2
Experimental and calculated logBB values for the test set compounds and their descriptors

	Compound	M_W	n_{pol}	logBB				
				Exp.	Pred.[a]	Pred.[b]	Pred.[c]	Pred.[d]
87	Theophylline	180.2	7	−0.29	−0.70	−1.43	−0.512	−0.86
88	Caffeine	194.2	6	−0.06	−0.42	−1.03	−0.219	−0.40
89	Antipyrine	188.2	3	−0.10	0.24	−0.03	0.474	0.39
90	Ibuprofen	206.3	3	−0.18	0.30	−0.09	−0.555	0.20
91	Codeine	299.4	5	0.55	0.06	−0.75	0.271	0.12
92	Pentobarbital	226.3	7	0.12	−0.55	−0.77	−0.191	−0.81
93	Alprazolam	308.8	4	0.04	0.30	−0.58	0.332	0.16
94	Indomethacin	357.8	6	−1.26	−0.11	−1.07	−1.032	0.52
95	Oxazepam	286.7	6	0.61	−0.18	−0.70	−0.476	0.39
96	Hydroxyzine	374.9	5	0.39	0.11	−0.20	0.128	0.10
97	Desipramine	318.9	2	1.20	0.77	0.77	0.426	0.99
98	Midazolam	327.8	3	0.36	0.55	−0.02	0.400	0.49
99	Verapamil	454.6	6	−0.70	−0.18	−1.32	−1.111	1.07
100	Promazine	284.4	2	1.23	0.72	0.78	0.832	1.02
101	Chlorpromazine	318.9	2	1.06	0.77	0.86	0.710	1.01
102	Trifluoroperazine	352.4	2	1.44	0.79	0.70	0.459	0.98
103	Thioridazine	370.6	2	0.24	0.79	0.89	1.062	0.92
104	BCNU	214.1	6	−0.52	−0.36	−0.56	−0.570	0.54
105	Phenserine	337.4	6	1.00	−0.12	−0.23	0.230	0.08
106	Physostigmine	275.4	6	0.08	−0.20	−0.50	0.007	0.05
107	Terbutylchlorambucil	360.3	3	1.00	0.57	0.28	−0.227	0.50
108	Didanosine	236.2	9	−1.30	−0.97	−1.95	−0.816	−1.19
109	Zidovudine	267.2	11	−0.72	−1.35	−2.37	−1.024	−1.92
110	Nevirapine	266.3	6	0.00	−0.22	−0.95	−0.285	−0.11
111	SB-222200	380.5	4	0.30	0.34	0.19	0.426	0.05

[a] Predicted from Eq. (1).
[b] Predicted from the model developed by Feher et al. [12].
[c] Predicted from the model developed by Luco [11].
[d] Predicted from the model developed by Fu et al. [20].

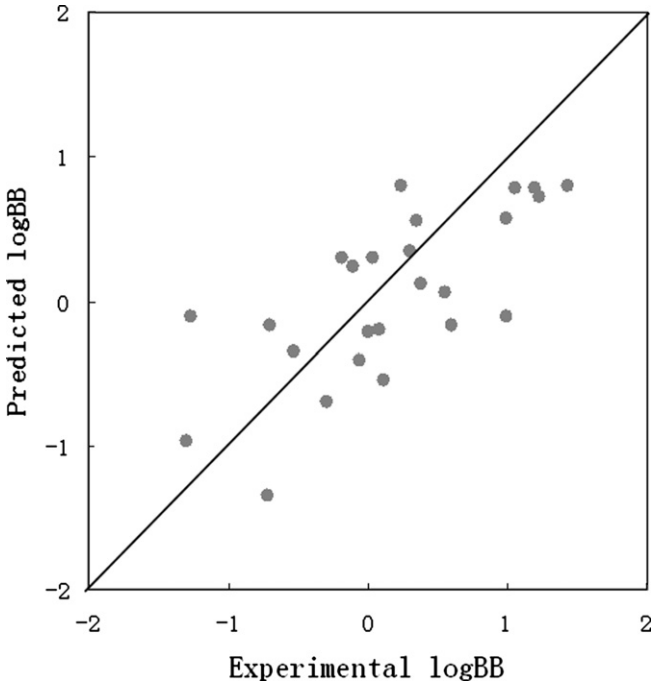

Fig. 3. Relationship between experimental and predicted logBB values for the test set.

molecular weight and number of polar atoms performs reasonably well.

This model compares favorably with others reported by Luco and Feher et al. [11,12]. A smaller training set of 61 compounds were used to develop their models. They predicted the logBB values for the same test set. As shown in Table 2, our prediction results are superior to the ones obtained from the model reported by Feher et al. (RMSE = 0.79) and as good as the three-component model based on 25 descriptors using the multivariate partial least-squares procedure (RMSE = 0.54). However, our model is much simpler than their models. Eq. (1) has similar predictive ability to our previous model (RMSE is 0.53 for the same test set) [20]. The descriptors in the older model are molecular volume and polar molecular surface area and in particular professional knowledge and software are needed to calculate them. The descriptors of the new model are molecular weight and number of polar atoms which can be obtained very easily. Therefore, the new model can be used for the first screening of BBB permeability more conveniently.

4. Conclusions

The model derived in this paper for the prediction of BBB penetration shows a good predictive power. It contains only three descriptors, namely number of polar atoms, molecular weight and square of molecular weight, which can be easy to interpret and compute. The model appears to be very simple and likely to give useful estimations of logBB values, so it is suitable for the rapid prediction of the BBB penetration of drug candidates. As in the case of most of the predictive models, our model is also only valid for passive diffusion processes across the BBB. Cellular BBB permeability might be totally different. There are non-passive transport systems such as P-glycoprotein efflux in the brain and compounds which are affected by these are not likely to be well predicted.

References

[1] W.M. Pardridrge, L.J. Mietus, Transport of steroid hormones through the rat blood–brain barrier, J. Clin. Invest. 64 (1979) 145–154.
[2] R.C. Young, R.C. Mitchell, T.H. Brown, C.R. Ganellin, R. Griffiths, M. Jones, K. K Rana, D. Saunders, I. R Smith, N. E Sore, T. J Wilks, Development of a new physicochemical model for brain penetration and its application to the design of centrally acting H_2 receptor histamine antagonists, J. Med. Chem. 31 (1988) 656–671.
[3] H.H.H. Sveigaard, L. Dalgaard, Evaluation of blood–brain barrier passage of a muscarine M1 agonist and a series of analogous tetrahydropyridines measured by in vivo microdialysis, Pharm. Res. 17 (2000) 70–76.
[4] M. Gumbleton, K.L. Audus, Progress and limitations in the use of in vitro cell cultures to serve as a permeability screen for the blood–brain barrier, J. Pharm. Sci. 90 (2001) 1681–1698.
[5] H. van de Waterbeemd, M. Kansy, Hydrogen bonding capacity and brain penetration, Chimia 46 (1992) 299–303.
[6] J.A.D. Calder, C.R. Ganellin, Predicting the brain-penetrating capability of histaminergic compounds, Drug Des. Discov. 11 (1994) 259–268.
[7] M.H. Abraham, H.S. Chadha, R.C. Michell, Hydrogen bonding. 33. Factors that influence the distribution of solutes between blood and brain, J. Pharm. Sci. 83 (1994) 1257–1268.
[8] F. Lombardo, J.F. Blake, W.J. Curatolo, Computation of brain–blood partitioning of organic solutes via free energy calculation, J. Med. Chem. 39 (1996) 4750–4755.
[9] U. Norinder, P. Sjoberg, T. Osterberg, Theoretical calculation and prediction of brain–blood partitioning of organic solutes using Molsurf parameterization and PLS statistics, J. Pharm. Sci. 87 (1998) 952–959.
[10] D.E. Clark, Rapid calculation of polar surface area and its application to the prediction of transport phenomena. 2. Prediction of blood–brain barrier penetration, J. Pharm. Sci. 88 (1999) 815–821.
[11] J.M. Luco, Prediction of the brain–blood distribution of a large set of drugs from structurally derived descriptors using partial least-squares (PLS) modeling, J. Chem. Inf. Comput. Sci. 39 (1999) 396–404.
[12] M. Feher, E. Sourial, J.M. Schmidt, A simple model for the prediction of blood–brain partitioning, Int. J. Pharm. 201 (2000) 239–247.
[13] P. Crivori, G. Cruciani, P. Carrupt, B. Testa, Predicting blood–brain barrier permeation from three-dimensional molecular structure, J. Med. Chem. 43 (2000) 2204–2216.
[14] Y.N. Kaznessis, M.E. Snow, C.J. Blankley, Prediction of blood–brain partitioning using Monte Carlo simulations of molecules in water, J. Comput. Aid. Mol. Des. 15 (2001) 697–708.
[15] K. Rose, L.H. Hall, Modeling blood–brain barrier partitioning using the electropological state, J. Chem. Inf. Comput. Sci. 42 (2002) 651–666.
[16] F. Ooms, T. Weber, P.A. Carrupt, B. Testa, A simple model to predict blood–brain barrier permeation from 3D molecular fields, Biochim. Biophys. Acta 1587 (2002) 118–125.

[17] M. Lobell, L. Molnar, G. Keseru, Recent advances in the prediction of blood–brain partitioning from molecular structure, J. Pharm. Sci. 92 (2003) 360–370.
[18] G. Subramanian, D.B. Kitchen, Computational models to predict blood–brain barrier permeation and CNS activity, J. Comput. Aid. Mol. Des. 17 (2003) 643–664.
[19] M.H. Abraham, The factors that influence permeation across the blood–brain barrier, Eur. J. Med. Chem. 39 (2004) 235–240.
[20] X.C. Fu, Z.F. Song, C.Y. Fu, W.Q. Liang, A simple predictive model for blood–brain barrier penetration, Pharmazie 60 (2005) 354–358.
[21] M.H. Abraham, H.S. Chadha, R.C. Michell, Hydrogen bonding. Part 36. Determination of blood brain distribution using octanol–water partition coefficients, Drug Des. Discov. 13 (1995) 123–131.
[22] N.H. Greig, A. Brossi, X.F. Pei, D.K. Ingram, T.T. Soncrant, in: J. Greenwood et al. (Eds.), New Concepts of a Blood–Brain Barrier, Plenum, New York, 1995, pp. 251–264.
[23] T. Salminen, A. Pulli, J. Taskinen, Relationship between immobilized artificial membrane chromatographic retention and the brain penetration of structurally diverse drug, J. Pharm. Biomed. Anal. 15 (1997) 469–477.
[24] A. von Sprecher, M. Gerpacher, G.P. Anderson, Neurokinin antagonists as potential therapies for inflammation and rheumatoid arthritis, Drugs 1 (1998) 73–91.
[25] M. Yazdanian, S.L. Glynn, In vitro blood–brain barrier permeability of nevirapine compared to other HIV antiretroviral agents, J. Pharm. Sci. 87 (1998) 306–310.
[26] J. Kelder, P.D.J. Grootenhuis, D.M. Bayada, L.P.C. Delbressine, J.P. Ploemen, Polar molecular surface as a dominating determinant for oral absorption and brain penetration of drugs, Pharm. Res. 16 (1999) 1514–1519.
[27] S. Wold, Cross-validatory estimation of the number of components in factor and principal component models, Technometrics 20 (1978) 397–406.