CODEN IEJMAT *Internet Electronic Journal of Molecular Design* **2005**, *4*, 737–750 ISSN 1538–6414
BioChem Press http://www.biochempress.com

Inter*net* Electronic Journal of Molecular Design

October 2005, Volume 4, Number 10, Pages 737–750

Editor: Ovidiu Ivanciuc

Proceedings of the Internet Electronic Conference of Molecular Design 2004
IECMD 2004, November 29 – December 12, 2004

A Predictive Model for Blood–Brain Barrier Penetration

Xuchun Fu,[1] Zhifang Song,[2] and Wenquan Liang [3]

[1] Department of Pharmacy, Zhejiang University City College, Hangzhou, P. R. China
[2] College of Medicine, Zhejiang University, Hangzhou, P. R. China
[3] College of Pharmaceutical Sciences, Zhejiang University, Hangzhou, P. R. China

Received: November 7, 2004; Revised: June 6, 2005; Accepted: July 1, 2005; Published: October 31, 2005

Citation of the article:
X. Fu, Z. Song, and W. Liang, A Predictive Model for Blood–Brain Barrier Penetration, *Internet Electron. J. Mol. Des.* **2005**, *4*, 737–750, http://www.biochempress.com.

Inter*net* Electronic Journal of Molecular Design

BioChem Press
http://www.biochempress.com

A Predictive Model for Blood–Brain Barrier Penetration

Xuchun Fu,[1,*] Zhifang Song,[2] and Wenquan Liang [3]

[1] Department of Pharmacy, Zhejiang University City College, Hangzhou, P. R. China
[2] College of Medicine, Zhejiang University, Hangzhou, P. R. China
[3] College of Pharmaceutical Sciences, Zhejiang University, Hangzhou, P. R. China

Received: November 7, 2004; Revised: June 6, 2005; Accepted: July 1, 2005; Published: October 31, 2005

Internet Electron. J. Mol. Des. **2005**, *4* (10), 737–750

Abstract

Motivation. It is important to determine whether a candidate molecule is capable of penetrating the blood–brain barrier in drug discovery and development. The aim of this paper is to develop a predictive model for blood–brain barrier penetration only using two simple descriptors, molecular volume (*V*) and 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).

Method. A data set of 100 compounds, which was studied by other research groups, is divided into a training set of 61 compounds and two test sets (14 and 25 compounds). Molecular volumes and polar surface areas are obtained from the molecular conformations optimized using the semiempirical self–consistent field molecular orbital calculation AM1 method. The model to predict blood–brain barrier penetration from molecular volume and polar surface area is derived on the training set using the stepwise multiple regression analysis and then cross–validated using leave–one–out procedure and tested on the external prediction.

Results. A log*BB* model is developed using the training set of 61 compounds (4 compounds are excluded as outliers): log*BB* = –16.79 (±3.28) V^2 + 11.24 (±2.06) *V* – 2.249 (±0.161) *PSA* – 0.6583 (±0.2326) (n = 57, r^2 = 0.832, q^2 = 0.804, s = 0.329, F = 87.2), where log*BB* is the logarithm of the ratio of the steady–state concentration of a compound in the brain to in the blood, n is the number of compounds, r is the correlation coefficient, q is the cross validation coefficient, s is the standard deviation, F is the Fisher F–statistic. The model is validated through two external test sets (14 compounds and 25 compounds). The root mean squared errors (RMSE) are 0.599 for test set 1 of 14 compounds and 0.551 for test set 2 of 25 compounds. The simple model performs as well as other log*BB* models developed using the same data set but different descriptors.

Conclusions. The model derived in this paper for the prediction of BBB penetration shows a good predictive power. It shows that the hydrogen–bonding potential, lipophilicity, and molecular size are important factors to affect BBB penetration. The model is one of the simplest log*BB* models and suitable for the rapid prediction of the BBB penetration for a wide range of drug candidates.

Keywords. Blood–brain barrier; predictive model; molecular volume; polar surface area.

Abbreviations and notations

BBB, blood–brain barrier	BB, brain/blood concentration ratio
CNS, central nervous system	V, molecular volume
PSA, polar surface area	RMSE, root mean square error

Presented in part at the Internet Electronic Conference of Molecular Design 2004, IECMD 2004.

* Correspondence author; phone: 86–571–8801–8711; fax: 86–571–8801–8442; E–mail: fuxc@zucc.edu.cn.

1 INTRODUCTION

It is important to determine whether a candidate molecule is capable of penetrating the blood–brain barrier (BBB) in drug discovery and development. Drugs that act in the central nervous system (CNS) need to cross the BBB to reach their molecular target. By contrast, for drugs with a peripheral target, little or no BBB penetration might be required 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 in the blood, usually expressed as $\log(C_{brain/blood})$ or log*BB*. The experimental determination of log*BB* is a time–consuming, expensive, and difficult technique, requiring animal experiments and the synthesis of the test compounds, usually in radiolabeled form [1–4]. It is of considerable value to predict log*BB* values of compounds from their physicochemical parameters or, ideally, from their molecular structures.

Young et al. [2] showed that log*BB* values of 20 H_2 receptor histamine antagonists were correlated with $\Delta\log P$ (octanol–cyclohexane):

$$\log BB = -0.485\ (\pm 0.160)\ \Delta\log P + 0.889\ (\pm 0.500)$$
$$n = 20\ r = 0.831\ s = 0.439\ F = 40.23 \quad (1)$$

where n is the number of compounds, r is the correlation coefficient, s is the standard deviation, F is the Fisher F–statistic. The quantities in parentheses are the standard deviations of the coefficients.

Van de Waterbeemd and Kansy [5] examined the same series of 20 compounds and found a significant correlation between log*BB* and the cyclohexane–water partition coefficient ($\log P_{cyc}$) when the molecular volume was included in the parameterization. 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):

$$\log BB = 0.338\ \log P_{cyc} - 0.00618\ V_M + 1.359$$
$$n = 20 \quad r = 0.934 \quad s = 0.29 \quad F = 58 \quad (2)$$

$$\log BB = -0.021\ (\pm 0.003)\ PSA + 0.003\ (\pm 0.001)\ V_M + 1.643\ (\pm 0.465)$$
$$n = 20 \quad r = 0.835 \quad s = 0.448 \quad F = 19.5 \quad (3)$$

where V_M is a calculated solute molar volume. The great advantage of Eq. (3) is that it only uses the calculated parameters. However, the model showed it to be poorly predictive when tested with compounds outside its training set [6], 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*. Thus Abraham *et al*. [7] constructed a larger training set of 65 compounds and derived a correlation between log*BB* and solvato–chromatic parameters for 57 compounds (8 compounds were excluded as outliers):

$$\log BB = -0.038\ (\pm 0.064) + 0.198\ (\pm 0.100)\ R_2 - 0.687\ (\pm 0.125)\ \pi_2^H - 0.715\ (\pm 0.334)\ \Sigma\alpha_2^H - 0.698\ (\pm 0.107)\ \Sigma\beta_2^H + 0.995\ (\pm 0.096)\ V_x \quad (4)$$
$$n = 57 \quad r = 0.952 \quad s = 0.197 \quad F = 99.2$$

where 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, and V_x is the characteristic volume of McGowan. From the data set above (57 compounds, of which some compounds were excluded as outliers), Lombardo [8], and Clark [9] utilized calculated parameters (free energy of solvation in water, ΔG_w^0 , *PSA*, and calculated octanol–water partition coefficient, *ClogP* or *MlogP*) to develop log*BB* models, respectively.

$$\log BB = 0.054\ (\pm 0.0005)\ \Delta G^0{}_W + 0.43\ (\pm 0.07) \quad (5)$$
$$n = 55 \quad r = 0.82 \quad s = 0.41 \quad F = 108.3$$

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

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

Luco [10] constructed a data set of 100 compounds which were divided into a training set (61 compounds) and two test sets (14 and 25 compounds) and developed a three–component model using the partial least–squares procedure (PLS) with 18 topological and constitutional descriptors. Both Feher [11] and Subramanian [12] examined the same data set and proposed the logBB models as Eq. (8) and (9):

$$\log BB = -0.0017\ A_{pol} + 0.1092\ ClogP - 0.3873\ n_{acc,solv} + 0.4275 \quad (8)$$
$$n = 61 \quad r^2 = 0.730 \quad q^2 = 0.688 \quad s = 0.424 \quad F = 51$$

where A_{pol} is the polar surface area estimated from the solvent–exposed area of the compound assuming a spherical solvent molecule with a radius of 1.4 Å and considering only those parts of the surface with the absolute value of the partial charge greater than 0.2, $n_{acc,solv}$ is the number of hydrogen–bond acceptors in an aqueous medium, q is the cross validation coefficient.

$$\log BB = 0.122\ S\_sssN - 0.114\ Rotlbonds + 0.0359\ Jurs\text{–}WNSA\text{–}3 - 0.0615\ S\_dsN + 0.313\ AlogP - 0.0959\ S\_sssCH + 0.108\ R_{og} - 0.0204 \quad (9)$$
$$n = 58 \quad r^2 = 0.845 \quad q^2 = 0.811 \quad s = 0.314 \quad F = 97.9$$

where *S_sssN* is the summation of the electrotopological indices for all N atoms connected by three single bonds, *Rotlbonds* is the number of rotatable bonds, *Jurs–WNSA–3* is the surface weighted charged partial surface area, *S_dsN* is the summation of the electrotopological indices for all N atoms connected by a double and single bond, *AlogP* is Ghose and Crippen logP, *S_sssCH* is the summation of the electrotopological indices for all CH groups connected by three single bonds, and R_{og} is the radius of gyration.

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 predictive model for BBB penetration only using two simple descriptors, molecular volume (*V*) and polar surface area.

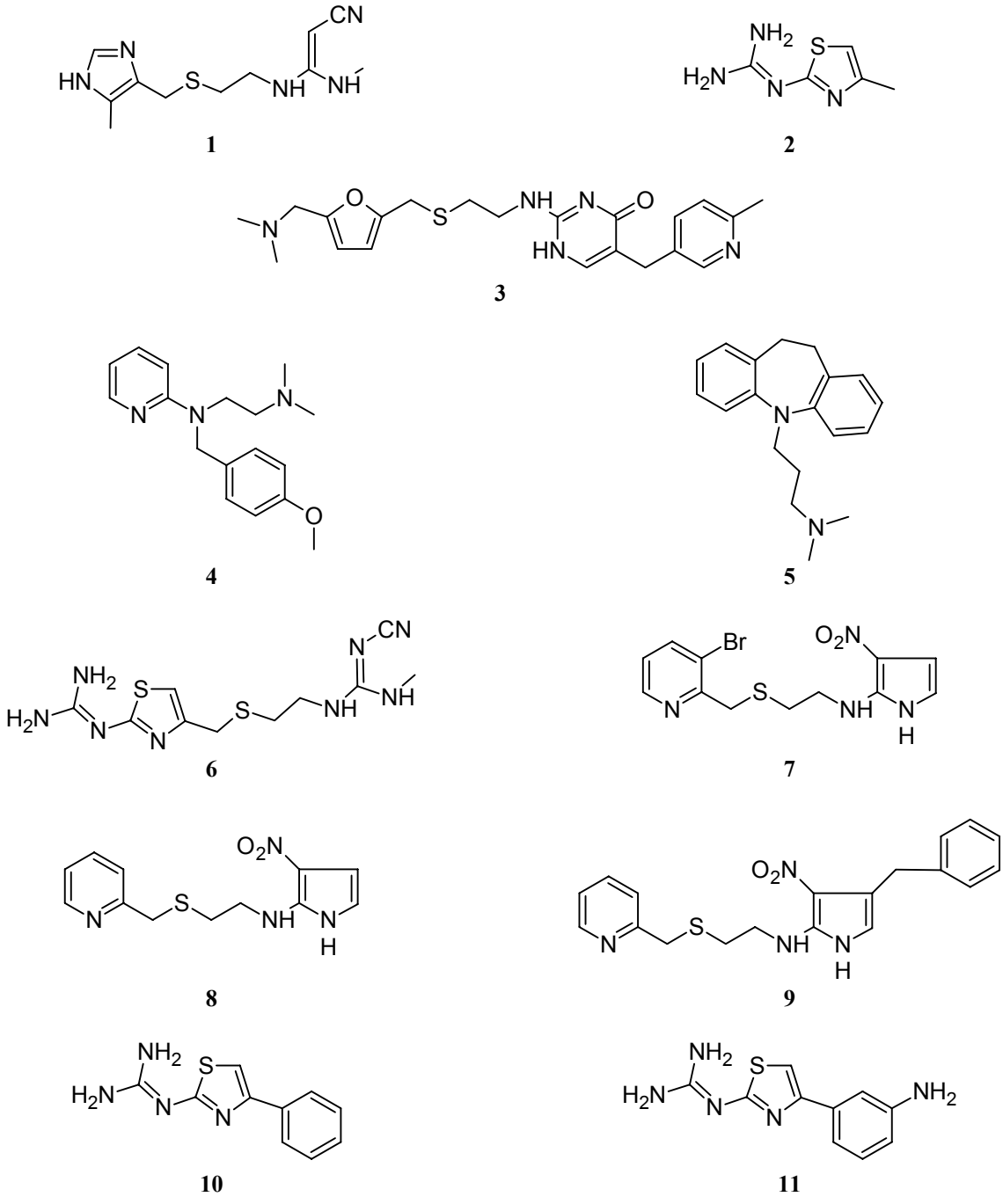
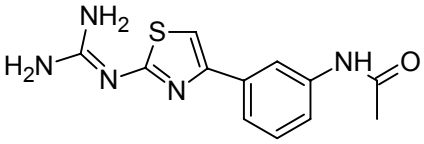
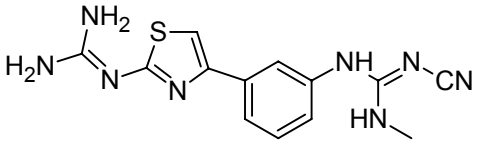


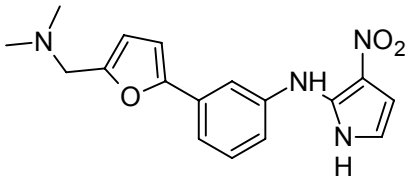
Figure 1. Compounds **1–31** and **62–75**.



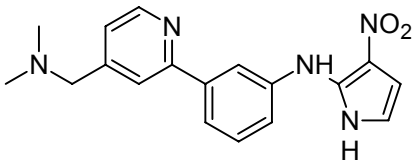
12



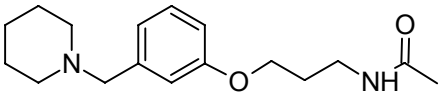
13



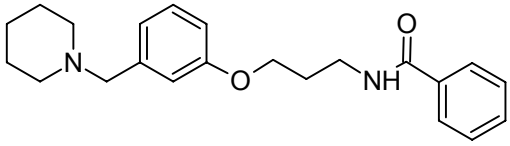
14



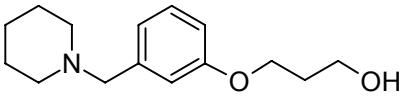
15



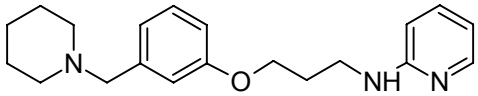
16



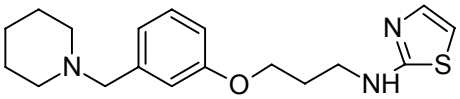
17



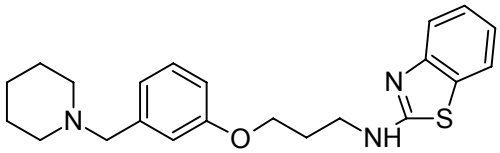
18



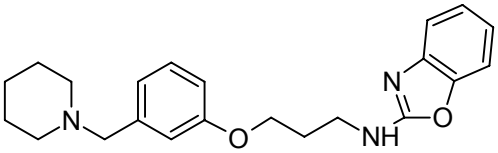
19



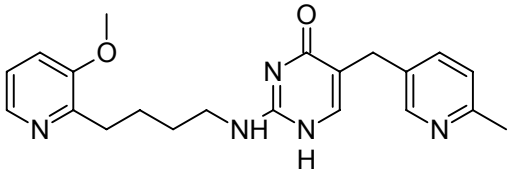
20



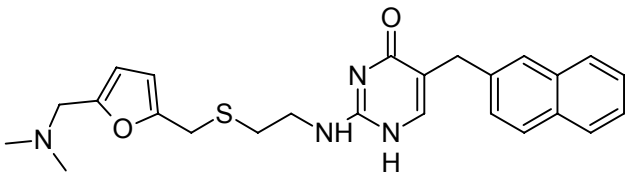
21



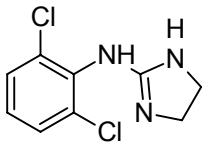
22



23

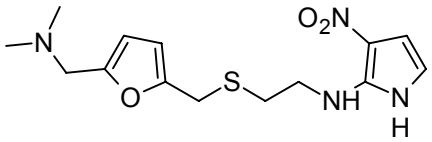


24

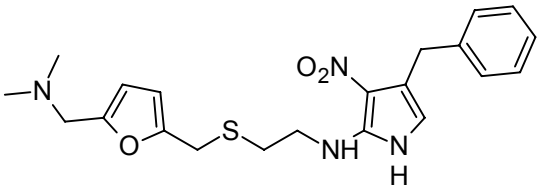


25

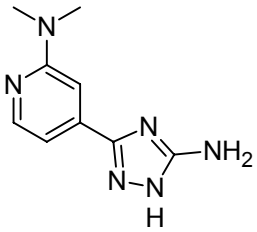
Figure 1. (Continued).



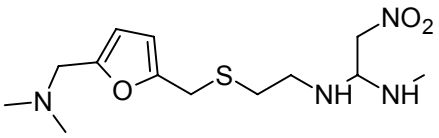
26



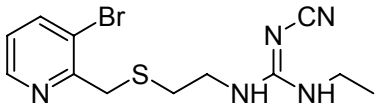
27



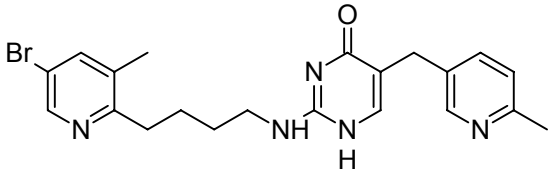
28



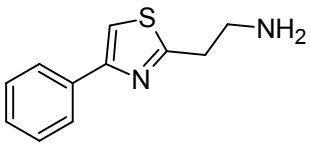
29



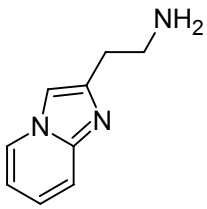
30



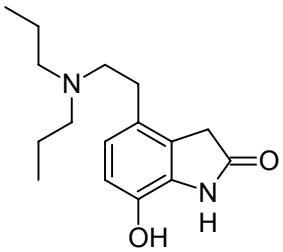
31



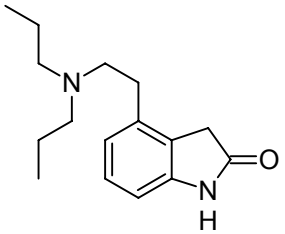
62



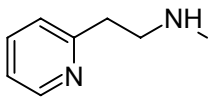
63



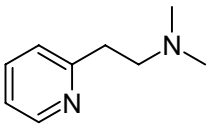
64



65



66



67

Figure 1. (Continued).

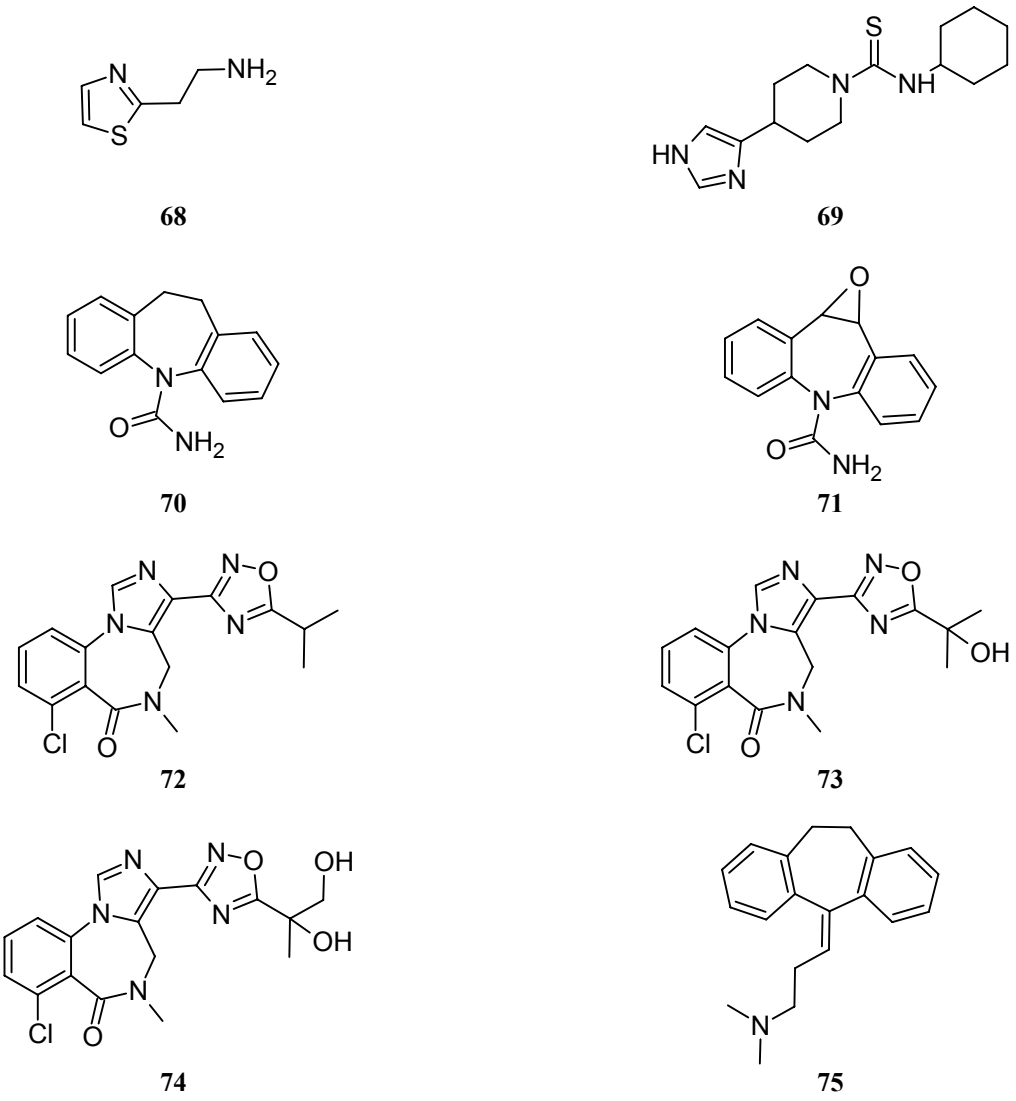


Figure 1. (Continued).

2 MATERIALS AND METHODS

The data set of 100 compounds and their corresponding log*BB* values is taken from the literature [10]. The division of compounds into a training set (61 compounds) and two test sets (14 and 25 compounds) is taken from the same source [10]. Molecular volumes and polar surface areas are selected as the structural descriptors to develop predictive model for BBB penetration. These structural descriptors are obtained from the molecular conformations optimized using the semiempirical self–consistent field molecular orbital calculation AM1 method [13] and the atomic radii used by Clark [9]. 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 [14] 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.

Table 1. Experimental and calculated log*BB* values for the training set compounds and their computed descriptors

No	Compound	V (nm^3)	PSA (nm^2)	log*BB*		
				Exp.[a]	Calc.[b]	Res.[c]
1		0.3097	0.9784	–1.42	–0.99	0.43
2		0.1735	0.7807	–0.04	–0.97	–0.93
3		0.5088	0.8774	–1.06	–1.26	–0.20
4		0.3812	0.3011	0.49	0.51	0.02
5		0.3828	0.0540	0.83	1.06	0.23
6		0.3488	1.4402	–0.82	–2.02	–1.20
7		0.3424	0.8425	–0.67	–0.67	–0.00
8		0.3169	0.8517	–0.66	–0.70	–0.04
9		0.4313	0.8171	–0.12	–0.77	–0.65
10		0.2418	0.7636	–0.18	–0.64	–0.46
11		0.2516	1.0403	–1.15	–1.23	–0.08
12		0.3016	1.0698	–1.57	–1.20	0.37
13		0.3420	1.3859	–1.54	–1.89	–0.35
14		0.3902	0.9170	–0.27	–0.89	–0.62
15		0.3897	0.9412	–0.28	–0.94	–0.66
16		0.3941	0.4831	–0.46	0.08	0.54
17		0.4633	0.4442	–0.24	–0.05	0.19
18		0.3383	0.3815	–0.02	0.36	0.38
19		0.4327	0.3664	0.69	0.24	–0.45
20		0.4219	0.3753	0.44	0.25	–0.19
21		0.4773	0.3608	0.14	0.07	–0.07
22		0.4654	0.5428	0.22	–0.28	–0.50
23		0.4736	0.9747	–2.00	–1.29	0.71
24		0.5482	0.7260	–1.30	–1.17	0.13
25		0.2404	0.4206	0.11	0.13	0.02
26		0.3875	0.8629	–1.12	–0.76	0.36
27		0.5010	0.8539	–0.73	–1.16	–0.43
28		0.2415	0.9040	–1.17	–0.96	0.21
29		0.3882	0.8955	–1.23	–0.84	0.39
30		0.3562	0.7315	–2.15	–0.43	1.72
31		0.4863	0.8364	–1.88	–1.04	0.84
32	butanone	0.1164	0.1998	–0.08	–0.03	0.05
33	benzene	0.1147	0.0000	0.37	0.41	0.04
34	3–methylpentane	0.1597	0.0000	1.01	0.71	–0.30
35	3–methylhexane	0.1828	0.0000	0.90	0.84	–0.06
36	2–propanol	0.0989	0.2311	–0.15	–0.23	–0.08
37	2–methylpropanol	0.1223	0.2201	–0.17	–0.03	0.14
38	2–methylpentane	0.1608	0.0000	0.97	0.71	–0.26
39	2,2–dimethylbutane	0.1587	0.0000	1.04	0.70	–0.34
40	1,1,1–trifluoro–2–chloroethane	0.1009	0.0000	0.08	0.31	0.23
41	1,1,1–trichloroethane	0.1237	0.0000	0.40	0.48	0.08
42	diethyl ether	0.1272	0.1052	0.00	0.26	0.26
43	enflurane	0.1446	0.0918	0.24	0.41	0.17
44	ethanol	0.0760	0.2421	–0.16	–0.45	–0.29
45	fluroxene	0.1311	0.1104	0.13	0.28	0.15
46	halothane	0.1273	0.0000	0.35	0.50	0.15
47	heptane	0.1857	0.0000	0.81	0.85	0.04
48	hexane	0.1630	0.0000	0.80	0.73	–0.07
49	isoflurane	0.1444	0.1003	0.42	0.39	–0.03
50	methylcyclopentane	0.1460	0.0000	0.93	0.63	–0.30
51	pentane	0.1388	0.0000	0.76	0.58	–0.18
52	propanol	0.0995	0.2417	–0.16	–0.25	–0.09
53	propanone	0.0932	0.2201	–0.15	–0.25	–0.10
54	teflurane	0.1141	0.0000	0.27	0.41	0.14
55	toluene	0.1389	0.0000	0.37	0.58	0.21

Table 1. (Continued)

No	Compound	V (nm^3)	PSA (nm^2)	log*BB*		
				Exp.[a]	Calc.[b]	Res.[c]
56	trichloroethene	0.1130	0.0000	0.34	0.40	0.06
57	acetylsalicylic acid	0.2048	0.6940	–0.50	–0.62	–0.12
58	valproic acid	0.2155	0.4233	–0.22	0.03	0.25
59	salicylic acid	0.1522	0.6312	–1.10	–0.76	0.34
60	*p*–acetamidophenol	0.1817	0.5959	–0.31	–0.51	–0.20
61	chlorambucil	0.3575	0.4884	–1.70	0.12	1.82

[a] From reference [10]; [b] Calculated from Eq. (10); [c] Residuals

3 RESULTS AND DISCUSSION

3.1 The Predictive Model of BBB Penetration only Including *V* and *PSA*

The 61 compounds of training set are illustrated in Figure 1 and listed in Table 1 along with their experimental log*BB* values.

Using *PSA* and *V* as regression variables, the following regression equation is obtained from the stepwise multiple regression analysis (including quadratic terms) for the training set of 61 compounds,

$$\log BB = -16.79\ (\pm 3.28)\ V^2 + 11.24\ (\pm 2.06)\ V - 2.249\ (\pm 0.161)\ PSA - 0.6583\ (\pm 0.2326)$$
$$n = 57 \quad r^2 = 0.832 \quad q^2 = 0.804 \quad s = 0.329 \quad F = 87.2 \qquad (10)$$

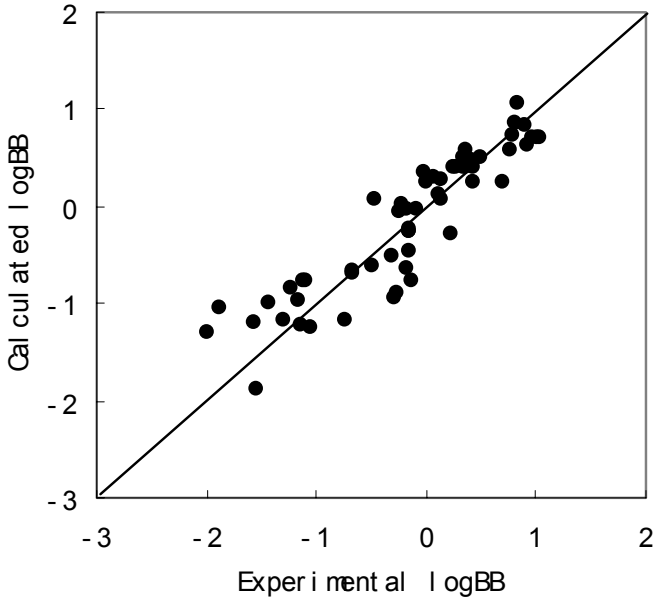


Figure 2. Relationship between experimental and calculated log*BB* values for the training set.

Compounds **2**, **6**, **30**, and **61** are found to have great residuals and removed from above equation as outliers. There appears to be no common structural similarity among the outliers that is significantly different from the rest of the training set. The outlier status of these compounds may

be due to inaccuracies in the measurement of the log*BB* value or may result from the influence of metabolic factors or active transport systems. The calculated log*BB* values for the training set are presented in Table 1 and the experimental and calculated log*BB* values are plotted in Figure 2.

Eq. (10) displays good statistical significance. As shown in Table 1 and Figure 2, the calculated log*BB* values are in good agreement with respective experimental ones. Because the polar surface area is a descriptor of hydrogen–bonding potential [15], Eq. (10) indicates that the log*BB* of a compound is inversely correlated with its hydrogen–bonding capacity. Eq. (10) shows the parabolic relation between log*BB* and molecular volume. The explicit descriptor for lipophilicity is absent from Eq. (10) and the molecular volume terms in the equation represent a combination of the impacts of molecular size and lipophilicity on BBB penetration. Increasing molecular volume decreases molecular diffusion through a lipid membrane and therefore decreases log*BB* value. On the other hand, bigger molecular volume also means higher lipophilicity which facilitates BBB penetration.

3.2 Model Validation Using the Leave–One–Out Procedure

Eq. (10) is validated using leave–one–out procedure. Its cross validation coefficient (q^2 = 0.804) is very close to its correlation coefficient (r^2 = 0.832).

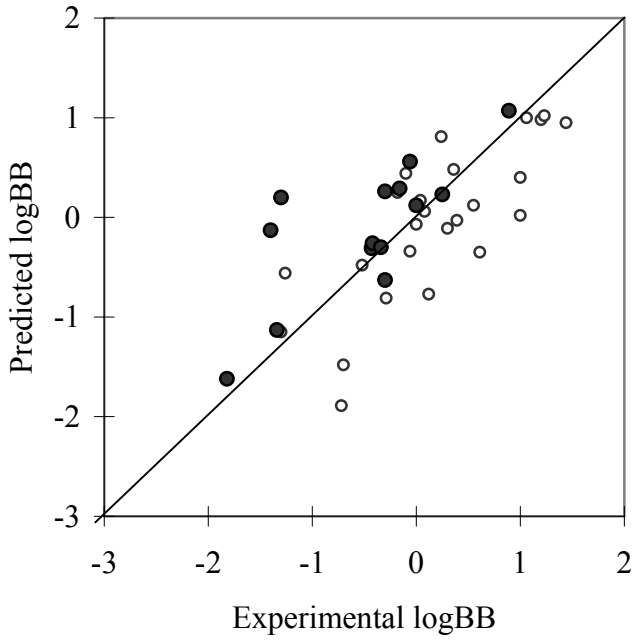


Figure 3. Relationship between experimental and predicted log*BB* values for the test sets. Test set 1 data are represented with ●, while test set 2 data are represented with ○.

3.3 Model Validation Using Test Set outside the Training Set

In order to assess the predictive power of Eq. (10) further, two test sets of log*BB* values are predicted. The experimental and predicted log*BB* values are listed in Table 2 and plotted in Figure 3, while compounds **62–75** of test set 1 are illustrated in Figure 1.

Table 2. Experimental and predicted log*BB* values for the test set compounds and their computed descriptors.

No	Compound	V (nm^3)	PSA (nm^2)	log*BB*				
				Exp.[a]	Pred.[b]	Pred.[c]	Pred.[d]	Pred.[e]
Test set 1								
62		0.2477	0.4004	–1.30	0.20	–0.036	–0.20	0.067
63		0.2051	0.4765	–1.40	–0.13	–0.292	–0.50	–0.159
64		0.3696	0.6736	–0.43	–0.31	–0.263	–0.28	–0.617
65		0.3624	0.4342	0.25	0.23	0.170	0.22	0.054
66		0.1936	0.2813	–0.30	0.26	–0.139	–0.07	–0.032
67		0.2164	0.1880	–0.06	0.56	0.016	0.01	0.350
68		0.1560	0.4216	–0.42	–0.26	–0.510	–0.44	–0.319
69		0.3755	0.4031	–0.16	0.29	–0.285	–0.39	0.342
70		0.2763	0.4667	0.00	0.12	–0.005	0.18	0.568
71		0.2858	0.6592	–0.34	–0.30	0.005	–0.46	0.064
72		0.3981	0.7959	–0.30	–0.63	–0.447	–1.63	0.204
73		0.4053	1.0088	–1.34	–1.13	–0.931	–2.19	–0.443
74		0.4124	1.2201	–1.82	–1.62	–1.308	–2.74	–1.289
75		0.3774	0.0560	0.89	1.07	0.966	1.00	1.414
Test set 2								
76	theophylline	0.1993	0.7688	–0.29	–0.81	–0.512	–1.43	0.011
77	caffeine	0.2253	0.6075	–0.06	–0.34	–0.219	–1.03	0.141
78	antipyrine	0.2357	0.2728	–0.10	0.44	0.474	–0.03	0.287
79	ibuprofen	0.2816	0.4133	–0.18	0.25	–0.555	–0.09	0.108
80	codeine	0.3596	0.4836	0.55	0.12	0.271	–0.75	–0.012
81	pentobarbital	0.2822	0.8646	0.12	–0.77	–0.191	–0.77	–0.545
82	alprazolam	0.3467	0.4675	0.04	0.17	0.332	–0.58	0.400
83	indomethacin	0.3988	0.7630	–1.26	–0.56	–1.032	–1.07	–1.633
84	oxazepam	0.3072	0.6951	0.61	–0.35	–0.476	–0.70	–0.743
85	hydroxyzine	0.4674	0.4264	0.39	–0.03	0.128	–0.20	–0.440
86	desipramine	0.3769	0.0932	1.20	0.98	0.426	0.77	0.943
87	midazolam	0.3677	0.3206	0.36	0.48	0.400	–0.02	–0.139
88	verapamil	0.5994	0.6787	–0.70	–1.48	–1.111	–1.32	–0.714
89	promazine	0.3607	0.0834	1.23	1.02	0.832	0.78	0.838
90	chlorpromazine	0.3788	0.0831	1.06	1.00	0.710	0.86	0.735
91	trifluoroperazine	0.3944	0.0948	1.44	0.95	0.459	0.70	0.311
92	thioridazine	0.4579	0.0698	0.24	0.81	1.062	0.89	0.708
93	BCNU	0.2258	0.6703	–0.52	–0.48	–0.570	–0.56	–0.877
94	phenserine	0.4191	0.4825	1.00	0.02	0.230	–0.23	1.000
95	physostigmine	0.3514	0.5167	0.08	0.06	0.007	–0.50	0.614
96	terbutylchlorambucil	0.4528	0.2624	1.00	0.40	–0.227	0.28	–0.939
97	didanosine	0.2625	1.0139	–1.30	–1.15	–0.816	–1.95	–1.115
98	zidovudine	0.2941	1.3735	–0.72	–1.89	–1.024	–2.37	–1.227
99	nevirapine	0.3132	0.5732	0.00	–0.07	–0.285	–0.95	–0.076
100	SB–222200	0.4817	0.4306	0.30	–0.11	0.426	0.19	0.723

[a] From reference [10]; [b] Predicted from Eq. (10); [c] Predicted from the model developed by Luco [10]; [d] Predicted from Eq. (8) [11]; [e] Predicted from Eq. (9) [12]

As may be seen from Table 2 and Figure 3, the predicted log*BB* values from Eq. (10) are in good agreement with the respective experimental ones for the test set 1 compounds except compounds **62** and **63**. The log*BB* values for the two compounds are strongly overestimated by the model. It can be seen that the log*BB* values of the two compounds are also overestimated by other models [10–12] and can be considered as outliers. For the remaining 12 non–outlier compounds, the root mean squared error (RMSE) is 0.314.

For test set 2 of 25 compounds, the predicted log*BB* values from Eq. (10) also agree well with the respective experimental ones and only three compounds (**84**, **94**, and **98**) are predicted above or near three standard deviations. The RMSE value calculated on the 25 validation compounds is 0.551, while the RMSE value for the reduced data set (excluding compounds **84**, **94**, and **98**) is 0.443. Considering the experimental difficulties and the varied experimental conditions under which the log*BB* values have been obtained, the predictive model for BBB penetration containing only molecular volume and polar molecular surface area performs reasonably well.

3.4 Comparison with Other Predictive Models

Luco [10], Feher [11], Subramanian [12] and their co–workers have analyzed the same data set. They developed logBB models from the same training data set of 61 compounds (excluding several outliers), then validated the models using leave–one–out procedure and external test sets. The model developed by Luco is a three–component model including 18 topological and constitutional descriptors [10]. Eq. (8) was developed by Feher *et al*. [11] and Eq. (9) was the best QSAR model derived by Subramanian *et al*. using constitutional, topological, and physicochemical descriptors [12]. The predicted log*BB* values for test sets from the models are listed in Table 2 and the properties of the models along with our model are presented in Table 3.

Table 3. Properties of four log*BB* models

	This work	Luco [10]	Feher *et al*. [11]	Subramanian *et al*. [12]
Training set	n = 57	n = 58 [a]	n = 61	n = 58[b]
r^2	0.832	0.850	0.730	0.845
s	0.329	0.318	0.428	0.314
q^2	0.804	0.752	0.688	0.811
Test set 1				
RMSE	0.599	0.499	0.628	0.659[c]
Test set 2				
RMSE	0.551	0.541	0.789	0.652[c]

[a] Compounds **2**, **6**, and **30** are excluded as outliers; [b] Compounds **6**, **28** and **30** are excluded as outliers; [c] RMSE for the reference [12] was calculated using the predicted and experimental values given there.

As shown in Table 2 and Table 3, our model performs as well as three–component PLS model [10] and seven–descriptor model, Eq. (9). They have similar outliers such as compounds **6** and **30** in

the training set, compounds **62** and **63** in test set 1, and compound **84** in test set 2. They also have similar correlation coefficient and standard deviation for the training set and similar RMSE values for both test set 1 and test set 2. However, our model is simpler than the other two models, so more suitable for the rapid prediction of the BBB penetration for a wide range of drug candidates.

Compared with Eq. (8) developed by Feher *et al*. [11], our model seems to perform better than Eq. (8), especially for the test sets. The predicted log*BB* values from Eq. (8) strongly deviate from the respective experimental ones for compounds **72** and **74** of test set 1 and compounds **76**, **77**, **80**, and **99** of test set 2 in addition to compounds **62**, **63**, **84**, **94**, and **98**.

4 CONCLUSIONS

The model derived in this paper for the prediction of BBB penetration shows a good predictive power. It contains only two descriptors, namely molecular volume and polar surface area which can be easy to interpret and compute. The model shows that the hydrogen–bonding potential, lipophilicity, and molecular size are important factors to affect BBB penetration. The model is one of the simplest log*BB* models and suitable for the rapid prediction of the BBB penetration for a wide range of drug candidates. However, it must be noted that our model is only valid for passive diffusion processes across the BBB. 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.

5 REFERENCES

[1] W. M. Pardridrge, L. J. Mietus, Transport of Steroid Hormones through the Rat Blood–Brain Barrier, *J. Clin. Invest*. **1979**, *64*, 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*. **1988**, *31*, 656–671.

[3] E. G. Chikhale, K. Y. Ng, P. S. Burton, R. T. Borchardt, Hydrogen Bonding Potential as a Determinant of the in vitro and in situ Blood–Brain Barrier Permeability of Peptides, *Pharm. Res*. **1994**, *11*, 412–419.

[4] 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*. **2000**, *17*, 70–76.

[5] H. van de Waterbeemd, M. Kansy, Hydrogen Bonding Capacity and Brain Penetration, *Chimia* **1992**, *46*, 299–303.

[6] J. A. D. Calder, C. R. Ganellin, Predicting the Brain–Penetrating Capability of Histaminergic Compounds, *Drug Des. Discov*. **1994**, *11*, 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*. **1994**, *83*, 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.* **1996**, *39*, 4750–4755.
[9] 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.* **1999**, *88*, 815–821.
[10] 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.* **1999**, *39*, 396–404.
[11] M. Feher, E. Sourial, J. M. Schmidt, A Simple Model for the Prediction of Blood–Brain Partitioning, *Int. J. Pharm.* **2000**, *201*, 239–247.
[12] G. Subramanian, D. B. Kitchen, Computational Models to Predict Blood–Brain Barrier Permeation and CNS Activity, *J. Comput. Aided Mol. Des.* **2003**, *17*, 643–64.
[13] M. J. S. Dewar, G.E. Zoebisch, E.F. Healy, J.J.P. Stewart, AM1: A New General Purpose Quantum Mechanical Molecular Model, *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.
[14] S. Wold, Cross–Validatory Estimation of the Number of Components in Factor and Principal Component Models, *Technometrics* **1978**, *20*, 397–406.
[15] P. Stenberg, U. Norinder, K. Luthman, P. Artursson, Experimental and Computational Screening Models for the Prediction of Intestinal Drug Absorption, *J. Med. Chem.* **2001**, *44*, 1927–1937.