Institute of Natural Medicine[1], Hefei University of Technology; Anhui Province Key Laboratory of R&D of Chinese Medicine[2], Anhui University of Traditional Chinese Medicine, Hefei, China

Comparison of prediction models for blood brain barrier permeability and analysis of the molecular descriptors

Ze-Yu Wu[1], Jian Pan[1], Yuan Yuan[1], Ai-Ling Hui[1], Yi Yang[1], An Zhou[2]

Received September 24, 2011, accepted October 28, 2011

Dr. Ze-Yu Wu, Institute of Natural Medicine, Hefei University of Technology, No. 193 Tunxi Road, Hefei, Anhui 230009, China
wu38331451@163.com

Pharmazie 67: 628–634 (2012) *doi: 10.1691/ph.2012.1726*

To develop an *in silico* model for predicting blood brain barrier (BBB) permeability, and evaluate whether incorporation of two new biological descriptors, high affinity P-glycoprotein substrate probability (HAPSP) and plasma protein binding ratio (PPBR), could result in a better model, four different multiple linear regression (MLR) models have been constructed and compared with each other. The optimized model demonstrated predictive ability and simplicity, not only suitable for passive but also for active transport. Moreover, the molecular descriptors used here are discussed.

1. Introduction

Predicting a drug's properties, such as absorption, distribution, metabolism, and excretion (ADME), is of growing interest among pharmaceutical researchers. This is especially true for those focusing on the discovery and development of central nervous system (CNS) drugs, where crossing the blood brain barrier (BBB) is an important step for distribution. As its name suggests, the BBB separates the brain and CNS from the bloodstream (Clark 2003). For drugs targeting the CNS, high BBB permeability is needed. Conversely, for non-CNS targets, low BBB permeability may be desirable in order to minimize CNS-related side effects. Hence, in drug discovery, it is very important to determine BBB permeability. A common measure of the degree of BBB permeability is logBB, which is defined as the ratio of the steady state concentration of a compound in the brain to that in the blood ($logBB = \log(C_{brain}/C_{blood})$). The experimental determination of logBB is a difficult, expensive, and time-consuming technique, requiring animal experiments and synthesis of test compounds, often in radiolabeled form (Fu et al. 2008).

In the past few decades, computational approaches have also been used to predict BBB permeability. Levin first reported the quantitative relationships between the logBB and the octanol/water partition coefficient (logP) and molecular weight (Mw) of 27 compounds (Levin 1980). Young et al. (1988) found a good linear correlation between logBB values of 20 H_2 receptor histamine antagonists and logP(octanol–cyclohexane). Abraham et al. enlarged this dataset and introduced a number of additional descriptors, such as molecular volume (Mv), dipolarity/polarizability, refraction, hydrogen bond acidity and hydrogen bond basicity (Abraham et al. 1994; Abraham et al. 2002). Kelder et al. (1999) and Clark (1999) found that logBB was correlated with the polar surface area (PSA), which is a very important molecular descriptor for the BBB. With the development of chemical theories and computational technology, more and more models have been established using various descriptors and different approaches (Crivori et al. 2000; Rose et al. 2002; Lobell et al. 2003; Abraham 2004; Adenot and Lahana 2004; Winkler and Burden 2004; Zhang 2004; Dureja and Madan 2006; Hemmateenejad et al. 2006; Wan et al. 2007; Bendels et al. 2008).

Although the models mentioned above contributed a lot to drug design and research, most of them were based almost exclusively on the assumption of passive diffusion, while ignoring carrier-mediated transport or efflux systems. It is obvious that a drug's distribution is affected not only by its physicochemical properties, but also by biological factors. Different from previous models, active transport of compounds has been taken into consideration in the form of the probability of a compound becoming a substrate for P-glycoprotein (P-gp) and its binding ratio to plasma proteins.

In this paper, four different multiple linear regression (MLR) models were constructed and compared with each other, to evaluate whether incorporation of these two new biological descriptors could result in a better model. In this way, an accurate and reliable BBB model more closely mimics the *in vivo* situation has been generated.

2. Investigations, results and discussion

2.1. Development of predictive model

All the models started with 86 compounds, but six compounds (**6**, **23**, **30**, **58**, **60** and **78**) in models 1, 2, 3 and five compounds (**6**, **23**, **30**, **60** and **78**) in model 4 had residuals greater than two standard deviations, so they were considered as outliers. A previous study (Fu et al. 2008) also identified these same compounds as outliers. This might be caused by experimental difficulties, metabolism or some other active transport systems not been considered in our models. Some structural features in outliers could be found through observing the chemical structures carefully: first, all of them contain aromatic ring structure; second, almost all these six compounds have a long chain except compound **58**, and **78** with an internal ring structure; third, there are

a relatively large number of heteroatoms in these compounds, especially nitrogen. In this way, NHD would increase.
After removing these outliers, four correlation equations were obtained:

$$\log BB = -0.0007 \times Mv + 0.3054 \times \log P - 0.0095 \times PSA - 0.0405 \times NRB - 0.0483$$
$$N = 80,\ R^2 = 0.80,\ S = 0.33,\ F = 74.4 \quad (1)$$

$$\log BB = -0.0007 \times Mv + 0.3109 \times \log P - 0.0095 \times PSA - 0.0412 \times NRB - 0.0355 \times PPBR - 0.0414$$
$$N = 80,\ R^2 = 0.80,\ S = 0.33,\ F = 58.7 \quad (2)$$

$$\log BB = 0.0004 \times Mv + 0.2966 \times \log P - 0.0102 \times PSA - 0.0433 \times NRB - 0.9090 \times HAPSP - 0.1267$$
$$N = 80,\ R^2 = 0.81,\ S = 0.33,\ F = 62.0 \quad (3)$$

$$\log BB = 0.0010 \times Mv + 0.3216 \times \log P - 0.0107 \times PSA - 0.0455 \times NRB - 1.0698 \times HAPSP - 0.2402 \times PPBR - 0.1141$$
$$N = 81,\ R^2 = 0.79,\ S = 0.34,\ F = 47.7 \quad (4)$$

where N is the number of compounds, R is the correlation coefficient, S is the standard error, and F is the Fisher F-statistic.
In order to optimize the models, leave one out (LOO) method, the simplest and most commonly used cross validation approach, was used in this study. In this approach, the property value for a given compound in the training set is predicted using the regression equation derived from the data of the remaining compounds. The predictive residual sum of squares statistic (PRESS) is calculated using the formula: $PRESS = \sum_{i=1}^{N}(y_i - y_i')^2$, where y_i' is the predicted logBB value calculated after eliminating the compound and y_i is the experimental logBB value. The cross validation coefficient square value (Q^2) is given by $Q^2 = 1 - PRESS/\sum_{i=1}^{N}(y_i - \bar{y})^2$. The calculated Q^2 values are listed in Table 4.
The predictive ability of the models was evaluated using R^2 and Q^2 values. R^2 is an estimate of the goodness of fit and Q^2 is an estimate of the goodness of prediction. Table 1 indicates that compared to the other three models, model 3 improves the predictive ability, but the increase in R^2 and Q^2 is small. Both of experimental and calculated (in model 3) logBB are listed in Table 1. A plot of experimental vs. calculated logBB values is shown in Fig. 2.

2.2. *Validation of the predictive model*

2.2.1. *Test set*

The predictive model was validated with a test set of 24 compounds. These compounds are listed in Table 2 and their experimental vs. calculated logBB values are shown in Fig. 1. Experimental and calculated values showed good correlation. The root mean square error (RMSE) value for the test set was 0.56, which indicated that our model had similar predictive ability compared to a previous model (Fu et al. 2008). Considering the different experimental conditions and the difficulties during the experiments, our model performed reasonably well. Moreover, not only passive transport but also efflux system has been taken into account in our model.

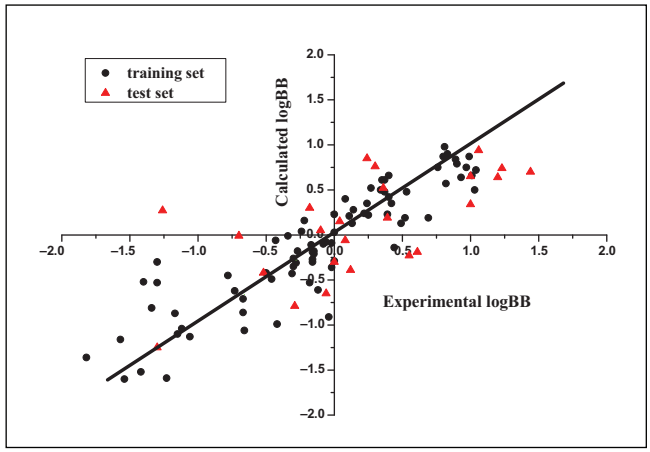

Fig. 1: Relationship between experimental and calculated logBB values for the training and test sets

2.2.2. *Comparison with other models*

In order to further assess the predictive power of our model, we chose to compare and validate it against other computational methods. The logBB values of 80 compounds of the training set were predicted using the published method (Fu et al. 2008) and the CSBBB predictor on www.chemsilico.com. The predicted values from these three methods were then compared, as shown in Table 2. Clearly, our model showed better results in comparison to the other two models. A good logBB prediction ($R^2 = 0.81$) was followed by CSBBB predictor ($R^2 = 0.78$), and the published paper's method ($R^2 = 0.74$).

2.3. *Discussion on the molecular descriptors*

Equation (3) shows that each molecular descriptor contributes to logBB. The significance of molecular descriptors could be evaluated by standard regression coefficient. From Table 3, it can be seen that the most important descriptor is HAPSP, followed by logP, NRB, PSA, Mv. Different descriptors affect the BBB permeability *via* different mechanisms. In Tables 2 and 3, the values of experimental logBB range from –1.82 to 1.44. Within this range, compounds with logBB > 0.3 readily cross the BBB, while those with logBB < −1.0 are only poorly distributed to the brain (Abraham et al. 1997). Based on these results, the way in which molecular descriptors affected the transport of compound across the BBB was discussed as follows.

2.3.1. *Molecular volume*

In our model, the Mv has a positive coefficient with logBB, consistent with previous models (Kaznessis et al. 2001; Platts et al. 2001; Abraham et al. 2002; Abraham 2004). This result would seem to conflict with our general understanding of diffusion. According to the Stokes-Einstein relation, the diffusion coefficient of an object is inversely related to its radius. However, since the Stokes-Einstein relation refers only to diffusion in a homogeneous solution, the validity of its use in terms of BBB permeability is not clear. Furthermore, most of the Mv values used in our model (Fig. 2A) are relatively small; there may not be any remarkable influence on BBB permeability. The values of 9 compounds (**3**, **21**, **24**, **27**, **83**, **85**, **86**, **99**, **110**) exceed 360, and all of their experimental logBB values are no greater than 0.3, even 78% of them (7 out of 9) are below 0. This indicates that when the Mv exceeds a certain value, the restriction function

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

Compound	Mv	logP	PSA	NRB	HAPSP	PPBR	LogBB				
							Exp.	Cal.[a]	Cal.[b]	Cal.[c]	Cal.[d]
1	235.61	–0.175	101.83	8	0.056	46.12%	–1.42	–1.52	–1.52	–0.93	–1.16
2	133.00	1.009	105.53	1	0.019	25.68%	–0.04	–0.91	–1.01	–	–0.12
3	381.09	2.237	108.06	10	0.307	32.42%	–1.06	–1.13	–1.12	–0.81	–1.61
4	285.94	2.626	28.60	7	0.042	78.34%	0.49	0.13	0.08	0.27	0.32
5	287.31	4.161	6.48	4	0.094	99.99%	0.83	0.90	0.89	0.71	0.76
6	266.46	0.775	191.04	9	0.072	24.86%	–0.82	–	–	–	–
7	258.44	2.818	114.84	7	0.055	51.95%	–0.67	–0.71	–0.71	–0.57	–0.78
8	240.55	1.671	114.84	7	0.059	34.64%	–0.66	–1.06	–1.08	–0.65	–0.54
9	328.76	3.559	114.84	9	0.115	79.77%	–0.12	–0.61	–0.64	–0.57	–0.35
10	187.85	2.461	105.53	2	0.053	69.23%	–0.18	–0.53	–0.57	–0.80	–0.19
11	199.13	1.513	131.55	2	0.084	59.66%	–1.15	–1.10	–1.09	–1.43	–0.76
12	235.79	1.655	134.63	3	0.130	65.38%	–1.57	–1.16	–1.11	–1.34	–1.07
13	269.33	1.762	165.74	6	0.167	88.25%	–1.54	–1.60	–1.62	–1.96	–1.64
14	291.86	3.811	93.03	6	0.103	70.69%	–0.27	–0.18	–0.17	–0.81	–0.44
15	306.13	3.236	92.78	6	0.068	67.99%	–0.28	–0.31	–0.31	–0.80	–0.56
16	294.13	2.099	41.57	7	0.417	48.36%	–0.46	–0.49	–0.50	0.05	0.26
17	348.98	3.771	41.57	8	0.356	90.72%	–0.24	0.04	0.08	0.11	0.06
18	254.20	2.320	32.70	6	0.177	60.98%	–0.02	–0.09	–0.09	0.19	0.53
19	325.84	3.535	37.39	8	0.143	95.55%	0.69	0.19	0.15	0.09	0.29
20	316.55	3.489	65.63	8	0.172	96.00%	0.44	–0.14	–0.16	0.10	–0.18
21	360.55	4.992	65.63	8	0.223	99.19%	0.14	0.28	0.30	0.11	0.11
22	351.40	4.350	50.53	8	0.224	98.85%	0.22	0.24	0.24	–0.11	0.34
23	356.88	2.387	88.50	9	0.138	55.10%	–2.00	–	–	–	–
24	412.68	4.603	95.17	10	0.326	89.07%	–1.30	–0.30	–0.17	–0.62	–0.99
25	181.94	2.612	36.42	2	0.061	60.41%	0.11	0.21	0.21	–0.08	–0.01
26	288.98	2.141	118.33	9	0.073	32.81%	–1.12	–1.04	–1.02	–0.81	–1.14
27	377.19	4.030	118.33	11	0.176	77.40%	–0.73	–0.62	–0.60	–0.81	–0.94
28	185.16	0.466	83.72	2	0.019	38.02%	–1.17	–0.87	–0.86	–1.07	–0.39
29	288.97	0.332	114.57	10	0.085	20.70%	–1.23	–1.59	–1.65	–0.82	–1.30
30	264.59	2.039	98.40	9	0.036	59.12%	–2.15	–	–	–	–
31	81.54	0.736	17.07	1	0.006	18.13%	–0.08	–0.10	–0.10	0.14	–0.06
32	84.04	1.937	0.00	0	0.006	59.91%	0.37	0.48	0.48	0.40	0.43
33	112.75	2.816	0.00	2	0.010	80.67%	1.01	0.66	0.65	0.44	0.69
34	129.55	3.375	0.00	3	0.010	88.64%	0.90	0.79	0.78	0.52	0.80
35	70.60	0.420	20.23	0	0.005	11.41%	–0.15	–0.18	–0.19	–0.16	–0.09
36	87.40	0.803	20.23	1	0.007	19.41%	–0.17	–0.11	–0.11	–0.08	–0.01
37	112.75	3.133	0.00	2	0.008	80.67%	0.97	0.75	0.74	0.44	0.70
38	112.18	2.894	0.00	1	0.013	80.46%	1.04	0.72	0.70	0.44	0.60
39	74.27	1.844	0.00	1	0.010	53.84%	0.08	0.40	0.41	0.62	0.14
40	86.31	2.537	0.00	0	0.005	65.03%	0.40	0.66	0.66	0.69	0.26
41	88.35	1.049	9.23	2	0.005	28.79%	0.00	0.03	0.04	0.15	0.05
42	109.97	2.360	9.23	3	0.053	72.00%	0.24	0.35	0.35	0.68	0.29
43	54.02	0.056	20.23	0	0.004	8.30%	–0.16	–0.30	–0.31	–0.25	–0.15
44	97.45	1.525	9.23	3	0.010	42.30%	0.13	0.13	0.13	0.43	0.14
45	92.18	2.555	0.00	1	0.014	69.30%	0.35	0.61	0.62	0.95	0.37
46	129.77	4.163	0.00	4	0.007	91.11%	0.81	0.98	0.74	0.52	0.83
47	112.96	3.657	0.00	3	0.006	84.57%	0.80	0.87	0.88	0.44	0.76
48	109.97	2.360	9.23	3	0.053	70.36%	0.42	0.35	0.34	0.68	0.30
49	102.39	2.471	0.00	0	0.005	75.23%	0.93	0.64	0.63	0.43	0.60
50	96.16	3.152	0.00	2	0.006	74.57%	0.76	0.75	0.75	0.36	0.63
51	70.82	0.559	20.23	1	0.004	14.47%	–0.16	–0.19	–0.19	–0.16	–0.07
52	64.74	0.234	17.07	0	0.004	10.60%	–0.15	–0.21	–0.21	0.05	–0.16
53	83.58	2.247	0.00	1	0.015	55.27%	0.27	0.52	0.52	0.89	0.27
54	100.61	2.386	0.00	0	0.007	71.87%	0.37	0.61	0.63	0.48	0.20
55	80.66	2.023	0.00	0	0.003	77.85%	0.34	0.50	0.51	0.68	0.39
56	155.57	1.434	63.60	3	0.002	62.22%	–0.50	–0.42	–0.41	–0.25	–0.60
57	156.79	2.804	37.30	5	0.009	67.44%	–0.22	0.16	0.18	0.06	–0.15
58	119.06	1.870	57.53	1	0.012	74.34%	–1.10	–	–	–	–
59	140.01	0.667	49.33	1	0.008	19.80%	–0.31	–0.43	–0.43	–0.36	–0.08
60	268.51	3.465	40.54	9	0.010	89.39%	–1.70	–	–	–	–
61	186.90	1.256	67.15	3	0.038	55.41%	–1.30	–0.53	–0.52	–	0.02
62	154.14	0.523	43.84	2	0.082	20.45%	–1.40	–0.52	–0.48	–	–0.13
63	276.12	2.760	52.57	7	0.025	46.68%	–0.43	–0.06	–0.02	–0.20	–0.16
64	268.11	3.027	32.34	7	0.024	74.77%	0.25	0.22	0.23	0.22	0.49

Table 1: (*Continued*)

Compound	Mv	logP	PSA	NRB	HAPSP	PPBR	LogBB				
							Exp.	Cal.[a]	Cal.[b]	Cal.[c]	Cal.[d]
65	142.45	0.731	24.92	3	0.026	23.03%	–0.30	–0.26	–0.26	0.02	0.02
66	159.40	0.976	16.13	3	0.008	26.43%	–0.06	–0.08	–0.07	0.31	0.06
67	115.49	–0.416	67.15	2	0.021	22.68%	–0.42	–0.99	–1.04	–0.24	–0.17
68	281.11	2.665	76.04	4	0.104	77.68%	–0.16	–0.27	–0.26	–0.17	–0.42
69	215.08	2.840	46.33	0	0.108	80.83%	0.00	0.23	0.24	–0.07	–0.32
70	219.47	2.538	68.09	0	0.037	40.61%	–0.34	–0.01	0.02	–0.25	–0.39
71	299.10	1.929	76.53	2	0.054	95.98%	–0.30	–0.35	–0.34	–0.34	–0.66
72	306.79	1.037	96.76	2	0.047	92.72%	–1.34	–0.81	–0.72	–0.79	–1.01
73	315.05	0.029	116.99	3	0.045	87.69%	–1.82	–1.36	–1.24	–1.25	–1.45
74	285.14	4.187	3.24	3	0.245	100.00%	0.89	0.84	0.84	0.93	0.72
75	259.93	3.615	6.48	0	0.120	89.41%	0.99	0.87	0.85	0.68	0.87
76	280.41	3.385	32.70	0	0.095	61.80%	0.82	0.57	0.53	0.28	0.58
77	263.61	3.114	32.70	0	0.079	80.64%	1.03	0.50	0.45	0.26	0.44
78	206.27	2.304	28.16	2	0.116	68.94%	1.64	–	–	–	–
79	263.47	2.789	41.49	0	0.216	63.15%	0.52	0.19	0.15	0.03	0.48
80	246.67	2.869	41.49	0	0.191	72.10%	0.39	0.23	0.22	0.01	0.31
81	255.78	2.715	19.37	0	0.114	74.78%	0.53	0.48	0.48	0.46	0.78
82	313.80	3.544	37.30	1	0.229	93.44%	0.40	0.42	0.42	0.53	0.48
83	367.98	2.790	71.16	5	0.391	68.24%	–0.78	–0.45	–0.41	–1.05	–0.99
84	249.51	1.758	52.05	4	0.089	67.09%	0.00	–0.29	–0.30	0.00	0.34
85	374.49	2.959	61.94	4	0.500	71.06%	–0.02	–0.36	–0.42	–0.13	–0.20
86	382.53	1.968	82.17	4	0.500	57.17%	–0.67	–0.86	–0.91	–0.60	–1.04

[a] Calculated from Eq. (3).
[b] Calculated using the LOO cross validation procedure.
[c] Calculated using the published paper's method (Fu et al. 2008).
[d] Calculated using CSBBB predictor.

may become obvious. Of the remaining 95 compounds, 38% (36 out of 95) that have a logBB > 0.3 may cross the BBB readily; whereas 14% (13 out of 95) with logBB < –1.0 may have very limited ability to cross the BBB.

2.3.2. Octanol/water partition coefficient

Molecular lipophilicity plays an important role in predicting BBB permeability. LogP is a frequently used descriptor

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

Compound	Mv	logP	PSA	NRB	HAPSP	logBB		
						Exp.	Cal.[a]	Cal.[b]
87	150.69	–0.005	69.30	0	0.013	–0.29	–0.79	–0.70
88	167.63	0.063	58.44	0	0.015	–0.06	–0.65	–0.42
89	178.29	1.398	23.55	1	0.024	–0.10	0.05	0.24
90	211.16	3.040	37.30	4	0.009	–0.18	0.30	0.30
91	274.21	1.410	41.93	1	0.181	0.55	–0.23	0.06
92	216.97	2.044	75.27	4	0.020	0.12	–0.39	–0.55
93	263.23	2.290	42.55	1	0.036	0.04	0.15	0.30
94	303.24	3.986	68.53	4	0.038	–1.26	0.27	–0.11
95	237.64	1.836	61.69	1	0.028	0.61	–0.19	–0.18
96	351.96	3.347	35.94	8	0.109	0.39	0.19	0.11
97	270.37	3.917	15.27	4	0.194	1.20	0.64	0.77
98	272.32	3.074	29.66	1	0.036	0.36	0.52	0.55
99	454.30	4.553	63.95	14	0.175	–0.70	–0.01	–0.18
100	271.83	4.379	31.78	4	0.044	1.23	0.74	0.72
101	285.37	5.033	31.78	4	0.045	1.06	0.94	0.77
102	355.72	5.141	35.02	5	0.288	1.44	0.70	0.79
103	346.35	5.683	57.08	4	0.101	0.24	0.85	0.79
104	165.91	1.665	61.77	5	0.009	–0.52	–0.42	–0.36
105	316.33	3.638	44.81	3	0.164	1.00	0.34	–0.12
106	261.48	1.940	44.81	2	0.078	0.08	–0.06	–0.20
107	334.70	5.099	40.54	10	0.024	1.00	0.65	0.57
108	199.17	–0.489	93.29	2	0.017	–1.30	–1.25	–0.97
109	236.62	1.385	58.12	1	0.043	0.00	–0.30	–0.22
110	364.83	5.192	41.99	5	0.169	0.30	0.76	0.34

[a] Calculated from Eq. (3).
[b] Calculated using the published paper's method (Fu et al. 2008).

Table 3: The relationship between molecular descriptors and logBB

Molecular descriptors (Variable)	Regression coefficient	Standard errors	Standard regression coefficient
Constant	–0.1267	0.1063	–
Mv	0.0004	0.0009	$3.5322*10^{-7}$
logP	0.2966	0.0417	0.0124
PSA	–0.0102	0.0014	$-1.4217*10^{-5}$
NRB	–0.0433	0.0167	–0.0007
HAPSP	–0.9090	0.4999	–0.4544

that relates to lipophilicity. As shown in Fig. 2B, logP values range from –0.489 to 5.683. In general, logP correlates positively with logBB; that is, increases in logP value tend to increase BBB permeability. This is very easily understood, since lipophilic compounds can traverse membranes more easily than can hydrophilic ones. When logP < 2.0, 8% (3 out of 39) of the logBB values are more than 0.3 and 28% (11 out of 39) are less than –1; when logP > 2.0, 52% (34 out of 65) are more than 0.3 and 5 % (3 out of 65) are less than –1. High Mv and PSA values of 3 and 24 may be responsible for their low logBB values despite having logP values within the acceptable range. These results agree with previous reports (Garg and Verma 2006; Bendels et al. 2008).

2.3.3. Polar surface area

PSA is another frequently used descriptor (Clark 1999; Kelder et al. 1999; Clark 2003) that relates to molecular polarity. It is defined as the area of the van der Waals surface that arises from oxygen atoms, nitrogen atoms, and hydrogen atoms attached to oxygen or nitrogen atoms.

In our model, PSA correlates negatively with logBB; in other words, increasing polarity leads to reduced BBB permeability. The values of PSA range from 0 to 165.74 (Fig. 2C). When PSA < 40.0, 60% (29 out of 48) of the logBB values are more than 0.3 and none is less than –1. When PSA > 70.0, none is more than 0.3 and 44% (11 out of 25) are less than –1. When PSA is between 40 and 70, 26% (8 out of 31) are more than 0.3 and 10% (3 out of 31) are less than –1.

2.3.4. The number of rotatable bonds

NRB is used to reflect flexibility of a molecule. The relationship between NRB and BBB is complex. The role of molecular flexibility in governing BBB permeability is not clear at present. The logBB increases for up to four to five rotatable bonds, while it decreases for more than six rotatable bonds (Fig. 2D). The reason for this phenomenon may be that when compounds cross the BBB, their shapes are restricted by the cell membrane. The more NRB they have, the more conformation changes will happen. Iyer et al. (2002) reported a positive correlation between molecular flexibility and BBB permeability, and proposed that there might be a parabolic relationship between them. Our result seems to be consistent with this hypothesis. Considering the small number of compounds with a high number of rotatable bonds in our model, statistically significant differences are not observed. Consequently, more data should be collected to perfect the model.

2.3.5. High affinity P-gp substrate probability

The four descriptors mentioned above are often used to describe passive transport. However, active transporters in the BBB should not be discounted. In our model, active transport has been considered in the form of HAPSP.

In the past few years, P-gp, a member of the large ATP binding cassette (ABC) superfamily of transport proteins, has become a hotspot in brain research (Ramakrishnan 2003; Adenot and Lahana 2004; Garg and Verma 2006). It is an ATP-driven efflux pump that is widely expressed in human tissues. The high level of expression of P-gp at the BBB restricts the passage of P-gp substrates into the brain, and therefore imposes a major challenge in the treatment of various brain diseases. The manner by which P-gp mediates the transport of various compounds is still unknown and a number of different models have been proposed to explain the mechanism of P-gp mediated drug transport (Ambudkar et al. 2003; Ramakrishnan 2003; Sharom et al. 2005; Garp and Verma 2006; Boccard et al. 2009). According to the theory established by Ambudkar et al. (2003), substrate first binds to the high affinity site of P-gp, then moves to the lower affinity site and is then extruded from the P-gp. Therefore, in the present model, HAPSP has been incorporated, which makes this quite different from other models.

From the theoretical analysis, logBB will decrease with an increase in the value of HAPSP because efflux of compound will increase. As shown in Fig. 2E, the value of HAPSP ranges from 0.002 to 0.500. When HAPSP < 0.3, about 39% (38 out of 97) of the logBB values are more than 0.3 and about 13% (13 out of 97) are less than –1. When HAPSP > 0.3, all seven compounds (**3**, **16**, **17**, **24**, **83**, **85** and **86**) have logBB values less than zero. This indicates that P-gp plays an unfavorable role in BBB permeability. For example, although their Mv (294.13 and 348.98), logP (2.099 and 3.771) and PSA (41.57 and 41.57) values are within the acceptable range, the logBB values of compounds **16** and **17** are –0.46 and –0.24, respectively. Their low logBB values may be due to their high HAPSP values (0.417 and 0.356).

2.4. Conclusions

A simple, reliable model based on MLR analysis for prediction of BBB permeability has been generated in this paper. Not only passive, but also active transport has been taken into consideration in the form of HAPSP. The molecular descriptors used in our model are very easily available from free online sources. Although the number of molecular descriptors is slightly greater than that used in previous models, we can more comprehensively find the relationship between BBB permeability and the molecular descriptors, especially for those compounds seriously influenced by efflux systems.

3. Experimental

3.1. Compound collection

A dataset of 110 compounds were taken from previously published papers (Young et al. 1988; Abraham et al. 1994; Kelder et al. 1999; Calder and Ganellin 1994; Abraham et al. 1995; Greig et al. 1995; Lombardo et al. 1996; Salminen et al. 1997; Glynn and Yazdanian 1998; Von Sprecher et al. 1998) with their experimental logBB values, then divided into training set (n = 86) and test set (n = 24). Details are available from the authors on request.

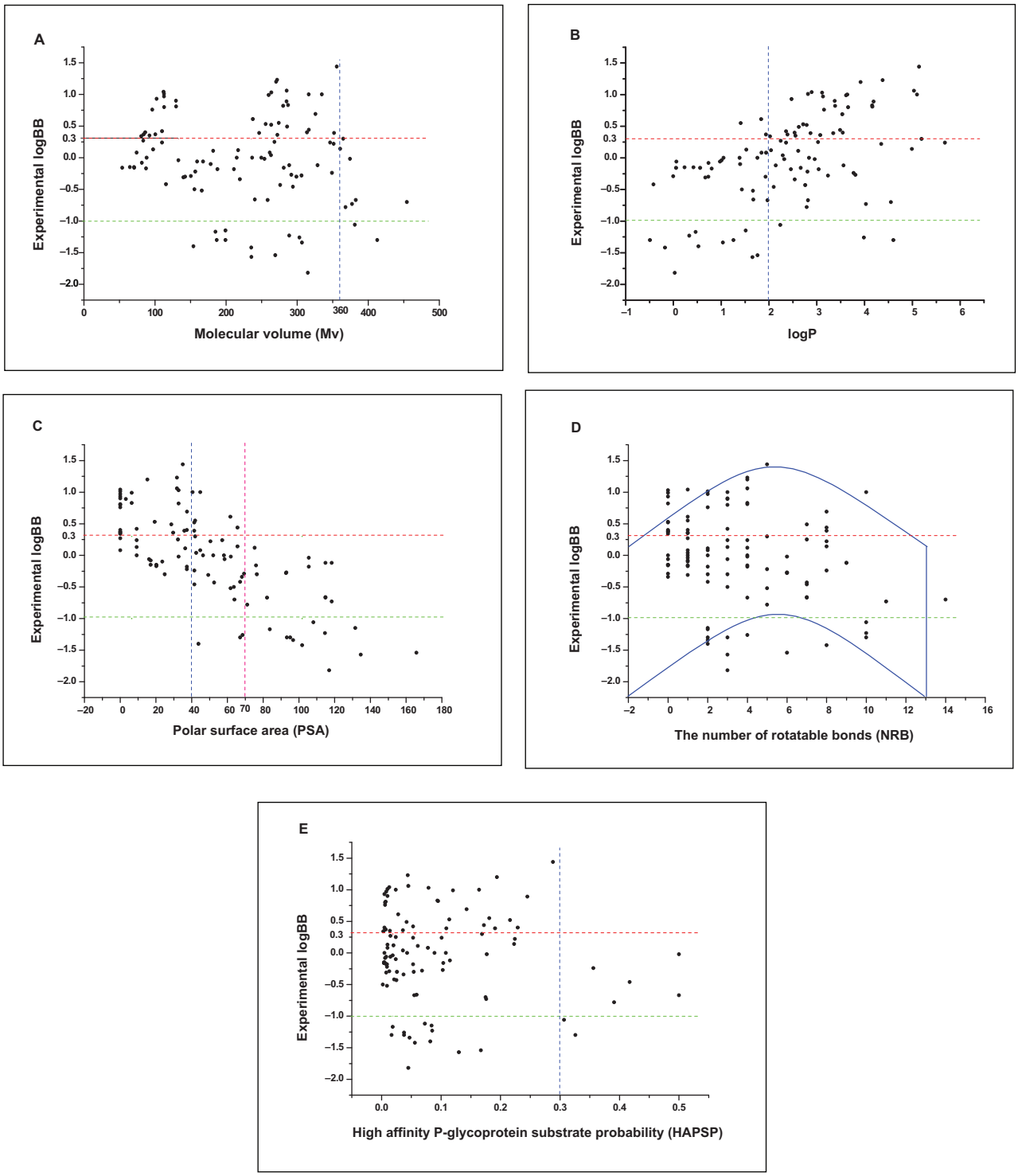

Fig. 2: Relationship between experimental logBB and molecular descriptors (A) Experimental logBB against Mv (B) Experimental logBB against logP (C) Experimental logBB against PSA (D) Experimental logBB against NRB (E) Experimental logBB against HAPSP

3.2. Descriptors collection and calculation

Initially, based on previous studies, several frequently used physicochemical descriptors such as Mw, Mv, logP, PSA, number of rotatable bonds (NRB), number of hydrogen bond donors (NHD), and number of hydrogen bond acceptors (NHA) were selected. Besides, two new biological descriptors, high affinity P-gp substrate probability (HAPSP) and plasma protein binding ratio (PPBR), were also added to predict BBB permeability. After using the data analysis tool in Microsoft Excel to calculate the correlation coefficient, strongly correlated descriptors (correlation coefficient > 0.8) were not selected simultaneously for the model development. Ultimately, six descriptors, including Mv, logP, PSA, NRB, HAPSP and PPBR, were used as regression variables to build the models.

The structures of the compounds were drawn using ChemBiooffice 2010. The Mv and logP values were obtained from the Calculation Molecular Properties and Drug-likeness tools on www.molinspiration.com. The rest of the descriptors were calculated by using the free software ADME/TOX Boxes v 4.0 on www.pharma-algorithms.com/webboxes. The methods used in the computation of HAPSP could be introduced briefly as follows: first, knowledge-based models classify compounds as P-gp substrates or non-substrates. This classification is made according to a set of rules based on the compound's physicochemical properties (such as ionization, molecular size), biological class, etc. Then, the statistical models, employing a fragment-based approach, calculate probability for a compound to be a P-gp substrate. In addition, PPBR is derived from automatic calculation of physicochemical properties, such as lipophilicity, ionization, and fragmental descriptors.

Table 4: Results of four different predictive models

Model	Number of compounds	R^2	Q^2	Molecular descriptors
1	80	0.80	0.77	Physicochemical
2	80	0.80	0.76	Physicochemical + PPBR
3	80	0.81	0.77	Physicochemical + HAPSP
4	81	0.79	0.75	Physicochemical + PPBR + HAPSP

3.3. Data analysis

The MLR analysis was applied to generate four predictive models with different descriptors. The number of compounds and molecular descriptors used in each model are listed in Table 4. The models were then compared and validated for their predictive ability.

Acknowledgements: This work was financially supported by the Special Program for New Drug Innovation of the Ministry of Science and Technology, China (No. 2008ZX09101-018). We would like to thank Li Zhang and Xiuheng Xue for their helpful assistance.

References

Abraham MH, Chadha HS, Mitchell RC (1994) Hydrogen bonding. 33. Factors that influence the distribution of solutes between blood and brain. J Pharm Sci 83: 1257–1268.

Abraham MH, Chadha HS, Mitchell RC (1995) Hydrogen bonding. Part 36. Determination of blood brain distribution using octanol-water partition coefficients. Drug Des Discov 13: 123–131.

Abraham MH, Takács-Novák K, Mitchell RC (1997) On the partition of ampholytes: application to blood-brain distribution. J Pharm Sci 86: 310–315.

Abraham MH, Ibrahim A, Zissimos AM, Zhao YH, Comer J, Reynolds DP (2002) Application of hydrogen bonding calculations in property based drug design. Drug Discov Today 7: 1056–1063.

Abraham MH (2004) The factors that influence permeation across the blood-brain barrier. Eur J Med Chem 39: 235–240.

Adenot M, Lahana R (2004) Blood-brain barrier permeation models: discriminating between potential CNS and non-CNS drugs including P-glycoprotein substrates. J Chem Inf Comput Sci 44: 239–248.

Ambudkar SV, Kimchi-Sarfaty C, Sauna ZE, Gottesman MM (2003) P-glycoprotein: from genomics to mechanism. Oncogene 22: 7468–7485.

Bendels S, Kansy M, Wagner B, Huwyler J (2008) In silico prediction of brain and CSF permeation of small molecules using PLS regression models. Eur J Med Chem 43: 1581–1592.

Boccard J, Bajot F, Di Pietro A, Rudaz S, Boumendjel A, Nicolle E, Carrupt PA (2009) A 3D linear solvation energy model to quantify the affinity of flavonoid derivatives toward P-glycoprotein. Eur J Pharm Sci 36: 254–264.

Calder JAD, Ganellin CR (1994) Predicting the brain-penetrating capability of histaminergic compounds. Drug Des Discov 11: 259–268.

ChemBioOffice Ultra Version 12.0, CambridgeSoft Inc, Cambridge, USA.

Clark DE (1999) Rapid calculation of polar molecular surface area and its application to the prediction of transport phenomena. 2. Prediction of blood-brain barrier penetration. J Pharm Sci 88: 815–821.

Clark DE (2003) In silico prediction of blood-brain barrier permeation. Drug Discov Today 8: 927–933.

Crivori P, Cruciani G, Carrupt PA, Testa B (2000) Predicting blood-brain barrier permeation from three-dimensional molecular structure. J Med Chem 43: 2204–2216.

Dureja H, Madan AK (2006) Topochemical models for the prediction of permeability through blood-brain barrier. Int J Pharm 323: 27–33.

Fu XC, Wang GP, Shan HL, Liang WQ, Gao JQ (2008) Predicting blood-brain barrier penetration from molecular weight and number of polar atoms. Eur J Pharm Biopharm 70: 462–466.

Garg P, Verma J (2006) In silico prediction of blood brain barrier permeability: an artificial neural network model. J Chem Inf Model 46: 289–297.

Glynn SL, Yazdanian M (1998) In vitro blood-brain barrier permeability of nevirapine compared to other HIV antiretroviral agents. J Pharm Sci 87: 306–310.

Greig NH, Brossi A, Pei XF, Ingram DK, Soncrant TT (1995) Designing drugs for optimal nervous system activity. In: Greenwood J, Begley DJ, Segal MB (eds.) New concepts of a blood-brain barrier, New York: Plenum, pp. 251–264.

Hemmateenejad B, Miri R, Safarpour MA, Mehdipour AR (2006) Accurate prediction of the blood-brain partitioning of a large set of solutes using ab initio calculations and genetic neural network modeling. J Comput Chem 27: 1125–1135.

Iyer M, Mishru R, Han Y, Hopfiger AJ (2002) Predicting blood-brain barrier partitioning of organic molecules using membrane-interaction QSAR analysis. Pharm Res 19: 1611–1621.

Kaznessis YN, Snow ME, Blankley CJ (2001) Prediction of blood-brain partitioning using Monte Carlo simulations of molecules in water. J Comput Aided Mol Des 15: 697–708.

Kelder J, Grootenhuis PD, Bayada DM, Delbressine LP, Ploeman JP (1999) Polar molecular surface as a dominating determinant for oral absorption and brain penetration of drugs. Pharm Res 16: 1514–1519.

Levin VA (1980) Relationship of octanol/water partition coefficient and molecular weight to rat brain capillary permeability. J Med Chem 23: 682–684.

Lobell M, Molnár L, Keserü GM (2003) Recent advances in the prediction of blood-brain partitioning from molecular structure. J Pharm Sci 92: 360–370.

Lombardo F, Blake JF, Curatolo WJ (1996) Computation of brain-blood partitioning of organic solutes via free energy calculations. J Med Chem 39: 4750–4755.

Platts JA, Abraham MH, Zhao YH, Hersey A, Ijaz L, Butina D (2001) Correlation and prediction of a large blood-brain distribution data set - an LFER study. Eur J Med Chem 36: 719–730.

Ramakrishnan P (2003) The role of P-glycoprotein in the blood-brain barrier. Einstein Q J Biol Med 19: 160–165.

Rose K, Hall LH, Kier LB (2002) Modeling blood-brain barrier partitioning using the electrotopological state. J Chem Inf Comput Sci 42: 651–666.

Salminen T, Pulli A, Taskinen J (1997) Relationship between immobilised artificial membrane chromatographic retention and the brain penetration of structurally diverse drugs. J Pharm Biomed Anal 15: 469–477.

Sharom FJ, Lugo MR, Eckford PD (2005) New insights into the drug binding, transport and lipid flippase activities of the p-glycoprotein multidrug transporter. J Bioenerg Biomembr 37: 481–487.

Von Sprecher A, Gerpacher M, Anderson GP (1998) Neurokinin antagonists as potential therapies for inflammation and rheumatoid arthritis. IDrugs 1: 73–91.

Wan H, Rehngren M, Giordanetto F, Bergström F, Tunek A (2007) High-throughput screening of drug-brain tissue binding and in silico prediction for assessment of central nervous system drug delivery. J Med Chem 50: 4606–4615.

Winkler DA, Burden FR (2004) Modelling blood-brain barrier partitioning using Bayesian neural nets. J Mol Graph Model 22: 499–505.

Young RC, Mitchell RC, Brown TH, Ganellin CR, Griffiths R, Jones M, Rana KK, Saunders D, Smith IR, Sore NE, Wilks TJ (1988) Development of a new physicochemical model for brain penetration and its application to the design of centrally acting H2 receptor histamine antagonists. J Med Chem 31: 656–671.

Zhang HB (2004) A new nonlinear equation for the tissue/blood partition coefficients of neutral compounds. J Pharm Sci 93: 1595–1604.