Department of Pharmacy[1], Zhejiang University City College, Department of Chemistry[2], College of Pharmaceutical Sciences[3], Ningbo Institute of Technology[4], Zhejiang University, Hangzhou, P.R. China

Predicting blood-brain barrier penetration of drugs using an artificial neural network

X. C. FU[1], G. P. WANG [2], W. Q. LIANG[3], Q. S. YU[4]

Received July 7, 2003, accepted July 22, 2003

Prof. Xuchun Fu, Department of Pharmacy, Zhejiang University City College, Hangzhou, 310015, P.R. China
Fuxc@zucc.edu.cn

Pharmazie 59: 126–130 (2004)

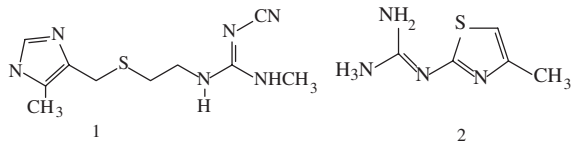
An artificial neural network model is developed to predict the ratios of the steady-state concentrations of drugs in the brain to those in the blood (log BB) from their molecular structural parameters. These molecular structural parameters are the molecular volume (V), the sum of the absolute values of the net atomic charges of oxygen and nitrogen atoms which are hydrogen-bond acceptors ($Q_{O,N}$), and the sum of the net atomic charges of hydrogen atoms attached to oxygen or nitrogen atoms (Q_H). For a training set of 56 compounds and a test set of 5 compounds, root mean squared errors (RMSE) between experimental log BB values and calculated/predicted log BB values were 0.236 and 0.258, respectively. These molecular structural parameters can be obtained easily from quantum chemical calculations. The model is suitable for the rapid prediction of the blood-brain barrier penetration of drugs.

1. Introduction

It is important in drug design to determine whether a candidate molecule is capable of penetrating the blood-brain barrier (BBB). High penetration is needed for drugs targeted at the central nervous system (CNS), while low penetration may be desirable for drugs aimed at other sites of action in order to minimize CNS-related 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 that in the blood, usually expressed as log ($C_{brain/blood}$) or log BB. However, the experimental determination of log BB is a time-consuming, expensive, and difficult technique, requiring animal experiments and the synthesis of the test compounds, often in radiolabeled form (Pardridrge and Mietus 1979; Young et al. 1988; Chikhale et al. 1994; Sveigaard and Dalgaard 2000). It is of very considerable value to predict the log BB of drugs from their physicochemical parameters or, ideally, from their molecular structural parameters.

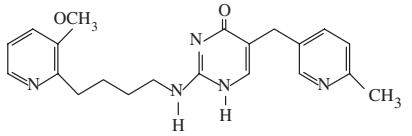
Young et al. (1988) showed that the log BB values of 20 H_2 receptor histamine antagonists were correlated with Δ log P (octanol-cyclohexane). van de Waterbeemd et al. (1992) examined the same series of 20 compounds and found a significant correlation between log BB and the cyclohexane-water partition coefficient when the molecular volume was included in the parameterization. They also found that log BB was correlated with the polar molecular surface area (PSA, defined as the area of the van de Waals surface arising from oxygen or nitrogen atoms or hydrogen atoms attached to oxygen or nitrogen atoms), but the model was shown to be poorly predictive when tested with compounds outside its training set (Calder and Ganellin 1994), 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. Abraham et al. (1994) therefore constructed a larger training set of 65 compounds and derived a correlation between log BB and solvatochromatic parameters for 57 compounds (8 compounds were excluded as outliers). Using a set of 57 compounds drawn from the Abraham training set (1994) mentioned above, Lombardo (1996), Norinder (1998), Clark (1999), and their co-workers developed models for log BB prediction using calculated molecular structural parameters such as free energy of solvation in water (ΔG_w^0) (Lombardo et al. 1996), Molsurf parameters (Norinder et al. 1998), PSA, and octanol-water partition coefficient (C log P or M log P) (Clark 1999) respectively. Some researchers have also derived prediction models for blood-brain barrier penetration with other data sets (Crivori et al. 2000; Feher et al. 2000).

We have reported that the permeability coefficients of β-adrenoreceptor antagonists in Caco-2 cell monolayers, excised rat ileum and excised rat colon are all well corre-



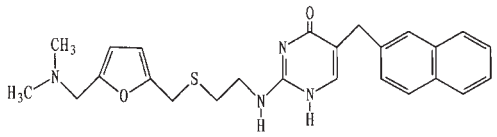
1

2

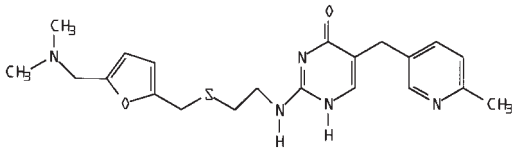


3

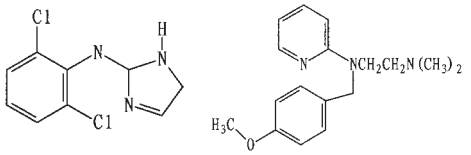
lated with certain net atomic charge parameters (Fu et al. 2001). In this paper, such parameters are selected to develop new models for the prediction of log BB.



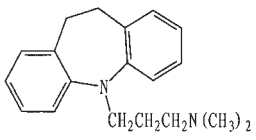
4



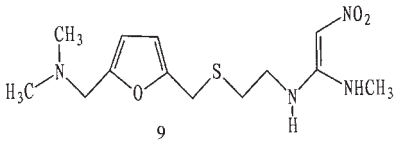
5



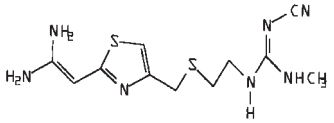
6 7



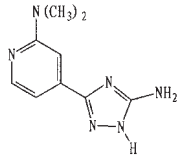
8



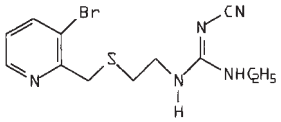
9



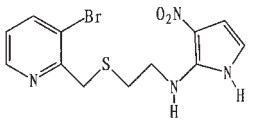
10



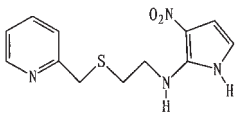
11



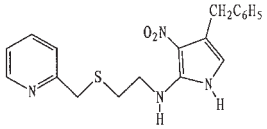
12



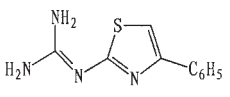
13



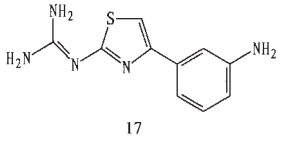
14



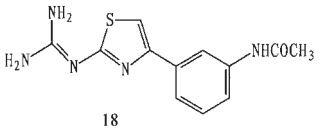
15



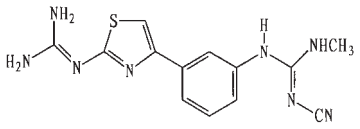
16



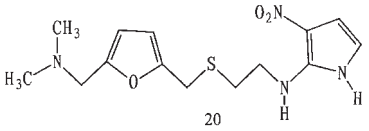
17



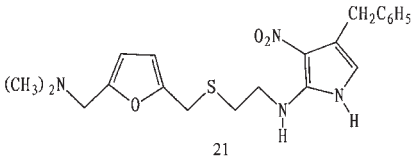
18



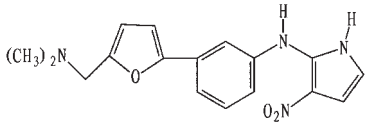
19



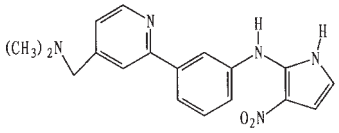
20



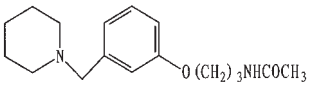
21



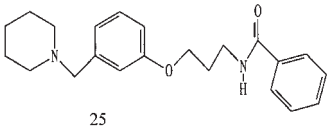
22



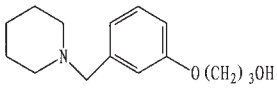
23



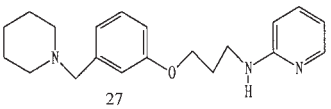
24



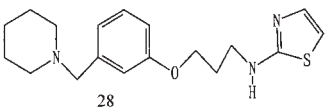
25



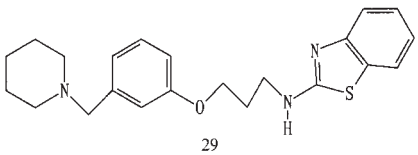
26



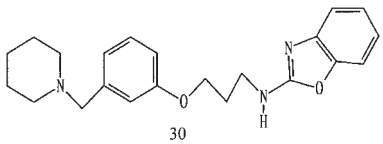
27



28



29



30

2. Investigations and results

The sum of the absolute values of the net atomic charges of oxygen and nitrogen atoms which are hydrogen-bond acceptors ($Q_{O,N}$), and the sum of the net atomic charges of hydrogen atoms attached to oxygen or nitrogen atoms (Q_H) are calculated by the semiempirical self-consistent field molecular orbital calculation CNDO/2 method, using the minimum energy conformation obtained from the optimization of the standard molecular geometry with the molecular mechanics MM+ method. The atomic radii used to calculate molecular volumes (V, nm^3) are those used by Clark (1999).

Table 1: log BB values of 57 compounds and their molecular structural parameters

Compound	V	Q_H	$Q_{O,N}$	log BB		
				Exp.[a]	Calc.[b]	Calc.[c]
1	0.3093	0.3709	1.2967	−1.42	−0.91	−1.27
2	0.1736	0.6227	1.1982	−0.04	−0.95	−0.06
3	0.4725	0.2276	1.5952	−2.00	−1.42	−1.99
4	0.5475	0.2449	1.3313	−1.30	−0.96	−1.22
5	0.5080	0.2029	1.3053	−1.06	−0.87	−0.93
6	0.2422	0.1798	0.5911	0.11	0.10	−0.10
7	0.3853	0.0000	0.7541	0.49	0.16	0.37
8	0.3838	0.0000	0.3258	0.83	1.03	1.05
9	0.3881	0.1786	1.2937	−1.23	−0.88	−0.84
10	0.3505	0.8205	1.8516	−0.82	−1.09	−1.26
11	0.2403	0.3722	1.2155	−1.17	−1.01	−0.85
12	0.3579	0.1637	0.7661	−2.15	–	–
13	0.3426	0.2585	1.1114	−0.67	−0.55	−0.58
14	0.3168	0.2609	1.1241	−0.66	−0.64	−0.61
15	0.4338	0.2537	1.1199	−0.12	−0.43	−0.52
16	0.2419	0.6253	1.2059	−0.18	−0.59	−0.29
17	0.2516	0.8584	1.4510	−1.15	−0.51	−0.96
18	0.3015	0.7313	1.7177	−1.57	−1.17	−1.46
19	0.3420	0.8366	2.1687	−1.54	−1.72	−1.34
20	0.3876	0.2569	1.1996	−1.12	−0.64	−0.78
21	0.5025	0.2451	1.1802	−0.73	−0.58	−0.57
22	0.3826	0.2429	1.1589	−0.27	−0.57	−0.65
23	0.3988	0.1537	1.3636	−0.28	−1.02	−0.95
24	0.3975	0.1087	0.9410	−0.46	−0.18	−0.09
25	0.4664	0.1074	0.9775	−0.24	−0.22	0.14
26	0.3433	0.1388	0.6428	−0.02	0.34	0.22
27	0.4373	0.1033	0.8036	0.69	0.13	0.41
28	0.4257	0.1257	0.8125	0.44	0.11	0.30
29	0.4793	0.1220	0.8409	0.14	0.06	0.54
30	0.4673	0.1222	1.1082	0.22	−0.48	−0.19
Butanone	0.1165	0.0000	0.2609	−0.08	−0.04	−0.01
Benzene	0.1145	0.0000	0.0000	0.37	0.48	0.50
3-Methylpentane	0.1597	0.0000	0.0000	1.01	0.80	0.74
3-Methylhexane	0.1827	0.0000	0.0000	0.90	0.94	0.84
2-Propanol	0.0993	0.1373	0.2659	−0.15	−0.15	−0.12
2-Methylpropanol	0.1221	0.1357	0.2749	−0.17	0.01	−0.07
2-Methylpentane	0.1612	0.0000	0.0000	0.97	0.81	0.74
2,2-dimethylbutane	0.1590	0.0000	0.0000	1.04	0.79	0.73
CF_3CH_2Cl	0.1009	0.0000	0.0000	0.08	0.38	0.42
CH_3CCl_3	0.1238	0.0000	0.0000	0.40	0.55	0.55
Diethyl ether	0.1276	0.0000	0.2248	0.00	0.12	0.10
CHF_2OCF_2CHFCl	0.1449	0.0000	0.2512	0.24	0.19	0.13
Ethanol	0.0760	0.1385	0.2559	−0.16	−0.32	−0.17
$CF_3CH_2OCH{=}CH_2$	0.1311	0.0000	0.1794	0.13	0.24	0.20
$CF_3CHClBr$	0.1272	0.0000	0.0000	0.35	0.57	0.57
Heptane	0.1858	0.0000	0.0000	0.81	0.96	0.85
Hexane	0.1624	0.0000	0.0000	0.80	0.81	0.75
$CHF_2OCHClCF_3$	0.1442	0.0000	0.2084	0.42	0.27	0.21
Methane	0.0442	0.0000	0.0000	0.04	−0.10	0.12
Methylcyclopentane	0.1459	0.0000	0.0000	0.93	0.70	0.67
Nitrogen	0.0394	0.0000	0.0000	0.03	−0.14	0.09
Pentane	0.1387	0.0000	0.0000	0.76	0.66	0.63
Propanol	0.0994	0.1375	0.2594	−0.16	−0.13	−0.12
Propanone	0.0931	0.0000	0.2616	−0.15	−0.22	−0.10
CF_3CHFBr	0.1141	0.0000	0.0000	0.27	0.48	0.49
Toluene	0.1375	0.0000	0.0000	0.37	0.65	0.62
Trichloroethene	0.1135	0.0000	0.0000	0.34	0.47	0.49

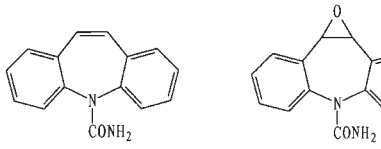
a From reference (Lombardo et al. 1996)
b From eq. (1)
c From neural network model

Using V, $Q_{O,N}$, and Q_H as regression variables, the following regression equation is obtained from a stepwise multiple regression analysis for the training set of 57 compounds (Table 1) studied by other researchers (Lombardo et al, 1996; Norinder et al, 1998; Clark 1999),

$$\log BB = -0.5088 - 10.98\,V^2 + 9.991\,V + 1.554\,Q_H^2 - 2.037 Q_{O,N} \quad (1)$$

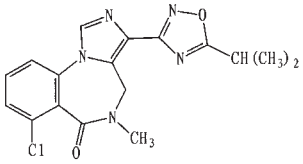
$$n = 56 \quad r = 0.9044 \quad s = 0.3300 \quad F = 57.28$$

where n is the number of compounds, r is the correlation coefficient, s is the standard error, and F is the Fisher value. As other investigators (Abraham et al. 1994; Lombardo et al. 1996; Norinder et al. 1998; Clark, 1999) have found, compound **12** is an outlier which was omitted from the original set of 57.

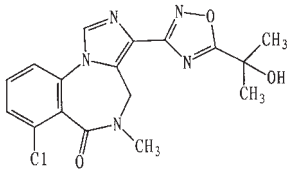


58

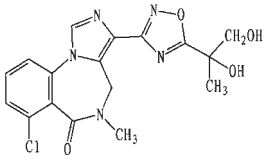
59



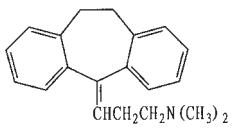
60



61



62



63

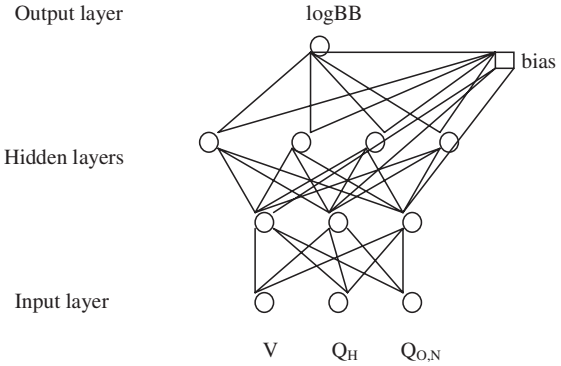


Fig.: Structure of the neural network

Artificial neural networks were established to predict log BB. The back-propagation algorithm with a modified learning rule and normalized cumulative delta was used to train the network. A than function was used as the transfer function. The neural network model is a four-layer network that includes an input layer, two hidden layers, and an output layer. The initial learning coefficients are 0.3, 0.25, and 0.15 for the first hidden layer, second hidden layer, and output layer, respectively. The initial momentum is 0.4. Epoch size is 16. F′ offset is 0.1. Transition point is 10000 and learning coefficient ratio is 0.5. Inputs to the neural network consist of V, $Q_{O,N}$, and Q_H. The hidden layers consist of three neurons and four neurons, respectively. The output layer consists of a single neuron, log BB. The network architecture is shown in the Fig.
The calculated log BB values for the training set of 56 compounds from eq. (1) and from the neural network model obtained after 50000 training cycles are also listed in Table 1 along with experimental log BB values taken from the reference (Lombardo et al. 1996).

3. Discussion

Equation (1) indicates that the log BB value of a compound is correlated with its molecular structural parameters including V, $Q_{O,N}$, and Q_H. These molecular structural parameters strongly related to the molecular size, lipophilicity, and hydrogen bond-forming ability of a compound which were considered to be the important factors determining its blood-brain barrier permeability (Feher et al. 2000).

Table 2: Predicted log BB values for a test set

Compound	V	Q_H	$Q_{O,N}$	log BB						
				Exp.[a]	This work[b]	This work[c]	Lombardo[a]	Norinder[d]	Clark-1[e]	Clark-2[e]
58	0.2768	0.2630	0.8735	0.00	−0.27	−0.30	−0.14	−0.58	−0.25	−0.01
59	0.2847	0.2625	1.1365	−0.34	−0.77	−0.63	−0.28	−1.11	−0.75	−0.37
60	0.3988	0.0000	1.2705	−0.30	−0.87	−0.47	−0.46	−0.75	−0.70	−0.38
61	0.4045	0.1442	1.5002	−1.34	−1.30	−1.34	−0.64	−0.99	−1.26	−0.83
62	0.4131	0.2992	1.7584	−1.82	−1.71	−2.18	−0.82	−1.35	−1.77	−1.28
63	0.3813	0.0000	0.1441	0.76–0.98	1.40	1.14	0.28	1.03	0.76	0.80
RMSE [f]					0.343	0.258	0.555	0.543	0.283	0.334
RMSE [g]					0.368	0.236	0.401	0.301	0.386	0.395

[a] From Lombardo et al. (1996)
[b] From eq. (1)
[c] From neural network model
[d] From Norinder et al. (1998)
[e] From Clark (1999)
[f] RMSE of compounds **58**–**62**
[g] RMSE of the training set

As shown in eq. (1), there is a parabolic correlation between log BB and V. This comes from the dual effect of molecular size on BBB penetration. Increasing V decreases molecular diffusion through a lipid membrane and therefore decreases the log BB value. On the other hand, bigger molecular volume also means higher lipophilicity which facilitates BBB penetration when $Q_{O,N}$, and Q_H remain unchanged.

For the training set of 56 compounds, root mean squared errors (RMSE) between experimental log BB values and calculated log BB values from eq. (1) and from the trained network are 0.368 and 0.236, respectively. The calculated results obtained from the neural network model are better than those obtained from eq. (1). This means that there are more complicated nonlinear relationships between the log BB of a compound and its V, $Q_{O,N}$, and Q_H.

To assess the predictive abilities of eq. (1) and the neural network model further, we predicted the log BB values of 6 compounds outside the training set and compared them with other models (Table 2).

The results in Table 2 show that the log BB values of compounds **68**–**73** predicted from the neural network model agree well with their experimental log BB values, while the neural network model performs better than other models. Furthermore, the parameters used in the neural network model can be obtained very easily. It is thus suitable for the rapid prediction of blood-brain barrier penetration of drugs.

References

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

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

Chikhale EG, Ng KY Burton PS, Borchardt RT (1994) Hydrogen bonding potential as a determinant of the in vitro and in situ blood-brain barrier permeability of peptides. Pharm Res 11: 412–419.

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.

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.

Feher M, Sourial E Schmidt JM (2000) A simple model for the prediction of blood-brain partitioning. Int J Pharm 201: 239–247.

Fu XC, Liang WQ Yu QS (2001) Correlation of drug absorption with molecular charge distribution. Pharmazie 56: 267–268.

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

Norinder U, Sjoberg P Osterberg T (1998) Theoretical calculation and prediction of brain-blood partitioning of organic solutes using Molsurf parameterization and PLS statistics. J Pharm Sci 87: 952–959.

Pardridge WM, Mietus LJ (1979) Transport of steroid hormones through the rat blood-brain barrier. J Clin Invest 64: 145–154.

Sveigaard HH, Dalgaard L (2000) 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: 70–76.

van de Waterbeemd H, Kansy M (1992) Hydrogen bonding capacity and brain penetration. Chimia 46: 299–303.

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 H_2 receptor histamine antagonists. J Med Chem 31: 656–671.