Blackwell Publishing

Full-length article

Predictive model of blood-brain barrier penetration of organic compounds[1]

Xiao-lei MA, Cheng CHEN, Jie YANG[2]

State Key Laboratory of Pharmaceutical Biotechnology, School of Life Science, Nanjing University, Nanjing 210093, China

Abstract

Aim: To build up a theoretical model of organic compounds for the prediction of the activity of small molecules through the blood-brain barrier (BBB) in drug design. **Methods:** A training set of 37 structurally diverse compounds was used to construct quantitative structure-activity relationship (QSAR) models. Intermolecular and intramolecular solute descriptors were calculated using molecular mechanics, molecular dynamics simulations, quantum chemistry and so on. The QSAR models were optimized using multidimensional linear regression fitting and stepwise method. A test set of 8 compounds was evaluated using the models as part of a validation process. **Results:** Significant QSAR models (R=0.955, s=0.232) of the BBB penetration of organic compounds were constructed. BBB penetration was found to depend upon the polar surface area, the octanol/water partition coefficient, Balaban Index, the strength of a small molecule to combine with the membrane-water complex, and the changeability of the structure of a solute-membrane-water complex. **Conclusion:** The QSAR models indicate that the distribution of organic molecules through BBB is not only influenced by organic solutes themselves, but also relates to the properties of the solute-membrane-water complex, that is, interactions of the molecule with the phospholipid-rich regions of cellular membranes.

Introduction

The two barriers separating the central nervous system (CNS) from the periphery are the blood-brain barrier (BBB), located at the endothelial cells of the brain tissue capillaries, and the blood-cerebrospinal fluid barrier (B-CSF-B), at the choroids plexus and the circumventricular organs[1]. Formed by complex tight junctions of the brain capillary endothelial cells, the BBB segregates the circulating blood from interstitial fluid in the brain[2]. The capillary endothelial cells regulate the permeability property of the BBB. However, there are at least four kinds of cells that comprise the brain microvasculature, and all contribute to the regulation of the cerebral microvasculature and, indirectly, to the regulation of BBB permeability. The endothelial cell and the pericyte share a common capillary basement membrane. There is approximately one pericyte for every two to four endothe-

lial cells[3]. Also, immunohistochemistry is a widely used research technique in BBB research for the cellular localization of proteins of interest in normal vessels and the documentation of altered expression following disease states, for the identification of cultured cells and for the spatial localization of novel gene products[4].

Although the BBB was thought to act as a static wall protecting the brain, application of recent advanced methodologies to study the BBB has led to the new concept that the BBB acts as a dynamic regulatory interface. Using primary cultured bovine brain capillary endothelial cells, Tsuji *et al*[5,6] have found that P-glycoprotein (ABCB1) acts as an efflux pump for the anti-cancer drugs, vincristine, at the BBB. Schinkel *et al*[7–9] have developed the *mdr* 1*a* gene knockout mouse and proved that P-glycoprotein (ie, *mdr* 1*a* gene products) plays a key role in restricting the apparent cerebral distributed vinblastine (a substrate of P-glycoprotein)

across the BBB. However, several hydrophilic substrates such as metabolites of cerebral neurotransmitters are present in the brain, which reduce the cerebral concentration, and as a result, could play an important role in CNS detoxification.

To be effective, CNS therapeutic agents must have the ability to cross over the BBB whereas peripherally acting drugs should hardly be able to pass the BBB. The uptake of a compound into the brain is a complex process. The moderately lipophilic drugs can pass the BBB by passive diffusion and the hydrogen bonding properties of drugs significantly influence their particular CNS uptake profiles. Polar molecules are generally poor CNS agents unless they undergo active transport to pass the CNS. Size, ionization properties, and molecular flexibility are other factors observed to influence transport of an organic compound across the BBB[10–13]. Recently, there has been a surge in computational efforts to evaluate absorption, distribution, metabolism, excretion, and toxicity (ADME/T) properties of drugs. These new computational approaches remain to focus on modeling structurally diverse data sets by dealing with the properties of the solutes. These properties are limited to relative lipophilicity indices, solvation and hydrogen bond parameters, topological indices, and limited three-dimensional solute properties[14–19]. Iyer *et al*[20] have developed a methodology called membrane-interaction (MI)-QSAR analysis, where structure-based design methodology is combined with classic intramolecular QSAR analysis to model different compounds interacting with cellular membranes. They have also built a predictive model of BBB penetration of organic compounds by simulating the interaction of an organic compound with the phospholipide-rich region of cellular membranes. As a result, they indicated that BBB penetration of an organic compound depended upon PSA, ClogP, and the conformational flexibility of the compounds as well as the strength of their "binding" to the model biologic membrane. The BBB penetration process can be reliably described for structurally diverse molecules whose interactions with the phospholipide-rich regions of cellular membranes are explicitly considered. There are other important applications of MI-QSAR analysis. For example, it can be used as a computational approach to estimate ADME properties such as the transport of organic solutes through biological membranes; and by this way, we can forecast intestinal absorption[21] of drugs and construct MI-QSAR model for skin irritation[22], and eye irritation[23]. However, as a prediction measure, the MI-QSAR computational model is itself somewhat inconvenient. It is quite easily susceptible to manipulation during the prediction process, and so models with more reliability can still have a reasonable-looking

MI-QSAR model. Moreover, there are several non-MI-QSAR computational models including other descriptors besides PSA and ClogP reported to describe and predict BBB penetration. Lombardo and colleagues[17] reported about computing BBB penetration of organic solutes via free energy calculations. Keseru *et al*[24] clarified a high-throughput prediction of BBB penetration of organic molecules using a thermodynamic approach. Crivori *et al*[25] elucidated a method to predict BBB penetration from a three-dimensional molecular structure. Their research results showed that the BBB penetration only relies on one or two descriptors, which leads to the simplification of the prediction model of BBB.
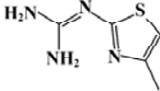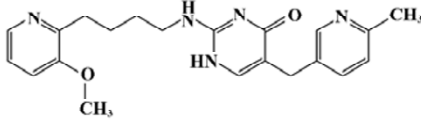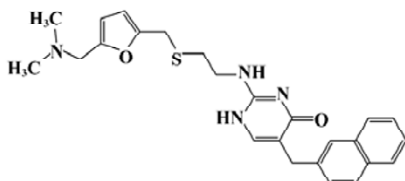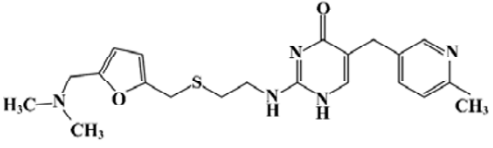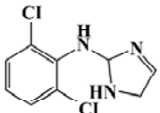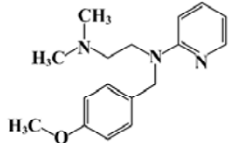
In this article, we focus on constructing the predictive models of BBB penetration of organic compounds on the basis of QSAR analysis and MI-QSAR analysis.

Materials and methods
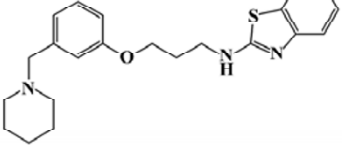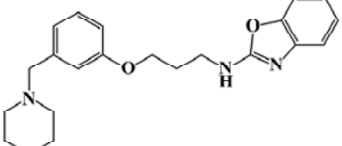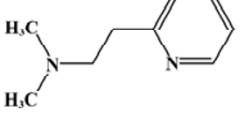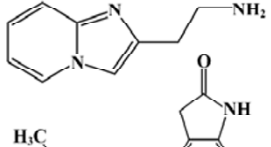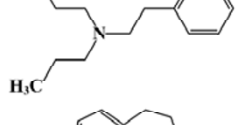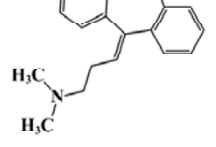
Building solute molecules A training set of 37 organic compounds[26,27] (Table 1) and a test set of 8 organic compounds (Table 2)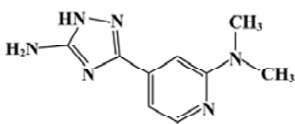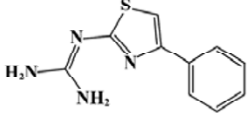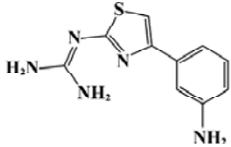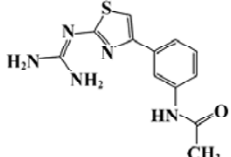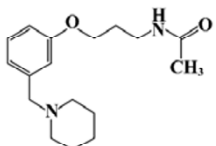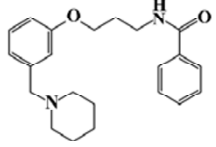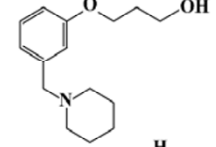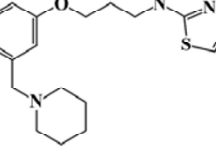 were selected. These compounds had ranges in molecular weights from 16.03 to 448.58, whose concentrations in blood and brain (C_{blood} and C_{brain}) were measured in units of mmol/L, and there were variations in net charge at pH 7.4[26]. The dependent variable used in this theoretical model was the logarithm of the BBB partition coefficient, log BB=log (C_{brain}/C_{blood})[27]. Experimental values of log BB published to date covered the range of about -2.00 to +1.04. Within this range, compounds with log BB> 0.30 cross the BBB readily while compounds with log BB< -1.00 are poorly distributed to the brain[20,27]. All these compounds were built on a PC computer using the Build module of the commercial software packages Hyperchem 6[28]. First, the geometry of these compounds were opitimized using the Amber 94 force field in gas state. Second, they were placed in a periodic solvent box whose volume was X=16Å, Y=10 Å, Z=18Å, which included 96 water molecules. Here, temperature was 300 K and pressure was 101.325 kPa. Then the compounds in water were minimized by the above method. Third, the compounds in water were simulated by the Monte Carlo method and minimized by the above method.

Molecular modeling of a DMPC membrane monolayer complex with a layer of water (DMPC-water model) A model of dimyristoylphosphatidylcholine (DMPC) membrane monolayer was constructed using the software Material Studio[29], and minimized for 200 steps with the smart minimizer. According to the work done by van der Ploeg and Berendsen[30], the DMPC monolayer was composed of 25 DMPC molecules (5×5×1).

Table 1. The structure, molecular weight, formula, and log BB of 37 organic compounds in the training set[20].

ID	Structure	M_r	Formula	Log BB
1		156.21	$C_5H_8N_4S$	-0.04
2		379.46	$C_{21}H_{25}N_5O_2$	-2.00
3		448.58	$C_{25}H_{28}N_4O_2S$	-1.30
4		413.54	$C_{21}H_{27}N_5O_2S$	-1.06
5		230.1	$C_9H_9C_{12}N_3$	0.11
6		285.39	$C_{17}H_{23}N_3O$	0.49
7		204.23	$C_9H_{12}N_6$	-1.17
8		218.28	$C_{10}H_{10}N_4S$	-0.18
9		233.29	$C_{10}H_{11}N_5S$	-1.15
10		275.33	$C_{12}H_{13}N_5OS$	-1.57

(Continued)

ID	Structure	M_r	Formula	Log BB
11		290.41	$C_{17}H_{26}N_2O_2$	-0.46
12		352.48	$C_{22}H_{28}N_2O_2$	-0.24
13		249.35	$C_{15}H_{23}NO_2$	-0.02
14		331.48	$C_{18}H_{25}N_3OS$	0.44
15		381.54	$C_{22}H_{27}N_3OS$	0.14
16		365.47	$C_{22}H_{27}N_3O_2$	0.22
17		150.22	$C_9H_{14}N_2$	-0.06
18		161.21	$C_9H_{11}N_3$	-1.40
19		260.38	$C_{16}H_{24}N_2O$	0.25
20		277.41	$C_{20}H_{23}N$	0.85

(Continued)

ID	Structure	M_r	Formula	Log BB
21		8.01	N_2	0.03
22		78.11	C_6H_6	0.37
23		60.1	C_3H_8O	-0.15
24		74.12	$C_4H_{10}O$	-0.17
25		86.18	C_6H_{14}	0.97
26		86.18	C_6H_{14}	1.04
27		118.49	$C_2H_2ClF_3$	0.08
28		133.41	$C_2H_3Cl_3$	0.40
29		184.49	$C_3H_2ClF_5O$	0.24
30		46.07	C_2H_6O	-0.16
31		126.08	$C_4H_5F_3O$	0.13
32		211.41	$C_3H_3BrClF_3$	0.35

(Continued)
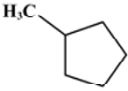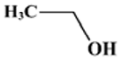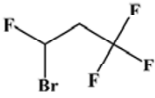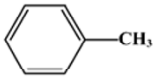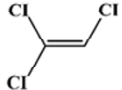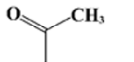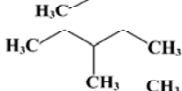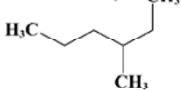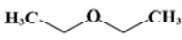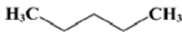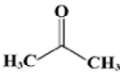
ID	Structure	M_r	Formula	Log BB
33		84.16	C_6H_{12}	0.93
34		46.07	C_2H_6O	-0.16
35		194.95	$C_3H_3BrF_4$	0.27
36		92.14	C_7H_8	0.37
37		31.39	C_2HCl_3	0.34

Table 2. The structure, molecular weight, formula, and log BB of 8 organic compounds in the test set[20].

ID	Structure	M_r	Formula	Log BB
T1		72.11	C_4H_8O	-0.08
T2		86.18	C_6H_{14}	1.01
T3		100.20	C_7H_{16}	0.90
T4		74.12	$C_4H_{10}O$	0.00
T5		100.20	C_7H_{16}	0.81
T6		16.04	CH_4	0.04
T7		72.15	C_5H_{12}	0.76
T8		58.08	C_3H_6O	-0.15

Here, the unit cell parameters used for building the DMPC monolayer were a=8Å, b=8Å, and γ=96.0º, which yield an average surface area per phospholipid of 64Å^2 that was similar to Stouch's research results[31]. Moreover, we added a layer of water (40×40×10) including 529 water molecules to the polar side of the DMPC monolayer.

Molecular dynamic simulation (MDS) of a solute-membrane-water complex In order to prevent unfavorable van der Waals interactions between a solute molecule and the membrane DMPC molecules, one of the "center" DMPC molecules was removed from the DMPC-water model and an organic compound (solute) was inserted in the space cre-
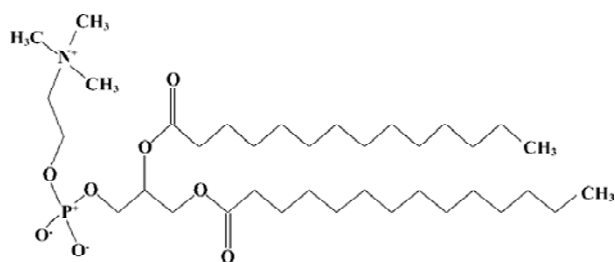
Structure of a single DMPC molecule

ated by the missing DMPC molecule to form a solute-membrane-water complex. The solute was inserted at three different positions in the DMPC-water model and three corresponding MDS models were generated for each compound. MDS of the complex was performed for 1000 steps using Material Studios[29] with a Compass force field. Here, the three-dimensional volume was restricted to a boundary of X=40Å, Y=40Å, Z=91.76Å, and γ=96.0°.

Calculation of descriptors Most of the intramolecular solute descriptors were calculated by the commercial software packages CS Chem3D Ultra7.0[32], which included molecular mechanism (MM) parameters, quantum chemistry parameters, hydrophobic parameters (such as ClogP), stereo parameters (such as Es and Balaban Index) and so on. The MM parameters comprised bending energy, stretch-bend energy, torsion energy, total energy, van der Waals energy and others. The quantum chemistry parameters consisted of electronic energy, HOMO energy, LUMO energy, total energy, and so on. The data of molecular PSA (polar surface area) came from Iyer and colleagues[20].

The intermolecular solute-membrane interaction descriptors were extracted directly from the MDS trajectories in which the solute-membrane-water complex had the lowest energy geometry. These descriptors were mainly energy parameters. The total energy of a system could be expressed as follows[20]:

$$E_{total}=E_{valence}+E_{crossterm}+E_{nonbond}$$

Here, the valence interactions include bond stretching (bond), valence angle bending (angle), dihedral angle torsion (torsion), and inversion, also called out-of-plane interactions (oop) terms, which are part of nearly all forcefields for covalent systems. In addition, a Urey-Bradley term (UB) may be used to account for interactions between atom pairs involved in 1, 3 configurations (ie, atoms bound to a common atom). $E_{valence}=E_{bond}+E_{angle}+E_{torsion}+E_{oop}+E_{up}$. Modern (second-generation) forcefields generally achieve higher accuracy by including crossterms to account for such factors as bond or angle distortions caused by nearby atoms. Crossterms can include the following terms: stretch-stretch,

stretch-bend-stretch, bend-bend, torsion-stretch, torsion-bend-bend, bend-torsion-bend and stretch-torsion-stretch. The interaction energy between non-bonded atoms is accounted by van der Waals, electrostatic (Coulomb), and hydrogen bond terms in some older forcefields. $E_{non\text{-}bond}=E_{van\text{-}der\text{-}Waals}+E_{Coulomb}+E_{hydrogen\text{-}bond}$. Restraints that can be added to an energy expression include distance, angle, torsion, and inversion restraints. Restraints are useful if you, for example, are interested in only part of a structure for information on restraints and their implementation, use, and also the documentation for the particular simulation engine.

Construction and testing of MI-QSAR models MI-QSAR models of some organic compounds through BBB were constructed by partial sum of squares for regression using software SPSS. A training set of 37 structurally diverse compounds whose BBB partition coefficients had been measured was used to construct MI-QSAR models. MDSs were used to determine the explicit interaction of each test compound with the DMPC-water model. An additional set of intramolecular solute descriptors were computed and considered in the trial pool of descriptors for building MI-QSAR models. The QSAR models were optimized using multidimensional linear regression fitting and stepwise method. A test set of 8 compounds was evaluated using the MI-QSAR models as part of a validation process.

Results

Building solute molecules A training set of 37 organic compounds and a test set of 8 organic compounds were built and minimized, dissolved in liquid, and optimized by Monte Carlo method and MM. Finally, the dominant conformations of these compounds were obtained (Figure 1).

MDS of a solute-membrane-water complex The result revealed that the energy of a solute inserted at the middle position of the DMPC-water model was lower than that of the other two positions. Figure 2 shows the dominant conformation of a solute-membrane-water complex in the MDS.

Construction and testing of MI-QSAR models MI-QSAR analysis was used to develop predictive models of some organic compounds through BBB and to simulate the interaction of a solute with the phospholipide-rich regions of cellular membranes surrounded by a layer of water. Molecular descriptors of 37 compounds in the training set were listed in Table 3. Six MI-QSAR equations were constructed based on Table 3 and were listed as follows:

log BB=0.552-1.73×10^{-2}PSA

n=37 R=0.835 S=0.398   (1)

log BB=0.229-1.70×10^{-2} PSA+0.131 Clog P
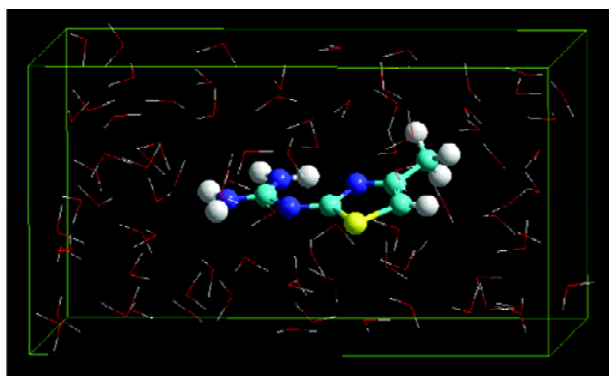
Figure 1. The dominant conformation of Number 1 compound labeled by balls and cylinders in water. The box denotes the water solvent box defined in Monte Carlo simulation.
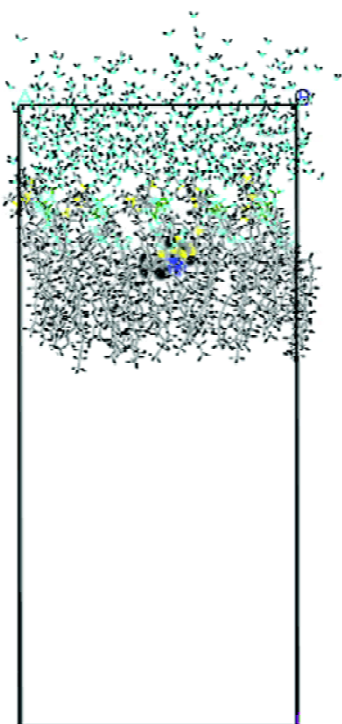
Figure 2. The dominant conformation of a solute-membrane-water complex in the MDS. The DMPC molecules appear as black stick. The upper molecules show water molecules. The center ball molecule expresses organic compound. The black box displays the border of volume.

$$n=37 \qquad R=0.878 \qquad S=0.352 \qquad (2)$$

$$\log BB=4.965\times10^{-2}-1.28\times10^{-2}\,PSA+0.211\,Clog\,P$$
$$-6.40\times10^{-7}BIndx$$

$$n=37 \qquad R=0.924 \qquad S=0.285 \qquad (3)$$

$$\log BB=6.262\times10^{-2}-1.36\times10^{-2}\,PSA+0.205\,Clog\,P$$
$$-7.11\times10^{-7}\,BIndx-0.185E_{sb}$$

$$n=37 \qquad R=0.938 \qquad S=0.264 \qquad (4)$$

$$\log BB=6.580\times10^{-2}-1.21\times10^{-2}\,PSA+0.206\,Clog\,P$$
$$-7.77\times10^{-7}BIndx-0.197E_{sb}+1.330\times10^{-3}\varDelta E_{total}$$

$$n=37 \qquad R=0.947 \qquad S=0.248 \qquad (5)$$

$$\log BB=8.730\times10^{-2}-1.04\times10^{-2}\,PSA+0.222\,Clog\,P$$
$$-9.60\times10^{-7}BIndx-0.183E_{sb}+1.364\times10^{-3}\varDelta E_{total}$$
$$-2.68\times10^{-3}\varDelta E_{torsion}$$

$$n=37 \qquad R=0.955 \qquad S=0.232 \qquad (6)$$

where PSA is the polar surface area, ClogP is the calculated logP (the logarithm of the partition coefficient for octanol/water), BIndx is the Balaban Index, namely connective index of molecular average total distance (relative covalent radius), and E_{sb} is the stretch-bend energy of a molecule. These intramolecular solute descriptors came from CS calculation. In addition, calculated from the MDS, intermolecular descriptors $\varDelta E_{total}$ and $\varDelta E_{torsion}$ are related to interactions between a solute and the DMPC-water model. They display the change in the total potential energy and the dihedral torsion energy of the solute-membrane-water complex comparing with that of the DMPC-water model, respectively. Here, n is the number of organic compounds, R is the correlation coefficient, and S is the standard deviation.

With the increase of the variable from one to six, the relativity of MI-QSAR equations was also improved, and the predictive ability of the models was enhanced. Figure 3 displays a diagnostic plot of the MI-QSAR models. Here, equation 6 was the most significant, which means that the capability of an organic compound through BBB depends upon PSA, ClogP, BIndx, E_{sb}, $\varDelta E_{total}$, and $\varDelta E_{torsion}$. Moreover, the potential of an organic compound through BBB is directly proportional to ClogP and $\varDelta E_{total}$, but inversely proportional to PSA, BIndx, E_{sb}, and $\varDelta E_{torsion}$. The observed and predicted log BB values of the training set are listed in Table

Figure 3. A diagnostic plot of the MI-QSAR models. R is the correlative coefficient and s is the standard deviation of the best x-term model, where x is plotted on the X-axis for the 37 compounds of the training set.

Table 3. The values of molecular descriptors of 37 organic compounds in the training set.

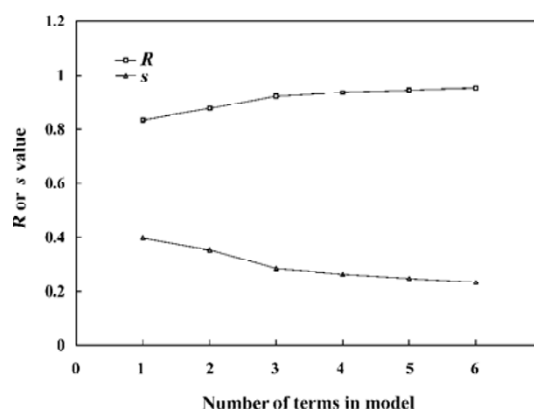| ID	PSA ($Å^2$)	ClogP	BIndx ($Å$)	E_{sb} (kcal/mol)	E_{total} (kcal/mol)	$E_{torsion}$ (kcal/mol)	ΔE_{total} (kcal/mol)	$\Delta E_{torsion}$ (kcal/mol)
1	78.90	1.20	12378.00	−1.3550	−298.30	−1713.11	42.46	11.30
2	94.00	1.98	1101758.00	−0.1560	−406.08	−1789.81	−65.32	−65.39
3	73.00	3.79	1738650.00	−1.4847	−256.30	−1703.14	84.46	21.27
4	87.00	1.63	1346396.00	−1.3911	−302.75	−1841.56	38.00	−117.15
5	39.00	1.02	41807.00	0.5813	−226.38	−1734.75	114.38	−10.33
6	26.80	3.23	305770.00	−0.0926	−228.29	−1679.46	112.47	44.96
7	88.80	1.01	58510.00	0.7104	−279.08	−1671.34	61.68	53.07
8	76.60	2.80	62216.00	−0.3833	−309.30	−1654.67	31.46	69.74
9	104.40	1.77	83798.00	−0.3560	−313.42	−1639.99	27.34	84.43
10	108.80	2.00	193593.00	−0.5217	−548.56	−1640.92	−207.80	83.49
11	47.90	2.51	352512.00	−0.0950	−312.12	−1656.75	28.64	67.67
12	45.20	4.27	779210.00	0.0048	−163.80	−1716.31	176.96	8.11
13	38.50	2.61	158640.00	−0.0949	−170.33	−1716.72	170.43	7.70
14	40.00	4.28	431722.00	−1.3051	−247.10	−1748.02	93.66	−23.61
15	39.20	5.88	766256.00	0.0991	−289.28	−1735.40	51.48	−10.98
16	54.90	5.14	766256.00	−0.1422	−181.06	−1743.61	159.70	−19.19
17	18.80	0.62	20863.00	0.1807	−331.70	−1695.70	9.05	28.72
18	46.70	0.27	20264.00	−1.3684	−209.47	−1644.68	131.29	79.74
19	44.10	2.80	190375.00	−2.9778	−311.92	−1713.89	28.84	10.52
20	5.40	4.85	210631.00	−0.0608	−235.72	−1704.34	105.03	20.08
21	0.00	−0.47	4.00	0.0000	−407.32	−1729.38	−66.56	−4.96
22	0.00	2.14	972.00	−0.0001	−239.88	−1675.18	100.88	49.23
23	23.40	0.07	213.00	0.0000	−160.13	−1672.39	180.63	52.03
24	22.60	0.69	712.00	0.0000	−319.07	−1742.70	21.69	−18.28
25	0.00	3.74	1899.00	0.0007	−282.37	−1751.62	58.39	−27.20
26	0.00	3.61	1661.00	0.0000	−285.71	−1731.95	55.05	−7.54
27	0.00	1.43	1661.00	−0.0001	−238.72	−1731.31	102.03	−6.89
28	0.00	2.48	633.00	0.0000	−291.56	−1725.74	49.20	−1.32
29	11.60	2.46	21380.00	−0.0001	−418.03	−1682.71	−77.27	41.70
30	24.40	−0.24	47.00	0.0000	−329.32	−1704.62	11.44	19.80
31	10.70	1.27	7864.00	−0.0000	−253.35	−1747.7	87.41	−23.29
32	0.00	2.37	7322.00	−0.0000	−268.83	−1714.25	71.93	10.17
33	0.00	3.31	931.00	0.0257	−353.84	−1739.77	−13.08	−15.35
34	24.40	−0.24	47.00	0.0000	−187.45	−1720.55	153.31	3.87
35	0.00	1.93	7322.00	−0.0000	−177.49	−1728.86	163.27	−4.45
36	0.00	2.64	2050.00	−0.0234	−220.39	−1681.15	120.36	43.26
37	0.00	2.63	712.00	−0.0000	−231.58	−1722.26	109.18	2.16

5 and plotted in Figure 4. The predicted log BB value of compound 18 in the training set was much higher than observed, and this molecule has also been identified as an outlier in other studies[20]. Protonation of the molecule could account for its low log BB value.

A test set of 8 organic compounds was constructed as a way to attempt to validate the MI-QSAR models given by the six equations mentioned. The compounds of the test set were selected so as to span almost the entire range in BBB penetration. The observed and predicted log BB values for this test set are given in Table 6 and plotted in Figure 5. It seems that the 5–6 terms MI-QSAR models could predict log BB for other compounds in drug design.

Discussion

We have built up a theoretical model of BBB penetration of organic compounds by simulating the interaction of an organic compound with the phospholipid-rich regions of cellular membranes. The family of these MI-QSAR models

Table 4. The values of molecular descriptors of 8 organic compounds in the test set.

ID	PSA (Å²)	ClogP	BIndx (Å)	E_{sb} (kcal/mol)	E_{total} (kcal/mol)	$E_{torsion}$ (kcal/mol)	ΔE_{total} (kcal/mol)	$\Delta E_{torsion}$ (kcal/mol)
T1	22.70	0.32	712.00	0.0000	−274.72	−1713.74	66.04	10.68
T2	0.00	3.74	1838.00	0.0000	−225.63	−1716.62	115.13	7.79
T3	0.00	4.27	4150.00	0.0000	−331.38	−1700.64	9.38	23.78
T4	11.30	0.87	791.00	0.0000	−181.6	−1700.84	159.16	23.57
T5	0.00	4.40	4650.00	0.0000	−404.29	−1741.24	−63.53	−16.83
T6	0.00	1.10	0.00	0.0000	−282.94	−1746.19	57.82	−21.77
T7	0.00	3.34	791.00	0.0006	−271.93	−1681.94	68.84	42.47
T8	22.70	−0.21	213.00	0.0000	−364.89	−1695.36	−24.13	29.06

Table 5. The experimental value of log BB and the predictive value of log BB of 37 organic compounds in the training set.

ID	Experimental value of log BB	Predictive value of log BB					
		TERM1	TERM2	TERM3	TERM4	TERM5	TERM6
1	−0.04	−0.81	−0.95	−0.71	−0.52	−0.33	−0.20
2	−2.00	−1.07	−1.11	−1.44	−1.56	−1.57	−1.39
3	−1.30	−0.71	−0.51	−1.20	−1.11	−0.98	−1.17
4	−1.06	−0.95	−1.04	−1.58	−1.49	−1.37	−1.13
5	0.11	−0.12	−0.30	−0.26	−0.40	−0.19	−0.06
6	0.49	0.09	0.20	0.19	0.16	0.34	0.28
7	−1.17	−0.98	−1.15	−0.91	−1.11	−0.90	−0.86
8	−0.18	−0.77	−0.71	−0.38	−0.38	−0.22	−0.22
9	−1.15	−1.25	−1.31	−0.97	−0.99	−0.79	−0.81
10	−1.57	−1.33	−1.36	−1.05	−1.05	−1.16	−1.20
11	−0.46	−0.28	−0.26	−0.26	−0.31	−0.21	−0.32
12	−0.24	−0.23	0.02	−0.13	−0.23	0.03	0.04
13	−0.02	−0.11	−0.08	0.01	−0.02	0.26	0.34
14	0.44	−0.14	0.11	0.17	0.33	0.51	0.64
15	0.14	−0.13	0.33	0.30	0.17	0.26	0.33
16	0.22	−0.40	−0.03	−0.06	−0.15	0.10	0.22
17	−0.06	0.23	−0.01	−0.07	−0.11	−0.07	−0.09
18	−1.40	−0.26	−0.53	−0.50	−0.28	−0.02	−0.14
19	0.25	−0.21	−0.15	−0.05	0.45	0.59	0.62
20	0.85	0.46	0.77	0.87	0.85	0.99	1.01
21	0.03	0.55	0.17	−0.05	−0.03	−0.12	−0.10
22	0.37	0.55	0.51	0.50	0.50	0.64	0.57
23	−0.15	0.15	−0.16	−0.23	−0.24	0.04	−0.03
24	−0.17	0.16	−0.06	−0.09	−0.10	−0.04	0.08
25	0.97	0.55	0.72	0.84	0.83	0.91	1.07
26	1.04	0.55	0.70	0.81	0.80	0.88	0.98
27	0.08	0.55	0.42	0.35	0.35	0.49	0.56
28	0.40	0.55	0.55	0.57	0.57	0.64	0.71
29	0.24	0.35	0.35	0.41	0.39	0.31	0.27
30	−0.16	0.13	−0.22	−0.31	−0.32	−0.26	−0.26
31	0.13	0.37	0.21	0.18	0.17	0.31	0.43
32	0.35	0.55	0.54	0.55	0.54	0.64	0.68
33	0.93	0.55	0.66	0.75	0.74	0.73	0.84
34	−0.16	0.13	−0.22	−0.31	−0.32	−0.07	−0.02
35	0.27	0.55	0.48	0.45	0.45	0.68	0.74
36	0.37	0.55	0.57	0.61	0.61	0.77	0.72
37	0.34	0.55	0.57	0.60	0.60	0.75	0.81

Table 6. The experimental value of log BB and the predictive value of log BB of 8 organic compounds in the test set.

ID	Experimental value of log BB	Predictive value of log BB					
		TERM1	TERM2	TERM3	TERM4	TERM5	TERM6
T1	−0.08	0.16	−0.11	−0.17	−0.18	−0.06	−0.02
T2	1.01	0.55	0.72	0.84	0.83	0.99	1.05
T3	0.90	0.55	0.79	0.95	0.93	0.95	0.98
T4	0.00	0.36	0.15	0.09	0.09	0.32	0.32
T5	0.81	0.55	0.81	0.97	0.96	0.88	1.02
T6	0.04	0.55	0.37	0.28	0.29	0.37	0.47
T7	0.76	0.55	0.67	0.75	0.75	0.84	0.81
T8	−0.15	0.16	−0.18	−0.28	−0.29	−0.28	−0.31
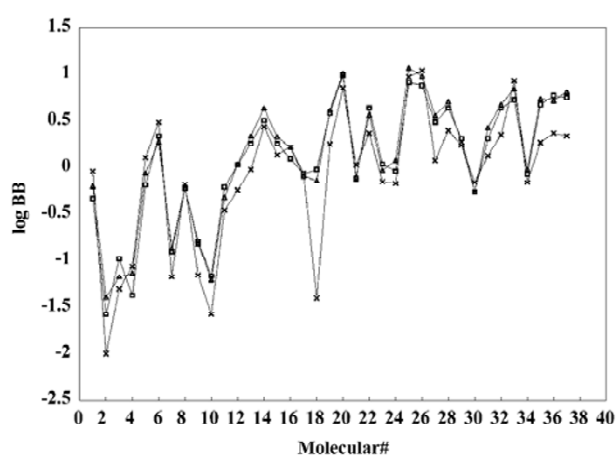
Figure 4. Comparison of the experimental log BB values (−×−) with the corresponding predicted log BB as predicted by the 5-term MI-QSAR model (– □ –) and by the 6-term MI-QSAR model (– △ –) for all the molecules of the training set.
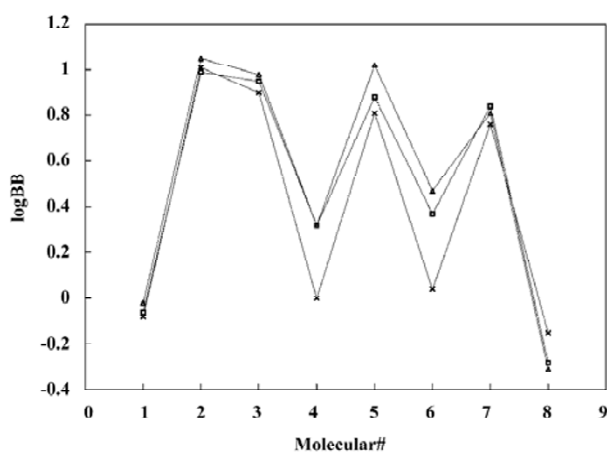
Figure 5. Comparison of the experimental log BB values (−×−) with the corresponding predicted log BB as predicted by the 5-term MI-QSAR model (– □ –) and by the 6-term MI-QSAR model (– △ –) for all the molecules of the test set.

reveal that the capability of an organic compound through BBB focus on six significant features, which are PSA, ClogP, BIndx, E_{sb}, ΔE_{total}, and $\Delta E_{torsion}$. The descriptor PSA is found as a dominant descriptor in these MI-QSAR models, which is related to the aqueous solubility of the solute and can be used as a direct lipophilicity descriptor[19]. When the value of PSA lessens within the range from 0 to 108.80 Å², the value of log BB will increase. This is consistent with the experimental result that the higher polarity it possesses, the more difficultly a molecule enters the hydrophobic environment of BBB[31]. Along with PSA, many descriptors can display various other kinds of the hydrophobic parameters of compounds, such as ClogP. In some cases it may be appropriate to display the parameter in the form of meshes, such as PSA. The higher the hydrophobic property, the greater the value of ClogP is. And the value of log BB of a molecule increases with the increase of ClogP. It means that the hydrophobic molecule can pass through BBB more easily than the hydrophilic molecule does, which is supported by the experimental results[33]. As stated, BIndx is the connective index of molecular average total distance, which pertains to the volume parameter. Our research result reveals that with the accretion of its bulk, a molecule becomes more and more difficult to cross over BBB by diffusion. The presence of the descriptor E_{sb} suggests that with the decrease of the stretch-bend energy the value of log BB increases.

Interestingly, the other two descriptors, ΔE_{total}, and $\Delta E_{torsion}$, reflect the interaction of the solute with the membrane and the behavior of the entire membrane-solute complex. Here, the more the value of ΔE_{total} changes, the more the value of log BB increases. This is because that small molecule acrossing the BBB membrane leads to the change in the structure of the complex, which therefore, results in a greater change of total potential energy, and the accretion of the energy change is the most important cause

for the increase of BBB penetration of small molecules. In contrast, the less the difference in the torsion energy is, the larger the value of log BB is. It displays that a small molecule tightly combined with the membrane-water complex has a higher value of log BB. The relationship suggests that as the solute becomes more flexible within the membrane-water complex, its log BB value would increase, which is in agreement with the research results by Iyer *et al*[20].

As an extra bonus of our work, we have developed an extension of traditional computational approaches that combines QSAR with solute-membrane-water complex. There is an obvious difference between our method and Iyer's method, which is the addition of a layer of water on the hydrophilic side of the DMPC membrane monolayer. So it is more analogous to the true BBB environment. Our results revealed that the distribution of organic molecules through BBB was not only influenced by the properties of organic solutes, but also related to the property of the solute-membrane-water complex. The former involves the polarity, hydrophobicity, size, and conformational freedom degree of organic molecules. The latter deals with the strength of an organic molecule to combine with the BBB membrane and the structural changeability of a solute-membrane-water complex. Moreover, the capability of a small molecule across BBB is mainly related to four physicochemical factors, which are the relative polarity of a small molecule, the molecular volume, the strength of a small molecule to combine with DMPC-water model, and the changeability of the structure of a solute-membrane-water complex. The MI-QSAR model shows that, relatively, less polar and more hydrophobic small molecules, which tightly combine with the membrane-water complex and are more flexible within the complex, can easily pass through BBB and enter the brain to be effective. In drawing conclusions from our model, it seldom matters if one or two parameters are slightly off (eg, *BIndx*), but it is often critical if some parameters are actually in the wrong value (eg, ClogP, *PSA,* or E_{sb}): that will change which molecules are in a position to penetrate through BBB.

Incidentally, since the molecular structures in the training set are not very comparable, our MI-QSAR equations possess universal significance. Nevertheless, the precision of the MI-QSAR equation is so low that there is still some time before it can be applied. If a series of organic compounds with similar structures were chosen to construct a training set, the precision of MI-QSAR simulation may be largely increased, while the prediction of the analogues through BBB will be greatly improved.

In conclusion, we have developed an extension of traditional computational approaches that combines QSAR with solute-membrane-water complex to simulate the BBB environment, which resembles Iyer's method, but differs from it by adding a layer of water on the hydrophilic side of the DMPC membrane monolayer. Our modified MI-QSAR method is more approximate to the body condition than Iyer's MI-QSAR analysis and possesses higher ability to predict organic compounds across BBB. Moreover, while still applying the structural information in a two-dimensional, "structure-function relationship" fashion, this method also takes into account the powerful three-dimensional information displayed by membrane structures, and thus improves existing MI-QSAR method. The MI-QSAR models indicate that the distribution of organic molecules through BBB was not only influenced by organic solutes themselves, but also related to the properties of the solute-membrane-water complex, that is, interactions of the molecule with the phospholipid-rich regions of cellular membranes.

References

1   Bickel U, Yoshikawa T, Pardridge WM. Delivery of peptides and proteins through the blood-brain barrier. Adv Drug Deliv Rev 2001; 46: 247–79.

2   Hosoyo K, Ohtsuki S, Terasaki T. Recent advances in the brain-to-blood efflux transport across the blood/brain barrier. Int J Pharm 2002; 248: 15–29.

3   Pardridge WM. Molecular biology of the blood-brain barrier. Methods Mol Med 2003; 89: 385–99.

4   Nag S. Immunohistochemical detection of endothelial proteins. Methods Mol Med 2003; 89: 489–502.

5   Tsuji A, Tamai I. Carrier-mediated or specialized transport of drugs across the blood-brain barrier. Adv Drug Deliv Rev 1999; 36: 277–90.

6   Tamai I, Tsuji A. Drug delivery through the blood-brain barrier. Adv Drug Deliv Rev 1996; 19: 401–2.

7   Jette L, Murphy GF, Leclerc JM, Beliveau R. Interaction of drugs with P-glycoprotein in brain capillaries. Biochem Pharmacol 1995; 50: 1701–9.

8   Chen W, Mehta SC, Lu DR. Selective boron drug delivery to brain tumors for boron neutron capture therapy. Adv Drug Deliv Rev 1997; 26: 231–47.

9   Friden PM. Utilization of an endogenous cellular transport system for the delivery of therapeutics across the blood-brain barrier. J Control Release 1997; 46: 117–28.

10 Begley DJ. The blood-brain barrier: Principles for targeting peptides and drugs to the central nervous system. J Pharm Pharmacol 1996; 48: 136–46.

11 Anderson BD. Prodrugs for improved CNS delivery. Adv Drug Deliv Rev 1996; 19: 171–202.

12 Pardridge WM. CNS drug design based on principles of blood-brain barrier transport. J Neurochem 1998; 70: 1781–92.

13 Cornford EM, Hyman S. Blood-brain barrier permeability to small and large molecules. Adv Drug Deliv Rev 1999; 36: 145–63.

14 Liu R, Sun H, So SS. Development of quantitative structure-property relationship models for early ADME evaluation in drug

discovery. 2. Blood-brain barrier penetration. J Chem Inf Comput Sci 2001; 41: 1623–32.

15 Rogers D, Hopfinger AJ. Applications of genetic function approximation to quantitative structure-activity relationships and quantitative structure-property relationships. J Chem Inf Comput Sci 1994; 34: 854–66.

16 Abraham MH, Chadha HS, Mitchell RC. Hydrogen bonding factors that influence the distribution of solutes between blood and brain. J Pharm Sci 1994; 83: 1257–68.

17 Lombardo F, Blake JF, Curatolo WJ. Computation of brain-blood partitioning of organic solutes via free energy calculations. J Med Chem 1996; 39: 4750–5.

18 Luco JM. Prediction of 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.

19 Clark DE. 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 1999; 88: 815–21.

20 Iyer M, Mishra R, Han Y, Hopfinger AJ. Predicting blood-brain barrier partitioning of organic molecules using membrane-interaction QSAR analysis. Pharm Res 2002; 19: 1611–21.

21 Kulkarni A, Han Y, Hopfinger AJ. Predicting Caco-2 cell permeation coefficients of organic molecules using membrane-interaction QSAR analysis. J Chem Inf Comput Sci 2002; 42: 331–42.

22 Kodithala K, Hopfinger AJ, Thompson ED, Robinson MK. Prediction of skin irritation from organic chemicals using membrane-interaction QSAR analysis. Toxicol Sci 2002; 66: 336–46.

23 Kulkarni A, Hopfinger AJ, Osborne R, Bruner LH, Thompson ED. Prediction of eye irritation from organic chemicals using membrane-interaction QSAR analysis. Toxicol Sci 2001; 59: 335–45.

24 Keseru GM, Molnar L. High-throughput prediction of blood-brain partitioning: A thermodynamic approach. J Chem Inf Comput Sci 2001; 41: 120–8.

25 Crivori P, Cruciani G, Carrupt PA, Testa B. Predicting blood-brain barrier permeation using three-dimensional molecular structure. J Med Chem 2000; 43: 2204–16.

26 Abraham MH, Chadha HS, Mitchell RC. Hydrogen bonding. Part 36. Determination of blood-brain barrier distribution using octanol-water partition coefficients. Drug Des Discov 1995; 13: 123–31.

27 Abraham MH, Takacs-Novak K, Mitchell RC. On the partition of ampholytes: Application to blood-brain distribution. J Pharm Sci 1997; 86: 310–5.

28 Hyperchem HyperChem. Release 6.0 for MS Windows. Waterloo, Ontario, Hypercube Inc, 2001.

29 Materials Studios. San Diego, USA, Accelrys Inc, 2001

30 van der Ploeg P, Berendsen HJC. Molecular dynamics simulation of a bilayer membrane. J Chem Phys 1982; 76: 3271–6.

31 Stouch TR. Lipid membrane structure and dynamics studied by all atom molecular dynamics simulations of hydrated phosphatidylcholine vesicles. Mol Simulation 1993; 1: 335–62.

32 Chemoffice 2002 CS Chem3D Ultra7.0. Cambridge, USA, Cambridgesoft Inc, 2002

33 Kaliszan R, Markuszewski M. Brain-blood distribution described by a combination of partition coefficients and molecular mass. Int J Pharm 1996; 45: 9–16.