Prediction of Blood–Brain Distribution: Effect of Ionization

Ali Shayanfar,[a] Somaieh Soltani,[b] and Abolghasem Jouyban*,[c]

[a] *Biotechnology Research Center, Tabriz University of Medical Sciences;* [b] *Liver and Gastrointestinal Diseases Research Center, Tabriz University of Medical Sciences; and* [c] *Drug Applied Research Center and Faculty of Pharmacy, Tabriz University of Medical Sciences; Tabriz 51664, Iran.*

Received September 13, 2010; accepted November 3, 2010; published online November 29, 2010

Two simple multiple linear regression models were proposed to calculate the logarithm of the blood to brain concentration ratio (log *BB*) of drugs or drug-like compounds. The drugs were classified into two groups according to their ionization state in blood, and the significant parameters were selected using the train sets for each group. For un-ionizable compounds, the logarithm of distribution coefficient in octanol–water in pH 7.4 (log $D_{7.4}$) and molecular weight are the significant parameters, whereas for ionizable compounds, log $D_{7.4}$ and number of hydrogen bond acceptor are significant parameters. The developed models were validated and their prediction capabilities checked using an external dataset of 25 compounds. In addition to the acceptable prediction errors, comparison of the external data analysis results with previously proposed models confirmed superior prediction capability of newly developed models.

Key words blood–brain barrier; ionization; prediction; classification; drug discovery

The blood–brain barrier (BBB) is a complex cellular system that separates the central nervous system (CNS) from the bloodstream whose purpose is to maintain the homeostasis of the CNS and protection of the brain from xenobiotics. As a result, the efficacy of CNS-targeted drugs, is limited by insufficient delivering into brain, while for systemically targeted drugs, low BBB penetration is needed for decreasing CNS adverse effects. Accordingly, prediction of drug permeation across BBB is an important issue in drug discovery investigations.[1,2]

A method for representation of the BBB permeability degree is the logarithm of drug concentration in brain to blood (log *BB*) or distribution between brain and blood. However, log *BB* is affected by other properties such as plasma protein binding and brain distribution volume. Howerer log *BB* is often used for representing of BBB permeability in the literature. Similar to other experimental biological activities or properties, *in vivo* measurement of BBB penetration is difficult and time–consuming; therefore computational models to predict log *BB* have been developed in recent years. Lipophilicity, hydrogen bonding capacity, molecular charge, molecular size, molecular shape, and molecular flexibility have been proposed as effective parameters for determining log *BB*.[1)]

Recent advances in log *BB* modeling were reviewed.[2,3)] According to these studies, one of the more accurate models was proposed by Abraham *et al.*, which correlated log *BB* to the Abraham solvation parameters of drugs including excess molar refraction, solute polarity/polarizability, solute hydrogen bond acidity and basicity, and McGowan which characteristic molar volume.[4,5)] Feher *et al.*[6)] proposed quantitative structure–property relationship (QSPR) models based on calculated octanol–water partition coefficient, the number of hydrogen-bond acceptors in aqueous medium, and the polar surface area. Subramanian and Kitchen[7)] employed a linear regression and a multivariate genetic partial least squares (G/PLS) approach to predict log *BB*. They employed seven descriptors available in the Cerius package. Another QSPR model using multivariate partial least square (PLS) using 25 calculated descriptors was proposed by Luco.[8)] Recently, Fu *et al.*[9)] proposed a QSPR model using molecular weight and number of polar atoms and another QSPR model using molar volume and partial surface area.[10)] Wichmann *et al.*[11)] applied a set of 5 COSMO-RS σ-moments obtained from quantum chemical calculations as descriptors. Artificial neural networks (ANNs) were also applied for predicting log *BB*[2)] by Chen *et al.*[12)] Futhure details of reviewed *in silico* models are summarized in Table 1.

Table 1. *In Silico* Models for the log *BB* Prediction

Method	Reference	Details of model	Eq. No.
a	9	$\log BB = 9.880\times10^{-6}\,\mathrm{MW} + 7.339\times10^{-3}\,\mathrm{MW} - 0.2268 n_{pol} - 0.1143$	1
b	6	$\log BB = 0.4275 - 0.3873 n_{acc,\,solv} + 0.192 \log P - 0.0017 A_{pol}$	2
c	8	Molecular structure descriptors using PLS method	
d	10	$\log BB = -13.31 V^2 + 9.601 V - 2.231\mathrm{PSA} - 0.5290$	3
e	4	$\log BB = 0.044 + 0.511 E - 0.8865 - 0.724 A - 0.666 B + 0.861 V$	4
f	12	ANN 8-5-1 model using eight physicochemical properties	
g	7	$\log BB = -0.0204 + 0.122 S\_sssN - 0.114\mathrm{Rotlbonds} + 0.0359\mathrm{Jurs\_}WNSA\_3 - 0.0615 S\_dsN + 0.1313 A \log P - 0.0959 S\_SSSCH + 0.108\mathrm{Rog}$	5
h	11	$\log BB = 0.1878 + 0.0046 M_0 - 0.0173 M_2 - 0.0027 M_3$	6

a: MW: molecular weight and n_{pol}: number of polar atom. b: log *P*: calculated octanol–water partition coefficient, polar surface area (A_{pol}), number of hydrogen-bond acceptors in an aqueous medium. d: *V*: molecular volume, PSA: polar surface area. e: *E*: excess molar refraction, *S*: solute polarity/polarizability, *A* and *B*: solute hydrogen bond acidity and *V*: basicity and McGowan characteristic molar volume. g: *S_sssN*: N connected by three single bonds, Rotlbonds: number of rotatable bonds, Jurs: surface weighted charged partial surface area, *S-dsN*: N connected by a double and single bond, *A* log *P*: Ghose and Crippen log *P*, *S_SSSCH*: CH connected by three single bonds and Rog: radius of gyration. h: M_0, M_2 and M_3: the COSMO-RS σ-moments were obtained from quantum chemical calculations.

* To whom correspondence should be addressed. e-mail: ajouyban@hotmail.com

Classification is a common method in chemical disciplines that was used for BBB modeling along with chemometrics methods.[2] Molecular charge and ionization state are critical parameters in determination of pharmacokinetic properties such as $\log BB$. Some drugs are in their ionized form in pH 7.4, while others are in neutral forms. Classification according to ionization provides a useful method in pharmacokinetic data modeling. As an example Ghafourian *et al.*[13] developed a QSPR model for the prediction of apparent volume of distribution in which drugs were classified in two classes of acidic and basic groups.

Although attempts were made to predict BBB transport with several physicochemical parameters, in particular with the octanol–water partition coefficient ($\log P$),[2] however, in some cases poor correlation was found between $\log BB$ and $\log P$.[8] A possible reason for such a poor correlation might be related to the ionization state of drugs that affect $\log P$ and consequently $\log BB$, *e.g.* $\log P$ of aspirin is 1.22 where $\log D_{7.4}$ is -2.47.[14] $\log D_{7.4}$ is the overall ratio of drug, in ionized and non-ionized forms, between octanol and buffer in pH=7.4.[15]

The aims of this study are classification of compounds according to their ionization states in blood pH and proposing two simple QSPR models using multiple linear regression (MLR) analysis. Descriptors are calculated by Pharma-Algorithms software, selected by stepwise regression; the validity and accuracy of the proposed models are investigated and compared with previous models.

COMPUTATIONAL METHODS

$\log BB$ values of 122 drugs and chemical compounds were collected from the literature.[4,12,16,17] The dataset was classified according to their ionization state in blood pH using the logarithm of the distribution coefficient in octanol/water ($\log D$) at pH 1.7 ($\log D_{1.7}$) and pH 7.4 ($\log D_{7.4}$). Thus a compound was allocated to the ionized group at pH=7.4 if $\log D_{1.7}<\log D_{7.4}$ and was classified as the un-ionized group at pH 7.4 if $\log D_{1.7}\geq\log D_{7.4}$. Each group was divided to train and test sets. To do this, $\log BB$ values were sorted based on the ascending order, and from every five compounds one was assigned as test set (12 compounds in each group) and the remainder considered as train set (49 compounds in each group). Calculated Abraham solvation parameters, molecular weight (MW), topological polar surface area (TPSA), $\log D_{7.4}$, number of hydrogen bond donor (NHBD), number of hydrogen bond acceptor (NHBA), and calculated $\log P$ ($c\log P$) value using Pharma-Algorithms software were used as descriptors. A stepwise regression analysis was employed to select the most significant predictors. The selected descriptors for each dataset (ionized and un-ionized) were correlated with $\log BB$ values using MLR method. To validate the proposed models and to assess their prediction capability, the leave-one out (LOO) method was used in which one compound is left out from the training set and the trained model was used to predict the removed data point. The LOO results (q^2 values) are inadequate to assess the validity of a model, so the external validation method was used to establish a reliable QSPR model considering the following criteria taken from the literature[18—20]:

Criterion 1: $R^2>0.6$ and $q^2>0.5$ where R^2 is the correlation coefficient between the predicted and experimental values of compounds and q^2 is defined as following:

$$q^2=\frac{\sum_{i=1}^{n}(y_i-y_i')^2}{\sum_{i=1}^{n}(y_i-\bar{y}_{\text{Train set}})^2} \tag{7}$$

y_i and y_i' are experimental and predicted values, respectively, n is number of compounds in the validation set, and $\bar{y}_{\text{Train set}}$ indicates the means of the training and validation sets.

Criterion 2: $(R^2-R_0^2)/R^2$ or $(R^2-R_0'^2)/R^2<0.1$ where R^2 from a test set should be close to R_0^2 or $R_0'^2$ (R_0^2 is predicted *versus* experimental values and $R_0'^2$ experimental *versus* predicted values that are a quantity characterizing linear regression with Y-intercept set to zero).

Criterion 3: k or k' slope of regression line (predicted *versus* observed regression lines and observed *versus* predicted regression lines) through origin should be between 0.85 and 1.15.

In addition to the above validation process an external test set containing 25 drugs, which was proposed in the literature to compare the prediction power of the proposed model with previous models, was employed.

The accuracy of the predicted $\log BB$ values is calculated by average absolute error (AAE) and root mean square error (RMSE) criteria, which are defined as:

$$\text{AAE}=\frac{\sum\left|\log BB^{\text{experimental}}-\log BB^{\text{calculated}}\right|}{N}=\frac{\sum \text{AE}}{N} \tag{8}$$

$$\text{RMSE}=\sqrt{\frac{\sum(\log BB^{\text{experimental}}-\log BB^{\text{calculated}})^2}{N}} \tag{9}$$

RESULTS AND DISCUSSION

Experimental and predicted $\log BB$ values for the training and test sets along with the numerical values of the selected descriptors are listed in Table 2. The developed QSPR model for un-ionizable compounds in blood pH is:

$$\log BB=0.238(\pm0.115)+0.191(\pm0.031)\log D_{7.4}-0.003\text{MW}(\pm0.000)$$
$$N=49\quad R^2=0.702\quad s=0.689\quad F=53\quad q^2_{\text{LOO}}=0.601 \tag{10}$$

and for ionizable compounds is:

$$\log BB=1.045(\pm0.168)+0.138(\pm0.039)\log D_{7.4}-0.292(\pm0.034)\text{HBA}$$
$$N=49\quad R^2=0.635\quad s=0.535\quad F=41\quad q^2_{\text{LOO}}=0.588 \tag{11}$$

where N is the number of compounds, R^2 the regression coefficient, q^2_{LOO} the cross validation coefficient, s the standard deviation, and F the Fisher F-statistic.

For un-ionizable compounds in blood, $\log D_{7.4}$ and MW are the significant parameters for determining $\log BB$. $\log P$ has been reported in the literature as a critical parameter in $\log BB$ prediction.[1,3] However, it provides no information about the ionization state of drugs, so $\log D_{7.4}$ ($\log P$ in pH 7.4) can be replaced as a more efficient parameter in $\log BB$ modeling. Higher $\log D_{7.4}$ values are favorable for $\log BB$ whereas lower MW increases $\log BB$ and represents a combined effect of molecular size and lipophilicity on $\log BB$.[9] For ionizable compounds that are not ionized in blood pH,

Table 2. Experimental and Calculated log *BB* Values of Drugs and Drug-Like Compounds and Absolute Errors for Training and Test Sets and Selected Descriptors

No.	Drug	log $D_{1.7}$	log $D_{7.4}$	Group[b]	MW[c]	NHBA[d]	log $BB_{observed}$	log $BB_{calulated}$	AE[e]
Train set[a]									
1	1,1,1-Trifluloro-2-chloroethane	1.87	1.87	1	118.49	0	0.080	0.278	0.198
2	1,2,3,4-Tetrahydroquinoline	−1.43	−1.18	2	135.21	1	0.650	0.590	0.060
3	1-Hydroxymidazolam	−4.45	−1.05	2	349.83	4	−0.070	−0.269	0.199
4	2,2-Dimethylbutane	3.04	3.04	1	86.17	0	1.040	0.588	0.452
5	2-Methylpentane	3.05	3.05	1	86.17	0	0.970	0.590	0.380
6	2-Methylpropanol-1-ol	0.67	0.67	1	74.12	1	−0.170	0.167	0.337
7	2-Propanol	0.19	0.19	1	60.09	1	−0.160	0.113	0.273
8	3-Methylhexane	3.54	3.54	1	100.2	0	0.900	0.646	0.254
9	4-Hydroxymidazolam	−1.63	2.79	2	341.77	4	−0.300	0.262	0.562
10	Acetylsalicylic acid	1.21	−2.47	1	180.16	4	−0.500	−0.718	0.218
11	Albuterol	−2.87	−0.99	2	239.31	4	−1.030	−0.260	0.770
12	Alprazolam	2.36	3.49	2	308.76	4	0.040	0.359	0.319
13	Aminopyrine	−2.08	1.11	2	231.29	4	0.000	0.030	0.030
14	Amitriptyline	2.3	3.7	2	277.40	1	0.980	1.265	0.285
15	Amobarbital	1.88	1.76	1	226.27	5	0.040	−0.033	0.073
16	Amphetamine	−1.44	−0.58	2	135.21	1	0.930	0.673	0.257
17	Antipyrine	0.4	0.54	2	188.23	3	−0.097	0.243	0.340
18	Atropine	−1.77	−0.91	2	289.37	4	−0.060	−0.249	0.189
19	Benzene	2.21	2.21	1	78.11	0	0.370	0.451	0.081
20	Bishydroxy L-663581 metabolite	−2.92	0.17	2	407.85	9	−1.820	−1.561	0.259
21	Bromocriptine	3.88	6.97	2	658.62	10	−1.100	−0.913	0.187
22	Buspirone	−3.46	1.77	2	385.50	7	0.480	−0.755	1.235
23	Butanone	0.61	0.61	1	72.11	1	−0.080	0.161	0.241
24	Carbamazepine	2.58	2.58	1	236.27	3	0.000	0.097	0.097
25	Carbamazepine-epoxide	0.99	0.99	1	252.27	4	−0.340	−0.250	0.090
26	Cefotetan	0.50	−4.08	1	577.64	15	−1.890	−2.092	0.202
27	Chlorambucil	1.50	0.26	1	304.21	3	−1.700	−0.529	1.172
28	Chloroform	1.74	1.74	1	119.38	0	0.290	0.250	0.040
29	Cimetidine	−2.8	0.20	2	252.34	6	−1.420	−0.680	0.740
30	Clobazam	3.27	0.88	1	300.74	4	0.350	−0.401	0.751
31	Clonidine	−1.53	0.95	2	230.09	3	0.110	0.300	0.190
32	Codeine	−1.85	0.47	2	299.36	4	0.550	−0.058	0.608
33	Cyclohexane	2.79	2.79	1	84.160	0	0.920	0.546	0.374
34	Cyclopropane	1.62	1.62	1	42.08	0	0.000	0.435	0.435
35	Desflurane	1.72	1.72	1	168.04	1	0.110	0.116	0.006
36	Desmethylclobazam	3.61	1.22	1	286.71	4	0.360	−0.298	0.658
37	Desmethyldesipramine	0.56	1.68	2	252.35	2	1.060	0.693	0.367
38	Desmethyldiazepam	−1.36	1.20	2	274.74	3	0.500	0.335	0.165
39	Diazepam	1.34	2.84	2	284.74	3	0.520	0.562	0.042
40	Dichloromethane	1.35	1.35	1	84.93	0	−0.110	0.268	0.378
41	Divinyl ether	0.81	0.81	1	70.09	1	0.110	0.205	0.095
42	Domperidone	1.17	3.49	2	425.91	7	−0.780	−0.517	0.263
43	Enflurane	2.33	2.33	1	184.49	1	0.240	0.188	0.052
44	Etoposide	0.44	0.44	1	588.56	13	−2.000	−1.257	0.743
45	Flunitrazepam	1.59	2.06	2	313.28	6	0.060	−0.423	0.483
46	Fluphenazine	−0.34	4.09	2	437.52	4	1.510	0.442	1.068
47	Fluroxene	1.45	1.45	1	126.08	1	0.130	0.177	0.047
48	Gentisic acid	1.22	−0.37	1	154.12	4	0.100	−0.246	0.346
49	Haloperidol	0.21	2.18	2	375.86	3	1.340	0.470	0.870
50	Hexane	3.28	3.28	1	86.17	0	0.800	0.634	0.166
51	Icotidine	−1.06	2.23	2	379.45	7	−2.000	−0.692	1.308
52	Imipramine	1.48	2.61	2	280.41	2	1.060	0.822	0.238
53	Indinavir	0.01	4.69	2	619.84	9	−0.750	−0.936	0.186
54	Isoflurane	2.26	2.26	1	184.49	1	0.420	0.175	0.245
55	Levodopa	−2.8	−2.48	2	197.19	5	−0.770	−0.759	0.011
56	Lupitidine	−4.54	0.27	2	413.54	7	−1.060	−0.963	0.097
57	Mesoridazine	1.56	2.69	2	386.57	3	−0.360	0.541	0.901
58	Methamphetamine	−0.46	0.00	2	151.25	1	0.990	0.753	0.237
59	Methane	0.45	0.45	1	16.04	0	0.040	0.281	0.241
60	Methohexital	2.73	2.65	1	262.30	5	−0.060	0.041	0.101
61	Methoxyflurane	2.21	2.21	1	164.97	1	0.250	0.218	0.032
62	Methylcyclopentane	2.79	2.79	1	84.16	0	0.930	0.546	0.384
63	Mirtazapine	−2.34	2.31	2	265.35	3	0.530	0.488	0.042
64	Morphine	−2.32	0.00	2	285.34	4	−0.160	−0.123	0.037
65	*m*-Xylene	3.03	3.03	1	106.16	0	0.300	0.532	0.232
66	Nor-1-chlorpromazine	1.83	2.30	2	304.84	2	1.370	0.779	0.591
67	Nor-2-chlorpromazine	1.29	2.42	2	290.81	2	0.970	0.796	0.174

Table 2. Continued

No.	Drug	$\log D_{1.7}$	$\log D_{7.4}$	Group[b]	MW[c]	NHBA[d]	$\log BB_{observed}$	$\log BB_{calulated}$	AE[e]
68	*o*-Xylene	3.03	3.03	1	106.16	0	0.366	0.532	0.166
69	Paraxanthine	−2.48	−2.01	2	182.18	6	0.060	−0.986	1.046
70	Pentane	2.80	2.80	1	72.15	0	0.760	0.580	0.180
71	Pentobarbital	1.88	1.76	1	226.27	5	0.100	−0.033	0.133
72	Pergolide	−2.09	0.71	2	316.5	2	0.300	0.559	0.259
73	Phencyclidine	1.79	3.20	2	245.4	1	0.680	1.196	0.516
74	Phenylbutazone	3.20	0.15	1	308.37	4	−0.520	−0.561	0.041
75	Phenytoin	1.96	1.9	1	252.27	4	−0.040	−0.076	0.036
76	Propanone	0.12	0.12	1	58.08	1	−0.150	0.105	0.255
77	Propranolol	−0.06	0.89	2	259.34	3	0.640	0.292	0.348
78	*p*-Xylene	3.03	3.03	1	106.16	0	0.310	0.532	0.222
79	Quinidine	−2.71	1.06	2	324.42	4	−0.460	0.023	0.483
80	Ranitidine	−5.8	−1.00	2	314.4	7	−1.230	−1.138	0.092
81	Risperidone	−3.03	1.09	2	410.48	6	−0.020	−0.557	0.537
82	Salicylic acid	2.02	−1.89	1	138.12	3	−1.100	−0.494	0.606
83	Salicyluric acid	−0.7	−4.83	1	197.19	5	−0.440	−1.215	0.775
84	SKF101468	−0.01	1.40	2	260.37	3	0.250	0.362	0.112
85	SKF89124	0.15	1.56	2	276.37	4	−0.430	0.092	0.522
86	Sulforidazine	−2.44	−0.54	2	404.59	4	0.180	−0.198	0.378
87	Tacrine	−0.22	0.48	2	198.26	2	−0.130	0.527	0.657
88	Teflurane	1.72	1.72	1	180.93	0	0.270	0.081	0.189
89	Temelastine	−7.59	−3.41	2	446.38	6	−1.880	−1.179	0.701
90	Theobromine	−0.80	−0.80	1	180.16	6	−0.280	−0.399	0.119
91	Thiopental	2.81	2.55	1	242.34	4	−0.140	0.075	0.215
92	Thioperamide	−0.48	2.26	2	292.44	4	−0.161	0.189	0.350
93	Tibolone	3.37	3.37	1	312.44	2	0.400	0.044	0.356
94	Tiotidine	−4.54	0.12	2	312.42	8	−0.820	−1.276	0.456
95	Toluene	2.62	2.62	1	92.140	0	0.370	0.492	0.122
96	Triazolam	0.20	4.06	2	343.21	4	0.740	0.438	0.302
97	Trichloroethene	2.76	2.76	1	131.39	0	0.340	0.413	0.073
98	Valproic acid	2.29	−0.2	1	144.21	2	−0.220	−0.187	0.033
								AAE[f]	0.337
Test set[a]									
1	1,1,1-Trichloroethane	2.20	2.20	1	133.4	0	0.4	0.258	0.142
2	3-Methylpentane	3.05	3.05	1	86.17	0	1.01	0.562	0.448
3	9-OH rispridone	−3.92	0.20	2	426.48	1	−0.67	−0.971	0.301
4	Atenolol	−2.84	−1.89	2	266.34	4	−0.87	−0.676	0.194
5	Bretazenil	3.88	4.18	2	418.28	0	−0.09	−0.130	0.040
6	Bromperidol	1.14	2.55	2	410.32	1	1.38	0.521	0.859
7	Desmonomethylpromazine	1.10	1.57	2	270.39	1	0.59	0.678	0.088
8	Diethyl ether	1.11	1.11	1	74.120	0	0.00	0.228	0.228
9	Ethanol	−0.07	−0.07	1	46.07	1	−0.16	0.086	0.246
10	Ethylbenzene	3.10	3.10	1	106.160	0	0.20	0.512	0.312
11	Flumanezil	2.11	2.41	2	303.29	0	−0.29	−0.374	0.084
12	Halothane	2.18	2.18	1	197.38	0	0.35	0.062	0.288
13	Heptane	3.77	3.77	1	100.200	0	0.81	0.657	0.153
14	Hexobarbital	1.98	1.9	1	236.27	1	0.1	−0.108	0.208
15	Levorphanol	0.94	2.35	2	259.390	1	0.00	0.785	0.785
16	Mepyramine	−1.95	1.63	2	285.3	0	0.49	0.102	0.388
17	Methotrexate	−5.2	−5.48	1	454.44	7	−1.52	−2.172	0.652
18	Mianserin	−1.54	3.14	2	264.360	0	0.99	0.894	0.096
19	Northioridazine	2.29	2.41	2	356.45	1	0.75	0.794	0.044
20	Paracetamol	0.23	0.23	1	151.16	2	−0.31	−0.172	0.138
21	Primidone	1.00	1.00	1	218.25	2	−0.07	−0.226	0.156
22	R-hydroxyalprazolam	−1.91	0.09	2	330.81	2	−1.27	−0.403	0.867
23	Trichloromethane	1.74	1.74	1	119.38	0	0.29	0.212	0.078
24	Zolantidine	0.72	4.49	2	381.53	1	0.14	0.497	0.357
								AAE[f]	0.298

a) The data were sorted based on f log *BB* values and from every five compounds one was assigned as test set and the remained were determined as train set. *b*) A compound was allocated to the un-ionized group (group 1) if $\log D_{1.7} \geq \log D_{7.4}$ and to the ionized group (group 2) if the $\log D_{1.7} < \log D_{7.4}$. *c*) Molecular weight. *d*) Number of hydrogen bond acceptor. *e*) AE is absolute error and is calculated using $AE = |\log BB_{experimental} - \log BB_{calculated}|$. *f*) AAE is absolute average error and is computed by: $AAE = (\sum N\,AE)/N$.

NHBA is the most effective parameter in which drugs with more hydrogen bond acceptor sites have lower permeation to brain. The impact of hydrogen bonding on log *BB* has been recognized and some models for prediction of log *BB* have been reported in the literature.[1)]

Table 3 shows the correlations between selected parameters and calculated log *BB* values and their intercorrelation. The selected descriptors were not significantly correlated with each other. For un-ionizable compounds, both $\log D_{7.4}$ and MW are important, whereas for ionizable compounds

hydrogen bond is very important and $\log D_{7.4}$ has partial effect.

The overall AAE value for train sets was 0.337±0.291 (0.264±0.237 and 0.410±0.322 for ionizable and un-ionizable compounds, respectively). The corresponding value for test sets was 0.298±0.252 (0.254±0.160 and 0.342±0.321 for ionizable and un-ionizable compounds, respectively). Figure 1 shows the relative frequencies of absolute error for each compound sorted into three subgroups, such as ≤0.3, 0.3—0.8, and >0.8. The acceptable AE for $\log BB$ prediction is 0.3 log unit.[3] Figure 1 shows that >60% of data were predicted by AE values <0.3 log unit. Overall RMSE values of train set was 0.444 (0.353 and 0.519 for un-ionizable and ionizable compounds, respectively) and for test set 0.387 (0.296 and 0.460 for un-ionizable and ionizable compounds,

Table 3. Intercorrelation (R^2) between Selected Parameters and Correlation with $\log BB$ for Train Set

	$\log BB$ (Eq. 10)	$\log BB$ (Eq. 11)	$\log D_{7.4}$	MW	NHBA
$\log BB$ (Eq. 10)	1.00	—	0.49	0.44	—
$\log BB$ (Eq. 11)	—	1.00	0.06	—	0.54
$\log D_{7.4}$	0.49	0.06	1.00	0.11	
MW	0.44	—	0.11	1.00	0.00
NHBA	—	0.54	0.00	—	1.00

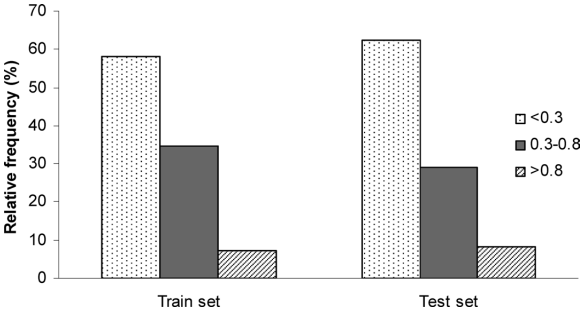

Fig. 1. Relative Frequency (%) of Absolute Errors (AE=|$\log BB_{observed}$ − $\log BB_{calculated}$|) for Train and Test Sets

Table 4. Statistical Parameters of Test Set

Statistical parameter	Group 1 (Eq. 1)	Group 2 (Eq. 2)
$R^2>0.6$	0.848	0.633
$q^2_{ext}>0.5$	0.764	0.636
$(R^2-R_0^2)/R^2$ or $(R^2-R_0'^2)/R^2<0.1$	$0.031\{(R^2-R_0^2)/R^2\}$	$0.027\{(R^2-R_0'^2)/R^2\}$
$0.85<k$ or $k'<1.15$	1.045 (k)	0.965 (k')

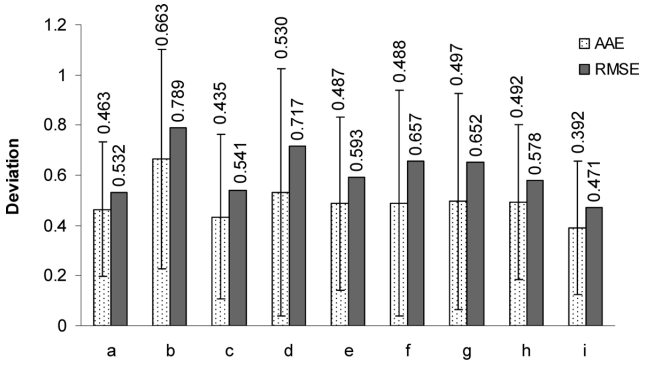

Fig. 2. AAE and RMSE of the Studied Models

a to h: predicted values using previous methods of Table 1, i: predicted using Eqs. 10 and 11, this study.

Table 5. Details of External Test Set, $\log D_{1.7}$, $\log D_{7.4}$, Class of Compounds, Significant Descriptors and Experimental (Observed) and Calculated $\log BB$ Values Using Various Models

No.	Drug	$\log D_{1.7}$	$\log D_{7.4}$	Class	MW	NHBA	$\log BB_{observed}$	$\log BB_{calulated}$								
								a	b	c	d	e	f	g	h	i
1	Alprazolam	2.36	3.49	2	308.8	4	0.04	0.30	−0.58	0.33	0.16	0.98	0.25	0.40	−0.33	0.38
2	Antipyrine	0.40	0.54	2	188.2	6	−0.10	0.24	−0.03	0.47	0.39	−0.36	0.09	0.29	−0.02	−0.62
3	BCNU	1.53	1.53	1	214.1	5	−0.52	−0.36	−0.56	−0.57	0.54	−0.79	−0.83	−0.88	−0.53	0.08
4	Caffeine	−0.45	−0.45	1	194.2	6	−0.06	−0.42	−1.03	−0.22	−0.40	−0.56	−0.76	0.14	−0.56	−0.33
5	Chlorpromazine	2.22	3.34	2	318.9	2	1.06	0.77	0.86	0.71	1.01	0.99	1.12	0.74	0.76	0.94
6	Codeine	−1.85	0.47	2	299.4	4	0.55	0.06	−0.75	0.27	0.12	0.13	0.36	−0.01	−0.15	−0.05
7	Desipramine	1.09	1.57	2	318.9	2	1.20	0.77	0.77	0.43	0.99	0.82	1.11	0.94	0.50	0.69
8	Didanosine	−1.19	−1.08	2	236.2	7	−1.30	−0.97	−1.95	−0.82	−1.19	−1.07	−0.41	−1.12	−1.57	−1.15
9	Hydroxyzine	−2.42	1.94	2	374.9	4	0.39	0.11	−0.20	0.13	0.10	0.50	0.11	−0.44	0.11	0.16
10	Ibuprofen	3.44	0.38	1	206.3	2	−0.18	0.30	−0.09	−0.56	0.20	0.33	0.05	0.11	0.15	−0.17
11	Indomethacin	3.49	0.71	1	357.8	5	−1.26	−0.11	−1.07	−1.03	0.52	0.02	−1.23	−1.63	−0.31	−0.40
12	Midazolam	−0.17	4.25	2	327.8	3	0.36	0.55	−0.02	0.40	0.49	1.13	0.65	−0.14	0.05	0.78
13	Nevirapine	1.46	3.66	2	266.3	5	0.00	−0.22	−0.95	−0.29	−0.11	−0.32	0.01	−0.08	−0.19	0.12
14	Oxazepam	2.11	2.32	2	286.7	4	0.61	−0.18	−0.70	−0.48	0.39	0.18	−0.99	−0.74	−0.79	0.21
15	Pentobarbital	1.88	1.76	1	226.3	4	0.12	−0.55	−0.77	−0.19	−0.81	−0.31	−0.52	−0.55	−0.55	0.10
16	Phenserine	−2.33	1.95	2	337.4	5	1.00	−0.12	−0.23	0.23	0.08	0.18	−0.14	1.00	0.19	−0.13
17	Physostigmine	−3.60	0.68	2	275.4	5	0.08	−0.20	−0.50	0.01	0.05	−0.02	0.43	0.62	−0.12	−0.31
18	Promazine	1.48	2.61	2	284.4	2	1.23	0.72	0.78	0.83	1.02	0.85	1.07	0.84	0.66	0.84
19	SB-222200	4.36	5.86	2	380.5	3	0.30	0.34	0.19	0.43	0.05	0.75	0.65	0.72	0.17	1.01
20	Terbutylchlorambucil	3.23	4.83	2	360.3	3	1.00	0.57	0.28	−0.23	0.50	1.13	−0.40	−0.94	0.49	0.87
21	Theophylline	0.12	0.10	1	180.2	6	−0.29	−0.70	−1.43	−0.51	−0.86	−1.03	−0.67	0.01	−0.92	−0.18
22	Thioridazine	3.04	4.17	2	370.6	2	0.24	0.79	0.89	1.06	0.92	1.46	1.32	0.71	0.77	1.06
23	Trifluoroperazine	0.17	4.37	2	352.4	3	1.44	0.79	0.70	0.46	0.98	1.11	1.21	0.31	0.86	0.80
24	Verapamil	1.76	3.17	2	454.6	6	−0.70	−0.18	−1.32	−1.11	1.07	0.29	0.18	−0.71	0.10	−0.24
25	Zidovudine	−0.80	−0.80	1	267.2	9	−0.72	−1.35	−2.37	−1.02	−1.92	−0.60	ND	−1.23	−1.19	−0.55

a to h: predicted using previous methods, i: predicted using Eqs. 10 and 11, this study.

respectively) which shows a similar accuracy pattern with AAE.

Model Validation Table 4 shows the results of model validation and statistical parameters of the test set. The proposed QSPR models were validated by LOO cross validation. The q^2 values of Eq. 10 and Eq. 11 are acceptable and did not differ substantially from the R^2 values.

Statistical characteristic of the test set is acceptable according to the above-mentioned criteria. R^2 and q^2 were >0.6 and >0.5, respectively; $(R^2-R_0^2)/R^2$ or $(R^2-R_0'^2)/R^2$ was <0.1, and k or k' values varied between 0.85 and 1.15.

Comparison of the Proposed Models with Previous Models A dataset of 25 data points proposed in the literature as external dataset was employed to assess the accuracies of the proposed models and to compare them with previous models.[8,9] The experimental and predicted values of $\log BB$ are listed in Table 5. Figure 2 shows AAE and RMSE of the models. As can be seen, the models produced the same accuracy pattern considering AAE and RMSE criteria. The results suggest that the proposed models can predict $\log BB$ accurately and a simple classification could improve prediction of $\log BB$.

CONCLUSION

MLR was used to propose simple and accurate $\log BB$ prediction model. Classification (according to ionization state) can be a useful method to improve $\log BB$ prediction. $\log D_{7.4}$ and MW or NHBA for un-ionizable and ionizable compounds, respectively, are critical parameters for $\log BB$ prediction. Proposed models are very simple and interpretative, and can be easily used the rapid prediction of $\log BB$ in drug discovery investigations.

Acknowledgment The authors would like to thank Biotechnology Research Center, Tabriz University of Medical Sciences for partial financial support under Grant No. 89313.

REFERENCES

1) Clark D. E., *Annu. Rep. Med. Chem.*, **40**, 403—415 (2005).
2) Mehdipour A. R., Hamidi M., *Drug Discov. Today*, **14**, 1030—1036 (2009).
3) Wang J., Hou T., *Annu. Rep. Comput. Chem.*, **5**, 101—127 (2009).
4) Platts J. A., Abraham M. H., Zhao Y. H., Hersey A., Ijaz L., Butina D., *Eur. J. Med. Chem.*, **36**, 719—730 (2001).
5) Abraham M. H., Ibrahim A., Zissimos A. M., Zhao Y. H., Comer J., Reynolds D. P., *Drug Discov. Today*, **7**, 1056—1063 (2002).
6) Feher M., Sourial E., Schmidt J. M., *Int. J. Pharm.*, **201**, 239—247 (2000).
7) Subramanian G., Kitchen D. B., *J. Comput. Aided Mol. Des.*, **17**, 643—664 (2003).
8) Luco J. M., *J. Chem. Inf. Comput. Sci.*, **39**, 396—404 (1999).
9) Fu X. C., Wang G. P., Shan H. L., Liang W. Q., Gao J. Q., *Eur. J. Pharm. Biopharm.*, **70**, 462—466 (2008).
10) Fu X., Song Z. F., Fu C. Y., Liang W. Q., *Pharmazie*, **60**, 354—358 (2005).
11) Wichmann K., Diedenhofen M., Klamt A., *J. Chem. Inf. Model.*, **47**, 228—233 (2007).
12) Chen Y., Zhu Q. J., Pan J., Yang Y., Wu X. P., *Comput. Methods Programs Biomed.*, **95** , 280—287 (2009).
13) Ghafourian T., Barzegar-Jalali M., Dastmalchi S., Khavari-Khorasani T., Hakimiha N., Nokhodchi A., *Int. J. Pharm.*, **319**, 82—97 (2006).
14) Pharma-Algorithms, ADME Boxes, Version 3.5, Pharma-Algorithms, Toronto, Canada, 2008.
15) Van De Waterbeemd H., Testa B., "Drug Bioavailability: Estimation of Solubility, Permeability, Absorption, and Bioavailability," Wiley-VCH, Weinheim, 2009.
16) Katritzky A. R., Kuanar M., Slavov S., Dobchev D. A., Fara D. C., Karelson M., Acree W. E. Jr., Solov'ev V. P., Varnek A., *Bioorg. Med. Chem.*, **14**, 4888—4917 (2006).
17) Guerra A., Páez J. A., Campillo N. E., *QSAR Comb. Sci.*, **27**, 586—594 (2008).
18) Golbraikh A., Shen M., Xiao Z., Xiao Y. D., Lee K. H., Tropsha A., *J. Comput. Aided Mol. Des.*, **17**, 241—253 (2003).
19) Golbraikh A., Tropsha A., *J. Mol. Graph. Model.*, **20**, 269—276 (2002).
20) Van Damme S., Langenaeker W., Bultinck P., *J. Mol. Graph. Model.*, **26**, 1223—1236 (2008).