Research Paper

QSAR Modeling of the Blood–Brain Barrier Permeability for Diverse Organic Compounds

Liying Zhang,[1] Hao Zhu,[1] Tudor I. Oprea,[2] Alexander Golbraikh,[1] and Alexander Tropsha[1,3]

Purpose. Development of externally predictive Quantitative Structure–Activity Relationship (QSAR) models for Blood–Brain Barrier (BBB) permeability.

Methods. Combinatorial QSAR analysis was carried out for a set of 159 compounds with known BBB permeability data. All six possible combinations of three collections of descriptors derived from two-dimensional representations of molecules as chemical graphs and two QSAR methodologies have been explored. Descriptors were calculated by MolconnZ, MOE, and Dragon software. QSAR methodologies included *k*-Nearest Neighbors and Support Vector Machine approaches. All models have been rigorously validated using both internal and external validation methods.

Results. The consensus prediction for the external evaluation set afforded high predictive power ($R^2 =$ 0.80 for 10 compounds within the applicability domain after excluding one activity outlier). Classification accuracies for two additional external datasets, including 99 drugs and 267 organic compounds, classified as permeable (BBB+) or non-permeable (BBB−) were 82.5% and 59.0%, respectively. The use of a fairly conservative model applicability domain increased the prediction accuracy to 100% and 83%, respectively (while naturally reducing the dataset coverage to 60% and 43%, respectively). Important descriptors that affect BBB permeability are discussed.

Conclusion. Models developed in these studies can be used to estimate the BBB permeability of drug candidates at early stages of drug development.

KEY WORDS: combinatorial QSAR; k-nearest neighbors; model validation; predictors of BBB permeability; support vector machines.

INTRODUCTION

The blood–brain barrier (BBB) separates the brain from the bloodstream and limits the transport of many substances from the systemic circulation into the brain tissue. The concept of the BBB and the mechanism of its functionality have been described in several reviews (1–3). The ability of an organic chemical to penetrate the brain can be estimated by measuring its brain-to-blood concentration ratio (BB),

Electronic supplementary material The online version of this article (doi:10.1007/s11095-008-9609-0) contains supplementary material, which is available to authorized users.

[1] The Laboratory for Molecular Modeling, School of Pharmacy, University of North Carolina at Chapel Hill, CB# 7360, Chapel Hill, North Carolina 27599-7360, USA.

[2] Division of Biocomputing, MSC11 6145, University of New Mexico School of Medicine, University of New Mexico, Albuquerque, New Mexico 87131-0001, USA.

[3] To whom correspondence should be addressed. (e-mail: alex_tropsha@unc.edu)

ABBREVIATIONS: AD, applicability domain; BBB, blood–brain barrier; Combi-QSAR, combinatorial QSAR; *k*NN, *k*-nearest neighbors; MAE, mean absolute error; NIH, National Institutes of Health; OECD, Organization for Economic Co-operation and Development; QSAR, quantitative structure–activity relationship; SVM, support vector machines.

which is defined as the ratio of drug concentration in brain tissue to the drug concentration in blood.

Brain penetration is one of the major parameters that are taken into consideration in chemical toxicological studies and in drug design. Any drug molecule targeting a receptor in the brain must first cross the BBB to realize its therapeutic potential. On the other hand, for a drug candidate not aimed at the Central Nervous System (CNS), passage across the BBB could induce undesirable side effects. The experimental determination of the brain-to-blood concentration ratio requires complex techniques that are expensive and time-consuming. Therefore, rapid and accurate computational methods for screening large chemical databases or virtual libraries are desirable to assist the experimental drug discovery process.

There have been many attempts to correlate the experimental brain-to-blood concentration ratio values, which is always represented by its logarithm (logBB), with physical–chemical parameters. Table I lists some of the previous QSAR studies of BBB permeability.

Young and co-workers conducted what we believe to be the first QSAR study of brain–blood distribution in 1988 (4). They reported the *in vivo* values in rats for a large number of H2 receptor histamine agonists, and found that the Seiler parameter, which is defined as $\log P(\text{o/w}) - \log P(\text{cycl/w})$ (where $P(\text{o/w})$ and $P(\text{cycl/w})$ are the octanol/water and cyclohexane/water partition coefficients, respectively), is

Table I. Summary of Previous QSAR Studies of the BBB Permeability

Group	N_{des}	Descriptors	N_{train}	R^2_{train}	q^2_{train}	S_{train}	N_{ext}	P_{test}
Young (4)	1	logP	20	0.69	–	0.439	–	–
Kansy (9)	2	PSA and molecular volume	20	0.697	–	0.448	–	–
Abraham (6)	5	Molecular property descriptors	57	–	0.91	0.197	–	–
Lombardo (7)	1	Free energy of solvation	55	0.67	–	0.41	6	–
Clark (9)	2	PSA and logP	55	0.787	–	0.354	5	MAE=0.24
Luco (10)	18	Topological and constitutional descriptors	58	0.850	0.752	0.318	22	RMSE=0.408
Feher (11)	3	Number of hydrogen-bond acceptors, logP, PSA	61	0.73	0.688	RMSE=0.424	25	RMSE=0.789
Kelder (12)	1	Dynamic PSA	45	0.841	–	–	–	–
Norinder (14,23)	14	Molecular property descriptors	28	0.862	0.782	0.311	28	RMSE=0.353
Platts (24)	6	Molecular property descriptors	148	0.745	0.711	0.343	–	–
Keseru (67)	1	Solvation free energies	55	0.72	–	0.37	5	MAE=0.14
Salminen (68)	3	Lipophilicity, molecular size and acid/base character	23	0.848	–	0.32	–	–
Ma (69)	8	Inter- and intra-molecular solute descriptors	37	0.912	–	0.232	8	–
Katritzky (70)	5	ClogP, etc.	113	0.781	0.752	0.351	19	0.032
Hou (71)	3	High-charged PSA, logP, MW360	72	0.785	–	0.358	–	–
Iyer (60)	5	PSA, clogP and membrane-solute descriptors	56	0.845	0.795	–	7	–
Pan (72)	2	TPSA and clogP	37	0.85	0.83	–	46	R^2=0.76
Subramanian (8)	8	AlogP	58	0.845	0.811	0.314	39	RMSE=0.463
Rose (60)	3	E-state index and molecular connectivity index	102	0.66	0.62	0.48	28	–[a]
Winkler (73)	7	Property-based descriptors	106	0.81	0.65	RMS=0.37	–	–

N_{des}: number of descriptors used, N_{train}: number of compounds in the training set, R^2_{train}: squared correlation coefficient between predicted and experimental logBB values for the model derived from the training set, q^2_{train}: the cross-validated R^2 value for the model derived from the training set, S_{train}: the standard deviation for the model derived from the training set, N_{ext}: number of compounds in the external evaluation set, P_{test}: test set predictive performance, PSA: polar surface area, MAE: mean absolute error of prediction for the compounds in the external evaluation set, RMS: root mean square, RMSE: root mean squared error
[a] Detailed prediction results such as R^2 or MAE weren't shown.

related to the logBB values for a series of 20 histamine antagonists. Following this first study, Abraham (5,6), Lombardo (7), Subramanian (8), Clark (9), Luco (10), Feher (11), Kelder (12), Brewster (13), Norinder (14) and their co-workers have added more data to the "Young dataset" and developed new logBB models that achieved higher statistical significance and better predictivity. The parameters used in these studies were lipophilicity, solvatochromic parameters, topological indices, and combinations of these parameters. Based on these reports, three properties, i.e., molecular volume, lipophilicity, and hydrogen bonding potential, were identified as contributing significantly to the transport through the BBB. However, the relationship between these properties and the brain–blood distribution remained obscure, and a relatively good correlation could be achieved only for small sets of molecules. Furthermore, some of the previous models were limited to retrospective analysis of historic data lacking external validation to prove their predictive ability. As we and others have shown (15–17) the internal measures of QSAR model accuracy are not indicative of a models' external predictive power. Furthermore, the so called OECD principles introduced by the QSAR Working Group of the OECD (18) explicitly recommend to include external validation as a mandatory component of model development. Thus, many previously published QSAR models of BBB permeability may in fact have unproven predictive power.

For this study, we have compiled the largest (to our knowledge) dataset of 159 compounds with quantitatively measured logBB using data from all previous publications. The dataset was randomly split into a modeling set of 144 compounds and an external evaluation set of 15 compounds. The former set was used to develop QSAR BBB models using different types of descriptors and modeling approaches in the context of a combinatorial QSAR modeling approach that we began to advocate for recently (19–21). All models were subjected to rigorous internal and external validation. The resulting validated models were used to predict the logBB values of the external evaluation set. Furthermore, two additional datasets containing 99 and 267 compounds, respectively, which were classified as BBB permeable (BBB+) or BBB non-permeable (BBB−), were also used for model validation. The results confirmed the high external prediction accuracy of our models, which led us to conclude that these models can be used reliably to evaluate BBB permeability of organic compounds. The successful validation allowed us to interpret our models in the context of chemical descriptors that were found to be significant and consequently responsible for the level of the BBB permeability of studied molecules. The models developed and validated in our studies can be used as reliable predictors of BBB permeability of organic molecules.

Table II. Compounds Included in the Modeling Set and the First External Evaluation Set

Comp. no.	Comp. name or id	Exp. logBB
1.	11[a]	−1.17
2.	12	−2.15
3.	19	−1.54
4.	Terbutylchlorambucil	1.00
5.	Org12692	1.64
6.	Org13011	0.16
7.	Org34167	0
8.	M2L-663581	−1.82
9.	RO19-4603	−0.25
10.	15	−0.12
11.	16	−0.18
12.	17	−1.15
13.	24	−0.46
14.	33	−0.3
15.	36	0.89
16.	SKF89124	−0.43
17.	Org4428	0.82
18.	Org5222	1.03
19.	Org32104	0.52
20.	Org30526	0.39
21.	9-OH risperidone	−0.67
22.	Compound 31	−0.67
23.	Compound 32	−0.66
24.	Compound 34	−1.57
25.	Compound 35	−1.12
26.	Compound 36	−0.73
27.	Compound 37	−0.27
28.	Compound 38	−0.28
29.	Compound 40	−0.24
30.	Compound 41	−0.02
31.	Compound 42	0.69
32.	Compound 43	0.44
33.	Compound 45	0.22
34.	Nor-1-chlorpromazine	1.37
35.	2-(3′-Iodo-4′-aminophenyl)-6-hydroxybenzothiazole	0.176
36.	SB-656104-A	−0.0457
37.	Alovudine	−0.605
38.	Granisetron	−0.687
39.	Zidovudine	−0.886
40.	Tamoxifen	0.924
41.	Desmethyldesipramine	1.06
42.	Phenylbutazone	−0.52
43.	Desipramine	1.2
44.	Imipramine	0.83
45.	Thioridazine	0.24
46.	Chlorpromazine	1.06
47.	Acetylsalicylic acid	−0.5
48.	Verapamil	−0.7
49.	Haloperidol	1.34
50.	Indomethacin	−1.26
51.	Hexobarbital	0.1
52.	Quinidine	−0.46
53.	Phenytoin	−0.04
54.	Amobarbital	0.04
55.	Caffeine	−0.055
56.	Aminopyrine	0
57.	Promazine	1.23
58.	Theophylline	−0.29
59.	Diethyl ether	0
60.	Antipyrine	−0.097
61.	Ethanol	−0.16

Table II. (continued)

Comp. no.	Comp. name or id	Exp. logBB
62.	2-Propanol	−0.15
63.	Propanone	−0.15
64.	Trichloromethane	0.29
65.	Hydroxyzine	0.39
66.	Fluphenazine	1.51
67.	Salicylic acid	−1.1
68.	Propanol	−0.16
69.	Benzene	0.37
70.	1,1,1-Trichloro-ethane	−0.6
71.	2,2-Dimethylbutane	1.04
72.	1,1,1-Trifluro-2-chloroethane	−0.92
73.	Pentobarbital	0.12
74.	Thiopental	−0.14
75.	2-Methylpropanol	−0.17
76.	Butanone	−0.08
77.	Theobromine	−0.28
78.	mepyramine	0.49
79.	3-Methylpentane	1.01
80.	Methylcyclopentane	0.93
81.	Valproic acid	−0.22
82.	Acetaminophen	−0.31
83.	2-Methylpentane	0.97
84.	Toluene	0.37
85.	Pentane	0.76
86.	Hexane	0.8
87.	Heptane	0.81
88.	Halothane	0.35
89.	Methohexital	−0.06
90.	BCNU	−0.52
91.	carbamazepine	0
92.	Chlorambucil	−1.7
93.	Fluroxene	0.13
94.	Teflurane	0.27
95.	Diazepam	0.52
96.	Salicyluric acid	−0.44
97.	Propanolol	0.64
98.	3-Methylhexane	0.9
99.	Oxazepam	0.61
100.	Paraxanthine	0.06
101.	Desmethyldiazepam	0.5
102.	Codeine	0.55
103.	Flunitrazepam	0.06
104.	Nor-2-chlorpromazine	0.97
105.	Desmonomethylpromazine	0.59
106.	Mesoridazine	−0.36
107.	Tibolone	0.4
108.	Y-G 14	−0.3
109.	Y-G 15	−0.06
110.	2	−0.04
111.	Bromperidol	1.38
112.	Northioridazine	0.75
113.	Sulforidazine	0.18
114.	Ibuprofen	−0.18
115.	Clobazam	0.35
116.	Mianserin	0.99
117.	Isoflurane	0.42
118.	Triazolam	0.74
119.	Atenolol	−1.42
120.	Zidovudine	−0.72
121.	Carbamazepine-10,11-epoxide	−0.35
122.	Y-G 20	−1.4
123.	Cimetidine	−1.42
124.	Domperidone	−0.78

Comp. no.	Comp. name or id	Exp. logBB
125.	Midazolam	0.36
126.	4-Hydroxymidazolam	−0.3
127.	1-Hydroxymidazolam	−0.07
128.	ranitidine	−1.23
129.	Tiotidine	−0.82
130.	Didanosine	−1.3
131.	SKF 93319	−1.3
132.	Icotidine	−2
133.	Flumazenil	−0.29
134.	Lupitidine	−1.06
135.	Bretazenil	−0.09
136.	Temelastine	−1.88
137.	SKF101468	0.25
138.	Phenserine	1
139.	Zolantidine	0.14
140.	Thioperamide	−0.161
141.	Risperidone	−0.02
142.	MIL-663581	−1.34
143.	Indinavir	−0.74
144.	SB-222200	0.3
145.	Lamotrigine	−0.081
146.	Morphine	−0.16
147.	Physostigmine	0.079
148.	Methoxyflurane	0.25
149.	Trichloroethene	0.34
150.	Trifluoperazine	1.44
151.	clonidine	0.11
152.	Enflurane	0.24
153.	Y-G 16	−0.42
154.	Desmethylclobazam	0.36
155.	Alprazolam	0.044
156.	Mirtazapine	0.53
157.	Y-G 19	−1.3
158.	Nevirapine	0
159.	SKF-93619	−1.3

Compounds no. 1–144 are in the modeling set and 145–159 are in the first external evaluation set.

[a] Compound names are the same as those in references (22–24).

METHODOLOGY

Datasets

A set of 159 unique compounds with known logBB values was compiled from several publications (22–24). This set included the Young, Salminen and Kelder data sets, as well as seven additional compounds (See Table II and Supplemental Materials). The range of logBB of the 159 compounds was −2.15 to 1.64. The dataset was divided randomly into two subsets: the modeling set ($n=144$) and the external evaluation set ($n=15$). Following our standard protocol (25), the modeling set was additionally divided into multiple training and test sets to allow for the development and validation of QSAR models, respectively. The second external evaluation set was derived from the WOMBAT-PK dataset (26) (cf. Supplemental Materials), which included 99 commercially available drugs. The third external evaluation set of 267 organic compounds was derived from a recent publication (27) (cf. Supplemental Materials). The duplicate compounds (identical to those in the

modeling set) in the two external evaluation datasets, as well as any inorganic compounds, were removed. Compounds in the second and the third evaluation sets were classified as permeable (BBB+) or non-permeable (BBB−) from their respective data sources. Thus, to afford the binary interpretation of predictions made with QSAR models developed using real logBB values for the modeling set, we followed the definitions in (27), i.e., compounds with experimental logBB<−1 were classified as relatively poor penetrators of the BBB barrier (i.e., BBB−), while compounds with logBB>−1 were classified as relatively good penetrators of the BBB (i.e., BBB+).

Computational Methods

A combinatorial QSAR (Combi-QSAR) approach (19–21), employing six combinations of three descriptor types and two types of optimization methods as described below, was used for QSAR studies of the BBB datasets.

Descriptors

Descriptor types used in this study are considered in the following sections. Each type of descriptors was used separately with each QSAR method in the context of our Combi-QSAR strategy (19–21).

MolConnZ 4.05 Descriptors

MolConnZ descriptors include valence, path, cluster, path/cluster and chain molecular connectivity indices (28–30), kappa molecular shape indices (31,32), topological (33) and electro-topological state indices (34–37), differential connectivity indices (29,38), graph's radius and diameter (39), Wiener (40) and Platt (41) indices, Shannon (42) and Bonchev-Trinajstić (43) information indices, counts of different vertices, counts of paths and edges between different types of vertices (http://www.edusoft-lc.com/molconn/manuals/400). Descriptors with zero values or zero variance were removed; the remaining descriptors were normalized by range-scaling so that their values were distributed within the interval between 0 and 1. The total number of descriptors used in model development was 346.

MOE Descriptors

2D MOE descriptors (Chemical Computing Group; http://www.chemcomp.com/software.htm) include physical properties, subdivided surface areas, atom counts and bond counts, Kier and Hall connectivity (28–30) and kappa shape indices (31,32), adjacency and distance matrix descriptors (39,40,44,45), pharmacophore feature descriptors and partial charge descriptors (Chemical Computing Group; http://www.chemcomp.com/software.htm) (45). MOE descriptors were range-scaled. 3D MOE descriptors were not used. Prior to QSAR studies, low variance descriptors (with variance less than 0.05) were eliminated. The final number of descriptors used in QSAR studies was 184.

Dragon Descriptors

A set of 929 theoretical molecular descriptors was computed using *DRAGON* software (46). The descriptors were

generated from the SMILES notation available for each compound. The typology of the included descriptors is: 0D-constitutional (atom and group counts); 1D-functional groups; 1D-atom centred fragments; 2D-topological descriptors; 2D-walk and path counts; 2D-autocorrelations; 2D-connectivity indices; 2D-information indices; 2D-topological charge indices; 2D-Eigenvalue-based indices; 2D-topological descriptors; 2D-edge adjacency indices; 2D-Burden eigenvalues; molecular properties. Constant and near to constant variables were deleted. If two descriptors were at least 99% correlated one of them was deleted. The final set used in QSAR studies included 324 descriptors. Dragon descriptors were range-scaled. The calculation procedures for these descriptors, with related literature references, are reported in (47).

QSAR Methods

Training set models were built using variable selection kNN and SVM approaches that were implemented in our group. The kNN QSAR method (48) employs the kNN classification principle and the variable selection procedure. Briefly, initially a subset of n_{var} (number of selected variables) descriptors is selected randomly. The n_{var} is set to different values to obtain the best q^2 possible. The models are optimized by leave-one-out cross-validation, where each compound is eliminated from the training set and its biological activity is predicted as the weighted average activity of k most similar molecules (k=1 to 5). The similarity is characterized by Euclidean distances between compounds in multidimensional descriptor space. A simulated annealing method with the Metropolis-like acceptance criteria is used to optimize the selection of variables.

The SVM regression approach (49) attempts to find the narrowest band in the descriptor-activity space containing most of the data points; it can be divided into linear or nonlinear SVM depending on the type of a kernel function used to construct the model. In this study, the linear SVM regression was used. The generalized performance of SVM depends on the selection of several internal parameters of the algorithm (C and ε). To find models with the highest accuracy for both training and test sets, the calculations were carried out for all combinations of C and ε with the C value varying from 0.1 to 100 with a step of 10, and ε varying from 0.0 to 0.5 with a step of 0.34. For example, if the total number of training/test sets generated for one type of descriptors was 36, $36 \times 10 \times 2 = 720$ models were constructed. Further details of the kNN and SVM method implementation are given elsewhere (20,21,50).

Validation of QSAR Models

As emphasized in one of our previous reports (17), training-set-only modeling is insufficient to achieve models with validated predictive power. For this reason, the 144-compound modeling set was divided into multiple training/test sets using the sphere-exclusion algorithm (25). For each collection of descriptors, the modeling set was divided into 36–50 training/test sets of different relative sizes. The kNN and SVM QSAR models were developed solely based on training sets that were part of the modeling set, and the resulting models were validated through predicting the BBB permeability of the compounds in the respective test sets. The

statistical significance of the kNN and SVM QSAR BBB models was characterized by the following parameters (17). (1) LOO cross-validated q^2; (2) square of the correlation coefficient R between the predicted and observed activities; (3) coefficients of determination (predicted *versus* observed activities R_0^2, and observed *versus* predicted activities $R_0'^2$); (4) slopes k and k' of regression lines (predicted *versus* observed activities, and observed *versus* predicted activities) through the origin. These criteria are calculated according to the previous paper (17). The importance of this procedure was discussed in previous publications (25,51).

We considered a QSAR model to have an acceptable predictive power if the following conditions were satisfied:

$$q^2 > 0.65; \tag{1}$$

$$R^2 > 0.65; \tag{2}$$

$$\frac{(R^2 - R_0^2)}{R^2} < 0.1 \text{ and } 0.85 \leq k \leq 1.15; \tag{3a}$$

or

$$\frac{(R^2 - R_0'^2)}{R^2} < 0.1 \text{ and } 0.85 \leq k' \leq 1.15; \tag{3b}$$

and

$$\left| R_0^2 - R_0'^2 \right| < 0.3. \tag{4}$$

The Mean Absolute Error (MAE) was also used to estimate the predictive power of QSAR models for self-correlation and to predict the first external evaluation set. For the second and third external evaluation sets, besides the overall prediction accuracy calculated as the fraction of correctly classified compounds, we also used sensitivity (SE), which is defined as (true positives)/(true positives+false negatives) and specificity (SP), which is defined as (true negatives)/(true negatives+false positives). SE and SP reflect the accuracy of predicting the compounds of BBB+ and the BBB− classes, respectively.

Applicability Domain (AD)

Formally, a QSAR model can predict the target property for any compound for which chemical descriptors can be calculated. However, if a compound is highly dissimilar to all compounds of the modeling set, the reliable prediction of its activity is unlikely. A concept of the AD was developed and used to avoid such an unjustified extrapolation of activity predictions. In this study, the AD was defined as a threshold distance D_T between a compound under prediction and its closest nearest neighbor of the training set, calculated as follows:

$$D_T = \bar{y} + Z\sigma. \tag{5}$$

Here, $\bar{y}$ is the average Euclidean distance between each compound and its k nearest neighbors in the training set (where k is the parameter optimized in the course of QSAR

modeling, and the distances are calculated using all descriptors and descriptors selected by the optimized model only), σ is the standard deviation of these Euclidean distances, and Z is an arbitrary parameter to control the significance level. We set the default value of this parameter Z to 0.5, which formally places the allowed distance threshold at the mean plus one-half of the standard deviation. Thus, if the distance of the external compound from its nearest neighbor in the entire descriptor space or the subspace of descriptors selected in the training set exceeds this threshold, the prediction is not made (52).

Robustness of QSAR Models

Y-randomization (randomization of response) is a widely used approach to establish model robustness. It consists of rebuilding the models using randomized activities of the training set and subsequent assessment of the model statistics. It is expected that models obtained from the training set with randomized activities should have significantly lower values of q^2 for the training set than the models built using training set with real activities, or at least these models should not have satisfied some of the validation criteria 1, 2, 3a, 3b, and 4 above. If this condition is not satisfied, real models built for this training set are not reliable and should be discarded. This test was applied to all data divisions considered in this study and it was repeated twice for each division.

RESULTS AND DISCUSSION

QSAR Models

For the development of QSAR models, two optimization methods, i.e., kNN and SVM were used, each with three types of descriptors, i.e., Dragon, MOE and MolConnZ. Consequently, six different types of QSAR models were developed for the modeling set. Table III presents the detailed information for all six types of models. It needs to be pointed out that the prediction of a compound's BBB permeability was made by taking the average predicted value using all models that satisfied the criteria discussed above (cf. Methodology). Clearly, the statistical parameters of models generated in our study with six types of QSAR methods for the extended modeling set appear similar or

Table III. Details of the Six Types of QSAR Models Developed with 144 Compounds in the Modeling Set

Model ID	Model type	N_{models}	N_{des}	R^2_{self}	MAE_{self}
1.	kNN-Dragon	53	324	0.92	0.18
2.	SVM-Dragon	37	324	0.86	0.27
3.	kNN-MOE	170	184	0.75	0.31
4.	SVM-MOE	12	184	0.82	0.24
5.	kNN-MolConnZ	234	346	0.95	0.15
6.	SVM-MolConnZ	24	346	0.87	0.25
Consensus self-correlation				0.91	0.21

N_{models}: number of accepted models (satisfying criteria 1, 2, 3a, 3b, and 4; see Methodology), N_{des}: number of descriptors used in the model, R^2_{self}: squared correlation coefficient between predicted and experimental logBB values of self-correlation, MAE_{self} : mean absolute error of prediction for the self-correlation
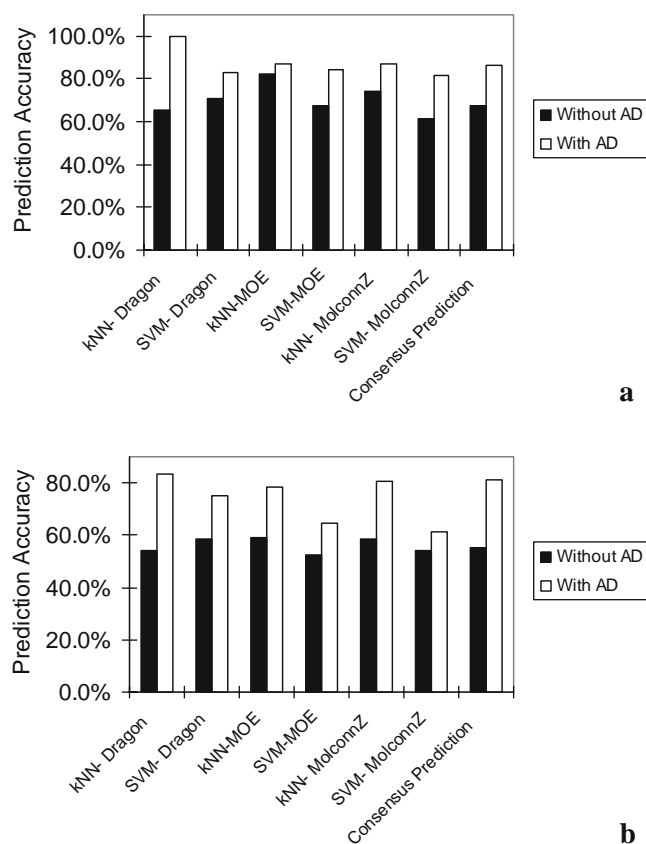
better than those obtained in previous studies (Table I). Of course, the direct comparison is difficult since different models were generated for different sets of compounds but we stress that our dataset was by far larger than any of the datasets studied previously (cf. Table I).

Self-correlation for the modeling set was carried out by predicting the logBB values of all 144 compounds in the modeling set that were used to develop QSAR models. The self-correlation R^2 values were higher than 0.75 for all six types of models (Table III). Not surprisingly, the results for the first external evaluation set were not as good as those for the self-correlation (Supplemental Materials). For example, the R^2 values of the first external evaluation set ($n=15$) prediction are much lower, even if the AD was applied. Without applying the AD to the external evaluation set, the R^2 and MAE of the predictions of these 15 external compounds range from 0.39 to 0.58 and 0.33 to 0.43, respectively, for the six types of models. However, when the AD is used, the values of R^2 become significantly higher, while MAE decreases (*vide infra*).

The Y-randomization test was performed, and the q^2 values for the modeling set with the experimental logBB values were shown to be significantly higher than those obtained from the same dataset with randomized logBB values. For the modeling set with real logBB values, there were 530 models that satisfied the criteria of $q^2>0.65$ and $R^2>0.65$ (Table III), whereas for the dataset with randomized logBB values, only seven models that had q^2 higher than 0.4 were generated, and the highest q^2 was only 0.53. These results indicate that our models are statistically robust.

Two additional datasets (described in Methodology) were then considered to test the predictive ability of the six types of models. They included 99 (the second external evaluation set) and 267 (the third external evaluation set) compounds. All compounds in both datasets were classified as either BBB+ or BBB−. Thus, the predicted logBB values for all compounds in these two datasets were converted into BBB+ or BBB− using the same threshold that was used in the analysis of the external evaluation sets (cf. Methodology). The best models that predict the second and the third external evaluation sets with the highest accuracy without applying AD were those built with the kNN method and MOE descriptors. The overall prediction accuracy, calculated as the ratio of the number of correctly predicted compounds to the total number of compounds in the external evaluation set, was 82.5% and 59.0%, respectively (cf. black columns in Fig. 1).

If we assess the predictive ability of models using SE and SP, the best SE is as high as 100% while the best SP is only 70.0%. The much higher predictive power for the BBB+ compounds as compared to that for the BBB− compounds can be explained by the biased distribution of BBB+/BBB− compounds in the modeling set, which includes significantly more BBB+ (logBB>−1) than BBB− (logBB<−1) compounds (Fig. 2). This problem arises from the fact that many of the BBB− compounds lack measurable logBB values because they can hardly go through the BBB. Taking into account that the kNN algorithm is based on similarity principle, it may not be surprising that these models are able to predict BBB+ compounds better than BBB− compounds.

Fig. 1. The effect of the AD on the model prediction accuracy for the second (**a**) and third (**b**) external evaluation sets.

The Effect of the Applicability Domain

The AD defines the area of the descriptor space in which QSAR models can accurately predict the target properties. If a compound is "too dissimilar" (beyond the defined distance cutoff value; see in Methodology) to all compounds of the modeling set in the descriptor space then we assume that we cannot predict its activity reliably. Thus, some compounds of
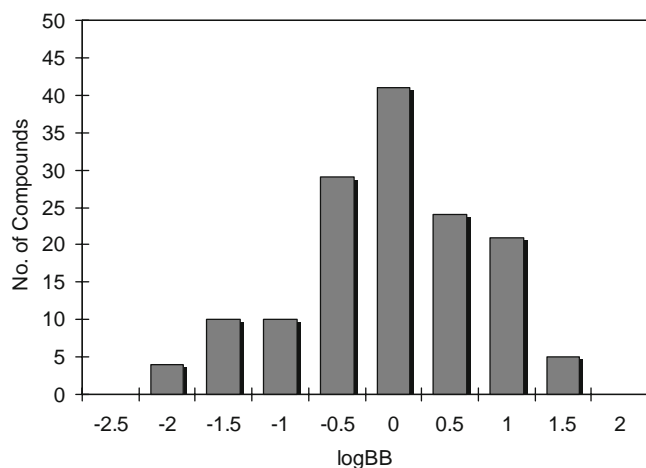
each external evaluation set were found outside of the AD and defined as "impossible to predict": Euclidean distances between these compounds and their closest nearest neighbors in the training set were exceeding the predefined similarity threshold (defined by rather conservative Z cutoff=0.5). For each external evaluation set, the number of compounds within the AD was lower than the total number of compounds, but the associated prediction accuracy increased significantly. After applying the AD to the first evaluation test set, the R^2 increases for all the six types of models while the coverage decreases (from 100% to 67–87%) compared to those obtained without the implementation of the AD. The R^2 values obtained for three model types (kNN-Dragon, SVM-Dragon and kNN-MolConnZ) were above 0.60, which is quite significant when considering the small size and high diversity of the modeling set used in this study.

Without applying the AD for the second and third external evaluation sets, the best overall accuracy was 82.5% and 59.0%, respectively (black columns in Fig. 1). This is similar to the classification BBB modeling results reported earlier (27). After applying the AD, the best overall prediction accuracy for the two additional external sets increased to 100% and 83.3% (white columns in Fig. 1). However, the increased accuracy came at the expense of reducing the number of compounds for which the prediction could be made. Thus, less than 60% of all compounds were within the AD for the second external evaluation set, and only 43% of all compounds in the third one. Notably, most of the BBB− compounds in the external evaluation sets were considered to be out of the AD and defined as "no prediction" unlike BBB+ compounds (Fig. 3). This discrepancy is caused by the biased distribution of BBB+/BBB− compounds in the modeling set.

We have investigated the effect of varying the threshold value of the AD on the interplay between chemical space coverage and prediction accuracy. We have found that for the first external evaluation set, the R^2 decreases when Z cutoff increases from 0.5 to 1 (data not shown). However, for the second and third external evaluation sets, increasing Z cutoff was found to increase the chemical space coverage while having relatively small effect on the prediction accuracy (Fig. 4). Furthermore, when Z cutoff increased from 0.5 to 8, the models were still predicting BBB+ compounds with
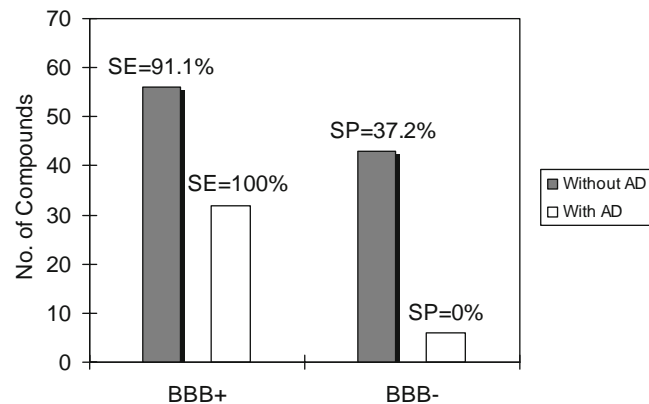

Fig. 2. The distribution of the logBB values for the modeling set (n=144).

Fig. 3. The effect of the AD on sensitivity (*SE*), specificity (*SP*), and chemical space coverage for the SVM-MOE model of BBB permeability for the second evaluation set of 99 compounds.
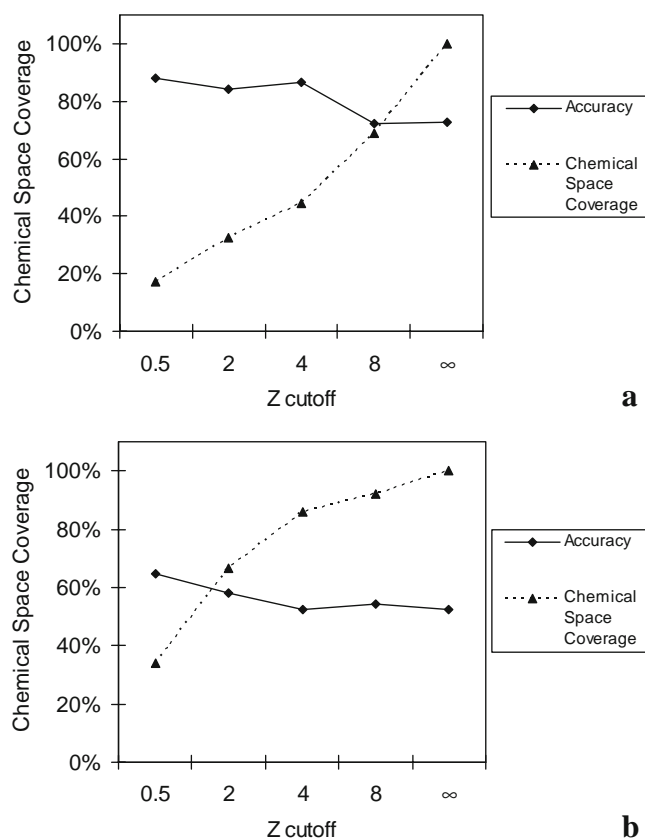
Fig. 4. The effect of the AD on the relationship between prediction accuracy and chemical space coverage for the SVM-MOE model of BBB permeability for the second (**a**) and third (**b**) external evaluation sets.

higher accuracy than BBB− compounds. It is generally hard to assign a standard AD threshold that should be used in all cases. Typically, we tend to use a conservative Z cutoff of 0.5 to ensure the high prediction accuracy for all (eligible) compounds in the external sets. However, results of this study illustrate that the external prediction accuracy was not affected significantly by increasing the AD, i.e., Z cutoff, up to values as high as 8, which allowed making accurate predictions for more than half of the compounds in the second and third external evaluation sets.

Consensus Prediction

An important question related to the above discussion is how to select the most predictive QSAR model from all available models for a given endpoint. If we use models that are considered acceptable using criteria 1, 2, 3a, 3b, and 4 and use R^2 for the self-correlation of the modeling set to evaluate the quality of the QSAR models, then the top-ranking model is kNN-MolConnZ. However, it was the SVM-Dragon model that gave the best prediction for the first external evaluation set. Thus, relying on the results obtained for the modeling set only could obfuscate the choice of the best modeling technique to achieve the most accurate external prediction. Our previous experience suggests that the consensus prediction that is based on the results obtained by all predictive models always provides the most stable solution.

In general, consensus prediction implies averaging the predictions for each compound made by individual models for continuous QSAR, or by majority voting for classification QSAR, using all models passing the validation criteria. Precision and stability of the consensus prediction is a consequence of the Central Limit Theorem (53). This approach naturally avoids the need for (the best) model selection based on the statistics for the training (or even training and test) set.

The ultimate consensus prediction was made by averaging the logBB values predicted by all six types of models. Thus, for the modeling set of 144 compounds we averaged the calculated logBB values resulting from the six types of models. The R^2 and MAE for the consensus self-correlation were 0.91 and 0.21, respectively, which is close to the best statistics for any individual model.

To ensure the reliability of the external prediction, we only predicted the logBB value for those compounds in the external evaluation sets that were within the AD of at least half of the six types of models. Four compounds were considered to be out of the AD for more than half of the six types of models. Therefore, we did not consider these four compounds in the consensus prediction (see Supplemental Materials). The R^2, MAE, and coverage of the consensus prediction were 0.73, 0.36, and 73%, respectively; these values are close to those obtained with the best individual model. If we only select compounds that are within the AD of all six types of models for consensus prediction, the R^2, MAE and coverage were 0.78, 0.37, and 60%, respectively.

QSAR models were found to predict accurately the BBB penetration for most of the compounds in the first external evaluation set which were within the AD. For example, compound 12 in this external set has a relatively small prediction error (Fig. 5a). It is interesting to notice that some of its nearest neighbors in the chemical space do not look similar to this compound in terms of their formal chemical structures. Mahar Doan *et al*. (54) found that drugs for the central nervous system (CNS) (CNS drugs), which are likely to cross BBB have fewer hydrogen bond donors, fewer positive charges, greater lipophilicity, and lower polar surface area (PSA) than non-CNS drugs, which cannot cross the BBB. These results are in agreement with our observations when comparing the related descriptors of compound 12 and its four nearest neighbors.

Compound 13 (Y-G19) has a relatively large error when comparing its predicted and experimental logBB values. In fact, compound 13 has been considered to be an outlier or has a large prediction error in other computational models as well (23). The values of the primary descriptors of this compound and its four nearest neighbors are close to each other (cf. Fig. 5b). However, the BBB permeability values for these five compounds vary greatly, which is the direct reason why the logBB values for compound 13 are predicted inaccurately (mind that in the kNN method the predicted value for compounds is the average of experimental measurements for compounds in the training set that are most similar to that compound). The large difference of logBB values among these five compounds may be because some of them are substrates of the efflux transporter, which pumps the drugs out of capillary endothelial cells and therefore prevents them from crossing the BBB. Y-G20, one of Y-G19's nearest
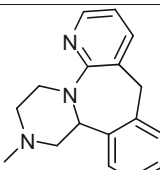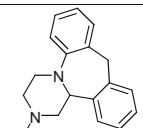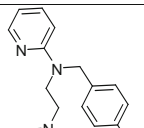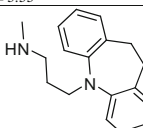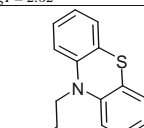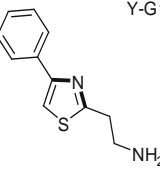
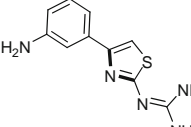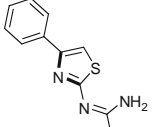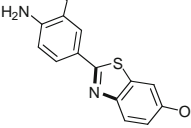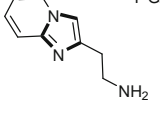| **a.** Compound 12	Nearest neighbors in the modeling set	
logBB_actr = 0.53 logBB_pred = 0.74 PEOE_VSA_POS = 160.40 nHBa = 2 TPSA = 19.37 logP = 2.50	logBB_actr = 0.99 PEOE_VSA_POS = 134.73 nHBa = 1 TPSA = 6.48 logP = 3.33	logBB_actr = 0.49 PEOE_VSA_POS = 199.07 nHBa = 3 TPSA = 28.60 logP = 2.62
	logBB_actr = 1.20 PEOE_VSA_POS = 124.78 nHBa = 1 TPSA = 15.27 logP = 3.84	logBB_actr = 0.59 PEOE_VSA_POS = 116.09 nHBa = 1 TPSA = 15.27 logP = 3.73
b. Compound 13	Nearest neighbors in the modeling set	
logBB_actr = -1.30 logBB_pred = -0.18 PEOE_VSA_POS = 152.49 nHBa = 2 TPSA = 38.91 logP = 1.82	Y-G19 logBB_actr = -1.15 PEOE_VSA_POS = 127.45 nHBa = 2 TPSA = 103.31 logP = 1.19	logBB_actr = -0.18 PEOE_VSA_POS = 136.56 nHBa = 2 TPSA = 77.29 logP = 1.82
	logBB_actr = 0.176 PEOE_VSA_POS = 177.96 nHBa = 2 TPSA = 59.14 logP = 4.44	Y-G20 logBB_actr = -1.4 PEOE_VSA_POS = 142.02 nHBa = 2 TPSA = 43.84 logP = 0.06

Fig. 5. Examples of compounds in the first external evaluation set with small (**a**) or large (**b**) prediction errors. Structural fragments in bold highlight the aromatic N atoms. *PEOE_VSA_POS*: total positive van der Waals surface area, *nHBa*: number of H-bond acceptors, *TPSA*: polar surface area, *logP*: log of the octanol/water partition coefficient.

neighbors, has been reported as an H1 receptor agonist (55). Meanwhile, antihistamine drugs have been proven to be P-gp substrates (56). Therefore it is possible that Y-G20 is also a substrate of P-gp. This hypothesis may explain why Y-G20 has much lower logBB value than its nearest neighbors. Similarly, Y-G19 may also be the substrate of P-gp. This analysis underlies the complexity of building global QSAR models of BBB penetration for diverse chemical compounds with multiple possible mechanisms of actions. Thus in statistical terms, Y-G19 appears as an activity outlier. After excluding this outlier, the statistical parameters for the consensus prediction of the first external evaluation set improved from $R^2 = 0.73$, MAE = 0.36 to $R^2 = 0.80$, MAE = 0.29.

The consensus prediction accuracy of the six types of models for the second and third external evaluation sets were 86.5% and 80.9%, respectively, which is a little lower than the accuracies of the best individual models (cf. Fig. 1). However, according to our experience the consensus prediction is more reliable and stable (57). The use of the AD can also help increasing the accuracy of the consensus prediction. Applying the consensus prediction without any AD threshold decreased the prediction accuracy for the second and third external evaluation sets to 67.7% and 55.1%, respectively (Fig. 1).

Interpretation of Predictive QSAR Models

Several descriptors were found to be most frequently used in acceptable models, suggesting that they may play a critical role in defining BBB permeability of organic compounds. The top ten descriptors of each type used in the *k*NN modeling approach are shown in Table IV, along with their frequencies of occurrence in acceptable models and their interpretation (Table V). Although many different types of descriptors were employed for model development, three descriptors were found to be used most frequently in the six types of acceptable models: PSA, the octanol/water partition coefficient (logP), and the number of hydrogen bond donor and acceptor atoms. PSA has been established as an important descriptor for drug transport properties such as BBB penetration (22,58). Van de Waterbeemd *et al.* (23,59) found that the upper limit for PSA in a molecule that is expected to penetrate into the brain is around 90 Å^2. Moreover, it is known that relatively lipophilic drugs can cross the BBB by passive diffusion, which is influenced by their H-bonding capacity (22). The Iyer group has reported that logBB increases if logP of the solute molecule increases (58). Furthermore, active transporters are known to facilitate the BBB penetration of polar molecules. Therefore the hydrogen bond donor and acceptor counts, which may influence the binding affinity between membrane transporters and organic compounds, have been also recognized as important factors affecting the BBB permeability.

Figure 6 shows how BBB+ and BBB− compounds can be distinguished based on the values of the most frequent descriptors. Obviously, descriptors with larger differences in their values for BBB+ *vs.* BBB− compounds should be underlined since they may influence the BBB permeability of drugs. Besides these three obviously important descriptors (discussed above), *E*-state indices in MolconnZ and van der Waals surface areas (VSA) descriptors in MOE do perform well in separating BBB+ from BBB− compounds. Some fragments descriptors also work well, e.g., nArNHR (number of secondary aromatic amines, positive contribution to logBB values) and C–040(R–C(=X)–X/R–C#X/X=C=X fragments count, negative contribution).

Five of the top ten most frequently used MolconnZ descriptors belong to the classes of atom-level *E*-state, atom type *E*-state and global *E*-state indices, indicating that *E*-state descriptors serve as structural features that define the BBB penetration. The *E*-state indices have been used to develop models for many types of biological activities and physical properties (60). It needs to be pointed out that the *E*-state descriptors are not just counts of fragments. They also reflect the influence of the intermolecular environment. Therefore they contain more information than those descriptors that are based on the presence/absence or simple count of a given fragment (61). The information encoded in the *E*-state value for an atom or an atom type is its electron accessibility, which is important in non-covalent intermolecular interaction, such

Table IV. Most Frequently Used Descriptors in Validated QSAR Models
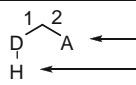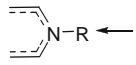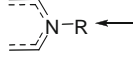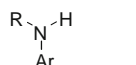
Descriptor name	Frequency rate (%)	Interpretation
MolConnZ descriptors		
Qsv	22.9	Average polarity descriptor
SHBint2*	21.2	Internal H-bond counts and *E*-states
Qv	19.9	A whole molecule polarity index that decreases in value as the polarity increases
nHBa	18.1	Number of strong H bond acceptors
SaaN*	16.7	Atom-type *E*-state sums
naaN*	16.5	Atom-type counts
Gmax	13.8	The maximum atom level *E*-state value in a molecule
nN	13.4	Number of element N
nHBd	13.3	Number of strong H bond donors
SssCH2*	12.4	Atom type E state indices
Dragon descriptors		
TPSA(Tot)	40.3	Topological polar surface area using N, O, S, P polar contributions
T (N..O)	34.1	Sum of topological distances between N..O
MLOGP2	34.0	Squared Moriguchi octanol–water partition coefficient ($logP^2$)
C-040	33.8	R–C(=X)–X/R–C#X/X=C=X fragments count
nArNHR*	33.2	Number of secondary amines (aromatic)
BLTF96	29.8	Verhaar model of Fish base-line toxicity from MLOGP (mmol/l)
nHAcc	27.7	Number of acceptor atoms for H-bonds (N, O, F)
nOHt	26.4	Number of tertiary alcohols
nCrs	25.4	Number of ring secondary C (sp3)
nHDon	24.5	Number of donor atoms for H-bonds (N and O)
MOE descriptors		
TPSA	57.1	Polar surface area
a_don	55.3	Number of hydrogen bond donor atoms
PEOE_VSA-1	50.6	Partial charge descriptors
a_nF	45.9	Number of fluorine atoms
SlogP_VSA6	42.9	Subdivided surface areas
lip_acc	37.7	The number of O and N atoms
PEOE_VSA+5	37.7	Partial charge descriptors
logP	32.9	Log of the octanol/water partition coefficient
GCUT_SLOGP_2	30.6	Adjacency and distance matrix descriptors
PEOE_VSA_POS	30.0	Fractional positive van der Waals surface area

Descriptors in italics are those used most frequently in all models. Descriptors with asterisk are illustrated in Table V.

as drug-receptor interactions, partition, vaporization, and solubility (60). Therefore, their significance may be related to the binding affinity of compounds with efflux transporters or the passive diffusion capability.

The importance of *E*-state descriptors in predicting BBB penetration has been reported in other studies. The Hall

Table V. Illustrations of Some Descriptors

|		
SHBint2	Sum of E-state of strength for potential hydrogen bonds if separated by 2 skeletal bonds.	Product of the E-state and HE-state values
SaaN	Sum of E-States for atoms of this type: aan=2 aromatic bonds to N.	Bond type E-state value
naaN	Number of atoms of this type: aan=2 aromatic bonds to N.	Bond type count
SssCH2	Sum of E-state for methylenes.	Bond type E-state value
nArNHR	Number of secondary amines (aromatic).	
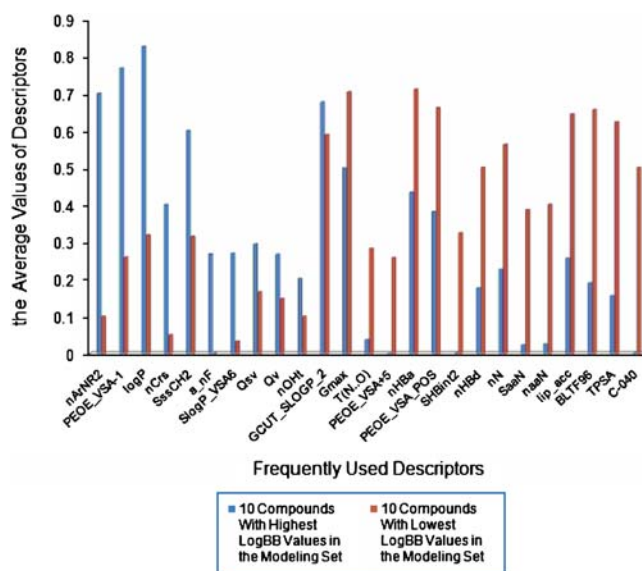
Fig. 6. Comparison of the (normalized) values for frequently used descriptors for compounds classified as BBB+ or BBB−.

group used MDL QSAR to model BBB partitioning and found that the SaaN descriptor (sum of *E*-states for N with 2 aromatic bonds) has a negative correlation with logBB values (62), which agrees with our observations (Fig. 6). It is possible that this type of atomic bond may influence the potential of the compound to be the substrate of P-gp, considering that many pharmacophore analyses of P-gp substrates mentioned the importance of nitrogen and amines (63). Hall *et al.* also identified the Qv descriptor, which is a molecule polarity index. It is reported to have a positive correlation with logBB values and this also agrees with our observations (Fig. 6) (62).

Five out of the ten most frequently used MOE descriptors are related to VSA. PSA (the sum of the VSA of oxygen atoms, nitrogen atoms, and attached hydrogen atoms in a molecule) was used in half of the QSAR models. Besides, PEOE_VSA-1, SlogP_VSA6, PEOE_VSA+5, and PEOE_VSA_POS, which are partial VSA descriptors, are also frequently used (cf. Table IV for details). These descriptors are the sum of the atomic VSA contributions of each atom within a certain range of a specific property. SlogP_VSA descriptors capture the hydrophobic and hydrophilic effects that are important for binding to the receptors. PEOE_VSA descriptors capture the direct electrostatic interactions. PEOE_VSA_POS is the total positive VSA (64). The Labute group developed a linear model of logBB as a function of PEOE_VSA, SlogP_VSA and SMR_VSA descriptors and the resulting R^2 was 0.83, indicating that VSA, not only PSA, should affect BBB permeability (64). VSA descriptors have several advantages: they are weakly correlated with each other; they are useful not only for physical property modeling but also in receptor affinity modeling; they are unlike the "whole molecule" properties that cannot distinguish the details of important substructural differences (64).

Litman *et al.* reported that the binding affinity of drugs to P-gp ATPase is highly correlated with VSA rather than log*P* (65), which suggests that binding between drugs and P-gp takes place across a wide interaction surface of the protein. Therefore we may conclude that compounds with higher VSA are more likely to be the substrates of P-gp; thus they probably have lower logBB values. The analysis of VSA descriptors identified in our QSAR models as significant has shown that not all partial VSA descriptors have a negative contribution to logBB values. Some of them, such as PEOE_VSA-1, have a positive contribution (Fig. 6).

Some other descriptors also have been proven to be important in predicting the BBB permeability of drugs, such as lip_acc (Lipinski's number of H-bond acceptors, such as N and O). Norinder *et al.* used the number of N and O atoms as an index of BBB penetration: if N+O is five or less in a molecule, it has a high chance of entering the brain; if log*P*−(N+O)>0, then logBB is positive (23). We can also find in Fig. 6 that lip_acc has a negative correlation with logBB values and log*P* has a positive correlation. Pharmacophore analysis of P-gp substrates indicated that the chlorine or fluorine substitutions on aromatic rings increase binding capability with P-gp (63, 66), but it disagrees with our observation that a_nF has a positive contribution to logBB values (Fig. 6). The reason may be that increasing the number of aromatic fluorine atoms would affect log*P*, VSA and other descriptors, which cannot be captured by pharmacophore models.

This analysis of frequent descriptors suggests that statistical QSAR models using descriptors derived from two-dimensional molecular topology are capable of providing meaningful interpretations concerning possible mechanisms of BBB penetration. Nevertheless, none of the descriptors could explain the observed distribution of biological data independently. Thus, we stress that each individual model was based on the use of many descriptors concurrently. We have uncovered the important contributions of several obvious descriptors of shape and surface area that are easy to interpret and that have been established as significant determinants of BBB penetration by other scientists. However, we have also established that less obvious (and more difficult to interpret) descriptors such as *E*-state indices also play a significant role in ensuring the high overall predictive power of our models.

CONCLUSIONS

We have applied a combinatorial QSAR approach to a dataset of 159 organic compounds with known logBB values. 144 of the 159 compounds were selected randomly as a modeling set and the remaining 15 were considered as the first external evaluation set. The resulting QSAR models were used further to evaluate BBB values for two external datasets of compounds classified experimentally as permeable (BBB+) or non-permeable (BBB−). The second external evaluation set included 99 commercially available drugs, and the third external evaluation set included 267 organic compounds. High prediction accuracy of training set models was demonstrated for all three external sets; however, the uneven distribution of BBB+/BBB− compounds in the modeling set led to the much higher prediction ability of the models for BBB+ compounds *vs*. BBB− compounds. The consensus prediction using all statistically significant training set models was found to provide a better balance between prediction accuracy and chemical space coverage than any individual model.

The analysis of the most frequent descriptors implicated in statistically significant and externally predictive statistical QSAR models afforded model interpretation in terms of chemical features influencing the BBB permeability. Some well-known descriptors such as PSA, log*P*, and the number of H-bond donor/acceptor atoms were found to dominate the models. However, additional molecular descriptors such as *E*-state indices and VSA descriptors were also found to contribute to statistically significant and externally predictive models. These descriptors could be associated with passive diffusion and bind affinity to efflux transporters.

In summary, we have compiled arguably the largest publicly available dataset of diverse organic molecules with experimentally available logBB values. We have shown that using the combinatorial QSAR approach and consensus prediction it is possible to build BBB permeability models with high external predictive accuracy. This analysis of resulting models in terms of significant chemical descriptors may facilitate the further exploration of the factors that influence the drug distribution between the bloodstream and the brain. These models can be used for estimating the BBB penetration of drug candidates at the early stages of drug discovery projects and exploring their intrinsic penetration mechanisms.

ACKNOWLEDGMENTS

REFERENCES

1. P. L. Golden, and G. M. Pollack. Blood–brain barrier efflux transport. *J. Pharm. Sci.* **92**:1739–1753 (2003).
2. U. Bickel, T. Yoshikawa, and W. M. Pardridge. Delivery of peptides and proteins through the blood–brain barrier. *Adv. Drug Deliv. Rev.* **46**:247–279 (2001).
3. C. L. Graff, and G. M. Pollack. Drug transport at the blood–brain barrier and the choroid plexus. *Curr. Drug Metab.* **5**:95–108 (2004).
4. R. C. Young, R. C. Mitchell, T. H. Brown, C. R. Ganellin, R. Griffiths, M. Jones, K. K. Rana, D. Saunders, I. R. Smith, N. E. Sore, and T. J. Wilks. Development of a new physicochemical model for brain penetration and its application to the design of centrally acting H2 receptor histamine antagonists. *J. Med. Chem.* **31**:656–671 (1988).
5. M. H. Abraham, H. S. Chadha, and R. C. Mitchell. Hydrogen bonding. 33. Factors that influence the distribution of solutes between blood and brain. *J. Pharm. Sci.* **83**:1257–1268 (1994).
6. M. H. Abraham, H. S. Chadha, and R. C. Mitchell. Hydrogen-bonding. Part 36. Determination of blood brain distribution using octanol–water partition coefficients. *Drug Des. Discov.* **13**:123–131 (1995).
7. F. Lombardo, J. F. Blake, and W. J. Curatolo. Computation of brain–blood partitioning of organic solutes via free energy calculations. *J. Med. Chem.* **39**:4750–4755 (1996).
8. G. Subramanian, and D. B. Kitchen. Computational models to predict blood–brain barrier permeation and CNS activity. *J. Comput. Aided Mol. Des.* **17**:643–664 (2003).
9. D. E. Clark. 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 (1999).
10. J. M. Luco. Prediction of the brain–blood distribution of a large set of drugs from structurally derived descriptors using partial least-squares (PLS) modeling. *J. Chem. Inf. Comput. Sci.* **39**:396–404 (1999).
11. M. Feher, E. Sourial, and J. M. Schmidt. A simple model for the prediction of blood–brain partitioning. *Int. J. Pharm.* **201**:239–247 (2000).
12. J. Kelder, P. D. Grootenhuis, D. M. Bayada, L. P. Delbressine, and J. P. Ploemen. Polar molecular surface as a dominating determinant for oral absorption and brain penetration of drugs. *Pharm. Res.* **16**:1514–1519 (1999).
13. M. E. Brewster, E. Pop, M. J. Huang, and N. Bodor. AM1-based model system for estimation of brain/blood concentration ratios. *Int. J. Quantum Chem.* **60**:51–63 (1996).
14. U. Norinder, P. Sjoberg, and T. Osterberg. Theoretical calculation and prediction of brain–blood partitioning of organic solutes using MolSurf parametrization and PLS statistics. *J. Pharm. Sci.* **87**:952–959 (1998).
15. A. Tropsha, P. Gramatica, and V. K. Gombar. The importance of being earnest: Validation is the absolute essential for successful application and interpretation of QSPR Models. *QSAR Comb. Sci.* **22**:69–77 (2003).
16. H. Kubinyi, F. A. Hamprecht, and T. Mietzner. Three-dimensional quantitative similarity–activity relationships (3D QSiAR) from SEAL similarity matrices. *J. Med. Chem.* **41**:2553–2564 (1998).
17. A. Golbraikh, and A. Tropsha. Beware of q2!. *J. Mol. Graph. Model.* **20**:269–276 (2002).
18. M. Vracko, V. Bandelj, P. Barbieri, E. Benfenati, Q. Chaudhry, M. Cronin, J. Devillers, A. Gallegos, G. Gini, P. Gramatica, C. Helma, P. Mazzatorta, D. Neagu, T. Netzeva, M. Pavan, G. Patlewicz, M. Randic, I. Tsakovska, and A. Worth. Validation of counter propagation neural network models for predictive toxicology according to the OECD principles: A case study. *SAR QSAR Environ. Res.* **17**:265–284 (2006).
19. L. P. de Cerqueira, A. Golbraikh, S. Oloff, Y. Xiao, and A. Tropsha. Combinatorial QSAR modeling of P-glycoprotein substrates. *J. Chem. Inf. Model.* **46**:1245–1254 (2006).
20. A. Kovatcheva, A. Golbraikh, S. Oloff, Y. D. Xiao, W. Zheng, P. Wolschann, G. Buchbauer, and A. Tropsha. Combinatorial QSAR of ambergris fragrance compounds. *J. Chem. Inf. Comput. Sci.* **44**:582–595 (2004).
21. A. Kovatcheva, A. Golbraikh, S. Oloff, J. Feng, W. Zheng, and A. Tropsha. QSAR modeling of datasets with enantioselective compounds using chirality sensitive molecular descriptors. *SAR QSAR Environ. Res.* **16**:93–102 (2005).
22. B. Hemmateenejad, R. Miri, M. A. Safarpour, and A. R. Mehdipour. Accurate prediction of the blood–brain partitioning of a large set of solutes using ab initio calculations and genetic neural network modeling. *J. Comput. Chem.* **27**:1125–1135 (2006).
23. U. Norinder, and M. Haeberlein. Computational approaches to the prediction of the blood–brain distribution. *Adv. Drug Deliv. Rev.* **54**:291–313 (2002).
24. J. A. Platts, M. H. Abraham, Y. H. Zhao, A. Hersey, L. Ijaz, and D. Butina. Correlation and prediction of a large blood–brain distribution data set—an LFER study. *Eur. J. Med. Chem.* **36**:719–730 (2001).
25. A. Golbraikh, M. Shen, Z. Xiao, Y. D. Xiao, K. H. Lee, and A. Tropsha. Rational selection of training and test sets for the development of validated QSAR models. *J. Comput. Aided Mol. Des.* **17**:241–253 (2003).
26. M. Olah, M. Mracec, L. Ostopovici, R. Rad, A. Bora, N. Hadaruga, I. Olah, M. Banda, Z. Simon, M. Mracec, and Y. I. Oprea. *WOMBAT: World of Molecular Bioactivity, in chemo-informatics in drug discovery.* Wiley-VCH, New York, 2004.
27. H. Li, C. W. Yap, C. Y. Ung, Y. Xue, Z. W. Cao, and Y. Z. Chen. Effect of selection of molecular descriptors on the prediction of blood–brain barrier penetrating and nonpenetrating agents by statistical learning methods. *J. Chem. Inf. Model.* **45**:1376–1384 (2005).
28. L. B. Kier, and L. H. Hall. *Molecular connectivity in structure–activity analysis.* Wiley, New York, 1986.
29. L. B. Kier, and L. H. Hall. *Molecular connectivity in chemistry and drug research.* Academic Press, New York, 1976.
30. M. Randic. Characterization of molecular branching. *J. Am. Chem. Soc.* **97**:6609–6615 (1975).
31. L. B. Kier. A Shape index from molecular graphs. *Quant. Struct.—Act. Relat.* **4**:109–116 (1985).
32. L. B. Kier. Inclusion of symmetry as a shape attribute in Kappa-Index analysis. *Quant. Struct.—Act. Relatsh.* **6**:8–12 (1987).
33. L. H. Hall, and L. B. Kier. Determination of topological equivalence in molecular graphs from the topological state. *Quant. Struct.—Act. Relat.* **9**:115–131 (1990).
34. L. H. Hall, B. K. Mohney, and L. B. Kier. The electrotopological state: An atom index for QSAR. *Quant. Struct.—Act. Relat.* **10**:43–51 (1991).
35. L. H. Hall, B. K. Mohney, and L. B. Kier. The electrotopological state: Structure information at the atomic level for molecular graphs. *J. Chem. Inf. Comput. Sci.* **31**:76–82 (1991).
36. G. E. Kellogg, L. B. Kier, P. Gaillard, and L. H. Hall. E-state fields: Applications to 3D QSAR. *J. Comput. Aided Mol. Des.* **10**:513–520 (1996).
37. L. B. Kier, and L. H. Hall. *Molecular structure description: The electrotopological state.* Academic Press, New York, 1999.
38. L. B. Kier, and L. H. Hall. A differential molecular connectivity index. *Quant. Struct.—Act. Relat.* **10**:134–140 (1991).
39. M. Petitjean. Applications of the radius–diameter diagram to the classification of topological and geometrical shapes of chemical compounds. *J. Chem. Inf. Comput. Sci.* **32**:331–337 (1992).
40. H. J. Wiener. Structural determination of paraffin boiling points. *J. Am. Chem. Soc.* **69**:17–20 (1947).
41. J. R. Platt. Influence of neighbor bonds on additive bond properties in paraffins. *J. Chem. Phys.* **15**:419–420 (1947).
42. C. Shannon, and W. Weaver. *In mathematical theory of communication.* University of Illinois, Urbana, Illinois, 1949.

43. D. Bonchev, O. Mekenyan, and N. Trinajstic. Isomer discrimination by topological information approach. *J. Comput. Chem.* **2**:127–148 (1981).

44. A. T. Balaban. Five new topological indices for the branching of tree-like graphs. *Theor. Chim. Acta*. **53**:355–375 (1979).

45. A. T. Balaban. Highly discriminating distance-based topological index. *Chem. Phys. Lett.* **89**:399–404 (1982).

46. Talete s.r.l. Dragon. [5.4.2006]. 2007. Milan (Italy).

47. R. Todeschini, and V. Consonni. *Handbook of molecular descriptors*. Wiley, Weinheim (Germany), 2000.

48. W. Zheng, and A. Tropsha. Novel variable selection quantitative structure–property relationship approach based on the k-nearest-neighbor principle. *J. Chem. Inf. Comput. Sci.* **40**:185–194 (2000).

49. V. N. Vapnik. *In the nature of statistical learning theory*. Springer, New York, 2000.

50. J. R. Votano, M. Parham, L. M. Hall, L. H. Hall, L. B. Kier, S. Oloff, and A. Tropsha. QSAR modeling of human serum protein binding with several modeling techniques utilizing structure-information representation. *J. Med. Chem.* **49**:7169–7181 (2006).

51. A. Tropsha, and A. Golbraikh. Predictive QSAR modeling workflow, model applicability domains, and virtual screening. *Curr. Pharm. Des.* 13:3494–3504 (2007).

52. M. Shen, C. Beguin, A. Golbraikh, J. P. Stables, H. Kohn, and A. Tropsha. Application of predictive QSAR models to database mining: Identification and experimental validation of novel anticonvulsant compounds. *J. Med. Chem.* **47**:2356–2364 (2004).

53. L. Sachs. *Applied statistics: A handbook of techniques*. Springer, New York, 1984.

54. K. M. Mahar Doan, J. E. Humphreys, L. O. Webster, S. A. Wring, L. J. Shampine, C. J. Serabjit-Singh, K. K. Adkison, and J. W. Polli. Passive permeability and P-glycoprotein-mediated efflux differentiate central nervous system (CNS) and non-CNS marketed drugs. *J. Pharmacol. Exp. Ther.* **303**:1029–1037 (2002).

55. G. J. Durant, J. M. Loynes, and H. B. Wright. Potential histamine H2-receptor autagonists. 1. Aminoethylimidayo(1,2-a)pyridines and -imidayo(1,5-a)pyridines. *J. Med. Chem.* **16**:1272–1276 (1973).

56. P. D. Hansten, and R. H. Levy. Role of P-glycoprotein and organic anion transporting polypeptides in drug absorption and distribution—Focus on H-1-receptor antagonists. *Clin. Drug Investig.* **21**:587–596 (2001).

57. H. Zhu, A. Tropsha, D. Fourches, A. Varnek, E. Papa, P. Gramatica, T. Oberg, P. Dao, A. Cherkasov, and I. V. Tetko. Combinatorial QSAR modeling of chemical toxicants tested against tetrahymena pyriformis. *J. Chem. Inf. Model.* in press (2008).

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

59. W. H. van de, G. Camenisch, G. Folkers, J. R. Chretien, and O. A. Raevsky. Estimation of blood–brain barrier crossing of drugs using molecular size and shape, and H-bonding descriptors. *J. Drug Target*. **6**:151–165 (1998).

60. K. Rose, L. H. Hall, and L. B. Kier. Modeling blood–brain barrier partitioning using the electrotopological state. *J. Chem. Inf. Comput. Sci.* **42**:651–666 (2002).

61. J. R. Votano, M. Parham, L. H. Hall, L. B. Kier, S. Oloff, A. Tropsha, Q. Xie, and W. Tong. Three new consensus QSAR models for the prediction of Ames genotoxicity. *Mutagenesis*. **19**:365–377 (2004).

62. L. H. Hall, and L. B. Kier. MDL QSAR modeling blood–brain barrier partitioning. http://www.mdl.com/products/pdfs/MDLQSARreprint.pdf. 2002.

63. T. R. Stouch, and O. Gudmundsson. Progress in understanding the structure–activity relationships of P-glycoprotein. *Adv. Drug Deliv. Rev.* **54**:315–328 (2002).

64. P. Labute. A widely applicable set of descriptors. *J. Mol. Graphics Modell.* **18**:464–477 (2000).

65. T. Litman, T. Zeuthen, T. Skovsgaard, and W. D. Stein. Structure–activity relationships of P-glycoprotein interacting drugs: Kinetic characterization of their effects on ATPase activity. *Biochim. Biophys. Acta*. **1361**:159–168 (1997).

66. T. Suzuki, N. Fukazawa, K. San nohe, W. Sato, O. Yano, and T. Tsuruo. Structure–activity relationship of newly synthesized quinoline derivatives for reversal of multidrug resistance in cancer. *J. Med. Chem.* **40**:2047–2052 (1997).

67. G. M. Keseru, and L. Molnar. High-throughput prediction of blood–brain partitioning: a thermodynamic approach. *J. Chem. Inf. Comput. Sci.* 41:120–128 (2001).

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

69. X. L. Ma, C. Chen, and J. Yang. Predictive model of blood–brain barrier penetration of organic compounds. *Acta Pharmacol. Sin.* 26:500–512 (2005).

70. A. R. Katritzky, M. Kuanar, S. Slavov, D. A. Dobchev, D. C. Fara, M. Karelson, W. E. Acree Jr., V. P. Solov'ev, and A. Varnek. Correlation of blood–brain penetration using structural descriptors. *Bioorg. Med. Chem.* 14:4888–4917 (2006).

71. T. J. Hou, and X. J. Xu. ADME evaluation in drug discovery. 3. Modeling blood–brain barrier partitioning using simple molecular descriptors. *J. Chem. Inf. Comput. Sci.* 43:2137–2152 (2003).

72. D. Pan, M. Iyer, J. Liu, Y. Li, and A. J. Hopfinger. Constructing optimum blood brain barrier QSAR models using a combination of 4D-molecular similarity measures and cluster analysis. *J. Chem. Inf. Comput. Sci.* 44:2083–2098 (2004).

73. D. A. Winkler, and F. R. Burden. Modelling blood–brain barrier partitioning using Bayesian neural nets. *J. Mol. Graph. Model.* 22:499–505 (2004).