Predicting Penetration Across the Blood-Brain Barrier from Simple Descriptors and Fragmentation Schemes

Yuan H. Zhao,[†] Michael H. Abraham,*[,†] Adam Ibrahim,[†] Paul V. Fish,[‡] Susan Cole,[||] Mark L. Lewis,[||] Marcel J. de Groot,[#] and Derek P. Reynolds[§]

Department of Chemistry, University College London, London WC1H 0AJ, U.K., Departments of Discovery Chemistry, Pharmacokinetics & Drug Metabolism, and Computational Chemistry, Pfizer Global Research and Development, Ramsgate Road, Sandwich, Kent CT13 9NJ, U.K., and REYTEK, 11C Rothsay Road, Bedford MK40 3PP, U.K.

Received July 25, 2006

The ability to cross the blood brain barrier (BBB), sometimes expressed as BBB+ and BBB−, is a very important property in drug design. Several computational methods have been employed for the prediction of BBB-penetrating (BBB+) and nonpenetrating (BBB−) compounds with overall accuracies from 75 to 97%. However, most of these models use a large number of descriptors (67−199), and it is not easy to implement the models in order to predict values of BBB±. In this work, 19 simple molecular descriptors calculated from Algorithm Builder and fragmentation schemes were used for the analysis of 1593 BBB± data. The results show that hydrogen-bonding properties of compounds play a very important role in modeling BBB penetration. Several BBB models based on hydrogen-bonding properties, such as Abraham descriptors, polar surface area (PSA), and number of hydrogen bonding donors and acceptors, have been built using binomial-PLS analysis. The results show that the overall classification accuracy for a training set is over 90%, and overall prediction accuracy for a test set is over 95%.

1. INTRODUCTION

Passive diffusion through the blood-brain barrier (BBB) is a very important process in ADME studies (Absorption, Distribution, Metabolism, and Excretion). Log BB is a parameter commonly used to express the extent of a drug passing through the blood-brain barrier. BB is defined as the brain-blood concentration ratio of a compound at steady state. Although the measurement of the blood-brain distribution ratio of a compound is a very direct approach, the necessary experiments are expensive and time-consuming and constitute a major bottleneck for high-throughput screening of large molecular libraries.[1] Therefore, there have been a large number of in silico models for log BB reported in the literature and reviewed several times.[2−5] In particular, Abraham and Hersey[5] concluded that a number of models can reproduce log BB values to about 0.3−0.35 log units, as also shown very recently by Abraham et al.[6]

Although many log BB models have been developed, the number of compounds used to train log BB models is very limited.[5,6] The largest data set published so far is that of Abraham et al.[6] which contains 300 compounds, but many data sets that have been used are much smaller.[5,6] Therefore, log BB models developed from these limited data sets and used for high-throughput screening of drug compounds may result in large prediction errors, since many groups of compound will be outside the chemical space of the compounds used to set up the models.

BBB± is another property used to study the ability of compounds to cross the blood-brain barrier (BBB). Compounds which are able to cross the blood-brain barrier are denoted as BBB+, and compounds that have little ability to cross the BBB are denoted as BBB−. Cruciani et al.[7] examined 97 molecules and defined compounds as BBB+ if log BB was greater than 0.0 (concentration in the brain greater than concentration in blood), as BBB± if log BB was between 0.0 and −0.3, and as BBB− if log BB was less than −0.3. On the other hand, Li et al.[8] divided compounds into BBB+ and BBB− groups according to whether the log BB ratio was ≥ -1 or < -1, respectively. Using this definition, they defined 276 BBB+ and 139 BBB− compounds. The biggest data set of BBB± values published so far is that of Adenot and Lahana.[9] They used data on 1337 BBB crossing drugs (BBB+) and 355 BBB noncrossing drugs (BBB−) based on the CNS activity (Central Nervous System). CNS activity (CNS+) implies BBB permeation (BBB+), but the converse is not necessarily true. Some drugs that are CNS inactive (CNS−) may still cross the BBB and show no activity because they do not interact with any CNS targets.[9] In the class of CNS inactive drugs, some compounds may actually be nonpermeable, while others do enter the brain but do not exhibit any receptor binding.[10] Thus the two types of behavior are not necessarily equivalent and do not encode the same level of information. Identifying BBB+ from CNS activity is much easier than identifying BBB−. Identifying BBB− compounds is by far

* Corresponding author phone: +44 (0)20 7679 4639; fax: +44 (0)20 7679 7463; e-mail: m.h.abraham@ucl.ac.uk.
[†] University College London.
[‡] Department of Discovery Chemistry, Pfizer Global Research and Development.
[||] Department of Pharmacokinetics & Drug Metabolism, Pfizer Global Research and Development.
[#] Department of Computational Chemistry, Pfizer Global Research and Development.
[§] REYTEK.

10.1021/ci600312d CCC: $37.00
Published on Web 12/07/2006

the most difficult task, for the simple reason that, in most cases, this information is not available.[9] Adenot and Lahana[9] give the most detailed account available of classifying compounds as BBB+ and BBB−.

Several BBB± models have now been developed.[7−9,11,12] The overall accuracies from these models range from 75 to 97%. The number of molecular descriptors used in the models, and calculated from different computer programs, varies from 67 to as large as 199. A number of models gave reasonably good classification of BBB± values, but details are often not available and consequently the models cannot easily be tested or extended. The first aim of the present work is to calculate 19 simple molecular descriptors for the BBB± data set of drugs of Adenot and Lahana[9] and to use recursive partitioning and binomial-PLS for SAR (Structure−Activity Relationships) and QSAR analysis of the BBB± set. A second aim is to use the fragmentation scheme approach developed by PharmaAlgorithm, Inc.[13] in order to build fragmentation schemes, again based on the data set of Adenot and Lahana.[9]

COMPUTATIONAL METHODS

BBB± Data Sets. The BBB± data set used is that of Adenot and Lahana[9] from a CNS library. They presented 1336 BBB crossing drugs (BBB+), 360 BBB noncrossing drugs (BBB−), and 91 P-gp substrates (either BBB+ or BBB−). The data set contains 26 permanently charged compounds, and eight compounds were given both BBB+ and BBB− assignments. Structures for some compounds could not be found. All of these were removed to leave 1593 compounds, comprising 1283 BBB+ and 310 BBB− compounds, for analysis.

The selection of the training and test sets was carried out using the Kennard−Stone method.[14] Five hundred compounds were selected as the test set 1 (451 BBB+ and 49 BBB−) to leave 1093 compounds as the training set (832 BBB+ and 261 BBB−). The Kennard−Stone method selects the test and training sets according to the compound descriptors, so that test set 1 occupies the same chemical space as the training set. The compounds used as test set 2 were obtained from Li et al.,[8] from their list of 415 BBB± compounds (276 BBB+ and 139 BBB−), separated into BBB+ and BBB− groups according to whether the log BB ratio was ≥ -1 or < -1, respectively. In this data set, three compounds have the same structures but were given different names, and 15 compounds are permanently charged. After removing these 18 compounds, there remained 397 BBB± compounds (267 BBB+ and 130 BBB−) which we used as test set 2. The details of the compounds in the training set and test sets 1 and 2 are listed in Tables S1 and S2 of the Supporting Information, together with SMILES and BBB± (given as 1 and 0 for BBB+ and BBB−, respectively).

Molecular Descriptors. The molecular descriptors chosen in this study were based on the descriptors used in various log BB studies.[2−6] Many of the descriptors are related to hydrogen-bonding properties. These descriptors were calculated using the Algorithm Builder (version 1.7) program developed by PharmaAlgorithms, Inc.[13] After importing SMILES strings for 1593 compounds from Adenot and Lahana[9] and 397 compounds from Li et al.,[8] the descriptors shown in Table 1 were calculated.

Table 1. Definition of Molecular Descriptors

no.	symbol	definition
1	E	excess molar refraction
2	S	polarizability/dipolarity
3	A	overall hydrogen-bond acidity
4	B	overall hydrogen-bond basicity
5	V	McGowan molecular volume in (mL/mol)/100
6	MW	molecular weight
7	PSA	polar surface area
8	logP	calculated octanol/water partition coefficient
9	logD(7.4)	octanol/water partition coefficient at pH = 7.4
10	NHD	number of hydrogen bonding donors
11	NHA	number of hydrogen bonding acceptors
12	pK_a (acid)	pK_a for acid
13	pK_a (base)	pK_a for base
14	Iv	indicator variable for carboxylic acid
15	F^+	positively charged form fraction at pH = 7.4
16	F^-	negatively charged form fraction at pH = 7.4
17	$F^\pm$	zwitterion form fraction at pH = 7.4
18	F^0	neutral form fraction at pH = 7.4
19	NRB	number of rotatable bonds

A number of the descriptors have the same, or similar, meaning. For example, both **A** and the number of H-bond donors express the ability of a molecule to act as a hydrogen bond acid. An analysis was carried out to see which descriptors were highly correlated. It was found that **A** is highly correlated to the number of H bond donors (R=0.89) and PSA (R=0.79), and **B** is highly correlated to the number of H bond acceptors (0.94) and PSA (0.89). **S** is correlated to the number of H bond acceptors (0.87). Multiple regression analysis showed that PSA is highly related to **A** and **B** (R=0.92) and to the number of H bond acceptors and donors (0.96). **V** is highly correlated to MW (R=0.98). Log D (7.4) is calculated from log P, and they are intercorrelated with R = 0.88. Descriptors 12−19 are not correlated to each other nor with descriptors 1−11, but descriptors 15−18 are calculated from descriptors 12−13. Because of the high correlations between Abraham descriptors (descriptors 1−5 in Table 1) and other hydrogen bond descriptors (PSA or number of H-Bond acceptors and donors), our statistical analysis is carried out either with the Abraham descriptors or with the other descriptors.

Fragmentation Schemes. Structures of the 1593 compounds were fragmented and listed in a matrix by the following methods implemented in Algorithm Builder.

- Atom-based: splits all structures in a data set into constituent atoms.
- IC-based: IC based fragmentation splits all structures in a data set into functional group fragments and isolating carbons. In addition, it allows the generation of tables that display interactions between various functional groups.
- Chain-based: splits all structures in a data set into Klopman-type scaffolds, i.e., heavy-atoms and branched chains as well as heterocycles.
- Scaffolding-based: splits all structures in a data set into scaffold fragments.

Statistical Analysis. There are several statistical methods available in the program of Algorithm Builder. If BBB+ is defined as 1 and BBB− is defined as 0, recursive partitioning and binomial-PLS methods can be applied for the analysis of BBB± in terms of molecular descriptors or fragment variables.

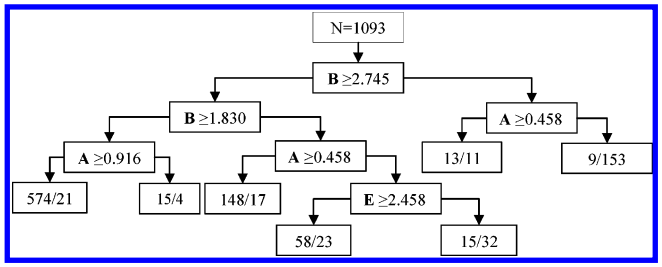

Figure 1. Decision tree with Abraham descriptors. Flow direction: BBB+ is to LHS; BBB− is to RHS. The number at the end of the tree is the number of BBB+/BBB−.

2. RESULTS AND DISCUSSION

2.1. BBB± Models from Abraham Descriptors (Descriptors 1−5 in Table 1). A linear free energy relationship (LFER) has been developed by Abraham et al.[6] for the correlation and prediction of log BB values for 302 compounds with a regression coefficient $R = 0.87$ and a standard error SD = 0.30; use of a test set and training set showed that the predictive capability of the LFER was around 0.30 log units.

Because compounds which can cross the blood-brain barrier (BBB) are defined as BBB+ and compounds that have little ability to across BBB are defined as BBB−, it is reasonable to believe that Abraham descriptors can be used in modeling BBB±. We first used a recursive partitioning (RP) approach for data classification using the Abraham descriptors. It is a tree-based method, can easily handle nonlinear stepwise variable selection and complexity reduction, and is extremely robust with respect to outliers.[13] In the RP method, the splitting decision was made according to the t-test value. Two termination criteria were applied: the Bonferroni probability and the node size; splitting stops if (a) for further splits, the adjusted Bonferroni probability exceeds 0.01 or (b) if the generated node size is less than five. The adjusted Bonferroni probability is the t-test *p*-value multiplied by the number of conditions tested.[15] Figure 1 shows the results from recursive partitioning approach based on Abraham descriptors for 1093 compounds in the training set. The BBB± values, as 1 and 0, and calculated molecular descriptors, together with the SMILES for all of compounds, are listed in Table S1 of the Supporting Information.

In the RP approach, four descriptors from the five Abraham descriptors were selected. The size descriptor, as the McGowan volume **V**, was not an important descriptor and was not selected from the decision tree. This result is in agreement with the finding of Crivori et al.[11] that size and shape descriptors have no marked impact on BBB permeation. Results of the RP method for the classification of BBB± are given in Table 2. The overall correctness of classification is very good, being 90.6% for the training set and 96.8% for the predictions of the test set 1. Of the other schemes shown in Table 2[8,9,11,12] only that of Adenot and Lahana[9] gives a comparable result, with overall correctness of classification being 97.0% for the training set and 91.0% for the test set. However, Adenot and Lahana[9] used a much larger number of descriptors (67) for some 1500 BBB± compounds, as compared to our analysis that uses only four descriptors for 1593 compounds. Hydrogen bond basicity and acidity are very important descriptors in the BBB decision tree (Figure 1), in agreement with results of other workers on the modeling of log BB.[5,6]

In order to study further the effect of Abraham descriptors on the BBB penetration, binomial-PLS analysis was carried out on the 1093 BBB± data set. The result is that the classification accuracy of binomial-PLS using the five Abraham descriptors (no. 6 in Table 2) is quite similar to the accuracy of recursive partition (no. 5 in Table 2), with an overall 90% classification accuracy. A similar result was found if the volume descriptor (**V**) was removed, as found also in the recursive partition analysis. On examination of the increments of the Abraham descriptors from the binomial-PLS, we find that the biggest effect on BBB± comes from the **A** and **B** terms. Binomial-PLS analysis shows that the overall classification accuracy from just the **A** and **B** terms is 90.5% (no. 7 in Table 2), only 1% less than the accuracy from the model that uses all five descriptors (no. 6 in Table 2). It should be noted that the equations in Table 2 obtained by the binomial-PLS method cannot be compared with reported equations for log BB.[6] The predicted values from the binomial-PLS equations can be converted into probabilities by use of the formula: probability = $1/[1+\exp(-x)]$ where x is the predicted value. The cutoff in the probability values to distinguish BBB+ and BBB− was taken as 0.7 for all the models.

In Figure 2 is shown the histogram for 1093 BBB± compounds in the training set for model 7 in Table 2. It clearly shows that BBB+ is on the right-hand side and BBB− is on the left-hand side. The cut off value that was used in the model is 0.7. If the cutoff is moved, the classification accuracy for BBB+ and BBB− will be affected but not the overall accuracy.

2.2. BBB± Models Using Hydrogen-Bonding Descriptors (Nos. 6−19 in Table 1). Descriptors related to hydrogen bonding, such as **PSA, log P, ND**, and **NA**, have been used in log BB predictions.[2−6] In Figure 3 is shown the recursive partitioning tree for descriptors nos. 6−19 in Table 1. From these descriptors, five were selected as the significant ones **(PSA, Iv, NRB, NHD**, and $\mathbf{F}^+$**).** Four of these descriptors are related to the hydrogen bond properties of the compounds, the other one being the number of rotatable bonds (**NRB**). Again, the size descriptor, this time as the molecular weight, is not selected by the decision tree.

Hou and Xu[16] found that molecular weight makes a negative contribution to log BB when the molecular weight is larger than 360. They attributed this to the tight junctions in the BBB that limit the size of molecules that can cross the membrane. However, other workers have not reported any similar observations, and for BBB± all workers are agreed that the size of the compounds does not play any important part. For the 1593 compounds that we have studied, the average MW for BBB+ (320) is much lower than that for BBB− (514). However, for compounds with molecular weight larger than 500 (Figure 4) there are still 47 compounds with BBB+. It seems reasonable, therefore, that the size descriptor was not selected by the RP decision trees.

The results shown in Figure 3 indicate that the fraction of positive form of a compound does not have a significant impact on the classification of BBB±. This can be demonstrated from a binomial-PLS analysis for these descriptors. The accuracy from PSA and NHD + NHA (nos. 10 and 11) is slightly lower than from the five descriptors selected in the recursive partitioning approach (no. 9). We checked the

Table 2. Classification Accuracy of Models for BBB± from Previous Studies and in This Work[a]

			training set accuracy (%)				test set accuracy (%)			
no.	method	des[b]	no. of BBB±	+	−	overall	no. of BBB±	+	−	overall
			Previous Work							
1	Trotter[12]	72	81/91	NA	NA	NA	256/48	78.9	60.4	76.0
2	Li[8]	199	276/139	88.6	75.0	83.7	NA	NA	NA	NA
3	Crivori[11]	72	46/64	NA	NA	NA	49/71	90.0	65.0	74.8
4	Adenot[9]	67	NA[c]	97.0	85.0	95.0	NA	90.0	92.0	91.0
			This Work, Total Set of 1593 Compounds (1093 in Training and 500 in Test Set)							
5	Abraham	R4	832/261	95.5	75.1	90.6	451/49	99.8	69.4	96.8
6	Abraham	5	832/261	94.4	81.6	91.3	451/49	98.2	75.5	96.0
7	A+B	2	832/261	93.6	80.5	90.5	451/49	96.5	73.5	94.3
8	PSA+Iv+NHD+F^++NRB	R5	832/261	92.7	85.8	91.0	451/49	97.8	77.6	95.8
9	PSA+Iv+NHD+F^++NRB	5	832/261	95.0	82.0	91.9	451/49	98.4	79.6	96.8
10	PSA	1	832/261	93.8	82.0	90.9	451/49	98.4	77.6	96.2
11	NHD+NHA	2	832/261	93.3	79.7	90.0	451/49	97.3	65.3	94.2
12	chain	69	832/261	98.0	94.3	97.1	451/49	98.2	87.8	97.2
			This Work, Total Set of 1990 Compounds (1593 in Training and 397 in Test Set)							
13	Abraham	5	1283/310	95.9	80.0	92.8	267/130	81.6	69.2	77.6
14	A+B	2	1283/310	94.9	77.7	91.6	267/130	80.5	79.2	80.1
15	PSA	1	1283/310	95.7	80.6	92.8	267/130	82.0	58.5	74.3
16	NHD+NHA	2	1283/310	94.7	77.4	91.3	267/130	81.6	73.1	78.8
17	chain	77	1283/310	99.3	87.4	97.0	267/130	80.1	63.1	74.6

[a] NA: no data available. R: recursive partitioning models. Other models are binomial-PLS, nos. 5−19. [b] Number of descriptors used in the models. [c] Adenot and Lahana used 1336 BBB+ and 360 BBB− compounds but excluded 201 compounds for which they were unable to compute electronic parameters. This leaves a total of 1495 compounds, but the exact number of BBB+ and BBB− compounds is not known.

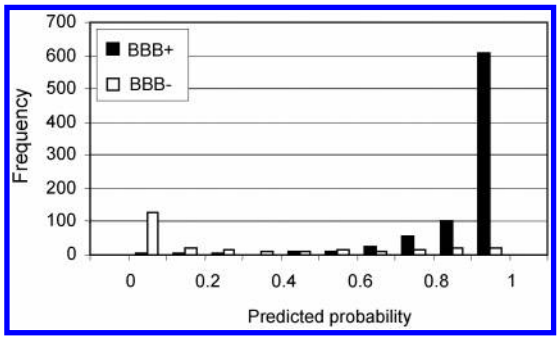

Figure 2. Histogram of the training set for BBB± from model no. 7 in Table 2.

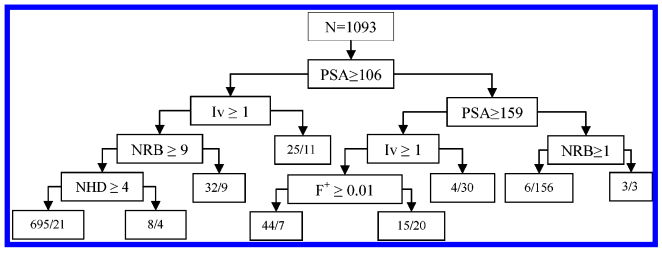

Figure 3. Decision tree from descriptors 6−19 in Table 1. Flow direction: BBB+ is to LHS; BBB− is to RHS. The number at the end of the tree is the number of BBB+/BBB−.

influence of charged species in compounds with ionizable groups by adding the descriptor $\mathbf{F}^+$ to model no. 10 (PSA). The classification accuracy is increased only from 90.9 to 91.3%, thus showing that $\mathbf{F}^+$ has no significant effect.

2.3. BBB± Models from Fragmentation Schemes. Fragmentation schemes are very popular methods for the prediction of physicochemical or biological activity. A typical example is the prediction of octanol/water partition coefficients by the ClogP program.[17] Usually, a fragmentation scheme program fragments structures based on functional groups, atoms or chains, and correlates the fragments with properties of chemicals by using multiple linear analysis (MLA) and Partial Least-Squares (PLS). The program,

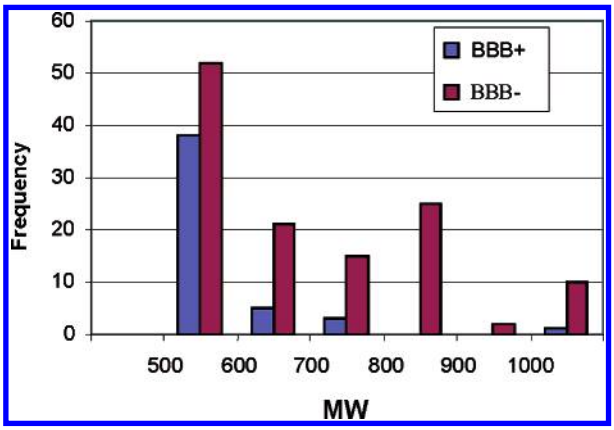

Figure 4. Histogram of molecular weight >500.

Algorithm Builder, released by PharmaAlgorithms, Inc.[13] not only implements the fragmentation but also considers the interactions between the functional groups in a molecule. However, fragmentation methods need to be trained on a large data set, especially if a large number of fragments are specified. A model that uses only a limited number of compounds in a training set can be overtrained and can lead to quite erroneous predictions for compounds outside the training set.

There are several fragmentation schemes available in the Algorithm Builder program: atom-based, chain-based, chain-based (advanced), IC-based (Isolation Carbon), scaffold-based, and custom-based fragmentations. Fragment databases and matrices of variables for 1093 compounds in a training set and 1593 compounds in the total set were created using all of the schemes, except the custom-based fragmentation scheme. These variables were then used to correlate the BBB± compounds with the fragments.

After binomial-PLS analysis between BBB± and matrices from the fragmentation schemes, we found that the chain-based scheme gave the best classification accuracy for the

Table 3. Some Simple Equations Used in the Models Listed in Table 2[a]

no.	method	des	models
6	Abraham	5	6.823 − 1.207**E** + 0.780**S** − 2.798**A** −1.928**B** + 0.250**V**
7	A+B	2	6.570 − 2.768**A** − 1.788**B**
9	PSA+Iv+NHD+F^++NRB	5	6.648 − 0.0408**PSA** − 2.644**Iv** +0.0310**NRB**− 2.095$\mathbf{F}^+$ −0.155**NHD**
10	PSA	1	5.378 − 0.0407**PSA**
11	NHD+NHA	2	5.773 − 0.530**NHD** − 0.512**NHA**
13	Abraham	5	6.908 − 1.259**E** + 0.844**S** − 2.857**A** −2.018**B** + 0.296**V**
14	A+B	2	6.662 − 2.834**A** − 1.787**B**
15	PSA	1	5.630 − 0.0425**PSA**
16	NHD+NHA	2	5.981 − 0.531**NHD** − 0.537**NHA**

[a] Predicted values calculated from these equations were converted into probabilities by the formula [probability = 1/(1+exp(−*x*)) where *x* is the predicted value]. The cutoff in the probability values to distinguish BBB+ and BBB− was taken as 0.7 for all the models.

1093 compounds in the training set. The chain matrix contains 2662 possible fragments. Examination of the contribution in the matrix for each fragment shows that not all of the fragments make significant contributions. Therefore, reduction in the number of fragments was carried out, based on the increments from each fragment. The final models listed in Table 2 (see no. 12) from the chain fragmentation used only 66 fragments. Most of the fragments are related to hydrogen bond acid and hydrogen bond basic groups; these make negative contributions to BBB±.

2.4. Testing the BBB± Models. To test the BBB± models that are summarized in Table 3, we selected two published BBB± data sets. One consisted of 500 BBB± compounds from the data set of Adenot and Lahana[9] (The remaining 1093 compounds comprises the training set we used to build BBB± models.). The 500 BBB± compounds have the same cut off value as before, that serves to define BBB+ and BBB−. Another data set that we used as test set 2 was taken from Li et al.[8] The definition of BBB+ and BBB− here refers to the cut off value of log BB ≥ -1 or < -1, respectively. Tables S1 and S2 in the Supporting Information list the details of compounds with BBB± and include the SMILES for the compounds.

In Table 2 are given results of predictions for the 500 BBB± compounds in test set 1 and 397 BBB± compounds in test set 2 (nos. 13−17 in Table 2). It appears that prediction of BBB± is quite satisfactory, with a correct overall classification of about 95% for test set 1 and 80% for test set 2. Because the test set 1 that we used comes from same source as the training set, it is expected that the prediction accuracy would be better than the prediction accuracy for test set 2. The prediction accuracy for the Adenot and Lahana test set 1 (overall accuracy 95%) is better than that achieved by Adenot and Lahana themselves (overall accuracy 91%); the number of compounds in the test set used by Adenot and Lahana is not known. The overall accuracy for test set 2 is reasonable, by comparison to the accuracy of Li et al. model, which used 199 descriptors and led to a 84% classification accuracy (see Table 2) for the training set; no test set was used by Li et al.

3. CONCLUSIONS

After analyzing over 1500 BBB± compounds using 19 simple descriptors and various fragmentation schemes, it was found that BBB penetration can be predicted from hydrogen-bonding descriptors, such as hydrogen bonding acidity and basicity, polar surface area, and number of hydrogen bond donors and acceptors. Several models have been built from these simple descriptors (Table 3) and a chain based fragmentation scheme. The models were able to separate BBB penetration and nonpenetration compounds with over 90% accuracy. The present methods are very simple and can easily be repeated, which suggests that they are very suited as part of fast screening methods in drug design.

ACKNOWLEDGMENT

We are grateful to Pfizer Global Research and Development for a fellowship (to Y.H.Z.). We are pleased to thank a number of Pfizer staff who have contributed to this work: Lourdes Cucurull-Sanchez, Jens Loesel, and Han van de Waterbeemd for their primary support and to Alexander Alex, Julian Blagg, Barry Jones, and Anthony Wood for their advice.

Supporting Information Available: BBB± values for the compounds together with SMILES notation for each compound in a spreadsheet containing Tables S1 and S2. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Subramanian, G.; Kitchen, D. B. Computational models to predict blood- brain barrier penetration and CNS activity. *J. Comput.-Aided Mol. Des.* **2003**, *17*, 643−664.
(2) Norinder, U.; Haeberlain, M. Computational approaches to the prediction of the blood-brain distribution. *Adv. Drug Delivery. Rev.* **2002**, *54*, 291−313.
(3) Lobell, M.; Molnar, L.; Keserü, M. Recent advances in the prediction of blood-brain partitioning from molecular structure. *J. Pharm. Sci.* **2003**, *92*, 360−370.
(4) Goodwin, J. T.; Clark, D. E. In silico predictions of blood-brain barrier penetration: considerations to "keep in mind". *J. Pharmacol. Exp. Ther.* **2005**, *315*, 477−483.
(5) Abraham, M. H.; Hersey, A. In *Comprehensive Medicinal Chemistry*, 2nd ed.; Testa, B., van de Waterbeemd, H., Eds.; 2006; Vol. 5: ADMET/Property-based approaches.
(6) Abraham, M. H.; Ibrahim, A.; Zhao, Y. H.; Acree, W. E., Jr. A data base for partition of volatile organic compounds and drugs from blood/plasma/serum to brain, and an LFER analysis of the data. *J. Pharm. Sci.* **2006**, *95*, 2091−2100.
(7) Cruciani, G.; Pastor, M.; Guba, W. A new tool for the pharmacokinetic optimization of lead compounds. *Eur. J. Pharm. Sci.* **2000**, *2*, S29−S39.
(8) Li, H.; Yap, C. W.; Ung, C. Y.; Xue, Y.; Cao, Z. W.; Chen, Y. Z. 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*. **2005**, *45*, 1376−1384.
(9) Adenot, M.; Lahana, R. J. Blood-brain barrier permeation models: Discriminating between potential CNS and non-CNS drugs including P- glycoprotein substrates. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 239−248.
(10) Waterbeemd, H. van de; Camenisch, G.; Folkers, G.; Chretien, J. R.; Raevsky, O. A. Estimation of blood-brain barrier crossing of drug using molecular size and shape, and H-bonding descriptors. *J. Drug Targeting* **1998**, *6*, 151−165.

(11) Crivori, P.; Cruciani, G.; Carrupt, P. A.; Testa, B. Predicting blood-brain barrier permeation from three-dimensional molecular structure. *J. Med. Chem.* **2000**, *43*, 2204−2216.
(12) Trotter, M. W. B.; Buxton, B. F.; Holden, S. B. Support vector machines in combinatorial chemistry. *Meas. Control* **2001**, *34*, 235−239.
(13) PharmaAlgorithms. *ADME Boxes, Version 3.0*; PharmaAlgorithms, Inc.: 591 Indian Road, Toronto, ON M6P 2C4, Canada.
(14) Kennard, R. W.; Stone, L. A. Computer aided design of experiments. *Technometrics* **1959**, *11*, 137−148.
(15) Chen, X.; Rusinko, A., III; Young, S. S. Recursive partitioning analysis of a large structure-activity data set using three-dimensional descriptors. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 1054−1062.
(16) Hou, T. J.; Xu, X. J. ADME Evaluation in drug discovery. 3. Modelling blood-brain barrier partitioning using simple molecular descriptors. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 2137 −2152.
(17) Leo, A. J. *ClogP for Windows, Version 4.0*; BioByte Corp.: 201 West 4th St., Suite 204, Claremont, CA 91711, U.S.A., 2001.

CI600312D