ELSEVIER

International Journal of Pharmaceutics 201 (2000) 239–247

international journal of pharmaceutics

www.elsevier.com/locate/ijpharm

A simple model for the prediction of blood–brain partitioning

Miklos Feher *, Elizabeth Sourial, Jonathan M. Schmidt

Nanodesign Inc., Suite 300, Research Park Centre, 150 Research Lane, Guelph, Ont., Canada, NIG 4Y5

Received 5 November 1999; received in revised form 3 March 2000; accepted 4 April 2000

Abstract

We derived a simple model for the prediction of blood–brain barrier penetration using three descriptors. The model contains the calculated octanol–water partition coefficient, the number of hydrogen-bond acceptors in an aqueous medium and the polar surface area. It was validated using an extensive dataset, comprising 100 diverse drug molecules. The descriptors are easily calculated and the model is suitable for the rapid prediction of the blood-brain barrier partitioning of drugs. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Blood–brain barrier; QSAR; Drug transport; Partition coefficients; Polar surface area; Hydrogen bonding

1. Introduction

The knowledge of the penetration of drugs through the blood–brain barrier (BBB) is central to the design of new central nervous system (CNS) active drugs and in improving the side effect profile of drugs with peripheral activity. The prediction of this property is important, as the experimental determination of BBB penetration is difficult and costly. In addition, a rapid method for making decisions is required for scoring or prioritizing large combinatorial databases or the solutions from de novo methods.

Several attempts to correlate BBB transport with physico–chemical parameters, have been recently reviewed (Waterbeemd et al., 1998). It was found that the octanol–water partition coefficient (log P) is an important factor, although in itself it correlates poorly with log BB. In addition to log P, the importance of a molecular size descriptor has been shown (Levin, 1980; Kaliszan and Markuszewski 1996; Salminen et al., 1997), as well as the necessity of incorporating a descriptor relating to hydrogen bond formation (Abraham et al., 1994; Waterbeemd et al. 1998). Unfortunately, these earlier models were based on small datasets and were not validated using an external prediction set. On a larger dataset, Lombardo established a correlation between log BB and the solvation free energy (Lombardo et al., 1996). Norinder et al. obtained a 3-component PLS model for a similar set of compounds, using prin-

* Corresponding author. Tel.: +1-519-8239088; fax: +1-519-8239401.

E-mail address: mfeher@nanodesign.com (M. Feher)

PII: S0378-5173(00)00422-1

cipal components based on 14 descriptors, including quantum chemical ones (Norinder et al., 1998). More recently, Luco published a PLS model for the most extensive dataset to date that included 100 compounds, using three principal components based on 25 descriptors (Luco, 1999).

In summary, the models reported in the literature are either insufficiently predictive (Levin, 1980; Kaliszan and Markuszewski, 1996; Salminen et al., 1997) or contain the principal components of numerous physicochemical descriptors (Norinder et al., 1998; Luco, 1999) that cannot be easily and rapidly calculated for an arbitrary compound. Our aim in this study was to derive a simple model based on all available experimental data but containing only a few relevant physico–chemical properties that can be calculated rapidly for large databases and for a wide range of compounds.

2. Methods

The dataset of 100 compounds and their corresponding biological activities was taken from the literature (Luco, 1999). The division of compounds into a training set (61 compounds) and two prediction sets (14 and 25 compounds) was also taken from the same source (Luco, 1999). The modeling work and the principle component regression (PCR) fits were performed with the Molecular Operating Environment (MOE) program (MOE, 1999). In case of our three-descriptor model, these PCR fits were found to be equivalent to multiple linear regression (MLR) models. These models, derived on the training set, were cross-validated using the leave-one-out procedure and then tested on the external prediction sets. The applied conformations were produced using a high quality molecular builder in MOE and optimized using the Merck force field (Halgren, 1996), applying the appropriate force field charges (Halgren, 1996). In order to check the quality of these vacuum conformations, a limited conformational search in vacuum was undertaken with the Random Incremental Pulse Search (RIPS) algorithm (Ferguson and Raber, 1989). In these random searches the first 100 conformers were produced. In all cases, the conformations from the builder were low energy ones, within 2 kcal/mol of the global energy minimum. Therefore, conformational searches were deemed unnecessary for the purposes of this QSAR.

The log P and log D values were calculated using the ACD suite (ACD/Labs, 1998). The number of hydrogen-bond acceptors was obtained using the Patty rules that take into account the solvation state of functional groups (Bush and Sheridan, 1993). The polar surface area was estimated from the solvent-exposed area of the molecule (MOE, 1999) assuming a spherical solvent molecule with a radius of 1.4 Å and considering only those parts of the surface with the absolute value of the partial charge greater than 0.2.

3. Results

Models involving only lipophilicity and a molecular size descriptor (Levin, 1980; Kaliszan and Markuszewski, 1996) have been shown to break down when tested on an alternative dataset (Salminen et al., 1997). For example, a direct fitting of log P and molecular mass or volume using 26 compounds produced an r^2 of approximately 0.3. On the other hand, according to the work done in the present study, this latter model (Salminen et al., 1997), which was based on log P, I_3 (where I_3 is an indicator variable for the presence of amino and carboxylic groups) and van der Waals volume predicted log BB inadequately for compounds in the Luco dataset (Luco, 1999). Although fitting the model of Salminen et al. on our training set produced a reasonable correlation ($r^2 = 0.77$, $rmse = 0.39$), this model predicted log BB poorly for molecules in our test set 1 ($r^2 = 0.26$, $rmse = 0.88$). Hence, it was necessary to search for a more general model.

The importance of including log P, a hydrogen-bond descriptor and the polar surface area to model the penetration of drugs across the blood–brain barrier has been documented (Waterbeemd et al., 1998). The simplest model obtained in the present work, derived for 61 compounds, was a three-descriptor model using the calculated oc-

tanol–water partition coefficient (log P) of the compound, the polar surface area (A_{pol}) and the number of hydrogen-bond acceptors in an aqueous medium ($n_{acc,solv}$). The regression model derived for the training set was:

$$\log BB = 0.4275 - 0.3873\ n_{acc,solv} + 0.1092\ \log P - 0.0017\ A_{pol} \quad (1)$$

($n = 61$, $r^2 = 0.730$, $q^2 = 0.688$, $rmse = 0.424$, $F = 51$, $P < 0.001$).

The properties of our three-descriptor model are compared to those of the 25-descriptor, 3 PCA-component PLS model (Luco, 1999) in Table 1. Our three-descriptor model has only slightly worse performance for the training set and the two prediction sets. The observed and calculated activities and the descriptors are presented in Table 2 for the training set and Table 3 for the two prediction sets. The overall fit is presented graphically for the three datasets in Fig. 1.

The statistics presented above do not consider any molecules in the dataset as outliers. Obviously, removing outliers from the prediction sets greatly improves the apparent performance of the model. For example, the removal of molecules 62 and 63 from prediction set 1 (the same outliers that were removed in the reference (Luco, 1999)) improves the r^2 value from 0.58 to 0.86. Similarly, removing the two worst outliers from the training set, compounds 30 and 61 (cf. three outliers removed (Luco, 1999)) improves the fit considerably ($n = 59$, $r^2 = 0.795$, $rmse = 0.346$). However, lacking specific information to explain why these molecules behave as outliers, their exclusion from the model was not justified in the present study. Furthermore, the removal of these two compounds had no major effect on the predictions made for either test set.

Table 1
Properties of the statistical model[a]

	Present work	Reference (Luco, 1999)
Training set ($n = 61$)		
r^2	0.730	0.764
rmse	0.424	0.399
q^2 (leave-one-out)	0.688	
Prediction set 1 ($n = 14$)		
r^2	0.576	0.651
rmse	0.628	0.500
Prediction set 2 ($n = 25$)		
r^2	0.616	0.577
rmse	0.789	0.522

[a] The statistics were obtained for both models with no molecules being considered as outliers. The statistics for the reference (Luco, 1999) were calculated using the predicted and experimental values given there. The root mean square error was obtained by dividing the residual sum of squares by n (i.e. ignoring the degrees of freedom) as was done in all of the quoted references.

4. Discussion

The model described above is robust and was obtained without the omission of any points as outliers. Nevertheless, it is important to check the validity and uniqueness of the descriptors in the model. This was achieved in three ways: studying the interdependence of the descriptors, determining the effect of removing any of them and searching for possible descriptors to replace them.

The independence of the applied descriptors was checked by calculating the correlation matrix of the parameters in the final model. The result is displayed in Table 4. It can be seen from Table 4 that some correlation between the polar surface area and the number of hydrogen bond acceptors exists ($r^2 = 0.30$), as both depend on the size of the molecule. However, this correlation is far less than the $r^2 = 0.8$ that would be required for the variables to substitute for each other in the regression equation (Martin, 1978). Therefore we can treat the three descriptors as being largely independent in the model. Not too surprisingly, the removal of any of the descriptors significantly reduces the quality of the fit, to an r^2 below 0.5 from 0.73 in the training set.

A systematic study showed that it is possible to replace some of the descriptors used in Eq. (1) without significantly reducing the predictive power of the model, provided the replacement descriptors are highly correlated with the original ones. The ease of calculation may be a rationale

Table 2
Experimental and calculated blood–brain barrier penetration (log BB) and the computed descriptors [a] for the training set molecules

Compound[b]	Log BB exp.[c]	Log BB calculated	$n_{acc,solv}$	log P	A_{pol}
1 (Cimetidine)	−1.42	−1.52	4	0.36	254.39
2	−0.04	−0.59	2	0.39	164.06
3 (Lupitidine)	−1.06	−1.32	4	1.94	236.44
4 (Pyrilamine)	0.49	−0.16	2	3.26	98.36
5 (Imipramine)	0.83	0.82	0	4.47	53.36
6 (Tiotidine)	−0.82	−1.75	4	−0.08	361.52
7	−0.67	−0.70	3	3.30	191.40
8	−0.66	−0.83	3	2.27	198.63
9	−0.12	−0.57	3	4.26	176.87
10	−0.18	−0.43	2	1.69	153.77
11	−1.15	−0.68	2	0.41	219.73
12	−1.57	−1.04	3	0.55	212.14
13	−1.54	−1.58	4	0.74	312.85
14	−0.27	−0.70	3	3.25	186.74
15	−0.28	−0.76	3	2.90	197.93
16	−0.46	−0.32	2	2.08	112.89
17	−0.24	−0.04	2	3.99	72.14
18	−0.02	−0.20	2	2.82	90.65
19	0.69	−0.10	2	3.25	62.35
20	0.44	−0.04	2	3.71	56.05
21 (Zolantidine)	0.14	0.13	2	5.36	62.31
22	0.22	−0.36	3	4.77	83.17
23 (Icotidine)	−2.00	−1.55	5	2.27	167.52
24	−1.30	−0.66	3	4.20	220.88
25 (Clonidine)	0.11	0.07	1	1.54	78.60
26	−1.12	−0.86	3	2.69	242.44
27	−0.73	−0.61	3	4.68	223.23
28	−1.17	−1.03	3	0.72	218.00
29 (Ranitidine)	−1.23	−1.04	3	1.31	262.60
30	−2.15	−0.77	3	2.27	162.81
31 (Temelastine)	−1.88	−0.99	4	3.41	137.14
32 (Butanone)	−0.08	0.00	1	0.37	44.04
33 (Benzene)	0.37	0.67	0	2.22	0.00
34 (3-methylpentane)	1.01	0.84	0	3.76	0.00
35 (3-methylhexane)	0.90	0.89	0	4.29	0.00
36 (2-propanol)	−0.15	−0.03	1	0.16	47.98
37 (2-methylpropanol)	−0.17	0.02	1	0.69	53.27
38 (2-methylpentane)	0.97	0.84	0	3.76	0.00
39 (2,2-dimethylbutane)	1.04	0.82	0	3.58	0.00
40 (1,1,1 trifluro-2-chloroethane)	0.08	0.23	0	1.11	187.63
41 (1,1,1 trichlorethane)	0.40	0.34	0	2.10	183.30
42 (diethyl ether)	0.00	0.10	1	0.98	28.56
43 (R-enflurane)	0.24	−0.16	1	2.10	249.63
44 (Ethanol)	−0.16	−0.09	1	−0.19	65.06
45 (Fluroxene)	0.13	−0.07	1	1.49	159.55
46 (R-halothane)	0.35	0.26	0	2.30	242.72
47 (Heptane)	0.81	0.91	0	4.47	0.00
48 (Hexane)	0.80	0.86	0	3.94	0.00
49 (R-isoflurane)	0.42	−0.04	1	2.79	225.88
50 (methylcyclopentane)	0.93	0.79	0	3.31	0.00
51 (Pentane)	0.76	0.80	0	3.41	0.00
52 (Propanol)	−0.16	−0.03	1	0.34	61.64
53 (Propanone)	−0.15	−0.06	1	−0.16	49.19

Table 2 (*Continued*)

Compound[b]	Log BB exp.[c]	Log BB calculated	$n_{acc,solv}$	log P	A_{pol}
54 (R-teflurane)	0.27	0.25	0	1.95	225.83
55 (Toluene)	0.37	0.72	0	2.68	0.00
56 (Trichloroethene)	0.34	0.66	0	2.26	7.75
57 (Acetylsalicylic acid)	−0.50	−1.20	4	1.19	119.86
58 (Valproic acid)	−0.22	−0.20	2	2.72	84.58
59 (Salicylic acid)	−1.10	−0.72	3	2.06	123.88
60 (Acetaminophen)	−0.31	−0.49	2	0.34	104.52
61 (Chlorambucil)	−1.70	−0.40	2	3.70	264.18

[a] The applied descriptors were the following: $n_{acc,solv}$ is the number of solvated hydrogen-bond acceptors, log P is the calculated octanol–water partition coefficient and A_{pol} is the polar surface area. See text for further details.
[b] The structure and numbering of the molecules, corresponds to that in the reference (Luco, 1999).
[c] Experimental log BB values taken from the reference (Luco, 1999).

for such replacement. The substitution of polar surface area with water-accessible surface area or even the total number of single bonds in the molecule leads to a moderate ($< 10\%$) increase of the rmse in the fit and predictions. The replacement of log P with log D (at pH 7) leaves the fit largely unchanged, presumably because log D is highly correlated with log P and $n_{acc,solv}$ ($r^2 = 0.70$). No replacement descriptor could be found for the number of solvated hydrogen-bond acceptors and hence this descriptor is both essential and unique to this model. Replacing $n_{acc,solv}$ with the number of hydrogen-bond acceptors for an isolated ligand produced only poor models. (Fitting the parameters log P, A_{pol} and $n_{acc,\ vacuum}$ on the training set produced $r^2 = 0.49$, $rmse = 0.59$).

An intriguing result of the above QSAR analysis is the fact that the number of hydrogen-bond donors (either solvated or vacuum) had little impact on the statistics of the model. This is clearly not the result of underrepresentation of molecules with hydrogen bond donor groups in the dataset. It may be partly due to the fact that a large proportion of the donor groups are already accounted for in $n_{acc,solv}$, namely those that can be simultaneously acceptors and donors such as OH, those that can tautomerize to act as either, such as the nitrogen in imidazole, or have pK_a values near neutrality, such as the nitrogen in aromatic sulphonamides (Bush and Sheridan, 1993). Recently it has been shown from Monte-Carlo simulations that hydrogen bond donors have no effect on octanol-water partitioning (Duffy and Jorgensen, 2000). In that work, this observation was explained by noting that the number of hydrogen bonds donated by solutes is the same in water and alcohol solvents, while the hydrogen-bond acceptor sites are more saturated in water. It appears likely that similar conclusions apply for blood–brain barrier penetration, which involves partitioning between the aqueous and the lipid phase. It must be noted that some non-linearity can be observed in the model for molecules with $\log BB < -1$. Molecules in this range are likely to build a number of strong hydrogen bonds and it is feasible that the neglect of H-bond donors is no longer justifiable in this range. However, molecules that fall in this range do not readily cross the blood-brain barrier and the model in this work clearly identifies them as such.

In order to assess the validity of the model, it is necessary to discuss the sources of errors. The statistical errors in the fit arise from three sources: the possible inadequacy of the model, inaccuracy of the experimental data and errors in the calculated parameters. The experimental log BB values were measured using different experimental procedures and many of them have large inter-animal variations (Bonate, 1995). In addition, the comparability of results obtained with different experimental techniques has not been established (Bonate, 1995). A further uncertainty arises due to the involvement of different transport mechanisms and binding to plasma proteins (Bonate,

Table 3
Experimental and calculated blood-brain barrier penetration and the computed descriptors [a] for the test set molecules

Test set 1 Compound[b]	Log BB exp.[c]	Log BB calculated	$N_{acc,solv}$	Log P	A_{pol}
62	−1.30	−0.20	2.00	2.29	61.66
63	−1.40	−0.50	2.00	0.10	95.10
64	−0.43	−0.28	2.00	2.95	146.35
65	0.25	0.22	1.00	3.31	103.48
66	−0.30	−0.07	1.00	0.10	68.65
67	−0.06	0.01	1.00	0.76	62.61
68	−0.42	−0.44	2.00	0.53	88.86
69	−0.16	−0.39	2.00	1.73	136.52
70 (Carbamazepine)	0.00	0.18	1.00	2.67	89.59
71 (Carbamazepine epoxide)	−0.34	−0.46	2.00	0.69	110.72
72	0.30	−1.63	5.00	0.90	127.09
73	−1.34	−2.19	6.00	−0.08	165.84
74	−1.82	−2.74	7.00	−0.73	216.47
75 (Amitriptyline)	0.89	1.00	0.00	6.14	55.08
Test set 2[a]					
76 (Theophylline)	−0.29	−1.43	4.00	0.05	180.69
77 (Caffeine)	−0.06	−1.03	3.00	−0.07	168.27
78 (Antipyrine)	−0.10	−0.03	1.00	0.27	58.32
79 (Ibuprofen)	−0.18	−0.09	2.00	3.72	87.87
80 (Codeine)	0.55	−0.75	3.00	1.83	126.67
81 (Pentobarbital)	0.12	−0.77	3.00	2.09	150.68
82 (Alprazolam)	0.04	−0.58	3.00	2.50	70.30
83 (Indomethacin)	−1.26	−1.07	4.00	3.10	165.64
84 (Oxazepam)	0.61	−0.70	3.00	2.31	124.45
85 (Hydroxyzine)	0.39	−0.20	2.00	3.00	104.85
86 (Desipramine)	1.20	0.77	0.00	3.97	52.29
87 (Midazolam)	0.36	−0.02	2.00	3.70	43.34
88 (Verapamil)	−0.70	−1.32	5.00	5.03	206.64
89 (Promazine)	1.23	0.78	0.00	4.63	90.27
90 (Chlorpromazine)	1.06	0.86	0.00	5.36	89.79
91 (Trifluorperazine)	1.44	0.70	0.00	5.11	164.67
92 (Thioridazine)	0.24	0.89	0.00	6.13	121.60
93 (BCNU)	−0.52	−0.56	2.00	1.30	206.30
94 (Phenserine)	1.00	−0.23	2.00	2.91	118.21
95 (Physostigmine)	0.08	−0.50	2.00	0.99	148.62
96 (Terbutylchlorambucil)	1.00	0.28	1.00	4.93	174.18
97 (Didanosine)	−1.30	−1.95	5.00	−0.92	198.99
98 (Zidovudine)	−0.72	−2.37	6.00	−0.58	238.10
99 (Nevirapine)	0.00	−0.95	3.00	−0.31	103.95
100 (Sb-222 200)	0.30	0.19	2.00	5.89	60.72

[a] The applied descriptors were the following, $n_{acc,solv}$ is the number of solvated hydrogen-bond acceptors, log P is the calculated octanol–water partition coefficient and A_{pol} is the polar surface area. See text for further details.
[b] The structure and numbering of the molecules, corresponds to that in the reference (Luco, 1999).
[c] Experimental log BB values taken from the reference (Luco, 1999).

1995). Hence the experimental data is highly heterogeneous and of fairly poor quality. With these errors in mind, it is questionable whether a model with a significantly higher number of parameters would not simply fit the model to errors in the experimental data.

Similarly, the applied descriptors are also potential sources of errors. Although the applied

conformations are likely to be near the global energy minimum, ideally the real conformational equilibrium should have been characterized using Boltzmann averaging. This could impact the calculated polar surface area or the water accessible surface area. The ACD log P values are obtained from a fragmental approach. Due to the high quality of the ACD model, it is likely to introduce only small errors, although this may become an issue with novel compound classes. It must be noted, however, that when atom-based log P values (MOE, 1999) were used instead of the fragment-based ones, the correlation coefficient of the fit did not change substantially. Finally, the number of hydrogen bond acceptors can also be a potential source of error, as partial hydrogen bonding is not considered in the Patty rules (Bush and Sheridan, 1993).

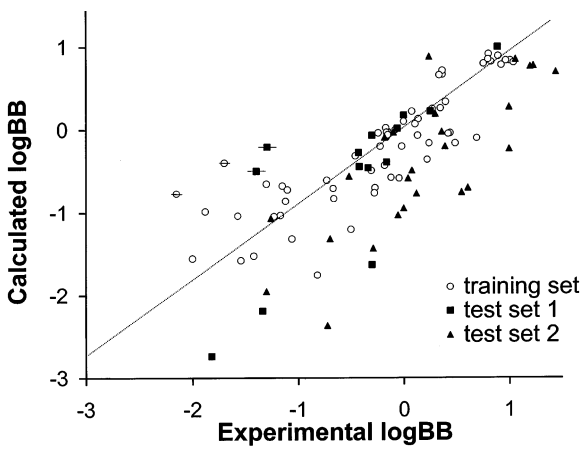

Fig. 1. Relationship between experimental and predicted blood-brain partitioning, using the three-descriptor model described in the text. Although the fit included all datapoints, each potential outliner is shown with a line through it.

Table 4
Correlation matrix (r^2) for the applied descriptors[a]

	Log P	A_{pol}	$n_{acc,solv}$
log P	1.00		
A_{pol}	0.04	1.00	
$n_{acc,solv}$	0.10	0.30	1.00

[a] The applied descriptors were the following, $n_{acc,solv}$ is the number of solvated hydrogen-bond acceptors, log P is the calculated octanol-water partition coefficient and A_{pol} is the polar surface area. See text for further details.

If the model in this work is used as a tool for screening large virtual libraries, it may be useful to define the point at which a molecule is accepted or refused. It is important to note that this point is completely arbitrary and has to be defined with the intended application in mind. In this work, the effect of choosing this cutoff limit at different values was studied. If a cutoff of -1 is selected similarly to an earlier PLS model (Luco, 1999), 67% of the inactives and 85% of the actives are predicted correctly in the test sets. Raising the cutoff increases the reliability of the prediction of inactives, whereas lowering it enhances the prediction of actives. At a log BB value of -0.5, for example, 78% of the inactives and 73% of the actives are predicted correctly. This effect is clearly caused by molecules with a BBB penetration close to the cutoff falling into the wrong group due to numerical inaccuracy (i.e. bin boundary effects). For this reason, it may be best to choose the cutoff with the specific pharmacology in mind and set the cutoff value depending on whether false positives or false negatives are more harmful for the intended application.

It is quite interesting to look at the set of 100 molecules and see how well their diversity is representative of molecules in a pharmaceutical setting. This was achieved by using the MACCS structural keys, as implemented in the Molecular Operating Environment (MOE, 1999). These MACCS keys contain information on the occurrences of small fragments in the molecules and were recently shown to be superior to other descriptors in capturing information on drug-size molecules for receptor binding (Brown and Martin, 1997). The diversity of structures was assessed by calculating the Tanimoto similarity coefficient (Downs and Willett, 1995) between all pairs of structures. The obtained unsorted similarity values, with the diagonal elements of the similarity matrix (self-similarity) removed, were summed and are shown in a histogram form in Fig. 2a. For comparison, the distribution of molecules was investigated in the Dictionary of Drugs (Elks and Ganellin, 1990), which contains 18 222 drug molecules. The cumulative normalized Tanimoto

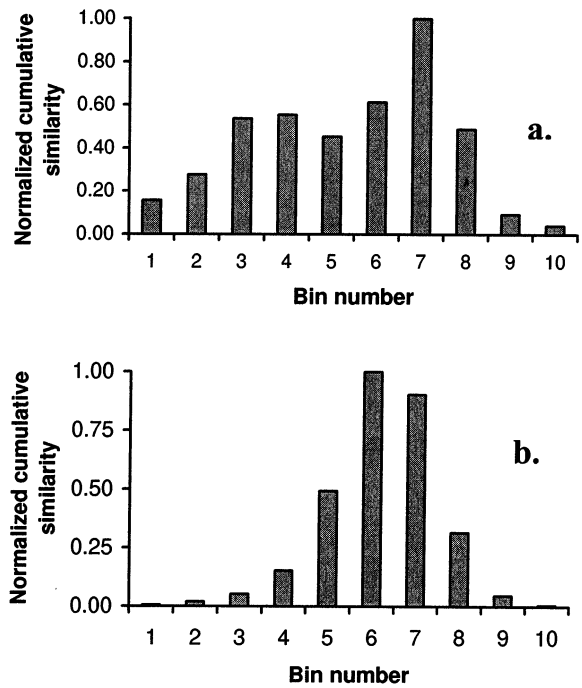

Fig. 2. Histogram displaying the distribution of the normalized cumulative Tanimoto similarity coefficients using MACCS structural keys. These cumulative coefficients were obtained by summing all the pairwise similarity values over all molecules and normalizing the result. The bin numbers correspond to the range of coefficient values between 0 and 1 with bin width 0.1, e.g. bin 1 represents the range 0.0–0.1, bin 2 contains 0.1–0.2, etc. (a) Molecules in this work (100 molecules). (b) Molecules in the Dictionary of Drugs, containing 18222 molecules (Elks and Ganellin, 1990).

coefficients for this database is displayed in Fig. 2b. A comparison of Fig. 2a and b reveals a significant difference in the similarity distributions. The similarity indices of molecules in the drug database are approximately normally distributed and the distribution peaks at bin 6 (Tanimoto similarity range of 0.5–0.6). The number of very dissimilar molecules (bin 1–bin 3) is practically negligible. In contrast, similarity distribution for the molecules in this study appears to be skewed towards higher dissimilarity, although there is a spurious peak in bin 7 (similarity range 0.6–0.7). Hence, on the basis of Fig. 2 we can conclude that the molecules in this study were at least as diverse as marketed drug molecules.

In summary, the need to include the number of solvated hydrogen-bond acceptors is a key finding of this study. The simple three-descriptor model developed in this work offers two significant advantages over previously developed models (Norinder et al., 1998; Luco, 1999). Firstly, the descriptors can be calculated easily and thus the model is applicable for large sets of molecules. Secondly, as can be seen from Fig. 1, a rapid yes/no decision on BBB-penetration can be given with reasonable certainty that could be useful for scoring and prioritizing large combinatorial libraries or solutions from de novo methods.

References

Abraham, M.H., Chadha, H.S., Mitchell, R.C., 1994. Hydrogen bonding. 33. Factors That Influence the Distribution of Solutes between Blood and Brain. J. Pharm. Sci. 83, 1257–1268.

ACD/Labs™, version 3.6, 1998. Advanced Chemistry Development, Toronto, ON, Canada.

Bonate, P.L., 1995. Animal models for studying transport across the blood–brain barrier. J. Neurosci. Methods 56, 1–15.

Brown, R.D., Martin, Y.C., 1997. The information content of 2D and 3D structural descriptors relevant to ligand–receptor binding. J. Chem. Inf. Comput. Sci. 37, 1–9.

Bush, B.L., Sheridan, R.P., 1993. PATTY: a programmable atom typer and language for automatic classification of atoms in molecular databases. J. Chem. Inf. Comput. Sci. 33, 756–762.

Downs, G.M., Willett, P., 1995. Clustering of chemical structure databases for compound selection. In: van de Waterbeemd, H. (Ed.), Advanced Computer-Assisted Techniques in Drug Discovery. Verlag Chemie, Weinheim.

Duffy, E.M., Jorgensen, W.L, 2000. Prediction of Properties from Simultations: Free energies of solvation in hexadecane, Octanol, and water. J. Am. Chem. Soc. 122, 2878–2888.

Elks, J., Ganellin, C.R., 1990. Dictionary of Drugs. Chapman and Hall, London.

Ferguson, D.M., Raber, D.J., 1989. A new approach to probing conformational space with molecular mechanics: random incremental pulse search. J. Am. Chem. Soc. 111, 4371–4378.

Halgren, T.A., 1996. Merck molecular force field. I: basis, form, scope, parametrization and performance of MMFF94. J. Comp. Chem. 17, 490–519.

Kaliszan, R., Markuszewski, M., 1996. Brain/blood distribution described by a combination of partition coefficient and molecular mass. Int. J. Pharm. 145, 9–16.

Levin, V.A., 1980. Relationship of octanol/water partition coefficient and molecular weight to rat brain capillary permeability. J. Med. Chem. 23, 682–684.

Lombardo, F., Blake, J.F., Curatolo, W.J., 1996. Computation of brain–blood partitioning of organic solutes via free energy calculations. J. Med. Chem. 39, 4750–4755.

Luco, J.M., 1999. 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.

Martin, Y.C., 1978. Quantitative Drug Design: A Critical Introduction. Marcel Dekker, New York, pp. 194–198.

Molecular Operating Environment, Version, 1999. Chemical Computing Group. Montreal, Quebec, Canada.

Norinder, U., Sjöberg, P., Österberg, T., 1998. Theoretical calculation and prediction of brain–blood partitioning of organic solutes using molsurf parametrization and PLS statistics. J. Pharm. Sci. 87, 952–959.

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

Waterbeemd, H., Camenisch, G., Folkers, G., Chretien, J.R., Raevsky, O.A., 1998. Estimation of blood–brain barrier crossing of drugs using molecular size and shape, and H-bonding descriptors. J. Drug Targeting. 6, 151–165.