Computational models to predict blood–brain barrier permeation and CNS activity

Govindan Subramanian[a,b] & Douglas B. Kitchen[a,*]
[a]*Medicinal Chemistry Department, Albany Molecular Research, Inc., 21 Corporate Circle, P.O. Box 15098, Albany, NY 12212-5098, USA;* [b]*Present address: Transtech Pharma, 41702 Mendenhall Oaks Parkway, High Point, NC 27265, USA*

Summary

The blood–brain permeation of a structurally diverse set of 281 compounds was modeled using linear regression and a multivariate genetic partial least squares (G/PLS) approach. Key structural features affecting the logarithm of blood–brain partitioning (logBB) were captured through statistically significant quantitative structure–activity relationship (QSAR) models. These relationships reveal the importance of logP, polar surface area, and a variety of electrotopological indices for accurate predictions of logBB. The best models reveal an excellent correlation (r > 0.9) for a training set of 58 compounds. Likewise, the comparison of the average logBB values obtained from an ensemble of QSAR models with experimental values also verifies the statistical quality of the models (r > 0.9). The models provide good agreement (r ∼ 0.7) between the predicted logBB values for 34 molecules in the external validation set and the experimental values. To further validate the models for use during the drug discovery process, a prediction set of 181 drugs with reported CNS penetration data was used. A >70% success rate is obtained by using any of the QSAR models in the qualitative prediction for CNS permeable (active) drugs. A lower success rate (∼60%) was obtained for the best model for CNS impermeable (inactive) drugs. Combining the predictions obtained from all the models (consensus) did not significantly improve the discrimination of CNS active and CNS inactive molecules. Finally, using the therapeutic classification as a guiding tool, the CNS penetration capability of over 2000 compounds in the Synthline® database was estimated. The results were very similar to the smaller set of 181 compounds.

Introduction

Recent advances in high throughput chemical synthesis and screening are generating large numbers of small-molecule leads for drug discovery research. Each lead requires significant resources to optimize the chemical and biochemical properties so that it can become a new drug candidate. These developments are increasing the pressure to prioritize lead compounds based on both their *in vitro* activity and their 'drug-like' properties [1]. A great deal of attention is therefore placed on the absorption, distribution, metabolism, and elimination (ADME) properties of molecules while identifying the most potent *in vitro* compound [2]. During the optimization process *in vitro* and *in vivo* screens are used to determine whether the molecules will be absorbed, delivered to the right organs, and whether there are likely metabolites. These tests are very rapid in comparison to many mammalian *in vivo* experiments but the screens are still time consuming. To aid chemists in the selection of target synthetic compounds, various computational approaches are being used to design 'drug-like' libraries so that molecules with poor physicochemical properties are removed early in the drug discovery process.

*To whom correspondence should be addressed. E-mail: douglask@albmolecular.com

A particularly useful measurement in the design of central nervous system (CNS) compounds is the partitioning or transport of a candidate drug molecule between the blood and the brain compartments [3]. *In vitro* experimental determination of the blood–brain permeation is often complex and depends heavily on the compound's stability, purity, assay condition, etc. while *in vivo* determinations are complicated by the need for radiolabeled compound in some cases [4].

$$BB = C_{brain}/C_{blood} \quad \text{(C is the concentration}$$
$$\text{of the compound)} \quad (1)$$

Although the laborious experimental measurement of the blood–brain distribution of a compound as described by Equation 1 is the ideal approach, the experiments are a major bottleneck for high-throughput screening of large molecular libraries. Therefore, it is desirable to focus synthetic efforts in drug design by using simple predictive tools so that 'candidate drug molecules' can be discriminated as CNS (in)actives (CNS$^{-/+}$) *a priori*.

Progress along these directions is evident from a variety of computational approaches reported in the literature for logBB prediction and CNS activity [5]. A few early studies used several computed molecular properties along with a few experimental parameters. For a diverse set of 20 H2-receptor histamine antagonists, Kansy and van de Waterbeemd [6], and Young et al. [7] showed a reasonable correlation between logP (cyclohexane-water) and the logBB values. Similarly, Levin correlated the octanol-water partition coefficient of a series of compounds to brain capillary permeability [8]. Additional studies reveal that there are other factors influencing logBB in addition to the partition coefficient [3d,9]. Using a set of 65 molecules, Abraham et al. [10] showed that logBB is related to donor–acceptor properties, polarizability, excess molar refraction, and molecular volume. There continues to be great usefulness to these equations because they clearly relate understandable physical properties to the ability of compounds to cross the blood–brain barrier. These models indicate that the barrier is at least due to the tight junctions and hydrophobic nature of the paracellular and transcellular pathways.

Recently several models were described that use only calculated quantities to predict logBB. Keserü and Molnár computed the logBB by employing the generalized Born/surface area continuum solvation model [11]. Using the semi-empirical AM1-SM2.1 solvation model, Lombardo et al. [12] correlated the computed solvation free energies of 57 molecules with experimental logBB. Similarly, Norinder and co-workers [13] used multivariate partial least squares (PLS) projection to latent variables [14] to determine the relationship between the BB partitioning and the computed molecular descriptors based on *ab initio* molecular orbital methods respectively. Both groups [12, 13] obtained good quality results by using the molecular descriptors derived from these quantum mechanical methods. The recent work by Crivori et al. [15] uses 3-dimensional grid-based parameters to discriminate blood–brain permeation from non-permeation. Such methods can be highly conformation dependent and as such offer limited applicability for virtual screening of large combinatorial libraries.

In contrast, Luco's [16] logBB model was developed using PLS with 18 topological and constitutional descriptors. This model is particularly advantageous due to the lack of any 3D terms or need for computer intensive methods. Recently, this blood–brain partitioning dataset was extensively re-examined and re-fit using electrotopological state indices [17]. These 3-term models are appealing because of their simplicity. Also, new stepwise regression models have been compared to other methods [18].

Most computational logBB estimations use some variation of logP as a descriptor [19]. Clark proposed a simple QSAR model for rapid prediction of logBB [20] using only the polar surface area (PSA) [21] and calculated logP [22] terms. These equations are easily interpreted and lend themselves to intuitive design of new molecules. Unfortunately, the use of only two terms introduces an over dependence on consistent logP estimates. Also, for compounds with many rotatable bonds, the PSA term may be influenced by the conformational choice.

Another avenue of investigation in CNS activity predictions has been to analyze the wealth of information found in databases of marketed drugs. Recent work has illustrated that it is possible to make qualitative predictions on CNS penetration based on 1D and 2D descriptors [23–25]. These models are very appealing because of the simple parameters such as the count of nitrogen and oxygen atoms that are used. For instance, the models developed using Bayesian neural networks, to distinguish CNS$^{-/+}$ for a large dataset retrieved from the CMC [26] and MDDR [27] databases were approximately 80% accurate [25].

The goal of the present study was to explore QSAR models [28] that are less dependent on 3D descriptors

Table 1. Compounds used in this study.

Training set compounds

31, temelastine	**39**, 2,2-dimethylbutane	**47**, heptane	**55**, toluene
32, butanone	**40**, F$_3$C-CH$_2$Cl	**48**, hexane	**56**, trichloroethene
33, benzene	**41**, 1,1,1-trichloroethane	**49**, isoflurane	**57**, acetylsalicylic acid
34, 3-methylpentane	**42**, diethyl ether	**50**, methylcyclopentane	**58**, valproic acid
35, 3-methylhexane	**43**, enflurane	**51**, pentane	**59**, salicyclic acid
36, 2-propanol	**44**, ethanol	**52**, propanol	**60**, p-acetamidophenol
37, 2-methylpropanol	**45**, fluroxene	**53**, propanone	**61**, chlorambucil
38, 2-methylpentane	**46**, halothane	**54**, teflurane	

Prediction set compounds

76, theophylline	**83**, indomethacin	**89**, promazine	**95**, physostigmine
77, caffeine	**84**, oxazepam	**90**, chlorpromazine	**96**, tertbutylchlorambucil
78, antipyrine	**85**, hydroxyzine	**91**, trifluoroperazine	**97**, didanosine
79, ibuprofen	**86**, despramine	**92**, thioridazine	**98**, zidovudine
80, codeine	**87**, midazolam	**93**, BCNU	**99**, nevirapine
81, pentobarbital	**88**, verapamil	**94**, phenserine	**100**, SB-222200
82, alprazolam			

Qualitative prediction set compounds from Ref. 25

CNS Active

abecarnil	amantadine	amitriptyline	amobarbital
apomorphine	baclofen	biperiden	CGP-37849
CGP-39551	CGS-19755	chlordiazepoxide	chlorprothixene
clobazam	clomipramine	clonazepam	clonidine
clozapine	cyclobarbital	diazepam	diphenhydramine
dopamine	doxepin	felbamate	fentanyl
flumazenil	flunitrazepam	fluoxetine	flupentixol
flurazepam	GYKI-52466	gabapentin	haloperidol
hydroxypethidine	ifenprodil	imipramine	isradipine
ketamine	lamotrigine	lorazepam	lorcainide
mCPP	mefenidil	meprobamate	mequitazine
methadone	mexiletine	milacemide	MK-801
morphine	naloxone	naltrexone	nifedipine
nitrazepam	nortriptyline	noxiptiline	orphenadrine
oxcarbazepine	PD117302	perphenazine	phaclofen
phenobarbital	piracetam	primidone	procyclidine
progabide	promethazine	quinpirole	quipazine
raclopride	ralitoline	remacemide	roxindole
SCH-23390	saclofen	spiradoline	stiripentol
sulpiride	taltrimide	tamitinol	thiopental
tiagabine	tifluadom	topiramate	tranylcypromine
triazolam	U50,488	valproate_pivoxil	vigabatrin
zimeldine	zolmitriptan	zonisamide	

CNS inactive

acarbose	acrivastine	alclofenac	alprenolol
amiodarone	ampicillin	asimadoline	astemizole
atenolol	betaxolol	bufuralol	captopril
carbenicillin	carbidopa	carebastine	carmoxirole

Table 1 (continued).

cefetametpivoxil	ceftriaxone	cephalexin	cephalothin
cetirizine	chloramphenicol	cimetidine	cisapride
clofibric_acid	corticosterone	coumarin	demeclocycline
dexamethasone	diltiazem	domperidone	doxylamine
ebastine	estradiol valerate	ethinyl estradiol	felodipine
fenoprofen	fleroxacin	fluvastatin	furosemide
guanabenz	guanfacine	hydrocortisone	ICI-204448
isotretinoin	ketoprofen	labetalol	lidocaine
loperamide	loratadine	lupitidine	mannitol
metoprolol	mibefradil	miglitol	minoxidil
n-methyl naltrexone	nadolol	naproxen	napsagatran
neostigmine	nitrendipine	norfenefrine	olsalazine
oxprenolol	penicillin G	phenacetin	pheniramine
pirenzepine	practolol	pravastatin	prednisolone
prednisone	probenecid	propranolol	proscillaridin
quinidine	remikiren	salbutamol	sulfasalazine
sumatriptan	temafloxacin	terbutaline	terfenadine
testosterone	tiacrilast	tocainide	tolamolol
toliprolol	warfarin		

and that use a relatively few number of variables to predict the logBB values for diverse structural classes. We also anticipated that models that were not overly dependent on logP would provide consistent predictions well outside the training set. We compared approaches where we eliminated logP as a variable allowing the algorithm to choose additional terms to cases where logP was used as one of the variables. To obtain this added generality, we were willing to sacrifice some accuracy in the computed logBB predictions (e.g., permeable vs. non-permeable) as long as computational speed was improved and model complexity was reasonable (i.e. not too many additional terms are needed). Models that are sufficiently fast to calculate can be used in a high throughput mode to predict CNS activity of large libraries of compounds. If these models are accurate enough, they could also aid in individual analog design. Since various published models were linear and based on only two terms, logP and PSA models [19–22], we examined models with these two terms. Also, models derived using Bayesian neural nets [25] were found to be useful in discriminating between CNS active and CNS inactive drug molecules and these neural nets indicated that seven descriptors could be used in a nonlinear approach. Therefore, we decided to explore linear, PLS-based models of seven terms in order to determine if additional terms were helpful beyond logP and PSA.

For the seven term equations, a genetic/PLS algorithm [29] was used to select variables. Simple equations were developed that provide automated, robust, and predictive models for a variety of target molecules. PLS offers the advantage that intercorrelation of independent variables is eliminated. Therefore, while the final equations are shown as seven terms, only three, linearly independent components were actually used in the fitting. Descriptors were obtained from easily calculated topological descriptors and 3D structural parameters. The models have been developed using a training set of 58 molecules [13, 16, 20] where logBB has been determined. The predictive ability of the models was judged by computing the logBB values for a set of 39 unrelated molecules and comparing them with reported experimental data. The models have also been extended to predict the logBB values for a diverse set of 181 molecules compiled in the literature [23–25], and were used to filter CNS^+ compounds from the inactives (CNS^-). We also used these models to analyze over 2000 drugs from the Synthline$^®$ database [30]. We find that the predicted logBB based on models derived from only 58 quantitative measurements is able to discriminate between CNS^+ and CNS^- compounds. The models developed in this study can be applied to widely
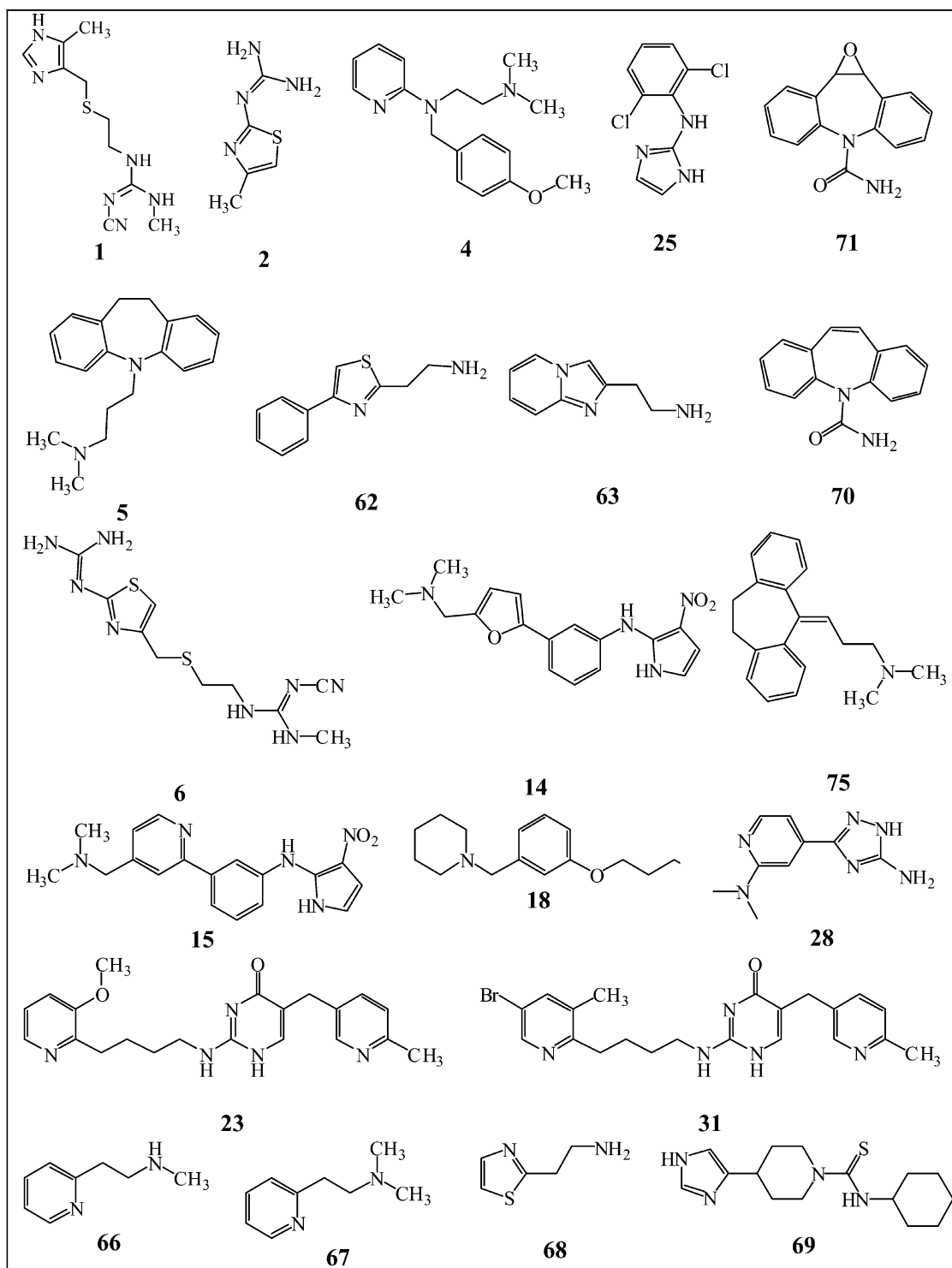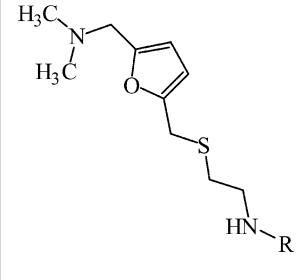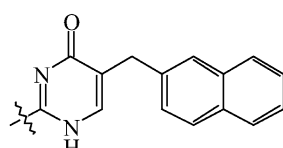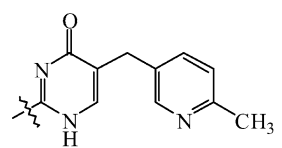
Chart 1.
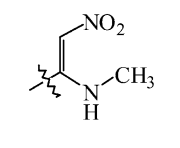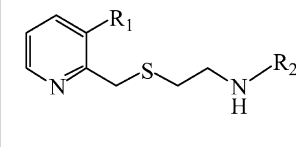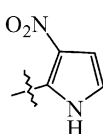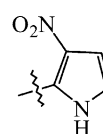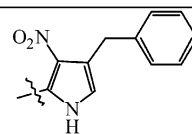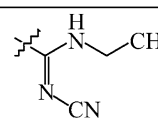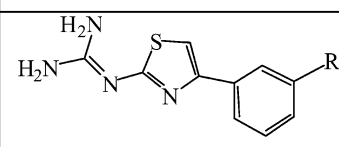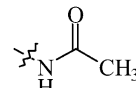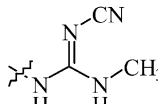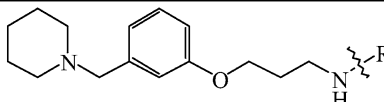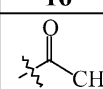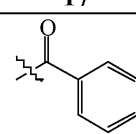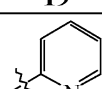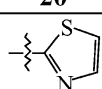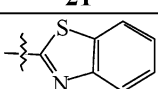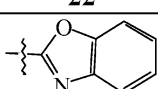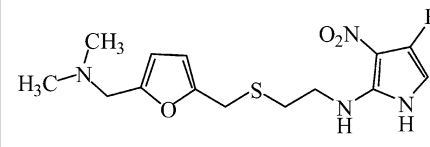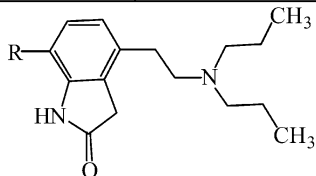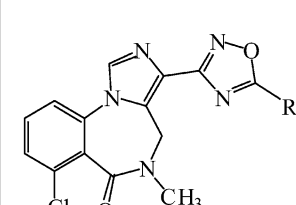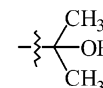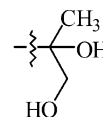
	3	**24**	**29**

		7	**8**	**9**	**30**
	R₁	Br	H	H	Br
	R₂				

	10	**11**	**12**	**13**
	H	NH₂		

16	**17**	**19**	**20**	**21**	**22**

	26	**27**		**64**	**65**
	H	Benzyl		OH	H

	72	**73**	**74**

Chart 1. (Continued).

Methods

A total of 281 molecules were energy minimized [31] employing the universal forcefield [32] implemented in the Cerius^2 molecular modeling package (Molecular Simulations Inc., San Diego, CA). Subsequently, constitutional, topological, and other physicochemical descriptors available in the Cerius^2 package were included. In addition, polar surface area (PSA) [33], cube root of the gravitational index [34], and the inverse square root of the molecular weight ($\sqrt{(1/MW)}$) [35] were added using Cerius^2 Tcl scripts. In all cases molecules were treated as neutral species. The 281 molecules were divided into a training set (**1–61**), a quantitative prediction set (**62–100**) and a qualitative prediction set of 181 compounds for which reliable CNS$^{-/+}$ classifications were obtained from literature (Chart 1 and Table 1) [23–25]. The division of training and prediction set and the experimental values of logBB were obtained from Ref. 16. The available quantitative logBB data spans four log units. Independent variables that did not have adequate variance and representation in the initial training set of 61 molecules were discarded. That is, any descriptors that were found to have a common value for all but a few molecules were eliminated. Previous studies identified compounds **6** [11, 16] and **30** [11–13, 16, 19, 20] as outliers and our initial studies found **28** also to be an outlier; so these were removed from the training set. Therefore, our final training set was composed of 58 molecules. We speculate that since others have found the compounds to be outliers, these outliers could be caused by experimental inconsistency or different mechanisms of transport into the brain or efflux out of it. There appears to be no common structural similarity among the outliers that is significantly different from the rest of the training set. For example, there are several compounds containing a nitrile group or bromo substitution that perform well as part of the training set.

The training set was initially subjected to linear regression (LR) and PLS methods. Using the training set of Clark [20], the LR method was used to fit the observed logBB to only PSA and AlogP [36] terms (Eq. 2). Equation 2 is based on 55 compounds and this training set is composed of the following compounds: **1–5**, **7–22**, **25–29**, and **32–60**. When the 58 compound training set was used and only these two terms, r^2 degraded to 0.635. For more complex models, the G/PLS method was used in the variable selection. G/PLS filters unrelated descriptors from the PLS model and unravels those descriptors that are most relevant for describing logBB. The genetic algorithm applies different combinations of the independent variables in the regression equation in order to optimize the fitness function. All the independent variables were scaled for the G/PLS run that did not include the AlogP term. The length of the QSAR equation was restricted to 7 independent variables and a constant with up to 4 PLS components. The PLS method has the advantage over LR methods in that each of the components are linearly independent and therefore contain no intercorrelation. Each generation of the PLS model started with 100 randomly generated G/PLS model equations. This initial population was allowed to evolve for 50,000 generations resulting in a final set of 100 model equations for each of the starting variable subsets. Equations 3 and 4 represent the best QSAR model derived for the 58 training set compounds as determined by highest r^2. The effect of the partition coefficient on logBB was studied by deriving a QSAR model without the AlogP term. Equation 3 represents the G/PLS run with logP and Eq. 4 represents the run without logP. The resulting Eq. 3, required only three PLS components and in the case of Eq. 4 only two components were required.

The fitness function for the genetic algorithm is measured as the square of the correlation coefficient (r^2). The internal cross-validation coefficient, r^2 (CV), using the leave-one-out method (LOO) and the F-statistic (F) are reported. The G/PLS algorithm produces a population of models (e.g., 100) involving different permutations of the descriptor variables. The importance of particular descriptors can be examined by monitoring the frequency of usage in the equation set. Variables that are used often are less likely to be the result of chance correlation. It should be noted that the frequency of usage only reflects the importance within the accumulated variables in each run. There remains some risk that a genetic algorithm can become inbred and certainly there could well be additional important variables that were eliminated early in the evolution of these models. Since there was no significant deviation in either the r^2 or in the coefficients on variables obtained through the 100 equations, only the best G/PLS equations were used in most of the analysis. Variable usage is reported in Table 4. The computed logBB values of the first validation set (**62–**

Table 2. Statistical results for the reported and derived models for the training (**1–61**) and prediction (**62–100**) sets[a].

Models	N-tr[b]	r^2 (tr)	R	r^2 (cv)	s[c]	rmse[d] (MAE)[k]	F[e]	N-pr[f]	r^2 (pr)[g]	rmse-pre[h] (MAE)[k]
Equation 2	55	0.821	0.906	0.798	0.297	0.292 0.311	119.0	39	0.689	0.490 0.461
Equation 3	58	0.845	0.919	0.811	0.314	0.308 0.231	97.9	39	0.614	0.463 0.516
Equation 4	58	0.833	0.913	0.786	0.325	0.320 0.264	137.4	39	0.383	0.708 0.769
GPLS 3[i]	58	0.846	0.920		0.313	0.308 0.232	98.6	39	0.617	0.413 0.499
GPLS 4[j]	58	0.834	0.914		0.325	0.262 0.262	137.8	34	0.387	0.686 0.758

[a]Outliers identified in Table 3 are not included in the statistical process.
[b]Number of molecules considered in the training set.
[c]Standard error.
[d]Root mean square error for the training set.
[e]Fischer value.
[f]Number of molecules considered in the prediction set.
[g]Square of the correlation coefficient for the prediction set.
[h]Root mean square error for the prediction set.
[i]Average of 50 equations derived using AlogP.
[j]Average of 50 equations derived without using AlogP.
[k]MAE is the mean absolute error and is in parentheses.

100) were compared with observed values in order to corroborate the predictive ability of the two-term LR equation and the two G/PLS equation sets. The equations were sorted by descending r^2 and the predictions for each compound averaged for the best 50 equations. The average predicted logBB from 50 equations of the G/PLS runs represented by Eq. 3 and Eq. 4 are presented in Table 2 and are labeled as GPLS 3 and GPLS 4, respectively.

There are a larger number of compounds available where a qualitative indication of CNS permeability has been determined. For this second validation set (181 molecules), we chose to label compounds with a predicted logBB less than −1 as CNS impermeable (CNS⁻) and the others as CNS permeable (CNS⁺) [16] as in reference 25. The structures of the 181 compounds are available upon request. The computed log BB values for the second prediction set are provided in the supporting information accompanying the paper. A final validation set was obtained by extracting more than 2000 compounds from the Synthline® database [30]. The therapeutic classification scheme from reference 25 (e.g., antidepressants, anxiolytics, etc.) was used to separate CNS active compounds from the CNS inactive compounds.

Prediction of $CNS^{+/-}$ was obtained by using individual equations. These results are summarized in Table 5. Each model (Eqs. 2, 3 and 4) was used to predict CNS penetration. When logBB is predicted to be less than −1 then the compound should be a poor brain penetrator and therefore not a CNS drug. However, when logBB is greater than −1 then the compound could be a CNS drug. The latter scenario is more ambiguous since a compound may penetrate the CNS but have no observable effect and therefore not marketed as a CNS drug. We also tested the choice of the logBB cutoff on prediction accuracy by using −0.75 and −0.40, similar to the choice used in Ref. 18.

Results

Statistical correlations for the training set

Good relationships between the experimental logBB values and the computed descriptors were obtained through the LR (Eq. 2) and the G/PLS methods (Eqs. 3 and 4). The results shown in Table 2 reveal that the statistical significance of the three models is similar.

$$logBB = -0.0213 - 0.0143 * PSA + 0.241 * AlogP \qquad (2)$$

Table 3. Experimental, reported, and computed average (GPLS 3 and 4) logBB values for **1–100**[a].

Molecule	Expt	Ref. 12	Ref. 13	Ref. 20	Ref. 16	Eq. 2	Eq. 3	Eq. 4	GPLS 3	GPLS 4
1	−1.42	−0.78	−1.21	−1.07	−1.183	−1.227	−1.334	−0.786	−1.237	−0.789
2	−0.04	−0.82	−0.46	−0.88	−0.969	−1.104	−0.711	−0.539	−0.659	−0.532
3	−1.06	−0.8	−1.05	−0.32	−1.420	−0.777	−1.662	−1.421	−1.752	−1.514
4	0.49	−0.02	0.30	0.18	−0.127	0.399	0.554	0.453	0.556	0.364
5	0.83	0.24	0.87	0.71	0.605	0.864	1.336	0.779	1.267	0.700
6	−0.82	−1.55	−1.46	−1.67	−1.990	−1.903	−2.520	−2.197	−2.329	−2.146
7	−0.67	−0.56	−1.00	−0.81	−0.515	−0.593	−0.999	−0.821	−0.920	−0.695
8	−0.66	−0.51	−1.05	−0.92	−0.534	−0.834	−0.910	−0.689	−0.873	−0.598
9	−0.12	−0.27	−0.50	−0.58	−0.297	−0.325	−0.686	−0.831	−0.678	−0.751
10	−0.18	−0.93	−0.27	−0.56	−0.699	−0.682	−0.417	−0.641	−0.372	−0.620
11	−1.15	−1.16	−0.60	−1.05	−1.364	−1.294	−0.945	−0.694	−0.924	−0.673
12	−1.57	−1.26	−0.92	−1.09	−1.361	−1.364	−1.390	−1.244	−1.375	−1.167
13	−1.54	−1.51	−1.41	−1.43	−1.811	−1.581	−1.619	−1.918	−1.494	−1.887
14	−0.27	−0.34	−0.87	−0.75	−0.436	−0.500	−0.398	−0.457	−0.521	−0.533
15	−0.28	−0.34	−0.70	−0.81	−0.364	−0.597	−0.075	−0.454	−0.127	−0.470
16	−0.46	−0.08	−0.53	−0.21	−0.082	−0.262	−0.326	−0.042	−0.306	−0.108
17	−0.24	−0.07	−0.30	0.03	−0.036	0.244	−0.154	−0.153	−0.215	−0.227
18	−0.02	0.03	−0.30	−0.07	0.138	0.012	−0.071	−0.204	−0.096	−0.202
19	0.69	0.00	0.19	0.12	0.008	0.201	0.352	0.401	0.384	0.360
20	0.44	−0.38	0.16	0.00	0.011	0.089	−0.104	0.351	−0.058	0.332
21	0.14	−0.35	0.32	0.23	0.132	0.424	0.190	0.441	0.181	0.380
22	0.22	−0.13	0.15	0.08	−0.057	0.202	0.169	0.453	0.149	0.384
23	−2.00	−0.68	−1.61	−0.84	−1.368	−0.709	−1.416	−1.409	−1.333	−1.418
24	−1.30	−0.66	−0.93	−0.77	−1.376	−0.028	−0.995	−1.472	−1.017	−1.502
25	0.11	−0.11	−0.53	0.09	−0.122	0.093	−0.376	−0.554	−0.384	−0.575
26	−1.12	−0.54	−1.03	−0.97	−0.651	−0.817	−0.971	−0.756	−0.974	−0.705
27	−0.73	−0.43	−0.38	−0.64	−0.569	−0.253	−0.643	−0.875	−0.577	−0.859
28	−1.17	−0.78	−0.77	−0.71	−1.123	−1.019	0.214	0.584	0.172	0.581
29	−1.23	−0.49	−1.04	−0.98	−0.828	−0.733	−1.150	−0.817	−1.025	−0.836
30	−2.15	−0.51	−0.61	0.00	−0.728	−0.524	−1.350	−1.057	−1.237	−0.976
31, temelastine	−1.88				−1.158	−0.297	−1.005	−1.366	−0.955	−1.424
32, butanone	−0.08	0.26	−0.17	−0.09	0.151	0.020	0.187	0.231	0.168	0.230
33, benzene	0.37	0.39	0.49	0.52	0.549	0.471	0.757	0.477	0.680	0.743
34, 3-methylpentane	1.01	0.54	0.81	0.73	0.876	0.656	0.701	0.777	0.741	0.746
35, 3-methylhexane	0.90	0.55	0.82	0.80	0.887	0.752	0.718	0.711	0.768	0.690
36, 2-propanol	−0.15	0.21	−0.28	−0.15	−0.240	−0.174	0.008	0.051	−0.053	−0.019
37, 2-methylpropanol	−0.17	0.21	−0.17	−0.06	0.023	−0.071	−0.036	0.036	−0.034	−0.006
38, 2-methylpentane	0.97	0.54	0.81	0.73	0.766	0.656	0.698	0.730	0.729	0.725
39, 2,2-dimethylbutane	1.04	0.53	0.80	0.73	0.704	0.665	0.880	0.786	0.861	0.786
40, $F_3C\text{-}CH_2Cl$	0.08	0.46	0.17	0.49	0.327	0.426	0.113	0.323	0.217	0.278
41, 1,1,1-trichloroethane	0.40	0.38	0.55	0.51	0.342	0.470	0.320	0.397	0.372	0.395
42, diethyl ether	0.00	0.41	0.09	0.10	−0.002	0.044	0.102	0.521	0.080	0.531
43, enflurane	0.24	0.47	0.12	0.27	0.437	0.447	0.193	0.073	0.035	0.191
44, ethanol	−0.16	0.16	−0.38	−0.19	−0.233	−0.276	−0.138	−0.080	−0.204	−0.143
45, fluroxene	0.13	0.42	0.00	0.19	0.188	0.247	−0.215	0.354	−0.152	0.304
46, halothane	0.35	0.45	0.39	0.58	0.557	0.453	0.301	0.300	0.346	0.262
47, heptane	0.81	0.55	0.83	0.80	0.859	0.768	0.760	0.599	0.769	0.608
48, hexane	0.80	0.54	0.82	0.73	0.826	0.672	0.718	0.665	0.715	0.663

Table 3 (continued).

Molecule	Expt	Ref. 12	Ref. 13	Ref. 20	Ref. 16	Eq. 2	Eq. 3	Eq. 4	GPLS 3	GPLS 4
49, isoflurane	0.42	0.46	0.15	0.27	0.528	0.467	0.266	0.110	0.124	0.240
50, methylcyclopentane	0.93	0.52	0.80	0.67	0.673	0.535	0.766	0.894	0.777	0.873
51, pentane	0.76	0.53	0.80	0.67	0.789	0.577	0.695	0.735	0.681	0.714
52, propanol	−0.16	0.18	−0.27	−0.16	−0.062	−0.167	−0.112	−0.016	−0.145	−0.068
53, propanone	−0.15	0.21	−0.29	−0.16	−0.081	−0.142	0.085	0.219	0.045	0.206
54, teflurane	0.27	0.50	0.31	0.55	0.609	0.371	0.251	0.311	0.239	0.311
55, toluene	0.37	0.40	0.58	0.59	0.694	0.584	0.912	0.587	0.832	0.638
56, trichloroethene	0.34	0.38	0.60	0.49	0.582	0.502	0.302	0.180	0.266	0.228
57, acetylsalicylic acid	−0.50				−1.181	−0.493	−0.677	−1.059	−0.649	−0.966
58, valproic acid	−0.22				−0.815	0.107	−0.152	−0.508	−0.102	−0.565
59, salicyclic acid	−1.10				−1.209	−0.407	−0.308	−1.064	−0.301	−1.056
60, p-acetamidophenol	−0.31				−0.458	−0.549	−0.473	−0.514	−0.525	−0.501
61, chlorambucil	−1.70				−1.156	0.445	−1.393	−1.099	−1.474	−1.006
62	−1.30			−0.17	−0.036	−0.061	0.067	0.384	0.063	0.429
63	−1.40			−0.40	−0.292	−0.548	−0.159	0.568	−0.179	0.551
64	−0.43			−0.40	−0.263	−0.294	−0.617	−0.813	−0.622	−0.799
65	0.25			−0.06	0.170	0.045	0.054	−0.057	0.063	−0.110
66	−0.30			−0.18	−0.139	−0.287	−0.032	0.511	−0.078	0.546
67	−0.06			0.02	0.016	−0.031	0.350	0.595	0.326	0.650
68	−0.42			−0.55	−0.510	−0.532	−0.319	0.346	−0.293	0.392
69	−0.16				−0.285	−0.024	0.342	0.761	0.354	0.686
70	0.00	−0.14	−0.58	−0.01	−0.005	0.033	0.568	0.167	0.304	0.212
71	−0.34	−0.28	−1.11	−0.37	0.005	−0.318	0.064	0.182	−0.020	0.227
72	−0.30	−0.46	−0.75	−0.38	−0.447	−0.079	0.204	0.219	0.179	0.241
73	−1.34	−0.64	−0.99	−0.83	−0.931	−0.653	−0.443	−0.558	−0.445	−0.525
74	−1.82	−0.81	−1.35	−1.28	−1.308	−1.067	−1.289	−1.345	−1.256	−1.344
75	0.89	0.27	1.03	0.80	0.966	1.002	1.414	0.795	1.274	0.776
76, theophylline	−0.29				−0.512	−0.999	0.011	−0.104	−0.069	−0.041
77, caffeine	−0.06				−0.219	−0.690	0.141	−0.111	0.054	−0.024
78, antipyrine	−0.10				0.474	−0.101	0.287	0.334	0.032	0.211
79, ibuprofen	−0.18				−0.555	0.402	0.108	−0.537	0.107	−0.507
80, codeine	0.55				0.271	−0.173	−0.012	−0.583	0.000	−0.539
81, pentobarbital	0.12				−0.190	−0.432	−0.545	−2.125	−0.424	−1.967
82, alprazolam	0.04				0.332	0.641	0.400	−0.555	0.425	−0.582
83, indomethacin	−1.26				−1.030	−0.068	−1.633	−1.350	−1.744	−1.238
84, oxazepam	0.61				−0.476	−0.219	−0.743	−1.640	−0.855	−1.586
85, hydroxyzine	0.39				0.128	0.292	−0.440	−0.380	−0.416	−0.378
86, despramine	1.20				0.426	0.585	0.943	0.689	0.906	0.691
87, midazolam	0.36				0.400	0.719	−0.139	−0.577	−0.166	−0.551
88, verapamil	−0.70				−1.111	0.401	−0.714	−0.206	−0.511	−0.193
89, promazine	1.23				0.83	0.643	0.838	0.624	0.765	0.678
90, chlorpromazine	1.06				0.710	0.767	0.735	0.497	0.689	0.517
91, trifluoroperazine	1.44				0.459	0.738	0.311	0.501	0.283	0.564
92, thioridazine	0.24				1.062	0.866	0.708	0.657	0.669	0.703
93, BCNU	−0.52				−0.57	−0.209	−0.877	−1.618	−0.761	−1.487
94, phenserine	1.00				0.230	0.361	1.000	0.191	0.901	0.214
95, physostigmine	0.08				0.007	−0.077	0.616	0.234	0.526	0.286
96, tertbutylchlorambucil	1.00				−0.227	0.909	−0.939	−0.573	−0.880	−0.552

Table 3 (continued).

Molecule	Expt	Ref. 12	Ref. 13	Ref. 20	Ref. 16	Eq. 2	Eq. 3	Eq. 4	GPLS 3	GPLS 4
97, didanosine	−1.30				−0.816	−1.282	−1.115	−2.790	−1.174	−2.714
98, zidovudine	−0.72				−1.024	−1.679	−1.227	−2.369	−1.323	−2.305
99, nevirapine	0.00				−0.285	−0.463	−0.076	0.286	−0.064	0.328
100, SB-222200	0.30				0.426	0.879	0.723	0.018	0.730	−0.001

[a]The computed logBB values correspond to the average of the first 50 equations among the population of 100 equations derived for each of the G/PLS QSAR models. The prediction set outliers are underlined.

$$logBB = -0.0204 + 0.122 * S\_sssN - 0.114 * Rotlbonds + 0.0359 * Jurs - WNSA - 3 - 0.0615 * S\_dsN + 0.313 * AlogP - 0.0959 * S\_sssCH + 0.108 * R_{og} \quad (3)$$

$$logBB = 0.182 - 0.0410 * S\_dO - 0.0896 * Rotlbonds - 0.0501 * Jurs - RPCS + 3.707 * Jurs - FPSA - 3 - 0.0645 * S\_sOH - 0.276 * S\_dsN + 0.000985 * Jurs - TASA \quad (4)$$

The results obtained using 7 variables with AlogP (Eq. 3) are better than the model derived without AlogP (Eq. 4). In general, Eqs. 2–4 provide good quality logBB predictions for many molecules (Table 2) but Eq. 3 performs in a balanced manner for small molecules ranging from ethanol (**44**) to compounds like verapamil (**88**) whose molecular weight exceeds 450. The coefficients and intercept of Eq. 2 are quite similar to those reported for a similar equation in Ref. 20. This is in spite of the fact that the details of the 3D structures and logP calculations differ.

To verify that the key descriptors (Table 4) for predicting logBB are captured by the best G/PLS models and that individual QSAR models represent a larger set, the logBB was computed for each of the first 50 model equations obtained from each G/PLS run. The statistical results of the average logBB values (GPLS 3 and 4) resulting from the 50 equations with the experimental data revealed similar prediction accuracies (Table 2) to the best G/PLS equations reported in the present study. These results also show that there are no significant deviations in the computed logBB among the 50 models for any of the molecules studied here. Also, we found no variation higher than about 10% for the ratio of the standard deviation of the coefficients of the independent variables to the average value of the

coefficients. Therefore the statistical validity of any of the 50 model equations taken individually should be similar to that of Eqs. 3 and 4. This suggests that the best G/PLS equations (3 and 4) are sufficient to represent the whole population of the 100 equations derived. It also demonstrates that there are multiple equations with similar prediction accuracies describing this dataset. In fact, the addition, deletion, or re-determination of logBB values for this dataset would likely change the order of the equations, but not necessarily the accuracy. For the prediction set, it is interesting that the average r^2 and the rms error decrease small amounts for Eq. 3. This may indicate that averaging the 50 equations may provide better accuracy because of the higher amount of information contained. However, the increased complexity is unlikely to be worth the small increase in accuracy.

Statistical results for the prediction set

Although the internal cross-validation results (cv-r^2, Table 2) are quite satisfactory, the success of a model should be judged on its predictive ability for molecules that were not part of the original training set. To achieve this, Eqs. 2–4 were used to compute the logBB values for 39 molecules (**62–100**) used as a validation set by Luco.

For each model we report the predictive ability on the 39 compounds of the prediction set. We obtain errors (Eq. 3) that are in good agreement with those in literature reports where the rms error was 0.235 and 0.541 for two collections of compounds of the 39 prediction set [16], mean absolute errors 0.13 and 0.24 for 5 compounds in each prediction set [20]. In each of Ref. 16 and Ref. 20, outliers were eliminated before calculation of errors. A further report of predictive quality found equations that produced mean absolute errors of 0.41 for 17 compounds [18]. Mean absolute errors for a variety of other models are reported in Ref. 18 and these models vary from 0.38 to 1.24. The statistical significance obtained for the average

Table 4. Frequency of the variables that appear in the 100 G/PLS Eqs. derived for the training set through Eqs. 3 and 4.

Descriptor	Family	Description	GPLS 3	GPLS 4
S_dsN (=N-)[a]	Electro_topological	N connected by a double and single bond	60	100
S_sssN (>N-)[a]	Electro_topological	N connected by three single bonds	65	12[d]
S_sOH (-OH)[a]	Electro_topological	OH connected by a single bond		100
S_dO (=O)[a]	Electro_topological	O connected by a double bond		99
S_sssCH (>CH-)[a]	Electro_topological	CH connected by three single bonds	45	
S_dsCH (=CH-)[a]	Electro_topological	CH connected by a double and single bonds	34[d]	
S_aaCH (..CH..)[a]	Electro_topological	CH connected by two aromatic atoms		16[d]
S_dssC (>C=)[a]	Electro_topological	CH connected by a double and two single bonds		16[d]
R_{og}[b]	Spatial	Radius of gyration	45	
Jurs-FPSA-3[c]	Spatial	Fractional charged partial surface areas		32
Jurs-WNSA-3[c]	Spatial	Surface weighted charged partial surface area	100	
Jurs-RPCS[c]	Spatial	Relative positive charge surface area	23[d]	35
Jurs-RNCG[c]	Spatial	Relative negative charge		43[d]
Jurs-RPCG[c]	Spatial	Relative positive charge		34[d]
Jurs-DPSA-1[c]	Spatial	Difference in charged partial surface areas		20[d]
Jurs-TASA[c]	Spatial	Total hydrophobic surface area		33
H-bond acceptor	Structural	Number of hydrogen bond acceptor atoms	29[d]	
Rotlbond	Structural	Number of rotatable bonds	99	99
AlogP	Thermodynamic	Ghose and Crippen logP	100	

[a]Electro_topological descriptors. The symbols for bond types are: single, s (−); double, d (=); aromatic, a (...). The initial S stands for the summation of the electrotopological indices for all atoms of that specified type.
[b]The radius of gyration, $R_{og} = \sqrt{(S((x_i^2 + y_i^2 + z_i^2)/N))}$ where x_i, y_i, and z_i are the atomic coordinates of the ith atom with respect to the center of mass and N is the number of atoms in the molecule.
[c]Component of the Jurs descriptors based on partial charges mapped on surface area.
[d]Additional descriptors observed in the population of the 7-variable 100 G/PLS equations.

Table 5. Qualitative prediction accuracy of CNS permeability ($CNS^{+/-}$, %) using Eqs. 2–4[a].

Statistical model	Training set		Prediction set		Ref. 25[b]		Synthline® [b]	
	CNS^+	CNS^-	CNS^+	CNS^-	CNS^+	CNS^-	CNS^+	CNS^-
Ref 25					91.2	73.3		
LogBB cutoff = −1								
N	47	14	33	6	91	90	389	1662
Equation 2	95.7	35.7	97.0	33.3	100	24.4	81.5	48.8
Equation 3	97.9	64.3	97.0	50.0	85.7	46.7	71.7	64.0
Equation 4	95.7	64.3	87.9	50.0	79.1	48.9	87.9	27.2
LogBB cutoff = −0.75								
N	46	15	33	6				
Equation 2	95.7	53.3	93.9	50.0	81.3	34.4	72.8	57.2
Equation 3	95.7	86.7	90.9	50.0	75.8	53.3	63.2	71.8
Equation 4	91.3	86.7	84.8	66.7	69.2	58.9	81.5	36.2
LogBB cutoff = −0.40								
N	41	20	28	11				
Equation 2	87.8	75.0	85.7	54.5	67.0	52.2	55.3	69.3
Equation 3	90.2	85.0	85.7	72.7	57.8	68.9	47.3	82.2
Equation 4	80.5	90.0	75.0	63.6	48.4	80.0	69.2	51.4

[a]Outliers in the training and prediction sets are also included.
[b]Classification as reported in the supporting information of Reference 25.

predicted logBB values (GPLS 3 and 4) is consistent with the r^2 (pr) values obtained by Eqs. 3 and 4. As evidenced from the r^2 (pr) values, the model that utilizes AlogP as one of the principal parameters results in better logBB prediction (Eq. 3) compared to the model without AlogP (Eq. 4). In fact, Eq. 4 has the lowest r^2 (pr) and is likely to be unreliable for quantitative logBB estimations.

Equation 3 best balances the number of variables in the equations and the accuracy of the equations in the present study.

Qualitative prediction of CNS activity

Even though an accurate prediction of logBB is desirable, it is often advantageous to evolve a qualitative means of discriminating CNS^- and CNS^+ compounds [37]. In this study, compounds with logBB values < -1 were treated as poor brain penetrators (CNS^-) and the remainder as CNS^+ molecules. Accordingly, Eq. 3 classified 90% of the molecules correctly with an improved overall performance than models reported earlier (52/63 [12], 54/63 [13], and 58/69 [20]) for qualitative CNS permeation prediction.

A second qualitative prediction set was obtained from Ref. 25. These results are reported in Table 5. In this case the 181 compounds are nearly equally divided between compounds known to penetrate the CNS and those that are not able to penetrate. The CNS actives are predicted more reliably. Similar to the smaller 39 compound prediction set we find that the simplest model, Eq. 2, has a tendency to poorly predict CNS inactives. A summary of qualitative predictions on the Synthline database is also provided in Table 5. In this case Eq. 3 provides significantly better results on non-CNS drugs. In general, the trend across the various validation sets is that CNS permeable or CNS drugs are predicted to be permeable 72–98% of the time by Eq. 3. Compounds that are non-CNS permeable or are not CNS drugs are predicted correctly 50–64% of the time by Eq. 3. The trend is similar for Eqs. 2 and 4; however, for CNS^- compounds the Eqs. 2 and 4 are significantly poorer predictors. Using less stringent criteria for $CNS^{+/-}$ of -0.4 or -0.75 are also reported. This cutoff significantly reduces the error on CNS^- compounds but increases the error on CNS^+ compounds. Equation 3 is significantly improved for CNS^- compounds.

Discussion

The G/PLS results presented in this study are consistent in reproducing the experimental logBB but with fewer terms in the equations as compared to some models. The G/PLS results are comparable to that obtained by Luco [16] and slightly to Norinder et al. [13], who used nearly the same set of molecules in their training set, but derived their statistical relationship with 18- and 14-variable equations, respectively. Many of the early quantitative models relied on fairly small prediction sets. While we find that about 20% of our prediction set are outliers, the remaining compounds are predicted to have an rms error of less than 0.5 log units. Our results are also improved slightly over Clark's [20] 2-term ((C or M) logP and PSA) logBB model. Although Eq. 2 and similar models are very simple, the results may be unreliable when extended to nonpolar compounds where PSA is 0. In a 2-term QSAR model [20] such as Eq. 2, the computed logBB is highly dependent on the PSA and the calculated logP terms. Inclusion of other variables in the model may allow models such as Eq. 3 to account for other effects in blood–brain transport. We speculate that some of these additional terms may indicate active transport or efflux for some compounds. Moreover, calculated logP's may be a liability for models used to predict new compounds where functionality may not be consistently parameterized. Under these circumstances, models such as those derived by Luco [16] and Eq. 4 prove to be advantageous. In these more complex equations the importance of rotatable bonds, molecular weight and other terms is indicated.

Interpretation of the physicochemical descriptors

The importance of the parameters obtained in the various models can be assessed from the frequency of their occurrence in the 100 model equations evolved during each G/PLS runs. In comparing the frequency of the descriptor variables (Table 4), it is very clear that there are terms that are essential but are not capable of explaining all the experimental logBB values. Attempts to obtain a QSAR model with descriptor variables that occurred in all of the 100 G/PLS equations resulted in poorer r^2 for the training set. Therefore, we believe that the additional terms involved in these equations encode some physicochemical information for predictions across diverse structural classes at a moderate statistical cost. Of significant interest are the descriptor variables that indicate a sizeable contribution from

AlogP, number of rotatable bonds, Jurs-terms [38], and the electrotopological connectivity indices [39] (Table 4). These variables exemplify the terms essential for promoting transport between the blood and brain compartments. For example, AlogP implies that hydrophobicity is important; while the number of rotatable bonds describes the conformational flexibility effects (shape) on logBB. The topological identity of the basic nitrogen in the regression equations (e.g.: S_sssN = tertiary nitrogen, Table 4) is similar to one of the pharmacophore elements suggested as a requirement for CNS activity by Lloyd and Andrews [40]. Likewise, components of the Jurs-descriptors identified in Eqs. 3 and 4 relate to the importance of PSA and H-bond donor/acceptor count [20, 23, 25] suggested for logBB computation and $CNS^{-/+}$ prediction. The derived QSAR models reveal some of the appropriate coding $(-/+)$ of the (de)stabilizing factors that affect the blood–brain barrier. For example, the positive coefficient on AlogP implies that compounds or substituents that enhance octanol-water partition (hydrophobic) also improve CNS penetration. In contrast, conformational flexibility (Rotlbonds) and polar substitutions (S_dsN, S_sOH, S_dO, and PSA) decrease BB permeability. The coefficients on the various electrotopological indices indicate that logBB is likely influenced by the nature of the hydrogen bonding group. The effect of hydrogen bond acceptors (e.g. S_dsN, a nitrogen with a double and single bond) is different than hydrogen bond donors and aliphatic amines (e.g. S_sOH, hydroxyl group and S_sssN, aliphatic amine).

The derived QSAR relationships also highlight that logBB depends on other molecular properties in addition to the partition coefficients. In agreement with earlier studies [7, 9, 41], a plot of the computed AlogP with the experimental logBB for the training set does not reveal any correlation ($r^2 = 0.030$) between these two entities. It has been determined [23] that the PSA of CNS^+ molecules are below 90 Å^2. A low PSA (or low polarity) may only be a necessary but not sufficient condition for CNS penetration. Other 1D physicochemical properties like the H-bond donor/acceptor counts and molecular weight did not play a role for quantitative logBB predictions when combined with other descriptors in this study. The Jurs descriptors, PSA, and electrotopological indices appear to replace atom counts found by others to be important [13, 16, 23, 25]. Replacement of atom counts and other discrete variables by continuous variables also allows the effect of substitution to modulate the BB permeation smoothly. The performance of these models was compared against $CNS^{-/+}$ molecules individually. In general, the CNS^- molecules are less accurately predicted compared to CNS^+ compounds (Table 5). This could be due to the poor representation of CNS^- molecules compared to CNS^+ compounds in the training set ($\sim$ a 1:4 ratio). Also, the QSAR models derived may not have adequate variables describing the physicochemical features of CNS^- molecules. For instance, the predictive ability of Eq. 2 or Clark's model [20] for CNS^- compounds is much inferior compared to the QSAR model expressed in Eqs. 3 and 4. This suggests that the PSA and calculated LogP (whether AlogP, ClogP or MlogP) may be limited to predicting CNS active compounds correctly, but inappropriate for CNS inactive molecules.

Predictions for the second validation set

Another result pertains to the predictive ability of the present model for the diverse set of molecules compiled by Ajay et al. [25]. Since Eq. 3 provides a balanced prediction for both $CNS^{-/+}$ classes of compounds (Table 4) it was used for qualitative predictions. The limitation of the present approach can be seen from the poor logBB prediction for molecules like neostigmine and minoxidil because the training set did not include compounds with quaternary amines or the N-oxide functional groups. In general, all the QSAR models reported here (Eqs. 2–4) predict the CNS^+ compounds with accuracies similar to Ajay et al. [25]. A single equation is used in the present study for $CNS^{-/+}$ prediction. This clearly justifies the appropriateness of the present models in being extended to unknowns. The discrimination of $CNS^{-/+}$ compounds based on therapeutic classes is not free from ambiguity, since CNS penetration may be dose dependent [23]. Another cause for misclassification could be the cut-off values used for differentiating $CNS^{-/+}$. For instance, using BB>1.0 (logBB>0) as the threshold for CNS^+ molecules, acetylsalicylic acid (**57**), ibuprofen (**79**), theophylline (**76**), verapamil (**88**), and zidovudine (**98**) were classified as CNS^- compounds [25], while they actually permeate the CNS based on their experimental logBB values and if a logBB >-1 cut-off is used. In the absence of experimental logBB values, it is difficult to identify misclassifications for the second validation set of 181 molecules. Consequently, a rule-of-thumb, similar to the rule-of-5 [42] and van de Waterbeemd et al.'s upper bound [23] was used to identify potential misclassifications. From an analysis of the 1D descriptors in **1**–

100, it is predicted that CNS$^-$ compounds should have at least one H-bond donor and two H-bond acceptor atoms [43]. Using this as a filter, 8 compounds are assumed misclassified [44]. Another potential source of misclassifications is analgesics and antihistamines. Although not all antihistamines are CNS permeable, some drugs like doxylamine, ebastine, and pheniramine cause drowsiness. Because drugs such as these may reach the CNS and induce CNS related pharmacological effects, it may be appropriate to classify these compounds as CNS$^+$. However, our false positive rate (CNS permeation predicted when no CNS activity is observed) remains at over 50% for Eqs. 3 and 4 using a cutoff of -1.0. Adjusting the CNS$^{+/-}$ cutoff improves the false positive rate. For this set and for the larger Synthline set of compounds -0.75 appears to be the best choice. Another view of this data can be obtained by examining the median predicted logBB for CNS$^+$ versus CNS$^-$. For Eq. 2, the median predicted logBB for CNS$^+$ is 0.00 while it is -0.40 for CNS$^-$ compounds. For Eq. 3, the medians are -0.30 (CNS$^+$) and -0.96 (CNS$^-$). Equation 4 provides medians of -0.42 (CNS$^+$) and -0.99 (CNS$^-$). All three equations properly predict CNS$^+$ compounds to have higher logBB and CNS$^-$ compounds to have lower logBB. Further, Eq. 3 provides the greatest separation in medians between the two activity groups. This in turn is consistent with observed effects of changing the logBB cutoff to estimate CNS activity. Overall this analysis confirms that Eq. 3 has validity outside of the original set of 58 compounds and that its predictions most nearly match the CNS classifications.

Consensus prediction of CNS activity

One of the limitations of the derived models (Eqs. 2–4) is the large probability of error in the classification of inactive compounds. For instance, given the 50% error in the discrimination of the CNS$^-$ compounds in the validation sets (Table 5), it is more likely that an incorrectly predicted CNS$^-$ molecule will be treated as CNS$^+$ (false positive). To overcome this limitation in applying only one model for computing logBB and to optimize the prediction performance, we investigated a consensus prediction scheme as an approach to improve the confidence level in the CNS$^{-/+}$ hit rates. In this process, the influence of a single descriptor variable on the computed logBB for a given model can be minimized and some errant predictions eliminated. Similarly, borderline cases (logBB >-0.95 and <-1.05) can be treated more appropriately such

that possible mispredictions are avoided. We define a prediction to be 100% likely (true positive or true negative) if all the models classify the molecule alike and vice versa. Another consensus approach is to choose the median of the three equations as the predicted logBB. Unfortunately, neither the median nor consensus estimation significantly improves the qualitative predictions of CNS impermeable compounds over Eq. 3. For some of the prediction sets we find Eqs. 2 and 4 particularly poor, however, the results are not improved by combining with predictions by Eq. 3.

Analysis of recent drugs and drug candidates database

The Synthline$^\circledR$ database contains approximately 2000 drugs that have been recently introduced into human clinical trials or launched as newly marketed drugs [30]. It took approximately 2 h of CPU time on a Silicon Graphics R12000 processor to generate the 3D structures and all necessary descriptors for all three equations. Interestingly, the prediction rates do not vary significantly if only a given clinical phase is analyzed (e.g., Phase III, Phase II, Phase I, or preclinical). When the consensus scheme is used, we obtain similar results to the 181 compound set. However, Eq. 3 performs best for all the qualitative prediction sets and achieves a 64% success rate on non-CNS drug compounds. Using a -0.75 cutoff for predicted logBB, we are more successful at predicting non-CNS drugs correctly. Also, the median logBB value for CNS$^+$ compounds is -0.30 and -0.97 for CNS$^-$ compounds. The results on this database clearly indicate that the predicted logBB using Eq. 3 can indicate CNS permeability for known CNS agents. It also predicts that the blood–brain permeability of non-CNS drugs is significantly lower.

Conclusions

In order for computer models to be useful in the early screening process they need to have well-understood accuracy and validity across a diverse set of compounds. The present study provides logBB estimates for a wide range of compounds that belong to different structural classes. Multiple QSAR equations derived through G/PLS regressions result in logBB values that are consistent with experimental data. The close correlations of the experimental logBB values with those computed from the average of 50 model equations

suggest that self-consistent QSAR models have been derived. As with any QSAR model, there is always a compromise between complexity and accuracy. In arriving at a practical model for quantitative logBB prediction, Eq. 3 is most effective.

The independent variables selected through the G/PLS calculations indicate properties that are reasonably expected to explain BB permeability. For instance, comparing the predictive ability of Eqs. 3 and 4 reveals the significance of AlogP for quantitative logBB prediction. Similarly, a limitation of using only the PSA and calculated logP terms for logBB prediction is suggested by the poor success rate for the CNS^- compounds. Unlike the choice of 1D descriptors for qualitative predictions ($CNS^{+/-}$), the use of continuous variables is probably more appropriate for predicting continuous measurements like logBB. Further, Eqs 3 and 4 show that the nature of polar atoms (hydrogen bond acceptors vs donors) is important. The result obtained through the use of simple descriptors is promising for screening large compound libraries at modest computational cost without sacrificing the prediction accuracy significantly. By including additional 2D and different 3D terms in Eq. 3 there is higher accuracy and probably lower dependence on the conformational choice of 3D structures for accuracy. Indeed, we have applied our equations to a 2000-member database of drug molecules to predict BB permeability with considerable success using only one 3D conformation.

When looking at the qualitative patterns of CNS activity, most of the derived QSAR models classify the CNS^+ molecules correctly. We found that the simplest models such as Eq. 2 are not predictive of CNS impermeable compounds and that more complicated models such as Eq. 3 are necessary. The large number of false positives in all cases is a weakness of these models. While adjusting the logBB cutoff for CNS activity is helpful in improving CNS^- predictions, it will be difficult to ascertain the true applicability of these models without further experimental measurements. We believe that there are active transport mechanisms that we have not identified in these models. We plan to investigate this area in future work.

Thus, the QSAR models examined here are expected to be a guiding tool for analog design to improve CNS penetration and for de novo modeling of $CNS^{-/+}$ libraries. Even though logBB is only one factor in determining CNS activity/inactivity we find that the predicted values correlate well with known CNS permeability (known CNS activity). The equations are very accurate for CNS permeable compounds. Eqs. 3 and 4 do much better at predicting CNS impermeability than two-term models, though still not better than 50%. The poorer relationship with CNS inactivity is likely due to many effects not accounted for in our models. To properly account for all the factors involved in BB permeability we will need many more measurements on impermeable compounds. The models we have derived indicate that there are many possible models comparable to Eqs. 3 and 4 and it can therefore be difficult to choose the best model. In practical uses, we examine predictions from all the models. We continue to examine other means of modeling the data including using new descriptors such as those used in Rose et al. [17] and including other data as it becomes available.

Acknowledgements

References

1. (a) Gallop, M.A., Barrett, R.W., Dower, W.J., Fodor, S.P.A. and Gordon, E.M., J. Med. Chem., 37 (1994) 1385. (b) Ajay, Walters, W.P. and Murcko, M.A., J. Med. Chem., 41 (1998) 3314. (c) Sadowski, J. and Kubinyi, H., J. Med. Chem., 41 (1998) 3325.
2. (a) Tarbit, M.H. and Berman, J., Curr. Opin. Chem. Biol., 2 (1998) 411 and references cited therein. (b) Clark, D.E. and Pickett, S.D., Drug Discov. Today, 5 (2000) 49 and references cited therein.
3. (a) Goldstein, G.W. and Betz, A.L., Sci. Am., 255 (1986) 74. (b) Greig, N.H., Brossi, A., Pei, X.F., Ingram, D.K., and Soncrant, T.T. (Eds.), In New Concepts of a Blood-Brain Barrier; Greenwood, J., et al., Plenum, New York, NY, 1995, pp. 251–264. (c) deVries, H.E., Kuiper, J., De Boer, A.G., van Berkel, T.J.C. and Breimer, D.D., Pharmacol. Rev., 49 (1997) 143. (d) Partridge, W.M., J. Neurochem., 70 (1998) 1781. (e) Pardridge, W.M., Drug Discov. Today, 6 (2001) 1.
4. (a) Tamai, I. and Tsuji, A., Adv. Drug. Deliv. Rev., 19 (1996) 401. (b) Eddy, E.P., Maleef, B.E., Hart, T.K. and Smith, P.L., Adv. Drug Deliv. Rev., 23 (1997) 185. (c) Reichel, A. and Begley, D.J., Pharm. Res., 15 (1998) 1270. (d) Cecchelli, R., Dehouck, B., Descamps, L., Fenart, L., Buee-Scherrer, V., Duhem, C., Landquist, S., Rentfel, M., Torpier, G. and Dehouck, M.P., Adv. Drug Deliv. Rev., 36 (1999) 165.
5. (a) Hansch, C., Björkroth, J.P. and Leo, A., J. Pharm. Sci., 76 (1987) 663. (b) Gupta, S.P., Chem. Rev., 89 (1989) 1765. (c) Basak, S.C., Gutte, B.D. and Drewes, L.R., Pharm. Res., 13 (1996) 775.
6. Kansy, M. and van de Waterbeemd, H., Chimia, 46 (1992) 299.

7. Young, R.C., Mitchell, R.C., Brown, T.H., Ganellin, C.R., Griffiths, R., Jones, M., Rana, K.K., Saunders, D., Smith, I.R., Sore, N.E. and Wilks, T.J., J. Med. Chem., 31 (1988) 656.

8. Levin, V.A., J. Med. Chem., 23 (1980) 682.

9. (a) Chadha, H., Abraham, M.H. and Mitchell, R.C., Proceedings of an International Conference on the New Concept of a Blood-Brain Barrier, London, 1994. (b) Seelig, A., Gottschlich, R., and Dervent, R.M., Proc. Natl. Acad. Sci. USA, 91 (1994) 68.

10. (a) Abraham, M.H., Chadha, H.S. and Mitchell, R.C., J. Pharm. Sci., 83 (1994) 1257. (b) Abraham, M.H., Chadha, H.S. and Mitchell, R.C., Drug Des. Discov., 13 (1995) 123.

11. Keserü, G.M. and Molnár, L., J. Chem. Inf. Comput. Sci., 41 (2001) 210.

12. Lombardo, F., Blake, J.F. and Curatolo, J.W., J. Med. Chem., 39 (1996) 4750.

13. Norinder, U., Sjöberg, P. and Österberg, T., J. Pharm. Sci., 87 (1998) 952.

14. (a) Wold, S., Johansson, E. and Cocchi, M. (Eds.), PLS – Partial Least Squares Projection to Latent Structures. In: 3D QSAR in Drug Design; Kubinyi, H. (Ed.) ESCOM, Leiden, The Netherlands, 1993, pp. 523–550. (b) Wold, S. (Ed.), In Chemometric Methods in Molecular Design; van de Waterbeemd, H., VCH, Weinheim, Germany, 1995, pp. 195–218.

15. Crivori, P., Cruciani, G., Carrupt, P. and Testa, B., J. Med. Chem., 43 (2000) 2204.

16. Luco, J.M., J. Chem. Inf. Comput. Sci., 39 (1999) 396.

17. Rose, K., Hall, L.H. and Kier, L.B., J. Chem. Inf. Comput. Sci., 42 (2002) 651.

18. Lobell, M., Molnár, L. and Keserü, G.M., J. Pharm. Sci., 92 (2003) 360.

19. Liu, R., Sun, H. and So, S.S., J. Chem. Inf. Comput. Sci., 41 (2001) 1623.

20. Clark, D.E., J. Pharm. Sci., 88 (1999) 815.

21. Clark, D.E., J. Pharm. Sci., 88 (1999) 807.

22. (a) ClogP. Daylight Chemical Information Software, version 4.51. Daylight Chemical Information Inc., Mission Viejo, CA. (b) Moriguchi, I., Hirono, S., Liu, Q., Nakagome, I. and Matsushita, Y., Chem. Pharm. Bull., 40 (1992) 127.

23. van de Waterbeemd, H., Camenisch, G., Folkers, G., Chretien, J. and Raevsky, O., J. Drug Target, 6 (1998) 151.

24. Fischer, H., Gottschlich, R. and Seelig, A., J. Membr. Biol., 165 (1998) 201.

25. Ajay, Bemis, G.W. and Murcko, M.A., J. Med. Chem., 42 (1999) 4942.

26. Comprehensive Medicinal Chemistry Release 94.1 is available from MDL Information Systems Inc., San Leandro, CA.

27. MACCS-II Drug Data Report is available from MDL Information Systems Inc., San Leandro, CA.

28. (a) Hansch, C., Leo, A. and Hoekman, D., Exploring QSAR, Vol. 1. Fundamentals and Applications in Chemistry and Biology; Vol. 2. Hydrophobic, Electronic, and Steric Constants; American Chemical Society, Washington, DC, 1995. (b) Structure-Property Correlations in Drug Research; van de Waterbeemd, H., VCH, Weinheim, Germany, 1995, Vol. 3.

29. (a) Rogers, D. and Hopfinger, A.J., J. Chem. Inf. Comput. Sci., 34 (1994) 854. (b) Rogers, D. (Ed.), In Genetic Algorithms in Molecular Modeling; Devillers, J., Academic, London, 1996.

30. Prous Science Synthline® Drug Synthesis Database on CD-ROM, Barcelona, Philadelphia, Feb. 2000 release.

31. Structures that resulted in inappropriate conformations (cis-amide bond or skewed conformation for the cyclohexyl ring) were identified and re-optimized.

32. Rappe, A.K., Casewit, C.J., Colwell, K.S., Goddard, W.A. and Skiff, W.M., J. Am. Chem. Soc., 114 (1992) 10024.

33. The PSA arising out of the oxygen, nitrogen, and halogen atoms and the OH, NH_2 groups was calculated using the solvent accessible surface area in Cerius^2 with the van der Waals radii reported in references 20 and 21. In this case, the van der Waals surface area is calculated and therefore the water probe radius is set to 0.0.

34. Wessel, M.D., Jurs, P.C., Tolan, J.W. and Muskal, S.M., J. Chem. Inf. Comput. Sci., 38 (1998) 726.

35. Jacobs, M.H. Diffusion Processes; Springer-Verlag, Berlin, Germany, 1935, p. 13.

36. Viswanadhan, V.N., Ghose, A.K., Revankar, G.R. and Robins, R.K., J. Chem. Inf. Comput. Sci., 29 (1989) 163 and references cited therein.

37. It is important to note that these models merely predict the permeation across the BB and is by no means a measure of the compounds potency at the CNS or other ligand binding sites. Many other determinants affect whether a drug would be a CNS agent, including metabolism in the CNS, efflux processes or failure to bind to receptor in the CNS, none of which would be described in logBB models. Further, some agents may permeate the CNS but have little noticeable effect and therefore their activities are not labeled in such a way as to be termed CNS active. A discussion of these caveats is found in the work of reference 15.

38. Stanton, D.T. and Jurs, P.C., Anal. Chem., 62 (1990) 2323.

39. Hall, L.H., Kier, L.B. and Brown, B.B., J. Chem. Inf. Comput. Sci., 35 (1995) 1074.

40. Lloyd, E.J. and Andrews, P.R., J. Med. Chem., 29 (1986) 453.

41. (a) Kaliszan, R. and Markuszewski, M., Int. J. Pharm., 145 (1966) 9. (b) Chikhale, E.G., Ng, K., Burton, P.S. and Borchardt, R.T., Pharm. Res., 3 (1994) 412.

42. Lipinski, C.A., Lombardo, F., Dominy, B.W. and Feeney, P.J., Adv. Drug Deliv. Rev., 23 (1997) 3.

43. Refer to Cerius^2 users manual for the types of atoms considered as hydrogen bond donor and acceptor. In version 4.0, a hydroxyl is both a hydrogen bond acceptor (HBA) and a hydrogen bond donor (HBD). A carbonyl oxygen is only an HBA; an amide N is a HBD and not an HBA. An N atom with less than 3 bonds is an HBA, a tetrahedral N with 3 bonds and sum of bond orders equal to 3 is an HBA and all N atoms with less than 3 bonds is an HBA.

44. Compounds Amiodarone, Coumarin, Diltiazem, Doxylamine, Ebastine, Loratadine, Neostigmine, and Pheniramine do not have any H-bond donor atoms.

Supplementary Information:

Table 1S.

Computed logBB values from Equations 2–4, experimental, reported, and predicted CNS activity ($CNS^+ = 1$; $CNS^- = 0$) for the 2nd validation set of 181 molecules.

$$logBB = -0.0213 - 0.0143*PSA + 0.241 * AlogP \tag{2}$$

$$logBB = -0.0204 + 0.122 * S\_sssN - 0.114 * Rotlbonds + 0.0359 * Jurs - WNSA - 3 \\ -0.0615 * S\_dsN + 0.313 * AlogP - 0.0959 * S\_sssCH + 0.108 * R_{og} \tag{3}$$

$$logBB = 0.182 - 0.0410 * S\_dO - 0.0896 * Rotlbonds - 0.0501 * Jurs - RPCS \\ +3.707 * Jurs - FPSA - 3 - 0.0645 * S\_sOH - 0.276 * S\_dsN + 0.000985 * Jurs - TASA \tag{4}$$

Molecule	logBB (2)	logBB (3)	logBB (4)	$CNS^{+/-}$ (Expt)	$CNS^{+/-}$ (Ref 25)	$CNS^{+/-}$ (2)	$CNS^{+/-}$ (3)	$CNS^{+/-}$ (4)
Abecarnil	−0.001	−0.323	−0.160	1	1	1	1	1
Acarbose	−5.308	−6.594	−9.487	0	0	0	0	0
Acrivastine	0.416	0.348	−0.579	0	1	1	1	1
Alclofenac	0.046	−0.726	−0.774	0	1	1	1	1
Alprenolol	0.110	−0.531	−0.469	0	0	1	1	1
Amantadine	−0.142	0.041	0.748	1	1	1	1	1
Amiodarone	1.419	−0.518	−0.520	0	0	1	1	1
Amitriptyline	1.005	1.373	0.800	1	1	1	1	1
Amobarbital	−0.623	−0.700	−0.725	1	1	1	1	1
Ampicillin	−1.511	−1.599	−1.867	0	0	0	0	0
Apomorphine	0.110	0.493	−0.581	1	1	1	1	1
Asimadoline	0.495	−0.199	−0.903	0	1	1	1	1
Astemizole	0.662	−0.064	0.498	0	1	1	1	1
Atenolol	−1.069	−1.524	−0.978	0	0	0	0	1
Baclofen	−0.567	−1.120	−0.765	1	1	1	0	1
Betaxolol	−0.101	−1.461	−0.646	0	0	1	0	1
Biperiden	0.540	0.533	−0.216	1	1	1	1	1
Bufuralol	0.169	0.087	−0.132	0	0	1	1	1
CGP-37849	−1.829	−2.105	−2.312	1	1	0	0	0
CGP-39551	0.646	1.045	0.964	1	0	1	1	1
CGS-19755	−0.450	−0.383	−2.074	1	1	1	1	0
Captopril	−0.700	−1.059	−1.195	0	1	1	0	0
Carbenicillin	−1.501	−2.199	−3.071	0	0	0	0	0
Carbidopa	−1.481	−1.540	−2.000	0	1	0	0	0
Carebastine	0.563	−1.029	−1.435	0	0	1	0	0
Carmoxirole	0.233	0.244	−0.716	0	1	1	1	1
Cefetametpivoxil	−2.268	−2.726	−2.981	0	0	0	0	0

Molecule	logBB (2)	logBB (3)	logBB (4)	CNS$^{+/-}$ (Expt)	CNS$^{+/-}$ (Ref 25)	CNS$^{+/-}$ (2)	CNS$^{+/-}$ (3)	CNS$^{+/-}$ (4)
Ceftriaxone	−2.498	−3.511	−3.226	0	0	0	0	0
Cephalexin	−1.733	−2.005	−1.851	0	0	0	0	0
Cephalothin	−1.800	−3.084	−2.554	0	0	0	0	0
Cetirizine	0.091	−0.705	−0.865	0	1	1	1	1
Chloramphenicol	−1.222	−1.855	−2.688	0	1	0	0	0
Chlordiazepoxide	0.488	0.130	−1.128	1	1	1	1	0
Chlorprothixene	0.938	0.295	0.676	1	1	1	1	1
Cimetidine	−1.232	−1.332	−0.808	0	0	0	0	1
Cisapride	−0.544	−2.673	−0.498	0	0	1	0	1
Clobazam	0.093	−0.112	−0.314	1	1	1	1	1
Clofibric_acid	0.010	−0.440	−0.607	0	0	1	1	1
Clomipramine	0.994	0.790	0.639	1	1	1	1	1
Clonazepam	−0.557	−0.849	−2.042	1	1	1	1	0
Clonidine	0.065	−0.302	−0.459	1	0	1	1	1
Clozapine	0.348	0.515	−0.468	1	1	1	1	1
Corticosterone	−0.178	−0.369	−1.696	0	0	1	1	0
Coumarin	0.134	0.405	0.156	0	1	1	1	1
Cyclobarbital	−0.751	−0.627	−0.723	1	1	1	1	1
Demeclocycline	−2.747	−3.519	−4.912	0	0	0	0	0
Dexamethasone	−0.464	−0.678	−2.510	0	0	1	1	0
Diazepam	0.293	−0.167	−1.010	1	1	1	1	0
Diltiazem	−0.096	−0.725	−0.599	0	0	1	1	1
Diphenhydramine	0.680	0.656	0.488	1	1	1	1	1
Domperidone	−0.623	−1.058	−0.483	0	1	1	0	1
Dopamine	−0.760	−0.680	−0.811	1	1	1	1	1
Doxepin	0.664	0.905	0.772	1	1	1	1	1
Doxylamine	0.352	0.451	0.495	0	1	1	1	1
Ebastine	1.248	0.442	−0.166	0	1	1	1	1
Estradiol valerate	0.142	−0.419	−0.589	0	0	1	1	1
Ethinyl estradiol	0.411	0.173	−0.677	0	0	1	1	1
Felbamate	−1.264	−1.500	−0.847	1	1	0	0	1
Felodipine	−0.364	−1.301	−0.651	0	1	1	0	1
Fenoprofen	0.282	−0.035	−0.617	0	0	1	1	1
Fentanyl	0.613	0.663	−0.070	1	1	1	1	1
Fleroxacin	−0.418	−0.976	−1.114	0	0	1	1	0
Flumazenil	−0.382	−0.614	−0.422	1	1	1	1	1
Flunitrazepam	−0.394	−0.616	−1.973	1	1	1	1	0
Fluoxetine	0.650	−0.705	0.367	1	1	1	1	1

Molecule	logBB (2)	logBB (3)	logBB (4)	CNS$^{+/-}$ (Expt)	CNS$^{+/-}$ (Ref 23)	CNS$^{+/-}$ (2)	CNS$^{+/-}$ (3)	CNS$^{+/-}$ (4)
Flupentixol	0.538	−1.102	−0.130	1	1	1	0	1
Flurazepam	0.472	−0.833	−1.304	1	1	1	1	0
Fluvastatin	−0.019	−1.739	−2.298	0	0	1	0	0
Furosemide	−1.349	−2.037	−1.827	0	0	0	0	0
GYKI-52466	−0.262	−0.298	−1.665	1	1	1	1	0
Gabapentin	−0.717	−0.657	−0.471	1	1	1	1	1
Guanabenz	−0.504	−0.729	−0.707	0	0	1	1	1
Guanfacine	−0.737	−0.973	−0.191	0	0	1	1	1
Haloperidol	0.260	−1.487	−0.983	1	1	1	0	1
Hydrocortisone	−0.592	−1.103	−2.560	0	0	1	0	0
Hydroxypethidine	−0.163	−0.178	−0.541	1	1	1	1	1
ICI-204448	−0.032	−1.609	−1.635	0	1	1	0	0
Ifenprodil	0.435	0.009	−0.928	1	1	1	1	1
Imipramine	0.844	1.389	0.674	1	1	1	1	1
Isotretinoin	0.670	−0.073	−0.418	0	0	1	1	1
Isradipine	−1.288	−1.321	−0.578	1	0	0	0	1
Ketamine	0.365	0.241	0.185	1	1	1	1	1
Ketoprofen	0.125	−0.220	−1.131	0	0	1	1	0
Labetalol	−0.694	−1.660	−1.657	0	0	1	0	0
Lamotrigine	−0.895	−0.583	0.411	1	1	1	1	1
Lidocaine	0.124	0.124	−0.008	0	1	1	1	1
Loperamide	0.665	−0.241	−0.993	0	1	1	1	1
Loratadine	0.469	−0.104	0.144	0	0	1	1	1
Lorazepam	−0.066	−0.829	−1.706	1	1	1	1	0
Lorcainide	0.721	0.095	−0.056	1	1	1	1	1
Lupitidine	−1.246	−2.172	−1.490	0	0	0	0	0
Mannitol	−2.129	−2.090	−3.464	0	0	0	0	0
MCPP	0.074	−0.340	−0.615	1	1	1	1	1
Mefenidil	−0.231	0.298	0.530	1	1	1	1	1
Meprobamate	−1.236	−1.396	−0.825	1	1	0	0	1
Mequitazine	0.692	1.105	0.804	1	1	1	1	1
Methadone	1.025	0.618	−0.040	1	1	1	1	1
Metoprolol	−0.279	−0.944	−0.455	0	0	1	1	1
Mexiletine	0.050	0.183	0.580	1	1	1	1	1
Mibefradil	0.388	−1.673	−0.456	0	0	1	0	1
Miglitol	−1.887	−2.179	−2.781	0	0	0	0	0
Milacemide	−0.864	−0.959	−0.101	1	1	1	1	1
Minoxidil	−0.805	0.191	−0.926	0	1	1	1	1

Molecule	logBB (2)	logBB (3)	logBB (4)	CNS$^{+/-}$ (Expt)	CNS$^{+/-}$ (Ref 23)	CNS$^{+/-}$ (2)	CNS$^{+/-}$ (3)	CNS$^{+/-}$ (4)
MK-801	0.545	0.961	0.807	1	1	1	1	1
Morphine	−0.374	−0.093	−0.587	1	1	1	1	1
N-CH$_3$ Naltrexon	−0.670	−0.214	−1.177	0	0	1	1	0
Nadolol	−0.787	−1.283	−1.658	0	0	1	0	0
Naloxone	−0.477	−0.507	−1.427	1	1	1	1	0
Naltrexone	−0.486	−0.524	−1.400	1	1	1	1	0
Naproxen	0.113	−0.016	−0.507	0	0	1	1	1
Napsagatran	−2.463	−3.736	−3.606	0	0	0	0	0
Neostigmine	−0.076	0.005	0.258	0	0	1	1	1
Nifedipine	−1.323	−1.329	−1.535	1	0	0	0	0
Nitrazepam	−0.708	−0.566	−1.992	1	1	1	1	0
Nitrendipine	−1.227	−1.511	−1.625	0	0	0	0	0
Norfenefrine	−0.835	−0.665	−0.853	0	0	1	1	1
Nortriptyline	0.734	0.647	0.758	1	1	1	1	1
Noxiptiline	0.728	0.875	−0.521	1	1	1	1	1
Olsalazine	−1.109	−2.032	−5.214	0	0	0	0	0
Orphenadrine	0.810	0.685	0.495	1	1	1	1	1
Oxcarbazepine	−0.437	−0.065	−0.310	1	1	1	1	1
Oxprenolol	−0.118	−0.957	−0.382	0	0	1	1	1
PD117302	0.513	0.275	0.131	1	1	1	1	1
Penicillin G	−0.977	−1.145	−1.738	0	0	1	0	0
Perphenazine	0.259	−0.360	−0.267	1	1	1	1	1
Phaclofen	−0.961	−1.284	−1.392	1	1	1	0	0
Phenacetin	−0.289	−0.416	0.038	0	1	1	1	1
Pheniramine	0.465	0.639	0.499	0	1	1	1	1
Phenobarbital	−0.695	−0.435	−0.841	1	1	1	1	1
Piracetam	−1.310	−0.988	−0.332	1	1	0	1	1
Pirenzepine	−0.820	0.010	−0.325	0	1	1	1	1
Practolol	−0.846	−1.305	−0.893	0	0	1	0	1
Pravastatin	−1.063	−2.906	−3.712	0	1	0	0	0
Prednisolone	−0.549	−0.794	−2.498	0	0	1	1	0
Prednisone	−0.399	−0.690	−2.372	0	0	1	1	0
Primidone	−0.520	−0.194	−0.255	1	1	1	1	1
Probenecid	−0.418	−1.151	−1.838	0	0	1	0	0
Procyclidine	0.565	0.610	−0.152	1	1	1	1	1
Progabide	−0.254	−1.738	−2.164	1	0	1	0	0
Promethazine	0.752	1.044	0.696	1	1	1	1	1
Propranolol	0.078	−0.107	−0.181	0	0	1	1	1
Proscillaridin	−0.877	−2.320	−2.829	0	0	1	0	0

Molecule	logBB (2)	logBB (3)	logBB (4)	CNS$^{+/-}$ (Expt)	CNS$^{+/-}$ (Ref 23)	CNS$^{+/-}$ (2)	CNS$^{+/-}$ (3)	CNS$^{+/-}$ (4)
Quinidine	−0.066	−0.255	−0.187	0	0	1	1	1
Quinpirole	−0.133	0.609	0.734	1	1	1	1	1
Quipazine	0.064	0.926	0.852	1	1	1	1	1
Raclopride	−0.218	−0.973	−1.042	1	1	1	1	0
Ralitoline	−0.145	−0.492	−0.433	1	1	1	1	1
Remacemide	−0.241	−0.410	−0.101	1	1	1	1	1
Remikiren	−1.374	−3.058	−4.183	0	0	0	0	0
Roxindole	0.446	0.779	−0.127	1	1	1	1	1
SCH-23390	0.590	0.435	−0.019	1	1	1	1	1
Saclofen	−0.877	−1.427	−0.995	1	1	1	0	1
Salbutamol	−0.730	−1.273	−1.624	0	0	1	0	0
Spiradoline	0.395	−0.737	0.070	1	1	1	1	1
Stiripentol	0.314	0.119	−0.154	1	0	1	1	1
Sulfasalazine	−1.089	−1.622	−4.648	0	0	0	0	0
Sulpiride	−1.389	−1.639	−1.211	1	1	0	0	0
Sumatriptan	−0.856	−0.568	−0.606	0	0	1	1	1
Taltrimide	−1.001	−1.254	−1.517	1	1	0	0	0
Tamitinol	−0.522	−0.972	−0.275	1	1	1	1	1
Temafloxacin	−0.270	−1.512	−1.072	0	0	1	0	0
Terbutaline	−0.649	−0.972	−1.627	0	0	1	1	0
Terfenadine	1.070	0.051	−1.412	0	1	1	1	0
Testosterone	0.458	0.236	−0.283	0	0	1	1	1
Thiopental	−0.213	−0.272	−0.511	1	0	1	1	1
Tiacrilast	−0.652	−1.097	−2.236	0	0	1	0	0
Tiagabine	0.416	−0.233	−0.660	1	1	1	1	1
Tifluadom	0.383	−0.503	−1.312	1	1	1	1	0
Tocainide	−0.428	−0.203	0.102	0	1	1	1	1
Tolamolol	−0.837	−1.483	−1.181	0	0	1	0	0
Toliprolol	−0.055	−0.284	−0.278	0	0	1	1	1
Topiramate	−1.301	−0.856	−0.353	1	1	0	1	1
Tranylcypromine	−0.092	0.154	0.518	1	1	1	1	1
Triazolam	0.778	0.147	−0.592	1	1	1	1	1
U50,488	0.672	−0.190	−0.006	1	1	1	1	1
Valproate pivoxil	0.492	−0.193	−0.758	1	0	1	1	1
Vigabatrin	−0.828	−0.941	−0.653	1	1	1	1	1
Warfarin	−0.069	−0.656	−1.203	0	0	1	1	0
Zimeldine	0.522	0.178	0.441	1	1	1	1	1
Zolmitriptan	−0.527	−0.383	0.020	1	1	1	1	1
Zonisamide	−1.170	−0.527	−0.513	1	1	0	1	1