Journal of Computer-Aided Molecular Design, **15:** 697–708, 2001.
KLUWER/ESCOM

Prediction of blood-brain partitioning using Monte Carlo simulations of molecules in water

Yiannis N. Kaznessis[a,b,c,*], Mark E. Snow[b] & C. John Blankley[b]
[a]*Department of Chemical Engineering, University of Michigan, Ann Arbor, MI 48109, USA;* [b]*Pfizer Global Research and Development, Ann Arbor, MI 48105, USA;* [c]*Current address: Department of Chemical Engineering and Materials Science, University of Minnesota, Minneapolis, MN 55455, USA*

Accepted 29 May 2001

Key words: blood-brain partition coefficient, computer simulations, Monte Carlo, QSPR

Summary

The brain-blood partition coefficient ($\log BB$) is a determining factor for the efficacy of central nervous system acting drugs. Since large-scale experimental determination of $\log BB$ is unfeasible, alternative evaluation methods based on theoretical models are desirable. Toward this direction, we propose a model that correlates $\log BB$ with physically significant descriptors for 76 structurally diverse molecules. We employ Monte Carlo simulations of the compounds in water to calculate such properties as the solvent-accessible surface area (*SASA*), the number of hydrogen bond donors and acceptors, the solute dipole, and the hydrophilic, hydrophobic and amphiphilic components of *SASA*. The physically significant descriptors are identified and a quantitative structure-prediction relationship is constructed that predicts $\log BB$. This work demonstrates that computer simulations can be employed in a semi-empirical framework to build predictive QSPRs that shed light on the physical mechanism of biomolecular phenomena.

Introduction

The relative distribution of drugs in the blood and brain compartments is a determining criterion for screening of potential therapeutic agents in the early preclinical drug discovery stages. In the case of central nervous system (CNS) acting drugs, penetration of compounds from the blood circulation into the brain is a functional prerequisite, whereas for peripherally acting drugs penetration is undesirable due to potential CNS-related side-effects. This penetration is regulated by the blood-brain barrier (BBB), a complex physical and biochemical interface, which morphologically is based on tightly joined blood capillary endothelial cells [1, 2].

Since the BBB is devoid of transport pores, passive diffusion is the dominant physical process of cerebrovascular transport. At the molecular level, the principal component of the barrier is the lipid bilayer of the capillary endothelial cell membrane, through which compounds have to diffuse to reach the brain.

The relative affinity for the blood or brain tissue can be expressed in terms of the blood-brain partition coefficient, $\log BB = \log(C_{\text{brain}}/C_{\text{blood}})$, where C_{brain} and C_{blood} are the equilibrium concentrations of the drug in the brain and the blood respectively. Both *in vivo* [3] and *in vitro* [4, 5] experiments have been conducted that calculate $\log BB$. In *in vivo* experiments, peripheral application of radiolabeled compounds is followed by brain concentration level measurements. In *in vitro* experiments, the partition of the compound between an aqueous and an organic phase is measured and the results are used for relative $\log BB$ ranking of compounds, since it is generally believed that the drug permeation is largely determined by the molecule's relative affinity for the water/lipid interface.

Both these methodologies are costly and difficult, hence not amenable to high-throughput screening of

*To whom correspondence should be addressed. E-mail: yiannis@cems.umn.edu

therapeutic candidates. Consequently, there have been numerous attempts to employ theoretical and computational methodologies to predict the blood-brain partition coefficient.

Young and co-workers [3] proposed the correlation between $\log BB$ and $\Delta \log P$, shown in Equation 1. $\Delta \log P$ is known as the Seiler parameter [6], and is defined as $\Delta \log P = \log P_{\mathrm{ow}} - \log P_{\mathrm{cyclw}}$, where P_{ow} and P_{cyclw} are the octanol/water and cyclohexane/water partition coefficients, respectively

$$\begin{aligned} \log BB &= -0.485(\pm 0.160)\Delta \log P \\ &\quad + 0.889(\pm 0.500), \\ n &= 20, r = 0.831, s = 0.439, \\ F &= 40.23. \end{aligned} \tag{1}$$

In Equation 1, n is the number of compounds, r is the correlation coefficient, s is the standard error, and F is the Fischer value that gives a measure of the statistical significance of the relationship. $\Delta \log P$ is considered to be a measure of hydrogen-bonding potential [7].

Seelig and co-workers [8, 9] identified the molecular parameters governing the passive diffusion of the molecules through lipid membranes, using theoretical arguments. They suggested that the optimal characteristics for a molecule to penetrate the BBB are (i) amphiphilicity $\Delta\Delta G_{am} > -3$ kJ/mol (they defined amphiphilicity as $\Delta\Delta G_{am} = \Delta G_{aw} - \Delta G_{mic}$ where ΔG_{aw} and ΔG_{mic} are the free energies of partitioning into the air-water interface and of micelle formation respectively), (ii) a value for the air/water partition coefficient K_{aw} in the range of 10^5–10^3 M^{-1}, and (iii) a molecular cross-sectional area $A_D < 80$ Å^2.

Kelder and co-workers [10] examined the distribution of the polar surface area of 776 CNS and 1590 non-CNS drugs and deduced that penetration of molecules is possible only if their polar surface area is less than 120 Å^2. They also suggested that drugs can be tailored for brain penetration by decreasing the polar surface to less than 60 Å^2.

A number of more comprehensive computational approaches that resulted in the development of quantitative structure-property relationships (QSPR) have also been reported. Van de Waterbeemd and Kansy [11] established the following QSPR for 20 molecules:

$$\begin{aligned} \log BB &= -0.021 PSA(\pm 0.003) \\ &\quad - 0.003(\pm 0.001) V_{\mathrm{mol}} \\ &\quad + 1.643(\pm 0.465), \\ r &= 0.835, s = 0.448, F = 19.5, \end{aligned} \tag{2}$$

where PSA is the molecular polar surface area, and V_{mol} is the molecular volume. This relationship provides additional physical insight into the mechanism of the diffusion process, suggesting that hydrophilicity and volume negatively impact the permeability of the compounds. However, Calder and Ganellin [12] found that Equation 2 overestimates the experimental values of 5 compounds not in the initial set. This finding indicated the need for a larger set of compounds.

Abraham and co-workers [13–15] constructed the following model equations using a fragment-based scheme and a larger set of 57 compounds:

$$\begin{aligned} \log BB &= -0.038(\pm 0.064) \\ &\quad + 0.198(\pm 0.100) R_2 \\ &\quad - 0.687(\pm 0.125) \pi_2^H \\ &\quad - 0.715(\pm 0.334) \sum \alpha_2^H \\ &\quad - 0.698(\pm 0.107) \sum \beta_2^H \\ &\quad + 0.995(\pm 0.096) V_x, \\ n &= 57, r = 0.952, s = 0.197, \\ F &= 99.2, \end{aligned} \tag{3a}$$

$$\begin{aligned} \log BB &= 0.023 \log P_{\mathrm{ow}} - 0.507 \sum \alpha_2^H \\ &\quad - 0.500 \sum \beta_2^H + 0.055, \\ n &= 49, r = 0.949, s = 0.201, \\ F &= 136.1, \end{aligned} \tag{3b}$$

where R_2 is an excess molar refraction, π_2^H is the dipolarity/polarizability parameter, $\sum \alpha_2^H$ and $\sum \beta_2^H$ are the solute hydrogen-bond acidity and basicity, respectively, and V_x is the characteristic volume of McGowan [14]. These equations demonstrate the importance of hydrogen-bonding potential in the permeation of molecules through the BBB. However, they require the calculation of many parameters for fragments, and the additivity of those parameters to molecular level properties is problematic, since this scheme assumes no intramolecular interactions between these fragments.

Lombardo and co-workers [16] correlated $\log BB$ with the free energy of solvation ΔG_w with the following equation

$$\begin{aligned} &\log BB = 0.054(\pm 0.005)\Delta G_w + 0.43(\pm 0.07), \\ &n = 55, r = 0.82, s = 0.41, F = 108.3. \end{aligned} \tag{4}$$

This correlation provides an elegant means for successful $\log BB$ prediction. It, however, relies on calculating the energy of solvation from semi-empirical calculations in the gas phase. The solvent might play

an important role in the conformations of the solute, which will in turn lead to different values for a number of parameters, such as the solvent-accessible surface area (*SASA*), the molecular volume and the molecular dipole moment, all of which influence the energy of solvation.

Kaliszan and Markuszewski [17] re-established the correlation of $\log BB$ with $\log P_{ow}$ and refined it, employing the molecular mass as an additional descriptor of molecular bulkiness:

$$\begin{aligned} \log BB &= -0.088(\pm 0.051) \\ &\quad + 0.272(\pm 0.017)\log P \\ &\quad - 0.001116(\pm 0.00049)M_m, \\ n &= 33, r = 0.947, s = 0.126, \\ F &= 131.1. \end{aligned} \tag{5}$$

These authors indicated that a molecular bulkiness descriptor should be used to better account for non-specific dispersive properties of molecules.

More recently, Norinder and co-workers [18] used MolSurf [19] parametrization to calculate various properties related to the molecular valence region, and combined it with the Partial Least Squares to Latent Structures (PLS) method [20] to develop a QSPR with three statistically significant components (the components were obtained by means of the Principal Component Analysis, PCA, method [21]) and the following statistics: $n = 56$, $r = 0.913$, $s = 0.312$, $F = 86.95$. Again this method relies on gas phase properties. In addition, the PCA method generally appears to strip the QSPR from explicit physical insight.

Luco [22] employed the PLS technique to develop a QSPR based on several topological and constitutional descriptors. This analysis also resulted in a significant three-component model with the following statistics: $n = 58$, $r = 0.922$, $s = 0.318$ and $F = 102$.

Recently, Feher and co-workers [23] revisited the $\log BB/\log P_{ow}$ correlation building the following regression model

$$\begin{aligned} \log BB &= -0.1092\log P_{ow} - 0.3873 n_{acc,solv} \\ &\quad - 0.0017 PSA + 0.4275, \\ n &= 61, r = 0.854, s = 0.424, \\ F &= 51, \end{aligned} \tag{6}$$

where $n_{acc,solv}$ is the number of hydrogen-bond acceptors. The correlation coefficient errors were not given for Equation 6.

Finally, Clark [24] built a model for the prediction of $\log BB$ from a set of 55 compounds. His results also indicate the importance of the polar surface area for predicting $\log BB$ and the correlation between the blood/brain and the octanol/water partition coefficients.

Clearly, there has been considerable progress in the development of semi-empirical models for the relative affinity of compounds for the blood and brain compartments, with considerable predictive ability. Nevertheless, all these approaches are based on mechanistically chosen topological descriptors or the calculation of properties of stand-alone molecules in the lowest gas-phase energy conformation. The solvation of compounds in water and a lipid phase might be accompanied by conformational changes that in turn lead to changes in the molecular properties. These changes will be more pronounced for large flexible molecules. Ideally, one might consider the simulation of the diffusion of molecules through a lipid bilayer. Molecular dynamics simulations of various molecules in lipid bilayers embedded in an aqueous environment have been reported [25, 26]. Unfortunately, the time scales of diffusion of small molecules span scales of several microseconds and molecular dynamics can simulate atomistic systems for only a few nanoseconds. Therefore, one has to turn to semi-empirical approaches that address the issue of the solvent's influence on the molecular conformations.

Jorgensen and co-workers [27, 28] have demonstrated that Monte Carlo simulations of molecules in water can be successfully employed to predict the gas to liquid free energies of solvation in hexadecane, octanol and water, the $\log P_{ow}$ and the water solubility $\log S$. Constant temperature and pressure ensemble averages were obtained for such properties as the Coulomb and Lennard–Jones energies of interaction between the solute and the solvent, the *SASA*, the hydrophobic and hydrophilic components of *SASA* and the numbers of donor and acceptor hydrogen-bond sites. Using these descriptors, Duffy and Jorgensen [28] developed the following QSPR model for the octanol/water partition coefficient

$$\begin{aligned} MC\log P_{ow} &= 0.01448 SASA - 0.731 HBAC \\ &\quad - 1.064\ (\text{no. of amines}) \\ &\quad + 1.1718\ (\text{no. of nitro} + \text{acid groups}) \\ &\quad - 1.772. \end{aligned} \tag{7}$$

We used the term $MC\log P_{ow}$ to distinguish the octanol/water partition coefficient calculated by Equation 7. The corrections for the number of amines, nitro and acidic groups were deemed necessary due to im-

Table 1. Set of simulated compounds with experimental log *BB* values. For compounds 46–65 we retained the names used in [13]; numbers of the lu compounds were the ones adopted by Luco [22]; for compounds 72–78 we retained the names used in [10]; for compounds 79–85 the names used in [15] were retained

#	Name	log *BB*	#	Name	log *BB*
1	acetone[13]	−0.15	44	toluene[13]	0.37
2	benzene[13]	0.37	45	zolantidine[13]	0.14
3	butanone[13]	−0.08	46	cmpd_2[13]	−0.04
4	chloroform[13]	0.29	47	cmpd_4[13]	−1.30
5	carbamazepine[16]	0.00	48	cmpd_13[13]	−2.15
6	carbamazepineepoxide[16]	−0.34	49	cmpd_15[13]	−0.67
7	cimetidine[13]	−1.42	50	cmpd_16[13]	−0.66
8	clonidine[13]	0.11	51	cmpd_17[13]	−0.12
9	desipramine[10]	1.00	52	cmpd_19[13]	−0.18
10	2,2-dimethylbutane[13]	1.04	53	cmpd_20[13]	−1.15
11	domperidone[10]	−0.78	54	cmpd_22[13]	−1.57
12	enflurane[13]	0.24	55	cmpd_23[13]	−1.54
13	ether[13]	0.00	56	cmpd_24[13]	−1.12
14	ethanol[13]	−0.16	57	cmpd_25[13]	−0.73
15	fluroxene[13]	0.13	58	cmpd_26[13]	−0.27
16	halothane[13]	0.35	59	cmpd_29[13]	−0.28
17	heptane[13]	0.81	60	cmpd_30[13]	−0.46
18	icotidine[13]	−2.00	61	cmpd_31[13]	−0.24
19	imipramine[13]	0.83	62	cmpd_34[13]	−0.02
20	isoflurane[13]	0.42	63	cmpd_36[13]	0.69
21	lupitidine[10]	−1.06	64	cmpd_37[13]	0.44
22	mepyramine[13]	0.49	65	cmpd_42[13]	0.22
23	methane[13]	0.04	66	lu28[22]	−1.17
24	methoxyflurane[13]	0.25	67	lu69[22]	−0.16
25	methylcyclopentane[13]	0.93	68	lu72[22]	−0.30
26	3-methyhexane[13]	0.90	69	lu73[22]	−1.34
27	2-methylpentane[13]	0.97	70	lu74[22]	−1.82
28	3-methylpentane[13]	1.01	71	lu75[22]	0.89
29	mianserin[10]	0.99	72	Org4428[10]	0.82
30	mirtazepine[10]	0.53	73	Org5222[10]	1.03
31	pentane[13]	0.76	74	Org12962[10]	1.64
32	propanol-1[13]	−0.16	75	Org13011[10]	0.16
33	propanol-2[13]	−0.15	76	Org32104[10]	0.52
34	ranitidine[13]	−1.23	77	Org30526[10]	0.39
35	risperidone[10]	−0.02	78	Org34167[10]	0.00
36	risperidone9OH[10]	−0.67	79	skf101468[15]	0.25
37	teflurane[13]	0.27	80	skf89124[15]	−0.43
38	temelastine[13]	−1.88	81	yg14[15]	−0.30
39	tibolone[10]	0.40	82	yg15[15]	−0.06
40	tiotidine[13]	−0.82	83	yg16[15]	−0.42
41	1,1,1-trichloroethane[13]	0.40	84	yg19[15]	−1.30
42	trichloroethylene[13]	0.34	85	yg20[15]	−1.40
43	1,1,1-trifluoro-2-chloroethane[13]	0.08			

perfections of the CM1P partial charges assigned to the compounds of their training set.

In this paper, we employ this methodology for 85 structurally diverse molecules and develop a QSPR for the prediction of the blood-brain partition coefficient. Our goal is to demonstrate the utility of statistical mechanics simulations in QSPR building and their positive influence on the early stages of the drug design process.

In the following section we describe the simulation technique. Then, we describe the results and the developed model and discuss its predictive ability. We then calculate MC $\log P_{ow}$ for all the compounds, as predicted by Equation 7 and use the calculated octanol/water partition coefficient values to compare our approach with previous $\log BB$-predicting models that incorporate $\log P_{ow}$. Finally, we conclude assessing the methodology as a tool for high-throughput screening in an industrial setting.

Methodology

Biological data

A set of 85 molecules, shown in Table 1, was simulated in water. The set was compiled from structures presented in [10, 13, 15, 22]. They represent a wide set of structurally diverse groups of molecules with publicly available blood-brain partition coefficients.

Monte Carlo simulations

Details of the simulation protocol are provided in [28]. Here, we describe the main aspects of the methodology. Initially the molecules were represented in SMILES format [29]. The SMILES were then transformed in mol2 files using the Sybyl software package [30]. The simulations were performed using the BOSS4.2 software package [31], after transforming the mol2 files in the appropriate BOSS-input format. The force field used by BOSS accounts for intramolecular interactions using harmonic bond-stretching terms, angle-bending terms and a Fourier series for the torsion terms, whereas the intermolecular interactions are dealt with a 12-6 Lennard-Jones potential and Coulomb's law. The total potential energy function is given by:

$$E = \sum_i k_{b,i}(r_i - r_{0,i})^2 + \sum_i k_{\theta,i}(\theta_i - \theta_{0,i})^2 + \sum_i \left[\tfrac{1}{2}V_{1,i}(1+\cos\varphi_i) + \tfrac{1}{2}V_{2,i}(1+\cos 2\varphi_i) + \tfrac{1}{2}V_{3,i}(1+\cos 3\varphi_i)\right] + \sum_i \sum_{j>i} \left\{ \frac{q_i q_j}{r_{ij}} + 4\varepsilon \left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6} \right] \right\} \quad (8)$$

where the indices i and j run over all the atoms in the simulation. The force constants k, the distance and angle reference values, the Fourier coefficients V, and the Lennard-Jones parameters are all from the OPLS-AA force field [32]. The partial charges were obtained from AM1 calculations using the CM1A procedure [33].

Table 2. Ensemble averaged properties calculated from Monte Carlo simulations

Property
Coulomb energy between solute and solvent, ESXC
Lennard-Jones energy between solute and solvent, ESXL
Solvent accessible surface area, SASA
Hydrophobic component of SASA, FOSA
Hydrophilic component of SASA, FISA
Aromatic component of SASA, ARSA
Dipole moment of solute, DIPO
Number of solute-solvent interaction <-3.75 kcal mol^{-1}, INST
Number of solute-solvent interaction <-2.75 kcal mol^{-1}, INME
Number of solute hydrogen-bond donors, HBDN
Number of solute hydrogen-bond acceptors, HBAC
Molecular volume, MVOL

Each molecule was solvated in 500 water molecules. The TIP4P model[34] was used for the waters. The systems were simulated in the NPT ensemble at 25 °C and 1 atm and periodic boundary conditions were employed.

Each simulation consisted of 3×10^5 equilibration steps and 10^7 production steps. All internal degrees of freedom of the solutes were sampled, whereas water was allowed to undergo only translational and rotational moves. The motion ranges were tuned to allow for a 30% acceptance rate. Ensemble averages were calculated during the simulations for the properties shown in Table 2.

The cutoff distance used for calculating the acceptor and donor hydrogen bonds is 2.5 Å, which is the distance of the minimum of the first peak in X-H radial distribution functions [35].

The *SASA* is determined using a spherical probe with a 1.4 Å radius. All heteroatoms and their hydrogens are considered hydrophilic whereas carbon atoms

and their hydrogens are considered either aromatic or hydrophobic.

Duffy and Jorgensen [28] introduced *INME* and *INST* as counts of the total number of hydrogen bonds and the strong hydrogen bonds, respectively.

Statistical analysis

A linear regression model, correlating the ensemble-averaged properties with the blood-brain partition coefficients, was built employing the regression module of the JMP program [36]. In a stepwise fashion, the statistically less significant descriptor was eliminated from the set of descriptors used in the model. The significance of the descriptors was evaluated in term of the F' ratios, defined as the ratios of the regression model mean square over the error mean square. The optimal descriptor set was chosen, so that the correlation coefficient was maximized, the standard error minimized and requiring that the probability of a greater F' value occurring by chance was less than 0.001. As noted in [28], for as small a set as the one used herein the separation of data points to training and testing sets is not statistically meaningful. Hence, cross-validation of the resulting model was performed by a leave-one batch-out procedure.

Results and discussion

Correlation coefficients

In Table 3, the correlation coefficients for all descriptor pairs are reported. The correlation coefficient, r_{ij}, for any two variables x_i and x_j reflects the strength of the linear relationship between the two variables and is calculated as the Pearson product moment:

$$r_{ij} = \frac{s_{ij}^2}{\sqrt{s_{ii}^2 s_{jj}^2}}, \; s_{ij}^2 = \frac{\sum (x_i - \overline{x_i})\left(x_j - \overline{x_j}\right)}{N-1}. \quad (9)$$

The indices in Equation 9 run over all N observations.

High correlations are revealed between *ESXC, DIPO, INST, INME, HBDN*, and *HBAC* as expected. Also expected are the correlations between *SASA, ESXL, MVOL*, and *MWEI*. A correlation coefficient of 0.804 between *HBAC* and *SASA* indicates that for the set of compounds investigated there is a uniform distribution of hydrogen-bond accepting atoms on the surface of the molecules.

Blood/brain partition coefficient

In the statistical analysis, 9 strong outliers were identified and removed. The excluded molecules are icotidine, temelastine, tiotidine, cmpd_2, cmpd_4, cmpd_13, Org12962, yg_19 and yg_20. Of those, icotidine, temelastine, cmpd_4, cmpd_13, were removed as outliers by Abraham and co-workers in [13], Org12962 was identified as a serious outlier by Kelder and co-workers [10], and yg_19 and yg_20 were removed by Abraham and co-workers in [15]. As noted in the literature the deviation of these molecules cannot be attributed to discernible structural trends differentiating them from the rest. Hence, this deviation can be explained in terms of the difficulty of experimental measurements, or it can attributed to metabolism or possible active transport mechanisms.

In the statistical analysis, in addition to the descriptors described previously, we used the term HBACxHBDN$^{1/2}$/SASA. This cohesive index was introduced by Jorgensen and co-workers [27, 28], and can be viewed as an electrostatic surface tension. The fractional power in HBDN reflects possible saturation effects expected for molecules with a large number of acceptors and donors, in which case it is not likely that all of them will be simultaneously satisfied. This term was statistically significant in a solubility model proposed in [27].

Using 76 compounds and employing the stepwise statistical approach described in the previous section, we arrived at Equation 10, henceforth called model_1.

$$\begin{aligned} \log BB &= -0.2339(\pm 0.013)\text{HBAC} \\ &\quad + 0.00147(\pm 0.00011)\text{MVOL} \\ &\quad 31.6099(\pm 4.0837)\text{HBAC}\times\text{HBDN}^{1/2} \\ &\quad /\text{SASA} - 0.04579(\pm 0.05808), \\ n &= 76, r = 0.97, s = 0.173, \\ F &= 311.307. \end{aligned} \quad (10)$$

Model_1 predicts the blood-brain partition coefficient extremely well, using only a small number of physical descriptors. In addition, it provides a solid physical picture of the molecular mechanisms involved in cerebrovascular transport, indicating that hydrophilicity negatively impacts the blood-brain permeation. In Table 4, the predicted $\log BB$ values are presented. It should be noted that even when all the compounds are included in the calculations, the blood/brain partition coefficient predicting QSPR contains the same descriptors with Equation 10.

Table 3. Correlation coefficients for descriptor pairs

Variable	ESXC	ESXL	SASA	FISA	FOSA	ARSA	DIPO	INST	INME	HBDN	HBAC	MWEI	MVOL
ESXC	1.000	−0.111	−0.580	−0.567	0.043	−0.520	−0.833	−0.967	−0.968	−0.680	−0.757	−0.589	−0.563
ESXL	−0.111	1.000	−0.726	0.120	−0.688	−0.388	0.055	−0.073	−0.082	0.198	−0.383	−0.686	−0.739
SASA	−0.580	−0.726	1.000	0.250	0.558	0.691	0.464	0.693	0.698	0.342	0.804	0.950	0.997
FISA	−0.567	0.120	0.250	1.000	−0.347	0.031	0.587	0.538	0.563	0.383	0.534	0.439	0.230
FOSA	0.043	−0.688	0.558	−0.347	1.000	0.007	0.028	0.103	0.096	−0.097	0.325	0.359	0.558
ARSA	−0.520	−0.388	0.691	0.031	0.007	1.000	0.239	0.545	0.539	0.363	0.480	0.682	0.703
DIPO	−0.833	0.055	0.464	0.587	0.028	0.239	1.000	0.838	0.853	0.455	0.688	0.485	0.441
INST	−0.967	−0.073	0.693	0.538	0.103	0.545	0.838	1.000	0.995	0.653	0.835	0.687	0.675
INME	−0.968	−0.082	0.698	0.563	0.096	0.539	0.853	0.995	1.000	0.631	0.848	0.703	0.680
HBDN	−0.680	0.198	0.342	0.383	−0.097	0.363	0.455	0.653	0.631	1.000	0.401	0.323	0.323
HBAC	−0.757	−0.383	0.804	0.534	0.325	0.480	0.688	0.835	0.848	0.401	1.000	0.790	0.785
MWEI	−0.589	−0.686	0.950	0.439	0.359	0.682	0.485	0.687	0.703	0.323	0.790	1.000	0.954
MVOL	−0.563	−0.739	0.997	0.230	0.558	0.703	0.441	0.675	0.680	0.323	0.785	0.954	1.000

In Figure 1, we plot the predicted $\log BB$ versus the experimental ones. We also include the excluded compounds for comparison.

The F values for the descriptors in the model indicate that the most important ones are the HBAC and the volume. The third term can be replaced with HBDN with only a minor deterioration in the predictive quality:

$$\begin{aligned} \log BB &= -0.28902(\pm 0.012)\text{HBAC} \\ &\quad + 0.00171(\pm 0.00011)\text{MVOL} \\ &\quad - 0.07464(\pm 0.01543)\text{HBDN} \\ &\quad - 0.13956(\pm 0.06618), \\ n &= 76, r = 0.958, s = 0.203, \\ F &= 270.46. \end{aligned} \tag{11}$$

The small differences between Equations 10 and 11 point to the plasticity of the models.

The sum of squared errors for Equation 10 is 2.149. The press statistic option in JMP, which calculates the sum of squared residuals, where the residual for each row is computed after dropping that row from the computation, is 2.384. The small difference between these two numbers verifies the predictive ability of the model. After the submission of the present article, and at the request of one of the reviewers, we simulated four additional compounds and calculated the blood/brain partition coefficient predicted by model_1. The compounds were antipyrine, zidovudine, pentobarbital, and trifluoroperazine. The experimental values for $\log BB$ are −0.097, −0.72, 0.12, and 1.44 respectively, spanning a wide range of $\log BB$ values. The predicted values are 0.084, −0.77, −0.251, and 1.19 respectively. The mean rms error is 0.48, a value that is satisfactory. Hence, this small blind test verifies the predictive quality of our model. Here again we stress the inability to establish the significance of the model with rigorous statistical methods. The small number of compounds in the literature with known blood/brain coefficients renders the partitioning of the data set into training and test sets and the validation of the model QSPR difficult to justify with precise statistical terms.

The necessary CPU time for the simulations was an average of 1.61 h for each compound on a 600 MHz Intel-PentiumIII LINUX-based PC. This is a significantly larger CPU time than the ones necessary for other computational approaches for the prediction of $\log BB$. Hence, although the overall cost for the 85 simulated compounds for this study is negligible, this method is not readily amenable for high-throughput computational screening of drug-like molecules.

Nevertheless, the rigorous theoretical foundation of the methodology and its ability to sample the interaction of many solute conformations with the solvent in various ensembles make it particularly attractive. This becomes evident when we plot the recorded instantaneous property values for the duration of the simulation. For example in Figure 2, the solvent accessible surface area values of skf89124, a moderately flexible molecule with eight rotatable bonds, are plotted against the recording intervals, which occur every 2×10^5 Monte Carlo steps. The fluctuations of the SASA values are significant for the duration of the

Table 4. Experimental and predicted values for the blood/brain partition coefficient. log *BB*_1 is predicted by model_1, log *BB*_2 is predicted by model_2 and log *BB*_3 is predicted by model_3. The MC log P_{ow} is calculated using Equation 7

Compound	log *BB*	MC log P_{ow}	log *BB*_1	log *BB*_2	log *BB*_3	Compound	log *BB*	MC log P_{ow}	log *BB*_1	log *BB*_2	log *BB*_3
1	−0.15	0.52	0.06	0.06	0.00	44	0.37	2.55	0.60	0.80	0.72
2	0.37	2.08	0.51	0.70	0.58	45	0.14	3.42	0.17	0.12	0.11
3	−0.08	0.71	0.06	0.09	0.03	46	−0.04	0.54	−0.80	−0.53	−0.32
4	0.29	1.71	0.44	0.13	0.29	47	−1.30	3.62	−0.07	−0.14	−0.03
5	0.00	1.97	−0.13	0.01	0.00	48	−2.15	2.47	−0.34	−0.35	−0.10
6	−0.34	0.68	−0.66	−0.36	−0.58	49	−0.67	2.51	−0.64	−0.51	−0.18
7	−1.42	0.75	−1.13	−0.76	−0.55	50	−0.66	2.21	−0.63	−0.37	−0.04
8	0.11	1.99	−0.01	−0.06	0.03	51	−0.12	3.83	−0.23	0.04	0.32
9	1.00	3.52	0.73	0.72	0.53	52	−0.18	2.25	−0.26	−0.11	0.18
10	1.04	2.67	0.63	0.82	0.79	53	−1.15	0.73	−1.04	−0.83	−0.50
11	−0.78	0.86	−0.83	−1.14	−1.09	54	−1.57	−0.22	−1.52	−1.27	−1.03
12	0.24	1.83	0.33	0.05	0.12	55	−1.54	0.74	−1.41	−1.11	−0.76
13	0.00	1.23	0.16	0.33	0.24	56	−1.12	0.20	−1.25	−1.15	−1.02
14	−0.16	−0.06	−0.33	−0.08	−0.19	57	−0.73	3.24	−0.38	−0.34	−0.08
15	0.13	1.46	0.18	0.16	0.16	58	−0.27	2.16	−0.54	−0.33	−0.22
16	0.35	2.23	0.54	0.37	0.24	59	−0.28	3.19	−0.28	−0.04	0.16
17	0.81	3.53	0.78	0.99	1.10	60	−0.46	1.20	−0.41	−0.43	−0.50
18	−2.00	2.47	−0.66	−0.50	−0.27	61	−0.24	1.92	−0.31	−0.35	−0.41
19	0.83	4.06	0.99	0.89	0.71	62	−0.02	1.96	0.01	0.05	−0.05
20	0.42	2.31	0.49	0.24	0.31	63	0.69	5.01	0.78	0.88	0.95
21	−1.06	−0.07	−1.26	−1.41	−1.43	64	0.44	4.06	0.48	0.46	0.54
22	0.49	2.21	0.36	0.18	−0.07	65	0.22	3.88	0.33	0.42	0.35
23	0.04	0.59	0.24	0.40	0.17	66	−1.17	−0.51	−1.33	−0.95	−0.91
24	0.25	1.10	0.13	−0.04	−0.12	67	−0.16	2.89	−0.12	0.05	0.19
25	0.93	2.57	0.61	0.80	0.75	68	−0.30	1.96	−0.26	−0.62	−0.41
26	0.97	2.89	0.69	0.86	0.88	69	−1.34	0.20	−1.16	−1.24	−1.18
27	1.01	2.76	0.73	0.84	0.83	70	−1.82	−0.12	−1.58	−1.46	1.37
28	0.90	3.25	0.76	0.94	0.98	71	0.89	4.04	0.95	0.86	0.71
29	0.99	3.43	0.96	0.80	0.51	72	0.82	3.33	0.65	0.64	0.36
30	0.53	2.37	0.58	0.39	0.07	73	1.03	3.15	0.83	0.52	0.32
31	0.76	2.59	0.66	0.80	0.80	74	1.64	1.17	0.01	−0.33	−0.42
32	−0.16	0.55	−0.17	0.06	0.01	75	0.16	2.54	0.25	−0.28	−0.21
33	−0.15	0.25	−0.26	0.03	−0.11	76	0.52	2.54	0.40	0.39	0.09
34	−1.23	−0.11	−1.36	−1.26	−1.11	77	0.39	1.58	0.27	−0.01	−0.28
35	−0.02	2.14	0.08	−0.49	−0.51	78	0.00	1.98	0.13	0.04	−0.09
36	−0.67	1.03	−0.68	−0.97	−1.02	79	0.25	2.92	0.40	0.30	0.31
37	0.27	2.06	0.50	0.29	0.22	80	−0.43	1.27	−0.40	−0.47	−0.42
38	−1.88	2.86	−0.47	−0.57	−0.32	81	−0.30	−0.13	−0.29	−0.27	−0.53
39	0.40	4.08	0.56	0.57	0.61	82	−0.06	−0.62	−0.25	−0.42	−0.77
40	−0.82	−0.18	−1.85	−1.70	−1.14	83	−0.42	−1.43	−0.73	−0.84	−1.03
41	0.34	2.02	0.49	0.20	0.37	84	−1.30	1.87	0.14	0.04	0.07
42	0.40	2.12	0.52	0.46	0.41	85	−1.40	−0.07	−0.42	−0.38	−0.58
43	0.08	1.87	0.46	0.34	0.35						

simulation, emphasizing the advantage of ensemble averaging over choosing a single conformation for the development of QSPRs.

The emergent important molecular properties are the potential of hydrogen bond formation and the size of the molecules. It is clear that the number of hy-

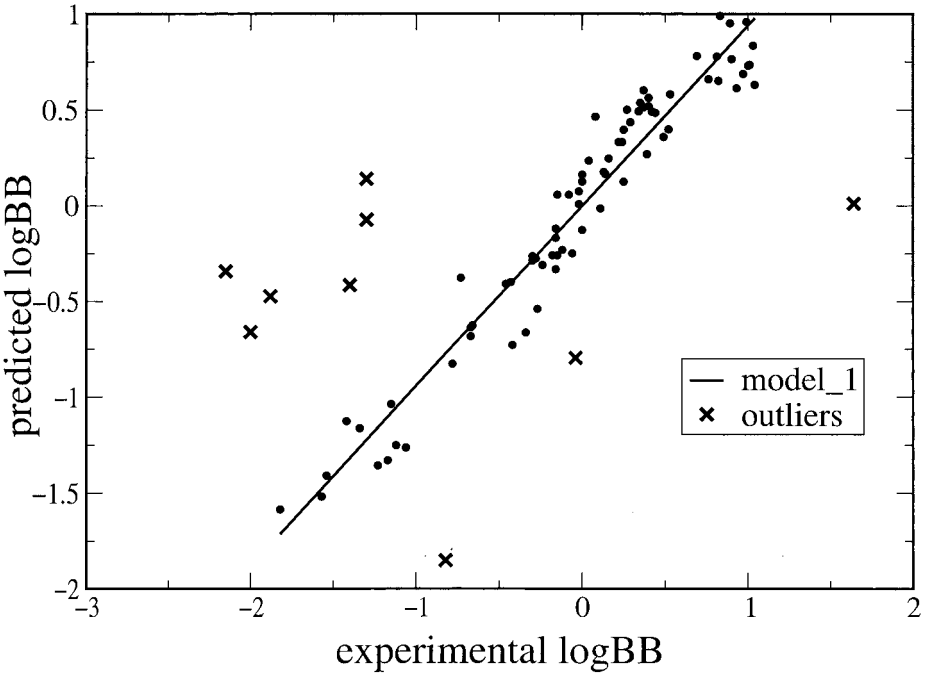

Figure 1. Predicted log *BB* (using model_1) versus experimentally determined log *BB*.

drogen bond acceptors and donors adversely affects the penetration of molecules through the blood/brain barrier. Hence, hydrophilic groups of designed CNS-acting drugs should be kept to a minimum, if the main mechanism of penetration is expected to be passive diffusion. A first reading of the model_1 equation would suggest that larger molecules penetrate the BBB with additional ease. We do feel, however, that this second term in model_1 is correcting for the relatively dominant importance of the HBAC term. There is a high positive correlation between HABC and MVOL (0.785) that justifies such an explanation. Hence, caution should be exercised in the development of rules for the rational design of drugs based on simple readings of QSPRs. Nevertheless, such an approach does give a robust qualitative picture of the important physicochemical molecular properties. Moreover, since the emergent important properties in our QSPR have been found to be important in previous works with independent modeling approaches, the usefulness and strength of our QSPR becomes more evident.

Comparison with literature log BB predictions

It is of interest to evaluate other models suggested in the literature, using the descriptors calculated by the simulations. The octanol/water partition coefficient was also calculated from the simulation results using the QSPR proposed by Duffy and Jorgensen [28]. The calculated MC log P_{ow} values are shown in Table 4.

We initially built a model employing the number of hydrogen-bond acceptors, the hydrophilic surface area and the octanol/water partition coefficient, in a way similar to the one adopted by Feher and co-workers. In this case 8 compounds were identified as outliers and removed. These compounds are the same ones considered as outliers in model_1, except cmpd_2, which fitted well. The resulting equation is 12 (model_2). The coefficients are of the same order of magnitude with the ones in Equation 6, but model_2 is significantly better in predictive ability.

$$\begin{aligned} \log BB &= -0.1096(\pm 0.0107)\text{HBAC} \\ &\quad - 0.00241(\pm 0.00046)\text{FISA} \\ &\quad + 0.20229(\pm 0.02408)MC \log P_{ow} \\ &\quad + 0.27961(\pm 0.08252) \\ n &= 79, r = 0.932, s = 0.256, \\ F &= 104.37. \end{aligned} \tag{12}$$

In Table 4, the predicted by model_2 log BB values are presented. In Figure 3, the predicted log BB is plotted versus the experimental log BB. Here again we present the excluded compounds.

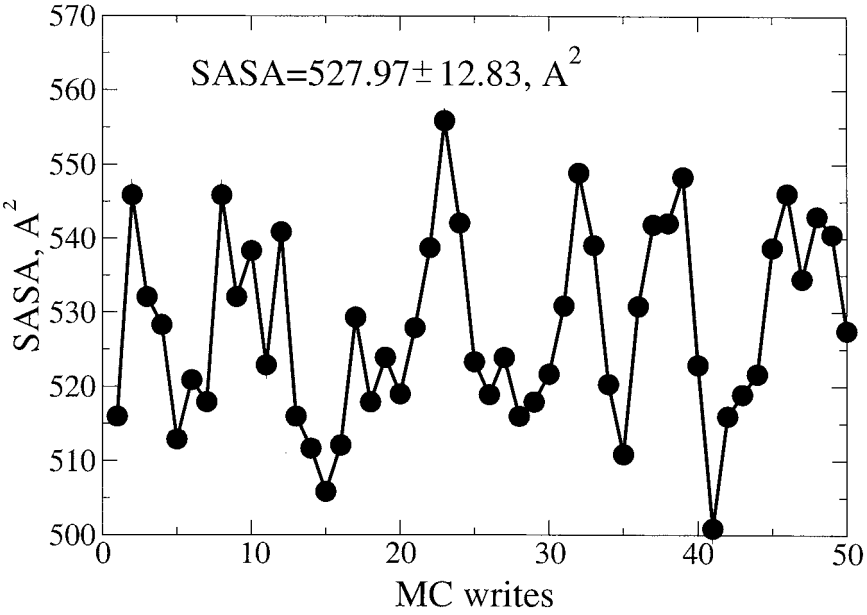

Figure 2. Solvent accessible surface area values of skf89124, recorded every 2×10^5 Monte Carlo steps

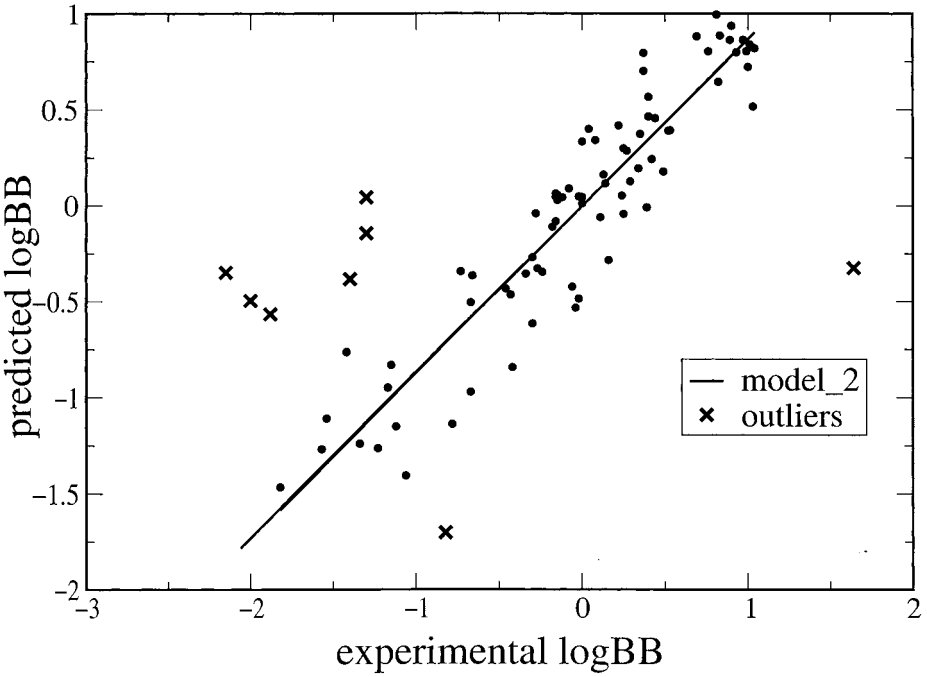

Figure 3. Predicted log BB (using model_2) versus experimentally determined log BB.

We also built a model (model_3) using just two descriptors, $\log P_{ow}$ and *MWEI*, as the one built by Kaliszan and Markuszewski [17]. In this case, only 6 compounds were chosen as outliers and removed, since tiotidine, cmpd_2 and yg_20 fitted well. The resulting Equation is 13 (model_3), which again outperforms the proposed one in [17]. The predicted blood/brain partition coefficients are presented in Table 4.

$$\begin{aligned} \log BB &= -0.00329(\pm 0.00038)\mathrm{MWEI} \\ &\quad + 0.41267(\pm 0.03071) MC \log P_{ow} \\ &\quad - 0.03688(\pm 0.101521) \\ n &= 77, r = 0.862, s = 0363, \\ F &= 180.59. \end{aligned} \tag{13}$$

The fact that in model_3 just six compounds were chosen as outliers, instead of the nine chosen in model_1, or the eight chosen in model_2 reflects the empirical nature of QSPR building. Nonetheless, the consistent classification of icotidine, temelastine, cmpd_4, cmpd_13, Org12962 and yg_19 as outliers in all of our models and in the literature strongly suggests that they might be the subject of possible active mechanisms of transport, or that simply there were difficulties with the experimental measurements.

Finally, if we employ the two descriptors used by van de Waterbeemd and Kansy in Equation 2, namely PSA (we use FISA instead) and MVOL a rather poor model results ($n = 76$, $r = 0.688$, $s = 0.52$, $F = 38.138$).

Conclusions

The present work has yielded quantitative structure-property relationships that accurately predict the blood/brain partition coefficient from properties calculated in statistical mechanics simulations. We have demonstrated that simulations can be employed successfully in a semi-empirical framework to elucidate the mechanisms of biomolecular phenomena. The traditional drawback of the simulations is their computationally intensive nature, necessary for efficient phase space sampling. One can however envision the development of model equations that would refine the information obtained by simulations and allow the prediction of pharmacologically important properties using the molecular structure as the sole input. On the other hand there is need for more detailed simulations that would investigate the diffusion of small drug-like molecules through lipid bilayers. Constant improvement of available computer power and the development of more accurate force-fields for lipids and organic molecules renders increasingly attractive such a research endeavor.

Acknowledgements

YNK is grateful to Pfizer Global Research and Development for the funding of his postdoctoral fellowship. We are grateful to Sangtae Kim and William Jorgensen for useful discussions.

References

1. De Vries, H.E., Kuiper, J., De Boer, A.G., Van Berkel, T.J.C., and Breimer, D.D., Pharmacol. Rev., 49 (1997) 143.
2. Tamai, I., Tsuji, A., Adv. Drug Delivery Rev.,19 (1996) 401.
3. 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.
4. Eddy, E.P., Maleef, B.E., Hart, T.K., and Smith, P.L., Adv. Drug Deliv. Rev., 23 (1997) 185.
5. Reichel, A., and Begley, D.J., Pharm. Res., 15 (1998) 1270.
6. Seiler, P., Eur. J. Med. Chem., 9 (1974) 473.
7. el Tayar, N., Tsai, R.S., Testa, B., Carrupt, P. A., and Leo, A., J. Pharm. Sci., 80 (1991) 590.
8. Fischer, H., Gottschlich, R., and Seelig, A., J. Membrane Biol., 165 (1998) 201.
9. Seelig, A., Gottschlich, R., and Devant, R.M.A., Proc. Natl. Acad. Sci. USA, 91 (1994) 68.
10. Kelder, J., Grootenhuis, P.D., Bayada, D.M., Delbressine, L.P., and Ploemen, J.P., Pharm Res., 16 (1999) 1514.
11. van de Waterbeemd, H., and Kansy, D., Chimia, 46 (1992) 299.
12. Calder, J.A., and Ganellin, C.R., Drug Des. Discov., 11 (1994) 259.
13. Abraham, M.H., Chadha, H.S., and Mitchell, R.C., J. Pharm. Sci., 83 (1994) 1257.
14. Abraham, M.H., and McGowan, J.C., Chromatographia, 23 (1987) 243.
15. Abraham, M.H., Chadha, H.S., and Mitchell, R.C., Drug Des. Discov., 13 (1995) 123.
16. Lombardo, F., Blake, J.F., and Curatolo, W.J., J. Med. Chem., 39 (1996) 4750.
17. Kaliszan, R., and Markuszewski, M., Int. J. of Pharmaceutics, 145 (1996) 9.
18. Norinder, U., Sjoberg, P., and Osterberg, T., J. Pharm. Sci., 87 (1998) 952.
19. Sjoberg, P., In van de Waterbeemd, H., Testa, B., and Folkers, G. (Eds): Computer-Assisted Lead Finding and Optimization, Verlag Helvetica Chimica Acta, CH - 4010 Basel, Switzerland, 1997, pp. 83-92.
20. Wold, S., Johansson, E., Cocchi, M., In Kubinyi, H. (ed.), 3D QSPR in Drug Design, ESCOM, Leiden 1993, pp. 523-550.
21. Jolliffe, I.T. Principal Component Analysis in Chemistry, Springer-Verlag, New York, 1986.
22. Luco, J.M., J Chem Inf Comput Sci., 39 (1999) 396.

23. Feher, M., Sourial, S., and Schmidt, J.M., Int. J. of Pharmaceutics, 201 (2000) 239.
24. Clark, D.E., J. Pharm. Sci. 88 (1999) 815.
25. Bassolino-Klimas, D., Alper, H.E., and Stouch, T.R., Biochemistry, 32 (1993) 12624.
26. Jin, B., and Hopfinger, A.J., Pharm Res. 13 (1996) 1786.
27. Jorgensen W.L, and Duffy E.M., Bioorg. Med. Chem. Lett., 10 (2000) 1155.
28. Duffy, E.M., and Jorgensen, W.L., J. Am. Chem. Soc., 122 (2000) 2878.
29. Weininger, D., J. Chem. Inf. Comp. Sci., 28 (1988) 31.
30. Sybyl 6.4.2. Tripos Inc. 1699 S, Hanley Rd., Suite 303, St. Louis, MO 63144-2913.
31. Jorgensen W.L, In Schleyer, V. R. (ed.), Encyclopedia of Computational Chemistry, Wiley, New York, 1998, Vol. 5, pp. 3281-3285.
32. Jorgensen, W.L., Maxwell, D.S., and Tirado-Rives, J., J. Am. Chem. Soc., 118 (1996) 11225.
33. Kaminski, G.A., and Jorgensen, W.L., J. Phys. Chem. B, 102 (1998) 1796.
34. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., and Klein, M.L., J. Chem. Phys., 79 (1983) 926.
35. Jorgensen, W.L., and Nguyen, T.B., J. Comput. Chem., 14 (1993) 195.
36. JMP, Version 3, SASA Institute Inc.: Cary, NC, 1995.