Predicting Blood–Brain Barrier Partitioning of Organic Molecules Using Membrane-Interaction QSAR Analysis

Manisha Iyer,[1] Rama Mishra,[1] Yi Han,[1] and A. J. Hopfinger[1,2]

Purpose. Membrane-interaction quantitative structure-activity relationship (QSAR) analysis (MI-QSAR) has been used to develop predictive models of blood–brain barrier partitioning of organic compounds by, in part, simulating the interaction of an organic compound with the phospholipid-rich regions of cellular membranes.

Method. A training set of 56 structurally diverse compounds whose blood–brain barrier partition coefficients were measured was used to construct MI-QSAR models. Molecular dynamics simulations were used to determine the explicit interaction of each test compound (solute) with a model DMPC monolayer membrane model. An additional set of intramolecular solute descriptors were computed and considered in the trial pool of descriptors for building MI-QSAR models. The QSAR models were optimized using multidimensional linear regression fitting and a genetic algorithm. A test set of seven compounds was evaluated using the MI-QSAR models as part of a validation process.

Results. Significant MI-QSAR models ($R^2 = 0.845$, $Q^2 = 0.795$) of the blood–brain partitioning process were constructed. Blood–brain barrier partitioning is found to depend upon the polar surface area, the octanol/water partition coefficient, and the conformational flexibility of the compounds as well as the strength of their "binding" to the model biologic membrane. The blood–brain barrier partitioning measures of the test set compounds were predicted with the same accuracy as the compounds of the training set.

Conclusion. The MI-QSAR models indicate that the blood–brain barrier partitioning process can be reliably described for structurally diverse molecules provided interactions of the molecule with the phospholipids-rich regions of cellular membranes are explicitly considered.

KEY WORDS: blood–brain barrier partitioning; QSAR; solute-membrane binding; conformational flexibility.

INTRODUCTION

Central nervous system (CNS) therapeutic agents must cross the blood–brain barrier (BBB) to be effective, whereas peripherally acting drugs must possess very limited ability to cross the BBB. The uptake of a compound into the brain is a complex process (1–4). However, it is seen that moderately lipophilic drugs can cross the BBB by passive diffusion and that the hydrogen bonding properties of drugs significantly influences their particular CNS uptake profiles. Polar molecules are generally poor CNS agents unless they undergo active transport across the CNS. Size, ionization properties,

and molecular flexibility are other factors observed to influence transport of an organic compound across the BBB (4).

It has been shown that in addition to unproved efficacy and toxicity, inadequate pharmacokinetic properties result in the withdrawal of a large proportion of drug leads from further development (5). BBB penetration is one of the most critical pharmacokinetic issues in the design of CNS active drugs and a toxicity concern in the development of other classes of drugs. Experimental measurement of BBB partitioning is difficult, tedious, time-consuming, and costly. Moreover, current experimental approaches to measure BBB partitioning are not amenable to high-throughput screening (HTS), as is becoming increasingly in demand in preclinical drug discovery. Some attempts at HTS of membrane permeation measurements and artificial membrane-based methods have been performed and show promise (6,7).

Recently there has been a surge in computational efforts to compute absorption, distribution, metabolism, excretion, and toxicity (ADMET) properties, including BBB partitioning, of structurally diverse compounds, including drugs (8–14). These new computational approaches remain focused on modeling structurally diverse (solute) data sets by dealing only with the properties of the solutes. Moreover, the solute properties have been largely limited to relative lipophilicity indices, solvation and hydrogen bond parameters, topological indices, and limited three-dimensional solute features. A philosophy has been adopted to get around the limitations of performing a QSAR analysis on a structurally diverse data set. The number of intramolecular solute properties computed is made as large as possible, and then some type of data reduction method is used as part of the data fitting process in constructing the QSAR model. The idea behind this philosophy is that if enough solute features are included, the key intramolecular solute properties for describing multiple mechanisms of action (BBB partitioning) will be captured and built into the QSAR model without doing data overfitting. Of course, there is no way to know 1) if the right set of intramolecular solute features are included in the QSAR descriptor pool and, if indeed, 2) any set of intramolecular solute descriptors exist that can capture the requisite mechanistic information by themselves.

Moreover, once data reduction is performed it becomes exceedingly difficult to interpret the resultant QSAR model and to gain insight into mechanisms of action. The QSAR is in a particular data reduction space and not in actual physicochemical property space. Clearly, some type of structure-based design QSAR approach is needed to meaningfully handle the chemical and structural diversity of the solutes of the training sets encountered in constructing absorption, distribution, metabolism, excretion, and toxicity properties (BBB partitioning) QSAR models.

We have developed a methodology called membrane-interaction (MI)-QSAR analysis, where structure-based design methodology is combined with classic intramolecular QSAR analysis to model chemically and structurally diverse compounds interacting with cellular membranes (15–17). In MI-QSAR analysis, the assumption is made that the phospholipid regions of a cellular membrane constitute the receptor required in structure-based design that permits the incorporation of structural and chemical diversity into a training

Table I. Molecular Weights and Log BB Values for the Molecules of the Training and Test Sets
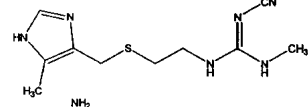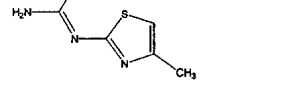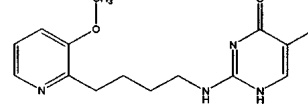
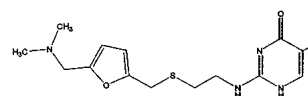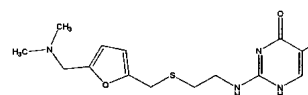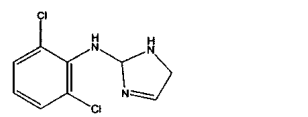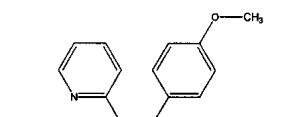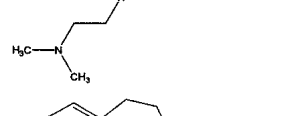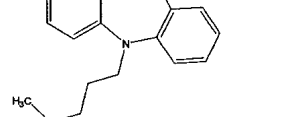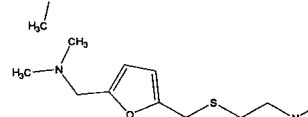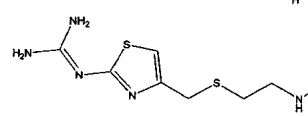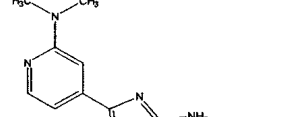| Molecule	MW (a.m.u.)	log BB
Training set		
1	252.34	−1.42
2	156.21	−0.04
3	379.46	−2.00
4	448.58	−1.30
5	413.54	−1.06
6	230.10	0.11
7	285.39	0.49
8	280.41	0.83
9	314.40	−1.23
10	312.41	−0.82
11	204.23	−1.17
12	342.26	−2.15

Table I. Continued
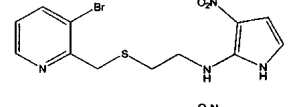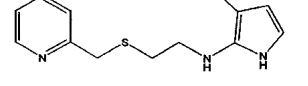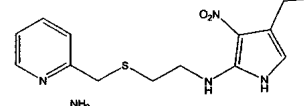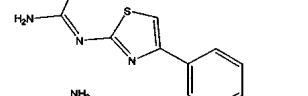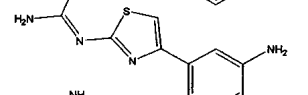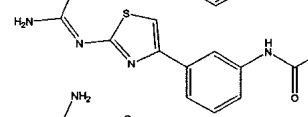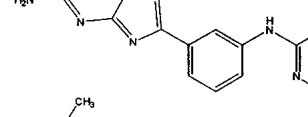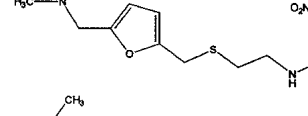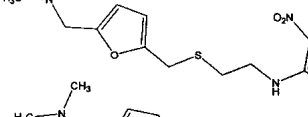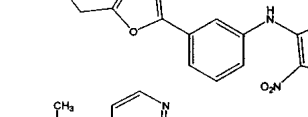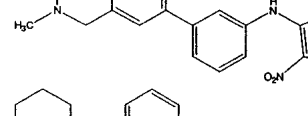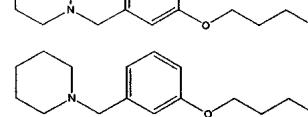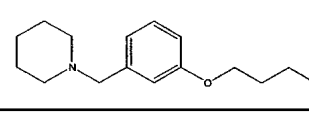
Molecule	MW (a.m.u.)	log BB
Training set		
13	357.22	−0.67
14	278.33	−0.66
15	368.45	−0.12
16	218.28	−0.18
17	233.29	−1.15
18	275.33	−1.57
19	314.37	−1.54
20	324.40	−1.12
21	414.52	−0.73
22	326.35	−0.27
23	337.38	−0.28
24	290.41	−0.46
25	352.48	−0.24
26	249.35	−0.02

Table I. Continued
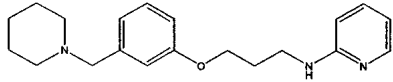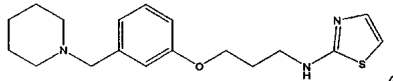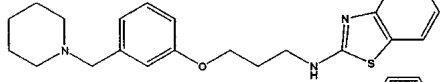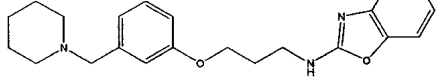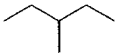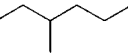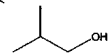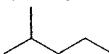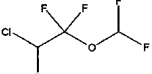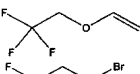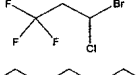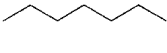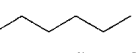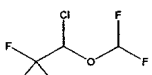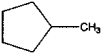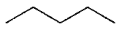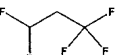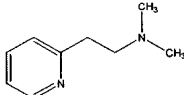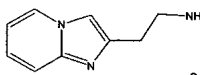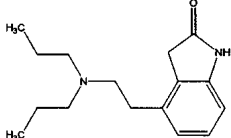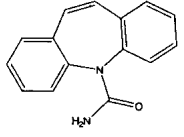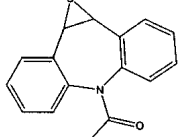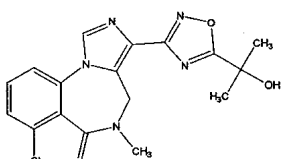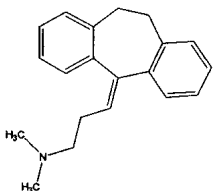
Molecule	MW (a.m.u.)	log BB
Training set		
27	325.45	0.69
28	331.48	0.44
29	381.54	0.14
30	365.47	0.22
31 butanone	72.11	−0.08
32 benzene	78.11	0.37
33 3-methylpentane	86.18	1.01
34 3-methylhexane	100.20	0.90
35 2-propanol	60.10	−0.15
36 2-methylpropanol	74.12	−0.17
37 2-methylpentane	86.18	0.97
38 2,2-dimethylbutane	86.18	1.04
39 1,1,1-trifluoro-2-chloroethane	118.49	0.08
40 1,1,1-trichloroethane	133.41	0.40
41 diethyl ether	74.12	0.00
42 enflurane	184.49	0.24
43 ethanol	46.07	−0.16
44 fluroxene	126.08	0.13
45 halothane	211.41	0.35
46 heptane	100.20	0.81
47 hexane	86.18	0.80
48 isoflurane	184.49	0.42
49 methane	16.04	0.04
50 methylcyclopentane	84.16	0.93

Molecule	MW (a.m.u.)	log BB
Training set		
51 pentane	72.15	0.76
52 propanol	46.07	−0.16
53 propanone	58.08	−0.15
54 teflurane	194.95	0.27
55 toluene	92.14	0.37
56 trichloroethene	131.39	0.34
Test Set		
T1	150.22	−0.06
T2	161.21	−1.40
T3	260.38	0.25
T4	236.27	0.00
T5	252.27	−0.34
T6	373.80	−1.34
T7	277.41	0.85

set. A set of membrane-solute intermolecular properties are determined and added to a set of comprehensive intramolecular solute QSAR descriptors to enlarge the trial QSAR descriptor pool and, ostensibly, to provide the information needed to incorporate chemical and structural diversity into the QSAR analysis.

MI-QSAR analysis has been successfully applied to construct robust models of both eye and skin irritation for structurally diverse training sets (15,16), and to predict the Caco-2 cell permeability of a diverse set of drugs (17). The goal of the study reported here is to demonstrate the applicability of MI-QSAR analysis to model and predict another ADME property, namely BBB partitioning.

METHODS

BBB Partition Coefficients

The dependent variable used in this MI-QSAR analysis is the logarithm of the BBB partition coefficient, log BB = log (C_{brain}/C_{blood}), where C_{brain} is the concentration of the test compound in the brain, and C_{blood} is the concentration of the test compound in blood. Experimental values of log BB published to date lie approximately between −2.00 to +1.00. Compounds with log BB values of >0.3 are readily distributed to the brain whereas compounds with values <−1 are poorly distributed to the brain (18).

Abraham and coworkers (19) have reported a BBB study based on a training set of 57 structurally and chemically diverse molecules. This popular data set has been used in a number of other reported BBB penetration studies (9,10). We omitted one molecule (N_2) from this data set for lack of sufficient computed properties, and used the remaining 56 compounds as the training set for the BBB MI-QSAR study. This training set has a range in molecular weights from 16.03 to 448.58 amu, the concentrations in blood and brain were measured in units of μM/mL, and there are variations in net charge at pH 7.4. Table I, Part A contains the chemical structures and log BB values of the training set compounds. Seven test compounds were selected from the literature whose log BB values were determined by the same protocol used to determine the log BB values of the training set compounds. These seven compounds were used as a validation set and their structures and log BB values are given in Table I, Part B. The test set was established by insisting that its members be representative of all members of the training set in terms of the range of log BB values.

Building Solute Molecules and a Dimyristoylphosphatidylcholine (DMPC) Monolayer

All the solute molecules of the training and test sets (Table I) were built using the Chemlab-II molecular modeling package (20). A single DMPC molecule was built using HyperChem from available crystal structure data (21,22). The AM1 Hamaltonian in Mopac 6.0 was used for the estimation of partial atomic charges on all molecules (23). DMPC was selected as the model phospholipid in this study.

An assembly of 16 DMPC molecules (4 × 4 × 1) in (x,y,z) directions, respectively, was used as the model membrane monolayer. The size of the monolayer simulation system was selected based on the work done by van der Ploeg and Berendsen (24). These workers performed a molecular dynamic simulation (MDS) study for two decanoate bilayers having (2 × 8 × 2) and (2 × 16 × 2) phospholipid molecules. It was found that the estimated order parameters for these two model bilayers agree with one another suggesting that the smaller assembly is adequate for modeling short-range properties. Other researchers have obtained similar geometric and energetic equilibrium property values with regard to the size of model simulation system permitting a minimum effective size (number of phospholipids) of the monolayer to be defined (25). Additional information regarding construction of the model monolayer used in this MI-QSAR analysis can be found in Refs. 15 and 16.

Molecular Dynamics Simulation

The conditions set for the MDS were established in the previous MI-QSAR analyses (15–17) and are only summarized here. An initial MDS on the model membrane, without a solute molecule present, was conducted to allow for structural relaxation and distribution of the kinetic energy over the monolayer. In order to prevent unfavorable van der Waals interactions between a solute molecule and the membrane DMPC molecules, one of the "center" DMPC molecules was removed from the equilibrated monolayer and a test solute molecule inserted in the space created by the missing DMPC molecule. Each of the test solute molecules of the training set was inserted at three different positions (depths) in the DMPC monolayer with the most polar group of the solute molecule "facing" toward the head group region of the monolayer. Three corresponding MDS models were generated for each solute molecule with regard to the trial positions of the solute molecule in the monolayer. The three trial positions were,

1. Solute molecule in the head group region.
2. Solute molecule in between the head group region and the aliphatic chains
3. Solute molecule in the tail region of the aliphatic chains.

The lowest energy geometry of the solute molecule in the monolayer was sought using each of the three trial solute positions. The three different initial MDS positions of ethanol (no. 43 of the training set solute molecules) are shown in Fig. 1a to illustrate this modeling procedure. The energetically most favorable geometry of this solute molecule in the model DMPC monolayer is shown in Fig. 1b.

MDS were performed using the Molsim package with an extended MM2 force field (26). The selection of the simulation temperature was based on the phase transition temperature for DMPC, which is 297°K (27). A simulation temperature of 311°K was selected because it is normal body temperature and also above the DMPC phase transition temperature. Temperature was held constant in the MDS by coupling the system to an external fixed temperature bath (28). The trajectory step size was 0.001 ps over a total simulation time of 12 ps for each test compound. Two-dimensional periodic boundary conditions, corresponding to the "surface plane" of the monolayer, were used (a = 32Å^2, b = 32Å^2, c = 80Å^2, and γ = 96.0°) for the DMPC molecules of the monolayer model but not the test solute molecule. The angle γ is the angle an extended DMPC molecule makes with the "planar surface" of the monolayer. Every membrane-solute system, for the solutes of the training and test sets used in the BBB study, reached equilibrium by 1500 trajectory steps, that is 1.5 ps.
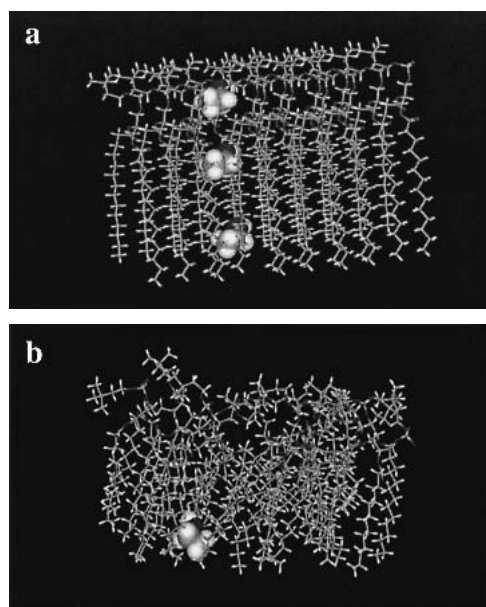
Fig. 1. (a) A "side" view of an ethanol molecule inserted at three different positions in the DMPC model monolayer before the start of each of the three corresponding MDS used in the MI-QSAR modeling. (b) The lowest energy geometry of a DMPC-ethanol complex in the MDS.

Calculation of Descriptors

Both intramolecular physicochemical properties and features of the training set solute molecules, and intermolecular solute-membrane interaction properties were calculated. "Properties" and "features" will both be referred to as descriptors from this point forward because they constitute the trial pool of independent variables used to build the QSAR models. The descriptors used in the MI-QSAR analysis can also be divided into 1) solute aqueous dissolution and solvation descriptors; 2) solute-membrane interaction descriptors; and 3) general intramolecular solute descriptors. The tables reporting the trial pool of descriptors used in the MI-QSAR modeling of BBB partitioning uses both classifications of the descriptors. Most of the general intramolecular descriptors used were calculated using Cerius2 (29) whereas ClogP was calculated using Daylight software (30) and polar surface area values of the training and test set molecules were taken from the study reported by Clark (10). The general intramolecular solute descriptors included as part of the trial descriptor pool are defined in Table II. The term general is used because solute descriptors in this class may be useful in describing different aspects of the bioavailability of a solute.

It should be noted that F(H$_2$O), F(OCT), and ClogP, the aqueous and 1-octanol solvation free energies of the solutes and the corresponding 1-octanol/water partition coefficient, respectively, are computed using intramolecular computational methods. This is also true for E(coh), T_M, and T_G, the cohesive energy and the hypothetical crystal-melt and glass transition temperatures of the solutes, which are used to estimate solute dissolution properties. However, all of these descriptors are intermolecular properties, the first three relating to solute solvation, and the last three to solute dissolution. Therefore, these descriptors are classified as solvation and dissolution intermolecular descriptors and are reported as Part B of Table III.

Table II. The General Intramolecular Solute Descriptors Used as Part of the Trial MIQSAR Descriptor Pool

HOMO	Highest occupied molecular orbital energy
LUMO	Lowest occupied molecular orbital energy
Dp	Dipole moment
Vm	Molecular Volume
SA	Molecular surface area
Ds	Density
MW	Molecular weight
MR	Molecular refractivity
N(hba)	Number of hydrogen bond acceptors
N(hbd)	Number of hydrogen bond donors
N(B)	Number of rotatable bonds
JSSA (X)	Jurs- Stanton surface area descriptors
PSA	Polar surface area
Chi-N, Kappa-M	Kier & Hall topological descriptors
Rg	Radius of Gyration
P_m	Principle moment of inertia
Se	Conformational entropy
Q(I)	Partial atomic charge densities)

E(coh) is a measure of the energy required to remove a molecule from being surrounded by other molecules like itself. T_M measures the crystal packing strength of a molecule, and T_G measures the amorphous packing strength of a molecule. The assumption is made here that E(coh), T_M, and T_G, taken in composite, can used to describe the dissolution behavior of any solute when developing an MI-QSAR model.

The intermolecular solute-membrane interaction descriptors were extracted directly from the MDS trajectories. These particular intermolecular descriptors were calculated using the most stable (lowest total potential energy) solute-membrane geometry realized from MDS sampling of the three initial positions. The intermolecular membrane-solute descriptors extracted from MDS trajectories are given in Table III part A. E(total) is the total potential energy of the membrane-solute complex at the most favorable position extracted from the MDS trajectory, whereas E_{INTER}(total) is the total intermolecular interaction energy between the DMPC monolayer and the solute at this position. E_{INTER}(total) is the sum of the van der Waal, electrostatic and hydrogen bonding interaction energy between the solute and the membrane. $E_{TT}(Z)$ denotes the various contributions to total system potential energy at the optimum (minimum energy) position, where Z could refer to the hydrogen bond energy, electrostatic interaction energy, 1,4-nonbonded interaction energy, general Van der Waals interaction energy, stretching, bending, torsion energy, or combinations thereof. $\Delta E_{TT}(Z)$ has the same meaning to its symbols as $E_{TT}(Z)$ but refers to the change in the particular energy term caused by the uptake of the solute into the membrane at the minimum total potential energy position of the membrane-solute system. $E_{MS}(Z)$ is the intermolecular interaction energy between the membrane and solute at total system minimum potential energy, where Z refers to intermolecular van der Waals interaction energy, intermolecular electrostatic energy, intermolecular hydrogen bonding energy, and combinations thereof. $E_{SS}(Z)$ is the intramolecular energy of the solute within the membrane at the most favorable energy position, where Z denotes 1,4 nonbonded, stretching, bending, torsion energies, and combinations thereof. $\Delta E_{SS}(Z)$ has the same meaning as $E_{SS}(Z)$, except that it refers to the change in the energy of the

Table III. The Intermolecular Interaction Descriptors Included in the Trial MI-QSAR Descriptor Pool[a]

Part A.

The membrane-solute descriptors: symbols	Description of the membrane-solute descriptors
$\langle E(total)\rangle$	Average total potential energy of the solute-membrane complex
$E_{\text{INTER}}(total)$	Total intermolecular interaction energy between the solute and the membrane at the total system minimum potential energy
$E_{\text{TT}}(Z)$	Z = 1,4-nonbonded, general Van der Waal, electrostatic, hydrogen bonding, stretching, bending, torsion and combinations thereof energies of the membrane-solute complex at the total system minimum potential energy.
$\Delta E_{\text{TT}}(Z)$	Change in the Z = 1,4-nonbonded, general Van der Waal, electrostatic, hydrogen bonding, stretching, bending, torsion and combinations thereof at the total [solute and membrane model] intermolecular system minimum potential energy
$E_{\text{MS}}(Z)$	Z = Intermolecular Van der Waal, electrostatic, hydrogen bonding interaction and combinations thereof energies between the solute and the membrane at the total system minimum potential energy
$E_{\text{SS}}(Z)$	Z = Intramolecular 1,4-nonbonded, general van der Waal, electrostatic, hydrogen bonding, stretching, bending, torsion and combinations thereof energies of the solute within the membrane at total system minimum potential energy
$\Delta E_{\text{SS}}(Z)$	Change in the Z = intramolecular 1,4-nonbonded, general van der Waal, electrostatic, hydrogen bonding, stretching, bending, torsion and combinations thereof energies of the solute due to its uptake to the total intermolecular system minimum potential energy.

Part B.

Dissolution and solvation solute descriptors: symbols	Description of the dissolution/solvation solute descriptors
F(H2O)	The aqueous solvation free energy
F(OCT)	The 1-octanol solvation free energy
ClogP	The 1-octanol/water partition coefficient
E(coh)	The cohesive packing energy of the solute molecules
T_{M}	The hypothetical crystal-melt transition temperature of the solute
T_{G}	The hypothetical glass transition temperature of the solute

[a] Part A includes the membrane-solute interaction descriptors, and Part B lists the intermolecular dissolution and solvation descriptors of the solute.

solute due to its uptake into the membrane at the total inter-molecular system minimum potential energy.

Construction and Testing of MI-QSAR Models

MI-QSAR models were built and optimized using multidimensional linear regression fitting and the genetic function approximation (GFA), which is a multidimensional optimization method based on the genetic algorithm paradigm (31,32). Both linear and quadratic representations of each of the descriptor values were included in the trial descriptor pool, and MI-QSAR models were built as a function of the number of descriptor terms in a model. Statistical significance in the optimization of an MI-QSAR model was judged jointly by the correlation coefficient of fit, R^2, and the leave-one-out cross-validation correlation coefficient, Q^2. In addition, GFA uses the Friedman's lack of fit measure to resist overfitting, which is a problem often encountered in constructing and optimizing statistical models (33). MI-QSAR model validation involved the following:

1. Random scrambling of the dependent variable (log BB) measures and attempted construction of statistically significant corresponding MI-QSAR models (34) and
2. Prediction of the log BB values of the test set and comparison to the observed values.

The absence of any significant correlation for each of the scrambled data sets is taken as evidence of the significance of the MI-QSAR models with respect to the original non-scrambled data set. Covariance among the descriptors in the optimized MI-QSAR models was evaluated by constructing the linear cross-correlation matrix of the descriptors, and by comparing relative descriptor usage in the crossover optimization process of the GFA analysis. No significant cross-correlations exist between the descriptors of the best MI-QSAR models.

RESULTS

The best log BB MI-QSAR models are constructed using a combination of general intramolecular solute, intermolecular solute-solvation, and intermolecular membrane-solute descriptors. The best MI-QSAR models are presented below as a function of the number of terms, that is, descriptors included in a given MI-QSAR model:

1. Term model

$$\log BB = 0.543 - 0.0161 PSA \tag{1}$$
$$n = 56 \ R^2 = 0.675 \ Q^2 = 0.647$$

2. Term model

$$\log BB = 0.133 - 0.0153 PSA + 0.1522 ClogP \tag{2}$$
$$n = 56 \ R^2 = 0.744 \ Q^2 = 0.713$$

3. Term model

$$\log BB = 0.122 - 0.0199 PSA + 0.1703 ClogP$$
$$- 0.0049 E_{\text{MS}}(chg + hbd) \tag{3}$$
$$n = 56 \ R^2 = 0.797 \ Q^2 = 0.759$$

4. Term model

$$\log BB = 0.124 - 0.0226 PSA + 0.1613 ClogP$$
$$- 0.0066 E_{MS}(chg + hbd) + 0.0338 E_{SS}(tor) \qquad (4)$$

$$n = 56 \; R^2 = 0.825 \; Q^2 = 0.784$$

5. Term model

$$\log BB = 0.0156 - 0.0231 PSA + 0.1591 ClogP$$
$$- 0.0071 E_{MS}(chg + hbd) + 0.0346 E_{SS}(tor)$$
$$+ 0.0075 \Delta E_{TT}(1-4) \qquad (5)$$

$$n = 56 \; R^2 = 0.845 \; Q^2 = 0.795$$

6. Term model

$$\log BB = -0.015 - 0.0235 PSA + 0.1673 ClogP$$
$$- 0.0076 E_{MS}(chg + hbd) + 0.0388 E_{SS}(tor)$$
$$+ 0.01 \Delta E_{TT}(1-4) - 0.0037 \Delta E_{TT}(stre + bend) \qquad (6)$$

$$n = 56 \; R^2 = 0.855 \; Q^2 = 0.792$$

n is the number of compounds, R^2 is the coefficient of determination, and Q^2 is the cross-validated coefficient of determination.

The descriptors found in the best MI-QSAR models are as follows;

1. ClogP is the calculated octanol-water partition coefficient.

2. PSA is the total polar surface area of a molecule.

3. $E_{MS}(chg + hbd)$ is the total intermolecular electrostatic and hydrogen bonding interaction energy between the solute and the DMPC monolayer when the solute molecule is in the optimum membrane-solute interaction state (minimum potential energy) within the membrane.

4. $E_{SS}(tor)$ is the torsion energy of the solute for the solute located at the position corresponding to the lowest solute-membrane interaction energy state of the model system.

5. $\Delta E_{TT}(1-4)$ is the *change* in the 1,4 nonbonded interaction energy of the system due to the uptake of the solute from free-space to the position corresponding to the lowest solute – membrane interaction energy state of the model system.

6. $\Delta E_{TT}(stre + bend)$ is the *change* in the sum of the total stretching and bending energy of the complex system due to the uptake of the solute from free space to the minimum potential energy state of the membrane-solute complex.

The values of the six descriptors found in the 1 to 6 term MI-QSAR models for each compound in the training and test sets are given in Table IV. The observed and predicted, using the 3 through 6 term MI-QSAR models, log BB values of the test and training set compounds are listed in Table V and plotted in Fig. 2. Compound 12 of the training set is predicted to have a much higher log BB than observed and has been identified as an outlier. This molecule has also been identified as an outlier in other studies, (10,13). Protonation and/or a charged tautomer form of the molecule could account for its low log BB value.

The 2- through 6-term MI-QSAR models are each successive refinements of the preceding smaller descriptor term model. That is the [n+1]-term MI-QSAR model can be viewed as essentially the [n]-term model with one new additional descriptor. The regression coefficients of corresponding descriptor terms across all of the MI-QSAR models are quite similar to one another indicating their respective roles in predicting log BB are about the same in each MI-QSAR model irrespective of the number of descriptor terms in the model.

A test set of seven solute compounds was constructed as one way to attempt to validate the MI-QSAR models given by Eqs. (1–6). The test set solute molecules were selected so as to span almost the entire range in BBB partitioning. The observed and predicted log BB values for this test set are given at the bottom of Table V and plotted as part of Fig. 2. The 3-6 term MI-QSAR models overpredict log BB for test molecule T2 and it has been identified as an outlier.

Figure 3 shows plots of R^2 and Q^2 for the training set, and R^2 for the combined training and test sets, the full set, as a function of the number of descriptor terms in the 3-6 term MI-QSAR models. The dip in value of Q^2 for the training set and R^2 for the combined training and test sets suggests that the 6-term model derived from the training set may be an overfit model.

In addition to the test set given in Table I, 6 sets of 10 compounds have been randomly selected from the training and original test sets as additional test sets with the constraint that each set of 10 compounds uniformly spanned the entire log BB training set range. Full cross-validation has been done on each of these 6 test sets using the 3- through 6-term MI-QSAR models given by Eqs. (1–6). The range in Q^2 over the six test sets is from 0.51 (3-term model) to 0.70 (5-term model). This procedure suggests that the (0.51 to 0.70) range in Q^2 is what should be expected for a typical test set (compound), and may be considered a relative measure of the range of predictability of the MI-QSAR models given by Eqs. (1–6).

DISCUSSION

The family of log BB MI-QSAR models found in this study (Eqs. 1–6) encompasses several significant features and statistical properties. Three of the six significant descriptors of the MI-QSAR log BB models have positive regression coefficients and the other three descriptors have negative regression coefficients. The constant of fit is near zero, and close to the mean in the log BB range of the training set, for all models except the one-term model. Despite allowing for quadratic descriptor terms in the model building process, all the descriptor terms in the family of MI-QSAR models for log BB are linear relationships between individual descriptors and log BB. Overall, these observations regarding the statistical structure of the MI-QSAR family of models strongly suggest they are very solid and reliable models.

Solute aqueous solvation free energy, F(H2O), which is a partial measure aqueous solute solubility, plays a major, if not sometimes dominant, descriptor role in MI-QSAR models for caco-2 cell permeation and eye irritation potency (15–17). F(H2O), and more generally, aqueous solvation of the solute, does not appear to be explicitly involved in the expression of log BB. The total polar surface area (PSA) descriptor, which is found as a dominant descriptor in all of the log BB MI-QSAR models, may include behavior related to the aqueous solubility of the solute. PSA has been shown to be highly useful in modeling intestinal absorption (35) and has yielded models for log BB along with a direct lipophilicity descriptor

Table IV. Values of the Six Descriptors Found to be the Significant MI-QSAR Terms in Eqs. 1–6[a]

Molecule no.	PSA(Å^2)	ClogP	E_{MS}(chg+hbd)	E_{SS}(tor)	ΔE_{TT}(1–4)	ΔE_{TT}(stre+bend)
1	92.100	0.351	−52.460	3.840	20.080	−5.820
2	78.900	0.952	−206.980	3.620	2.530	10.300
3	94.000	2.297	−65.450	4.850	−4.020	37.410
4	73.500	4.046	−10.290	−3.450	20.840	−19.230
5	87.000	1.874	−36.240	6.00	16.920	0.500
6	39.000	0.743	−7.530	5.050	56.590	−6.370
7	26.800	2.787	−7.140	−1.280	12.920	16.900
8	6.000	4.413	−16.520	−2.290	−4.750	26.790
9	84.500	1.327	−10.070	5.810	−18.390	32.430
10	139.200	0.844	−164.150	6.930	6.380	6.580
11	88.800	0.911	−99.350	−1.410	39.480	40.950
12	73.500	2.282	−4.110	1.540	−0.160	34.870
13	83.900	2.747	−6.270	9.100	26.360	−31.110
14	84.000	1.800	−62.080	10.190	12.680	16.780
15	78.000	3.637	−79.530	2.900	15.510	18.940
16	76.600	2.781	−188.830	−5.750	19.070	37.440
17	104.400	1.784	−166.100	−1.350	2.230	19.350
18	108.800	1.977	−139.010	2.380	−1.010	6.650
19	135.800	1.880	−126.240	−2.000	38.660	0.130
20	85.500	2.287	−74.040	11.220	1.250	66.130
21	79.500	4.124	−18.690	6.430	17.680	−34.110
22	82.700	3.849	−80.630	7.260	46.080	29.600
23	85.700	3.234	−13.330	6.880	36.310	−42.070
24	47.900	2.065	2.090	4.870	4.230	−16.610
25	45.200	4.004	−9.810	4.520	23.790	35.130
26	38.500	2.379	0.330	5.260	15.370	32.910
27	39.100	4.259	−7.100	6.650	14.120	21.480
28	40.000	4.165	−27.980	7.100	21.880	9.480
29	39.200	5.759	−10.460	4.420	39.480	23.560
30	54.900	5.029	−33.370	1.070	17.080	−18.340
31	22.700	0.834	−0.690	−6.000	13.120	26.320
32	0.000	2.142	−0.660	−2.900	25.780	15.340
33	0.000	3.738	−0.310	1.760	8.860	26.470
34	0.000	4.267	−1.060	2.180	8.070	17.530
35	23.400	0.074	−24.340	1.410	0.030	32.550
36	22.600	0.693	−35.690	0.570	12.320	22.260
37	0.000	3.738	2.020	1.720	5.200	9.250
38	0.000	3.608	3.450	1.130	18.720	24.630
39	0.000	1.714	1.200	0.040	11.950	−25.070
40	0.000	2.481	−0.720	0.540	16.730	−45.870
41	11.300	0.870	0.780	1.310	23.950	13.420
42	11.600	2.459	−4.180	1.740	25.280	44.850
43	24.400	−0.235	1.760	0.180	34.160	−46.950
44	10.700	1.765	−4.120	2.310	21.400	21.470
45	0.000	2.447	−0.100	1.130	−3.850	32.860
46	0.000	4.397	−4.160	2.650	−2.590	25.230
47	0.000	3.868	0.750	2.040	19.460	31.890
48	11.000	2.999	6.400	0.190	26.000	−28.970
49	0.000	1.103	0.440	0.000	16.290	18.610
50	0.000	3.314	−1.550	6.420	24.380	−1.590
51	0.000	3.339	−3.500	0.180	17.050	6.570
52	24.400	0.294	−27.810	0.680	13.190	−18.180
53	22.700	0.305	1.860	−0.390	9.080	−6.130
54	0.000	2.007	−4.510	0.380	22.980	10.620
55	0.000	2.641	−4.110	−4.480	4.040	16.020
56	0.000	2.627	1.240	0.940	24.370	29.080
Test Set						
T1	18.800	−0.338	−2.910	−1.500	11.400	19.510
T2	46.700	0.110	−3.630	9.130	7.370	20.070
T3	44.100	1.999	44.100	7.270	−14.000	16.830
T4	46.700	1.980	−29.300	9.850	8.960	36.580
T5	62.700	0.260	−16.030	13.570	7.430	2.870
T6	98.500	0.380	98.500	7.690	−4.070	72.840
T7	5.400	4.641	−6.780	9.070	−2.460	22.280

[a] See text and Tables II and III for definitions of the terms. All energies are in kcal/mol

Table V. Observed and Predicted Log BB Values for the 3–6 Term MI-QSAR Models

Molecule no.	Obs. Log BB	Predicted log BB			
		3-term	4-term	5-term	6-term
		Training set			
1	−1.420	−1.390	−1.419	−1.401	−1.349
2	−0.040	−0.261	−0.009	−0.048	−0.008
3	−2.000	−1.032	−1.024	−1.190	−1.335
4	−1.300	−0.599	−0.922	−0.928	−0.839
5	−1.060	−1.109	−1.089	−1.104	−1.070
6	0.110	−0.489	−0.413	−0.113	0.040
7	0.490	0.099	−0.021	−0.056	−0.107
8	0.830	0.836	0.742	0.581	0.471
9	−1.230	−1.282	−1.302	−1.591	−1.785
10	−0.820	−1.690	−1.558	−1.617	−1.589
11	−1.170	−0.997	−1.121	−0.939	−1.004
12	−2.150	−0.930	−1.082	−1.238	−1.403
13	−0.670	−1.047	−0.970	−0.927	−0.745
14	−0.660	−0.934	−0.721	−0.751	−0.756
15	−0.120	−0.416	−0.417	−0.428	−0.437
16	−0.180	0.007	−0.096	−0.031	−0.084
17	−1.150	−0.828	−0.887	−0.967	−1.009
18	−1.570	−1.017	−1.008	−1.125	−1.127
19	−1.540	−1.634	−1.865	−1.707	−1.621
20	−1.120	−0.822	−0.562	−0.674	−0.879
21	−0.730	−0.664	−0.654	−0.676	−0.497
22	−0.270	−0.468	−0.334	−0.114	−0.066
23	−0.280	−0.965	−0.960	−0.843	−0.597
24	−0.460	−0.489	−0.468	−0.577	−0.518
25	−0.240	−0.046	−0.023	0.014	−0.049
26	−0.020	−0.240	−0.180	−0.200	−0.289
27	0.690	0.105	0.210	0.177	0.152
28	0.440	0.175	0.328	0.363	0.415
29	0.140	0.375	0.399	0.551	0.588
30	0.220	0.052	−0.036	−0.051	0.072
31	−0.080	−0.184	−0.267	−0.293	−0.393
32	0.370	0.490	0.381	0.455	0.439
33	1.010	0.760	0.797	0.740	0.672
34	0.900	0.854	0.902	0.838	0.808
35	−0.150	−0.210	−0.183	−0.292	−0.434
36	−0.170	−0.033	−0.017	−0.031	−0.096
37	0.970	0.749	0.780	0.695	0.680
38	1.040	0.720	0.729	0.745	0.703
39	0.080	0.408	0.398	0.371	0.479
40	0.400	0.548	0.553	0.560	0.767
41	0.000	0.042	0.051	0.113	0.101
42	0.240	0.331	0.351	0.419	0.310
43	−0.160	−0.412	−0.470	−0.334	−0.114
44	0.130	0.231	0.277	0.320	0.285
45	0.350	0.540	0.563	0.416	0.278
46	0.810	0.892	0.960	0.817	0.735
47	0.800	0.777	0.820	0.843	0.783
48	0.420	0.383	0.330	0.396	0.558
49	0.040	0.308	0.302	0.311	0.261
50	0.930	0.694	0.893	0.960	1.052
51	0.760	0.708	0.699	0.706	0.725
52	−0.160	−0.175	−0.171	−0.182	−0.100
53	−0.150	−0.286	−0.363	−0.418	−0.412
54	0.270	0.486	0.495	0.553	0.562
55	0.370	0.592	0.431	0.340	0.266
56	0.340	0.564	0.577	0.641	0.589
Test set					
T1	−0.06	−0.295	−0.386	−0.418	−0.508
T2	−1.4	−0.770	−0.578	−0.648	−0.714
T3	0.25	−0.181	0.010	−0.206	−0.281

Table V. Continued

Molecule no.	Obs. Log BB	Predicted log BB			
		3-term	4-term	5-term	6-term
		Training set			
T4	0	−0.324	−0.083	−0.132	−0.223
T5	−0.34	−1.001	−0.683	−0.752	−0.733
T6	−1.34	−1.724	−1.713	−1.896	−2.210
T7	0.85	0.839	1.103	0.972	0.930

(10). Three of the descriptors found in the five-term log BB MI-QSAR model reflect the behavior of the solute in the membrane and/or the entire membrane-solute complex. Thus, the blood–brain partitioning process seems from the MI-QSAR models to largely involve interactions of the solute in a phospholipid-rich membrane medium.

A mechanistic interpretation of each of the MI-QSAR model descriptors is possible which, in turn, permits the realization of a composite self-consistent picture of the blood–brain partitioning process;

1. The PSA of the solute is the dominant descriptor of this training set. This descriptor ranges from 0 to 139.1Å^2 for the molecules of the training set. Overall, this descriptor suggests that as PSA of a solute decreases, log BB increases. This is in agreement with the observation that highly polar molecules do not easily enter the hydrophobic environment of the BBB.

2. As the octanol/water partition coefficient (ClogP) of the solute molecule increases, log BB also increases. Thus, lipophilic drugs have a greater BBB permeability than hydrophilic drugs, which again is in agreement with general observations.

3. As the composite intermolecular electrostatic and hydrogen bond interaction energy between the solute and membrane for the lowest energy "binding" state, $E_{MS}(chg + hbd)$, becomes *more* stable, log BB increases. This relationship suggests that strong binding between the membrane and solute increases BBB permeation. The values of $E_{MS}(chg + hbd)$ are negative for most solutes in the training set.

4. The higher the torsion energy of the solute, $E_{SS}(tor)$, within the membrane for the preferred binding state, the higher its log BB. This relationship would, in turn, suggest that as the solute becomes more flexible within the membrane, the greater would be its log BB value.

5. A positive regression coefficient for $\Delta E_{TT}(1\text{–}4)$ indicates that the greater the change in 1–4 nonbonded energy of the system upon uptake of the solute into the membrane, the higher its log BB value. $\Delta E_{TT}(1\text{–}4)$ is the difference between the 1–4 nonbonded energy of the complex and that of the free solute and membrane. The presence of $\Delta E_{TT}(1\text{–}4)$ in the MI-QSAR model suggests that an increase in the conformational flexibility of the entire DMPC-solute complex, caused by the uptake of the solute, will increase BBB penetration of the solute.

6. $\Delta E_{TT}(stre + bend)$ has a negative regression coefficient in the 6-term MI-QSAR model. This descriptor is not too significant since the 6-term model is likely overfit since its Q^2 value is less than that of the 5-term model. Overall, the presence of this descriptor suggests that if an increase in BBB partitioning can't be fully realized from torsional degrees of
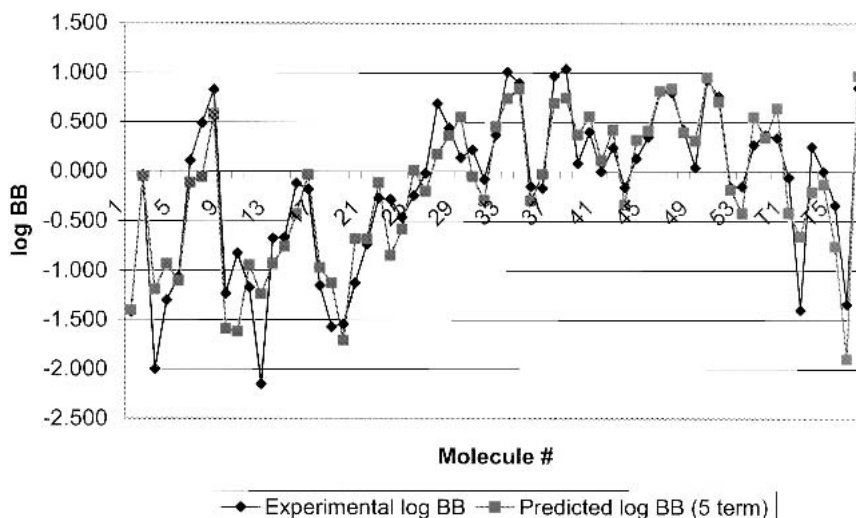
Fig. 2 . Comparison of the experimental log BB values for all the molecules of the training and test sets to the corresponding predicted log BB as predicted by the 5-term MI-QSAR model.

conformational freedom, valence geometry deformations will not facilitate transport, but rather actually diminish BBB penetration.

Overall, the family of log BB MI-QSAR models developed in this study suggests that the blood–brain barrier partitioning process depends upon three physicochemical factors;

1. The relative polarity of the solute as measured in combination by the total polar surface area and the octanol/water partition coefficient of the solute. These two properties indicate that, in general, less polar, more lipophilic compounds partition more readily into the brain.

2. The strength of interaction (binding) of the solute with the membrane. The greater the binding of the solute to the membrane, the higher the BBB partitioning. If the solute does not interact favorably with the membrane, it cannot cross the BBB and does not enter the brain.

3. The conformational flexibility of the solute in the membrane and the conformational flexibility of the membrane-solute complex (36). Increasing solute conformational flexibility within the membrane corresponds to increasing log BB. Solute conformational flexibility is likely related to the

Fig. 3. A diagnostic plot of the MI-QSAR models: R^2 is the correlation coefficient and Q^2 is the cross-validation correlation coefficient of the best x-term model, where x is plotted on the X-axis for the 56 compounds of the training set. R^2 *Full* is the correlation coefficient between the predicted and the observed values for all compounds, that is, the training and test compounds.

ease/difficulty of the solute to 'worm' through the membrane. Also, the more flexible the solute-membrane complex upon uptake of the solute, the greater is log BB. Overall solute-membrane flexibility is again related to the ease of the solute to traverse through the membrane.

Veber and coworkers (37) have found increasing solute molecular flexibility (as measured by the number of rotatable bonds) promotes a decrease in oral bioavailability. This is clearly opposite our findings. We think there may be three possible, and related, explanations to reconcile these seemingly opposite findings. a) The set of oral bioavailability measures used by Veber and coworkers may be governed by different physicochemical behavior than the log BB measures used in this work. b) Molecular flexibility is measured by Veber *et al.* in terms of the number of rotatable bonds. The number of rotatable bonds, in turn, is proportional to molecular weight in most molecules, and this seems to be the case for the Veber *et al.* data set. Therefore, it is possible that the number of rotatable bonds is a playing the role of a masked variable for molecular weight, and the real relationship is increasing molecular weight corresponds to a decrease in log BB. c) A plausible and, likely, global relationship between log BB and molecular flexibility is that some amount of flexibility enhances log BB, but too much flexibility will diminish log BB. That is, a parabolic relationship exists between log BB and molecular flexibility. The range in the number of rotatable bonds is larger for the Veber *et al.* data set than that of the data set used in this MI-QSAR study. Thus, the Veber *et al.* model 'sees' the decrease in log BB with increasing flexibility part of the parabola, while the MI-QSAR model explores the increase in log BB with an increase in flexibility side of the parabola.

Several non-MI-QSAR computational models to describe and/or predict BBB partitioning have been reported (8–14) that involve a large range of descriptors. There is little consistency among the descriptors used in these reported BBB penetration models. The most prevalent descriptors found among the reported models are the 1-octanol/water partition coefficient (log P), some measure of solute volume and some

measure of solute polarity. However, these three classes of descriptors are not all found in a single model among the reported blood–brain partitioning models. Moreover, some of these types of descriptors do not correlate with reported log BB values. For example, Young *et al.* (38) report for a series of 20 structurally diverse histamine H_2 antagonists that log BB does *not* correlate with log P, but does correlate with the both log P(cyclohexane) and the difference in log P and log P(cyclohexane). Hence, it is not feasible to explain BBB penetration with log P, or another polarity descriptor alone, and to attempt to do so would likely be an attempt at oversimplification of the mechanism of penetration.

Our view is that these lipophilicity, size and polarity descriptors of the reported non-MI-QSAR models are capturing glimpses, or pieces, of the three-dimensional process of blood–brain barrier partitioning. MI-QSAR analysis reveals blood–brain barrier partitioning to be governed by solute flexibility, solute-membrane flexibility, solute-membrane binding as well as solute lipophilicity and polarity.

ACKNOWLEDGMENTS

REFERENCES

 1. G. W. Goldstein and A. L. Betz. The blood-brain barrier. *Sci. Am.* **255**:74–83 (1986).
 2. W. M. Pardridge. CNS drug design based on principles of blood-brain barrier transport. *J. Neurochem.* **70**:1781–1792 (1998).
 3. D. J. Begley. The blood-brain barrier: Principles for targeting peptides and drugs to the central nervous system. *J. Pharm. Pharmacol.* **48**:136–140 (1996).
 4. O. G. Mouritsen and K. Jorgensen. A new look at lipid-membrane structure in relation to drug research. *Pharm. Res.* **15**:1507–1519 (1998).
 5. J. H. Lin and A. Y. Lu. Role of pharmacokinetics and metabolism in drug discovery and development. *Pharmacol. Rev* **49**:403–449 (1997).
 6. K. W. Otis, M. L. Avery, S. M. Broward-Partin, D. K. Hansen, H. W. Behlow, D. O. Scott, and T. N. Thompson. Evaluation of the BBMEC model for screening the CNS permeability of drugs. *J. Pharmacol. Toxicol. Methods* **45**:71–77 (2001).
 7. A. Reichel and D. J. Begley. Potential of artificial membranes for predicting drug penetration across the blood-brain barrier. *Pharm. Res.* **15**:1270–1274 (1998).
 8. G. M. Keseru and L. Molnar. High-throughput prediction of blood-brain partitioning: A thermodynamic approach. *J. Chem. Inf. Comput. Sci.* **41**:120–128 (2001).
 9. R. Liu, H. Sun, and S.-S. So. Development of quantitative structure-property relationship models for early ADME evaluation in drug discovery. 2. Blood-brain barrier penetration. *J. Chem. Inf. Comput. Sci.* **41**:1623–1632 (2001).
10. D. E. Clark. Rapid calculation of polar molecular surface and its application to the prediction of transport phenomena. 2. Prediction of blood-brain barrier penetration. *J. Pharm. Sci.* **88**:815–821 (1999).
11. P. Crivori, G. Cruciani, P.-A. Carrupt, and B. Testa. Predicting blood-brain barrier permeation using three-dimensional molecular structure. *J. Med. Chem.* **43**:2204–2216 (2000).
12. J. M. Luco. Prediction of the brain-blood distribution of a large set of drugs from structurally derived descriptors using partial least-square (PLS) modeling. *J. Chem. Inf. Comput. Sci.* **46**:299–303 (1999).
13. F. Lombardo, J. F. Blake, and W. J. Curatolo. Computation of brain-blood partitioning of organic solutes via free energy calculations. *J. Med. Chem.* **39**:4750–4755 (1996).
14. R. Kaliszan and M. Markuszewski. Brain-blood distribution described by a combination of partition coefficients and molecular mass. *Int. J. Pharm.* **45**:9–16 (1996).
15. A. S. Kulkarni, A. J. Hopfinger, R. Osborne, L. H. Bruner, and E. D. Thompson. Prediction of eye irritation from organic chemicals using membrane-interaction QSAR analysis. *Tox. Sci* **59**:335–345 (2001).
16. A. S. Kulkarni and A. J. Hopfinger. Membrane-interaction QSAR analysis: application to the estimation of eye irritation by organic compounds. *Pharm. Res.* **16**:1244–1252 (1999).
17. A. Kulkarni, Y. Han, and A. J. Hopfinger. Predicting caco-2 cell permeation coefficients of organic molecules using membrane-interaction QSAR Analysis. *J. Chem. Inf. Comput. Sci.* **42**:331–342 (2002).
18. M. H. Abraham, K. Takacs-Novak, and R. C. Mitchell. On the partition of ampholytes: Application to blood-brain distribution. *J. Pharm. Sci.* **86**:310–315 (1997).
19. M. H. Abraham, H. S. Chadha, and R. C. Mitchell. Hydrogen bonding. 36. Determination of blood-brain barrier distribution using octanol-water partition coefficients. *Drug Des. Discov.* **13**:123–131 (1995).
20. R. A. Pearlstein. *CHEMLAB-II Users Guide.* CHEMLAB Inc., Chicago, Illinois 1988.
21. Hyperchem HyperChem. *Release 4.5 for MS Windows.* Hypercube Inc, Waterloo, Ontario, 1998.
22. H. Hauser, I. Pascher, R. H. Pearson, and S. Sundell. Preferred conformation and molecular packing of phosphatidylethanolamine and phosphatidylcholine. *Biochim. Biophys. Acta* **650**:21–51 (1981).
23. *Mopac Mopac 6.0 release notes.* Frank J. Seiler Research laboratory, United States Air Force Academy, Colorado Springs, Colorado 1990.
24. P. van der Ploeg and H. J. C Berendsen. Molecular dynamic simulation of a bilayer membrane. *J. Chem. Phys.* **76**:3271–3276 (1982).
25. T. R. Stouch. Lipid membrane structure and dynamics studied by all atom molecular dynamics simulations of hydrated phospholipid bilayer. *Mol. Simulation* **1**:335–362 (1993).
26. D. C. Doherty. *Molsim Version 3.0 User's Guide.* Chicago, The Chem 2 Group Inc., 1994.
27. M. Bloom, E. Evans, and O. Mouritsen. Physical properties of the fluid lipid-bilayer component of cell membranes. A perceptive (review). *Quarterly Rev. Biophys.* **24**:293–397 (1991).
28. H. J. C. Berendsen, J. P. M. Postman, W. F. v. Gunsteren, A. D. Nola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**:3684–3690 (1984).
29. MSI. *Cerius2, Molecular Simulations Users Guide, Ver 3.0.* MSI, San Diego, California 1997.
30. *ClogP Daylight Chemical Information Software, version 4.51.* Daylight Chemical Information Inc., 1998, Los Altos, CA.
31. D. Rogers and A. J. Hopfinger. Applications of genetic function approximation to quantitative structure-activity relationships and quantitative structure-property relationships. *J. Chem. Inf. Comput. Sci.* **34**:854–866 (1994).
32. D. Rogers. WOLF 6.2 GFA Program, 1994.
33. J. Friedman. *Multivariate Adaptive Regression Splines; Department of Statistics.* Stanford University, Stanford, Connecticut 1988.
34. H. V. D. Waterbeemd. *Chemometric Methods in Molecular Design.* VCH Publishers, Inc., New York, 1995.
35. D. E. Clark. Rapid calculation of polar molecular surface area and its application to prediction of transport phenomena. 1. Prediction of Intestinal Absorption. *J. Pharm. Sci.* **88**:807–814 (1999).
36. A. J. Hopfinger. *Conformational Properties of Macromolecule.*, Academic Press, New York, 1973.
37. D. F. Veber, S. R. Johnson, H.-Y. Cheng, B. R. Smith, K. W. Ward, and K. D. Kopple. Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.* **45**:2615–2623 (2002).
38. R. C. Young, R. C. Mitchell, T. H. Brown, C. R. Ganellin, R. Griffiths, M. Jones, K. K. Rana, D. Saunders, I. R. Smith, N. E. Sore, and T. J. Wilks. Development of a new physicochemical model for brain penetration and application to the design of centrally acting H2 receptor histamine antagonists. *J. Med. Chem.* **31**:656–671 (1988).