# A Mathematical Model for Prediction of Drug Molecule Diffusion Across the Blood-Brain Barrier

*Jonathan Burns, Donald F. Weaver*

**ABSTRACT:** ***Background:*** Predicting the ability of drugs to enter the brain is a longstanding problem in neuropharmacology. The first step in creating a much-needed computational algorithm for predicting whether a drug will enter brain is to devise a rigorous mathematical model. ***Methods:*** Employing two experimental measures of blood-brain barrier (BBB) penetrability (brain/plasma ratio and the brain-uptake index) and 14 theoretically derived biophysical predictors, a mathematical model was developed to quantitatively correlate molecular structure with ability to traverse the BBB. ***Results:*** This mathematical model employs Stein's hydrogen bonding number and Randic's topological descriptors to correlate structure with ability to cross the BBB. The final model accurately predicts the ability of test molecules to cross the BBB. ***Conclusions:*** A mathematical method to predict blood-brain barrier penetrability of drug molecules has been successfully devised. As a result of bioinformatics, chemoinformatics and other informatics-based technologies, the number of small molecules being developed as potential therapeutics is increasing exponentially. A biophysically rigorous method to predict BBB penetrability will be a much-needed tool for the evaluation of these molecules.

**RÉSUMÉ: Un modèle mathématique pour prédire la diffusion de molécules à travers la barrière hémato-encéphalique.** ***Introduction :*** En neuropharmacologie, il est difficile de prédire quels médicaments pourront pénétrer dans le cerveau. La première étape dans la création d'un algorithme pour prédire si un médicament pénétrera dans le cerveau est d'élaborer un modèle mathématique rigoureux. ***Méthodes :*** Un modèle mathématique a été développé en utilisant deux mesures expérimentales de la perméabilité de la barrière hémato-encéphalique (BHE) [le ratio cerveau/plasma (RCP) et l'indice de captation du cerveau (ICC)] et 14 prédicteurs biophysiques théoriques, afin de corréler quantitativement la structure moléculaire d'une substance et sa capacité à pénétrer la BHE. Ce modèle mathématique utilise le nombre de liaisons hydrogène de Stein et les indices topologiques de Randic pour corréler la structure de la molécule à sa capacité à pénétrer la BHE. ***Conclusions :*** Une méthode mathématique pour prédire la capacité d'une substance à pénétrer la BHE a été élaborée avec succès. Conséquemment, le nombre de petites molécules en développement a augmenté de façon exponentielle grâce à la bio-informatique, la chimie-informatique et les autres technologies informatiques. Une méthode rigoureuse au point de vue biophysique pour prédire la perméabilité de la BHE sera très utile pour l'évaluation de ces molécules.

Can. J. Neurol. Sci. 2004; 31: 520-527

A robust numerical algorithm to computationally predict the ability of drug molecules to cross the blood-brain barrier (BBB) is of relevance to basic neuroscience and to the pharmacology of drug design.[1-4] A molecule can cross the BBB by either active transport or passive diffusion;[4] passive diffusion remains the most important method for the greatest structural diversity of drug molecules. The two most widely recognized principal physical properties that influence passive diffusion across the BBB (with subsequent entry into the brain) are molecular size and lipophilicity.[4-7] Although equations that quantitatively relate trans-BBB diffusion to these two properties have been proposed,[8-10] these models use only one predictor to encode each of the factors of size and lipophilicity.

This study endeavours to develop a rigorous theoretical prediction algorithm to assess ability to cross the BBB through an analysis of a comprehensive set of molecular predictors reflecting a wider range of physical properties for molecules known either to cross or not to cross the BBB. Such an algorithm will have utility in the development of a computer program for predicting the ability of any clinically employed drug (whether

From the Department of Mathematics, University of Toronto, Ontario (JB); Departments of Medicine (Neurology), Chemistry and Biomedical Engineering, Dalhousie University, Nova Scotia (DFW), Canada.

RECEIVED MAY 16, 2003. ACCEPTEDINFINALFORM MAY 31, 2004.

*Reprint requests to:* D.F. Weaver, Depts. of Medicine (Neurology) and Chemistry, Chemistry Building, Dalhousie University, Halifax, Nova Scotia, Canada, B3H 4J3.

for neurological indications or not) to penetrate the central nervous system and to elicit a biological response, toxic or therapeutic.

## METHODS

The general strategy employed in developing the prediction algorithm was as follows. First a group of numerical values ("predictors") was assembled which would comprehensively reflect the physical properties of the molecules being studied. Next, to develop a prediction algorithm (for ability to cross the BBB), the predictors were calculated for a group of compounds with known biological activity (designated as the "training set"). The predictors were then statistically correlated with the biological properties for the molecules of the training set, thereby permitting the development of the prediction model. The validity of this model was then verified by application to a second group of independent compounds (designated as the "test set") also with known bioactivity.

### Predictor Selection

Twelve predictors were initially selected to describe the physiochemical, topological, electronic and geometric/bulk properties of molecules diffusing across the BBB (later in the study, two additional composite predictors were added). These predictors were chosen to represent the diverse structural properties of molecules in an unbiased, yet comprehensive manner. These twelve predictors have been used extensively in the drug design literature over the past 20 years and have a demonstrated ability to "capture" molecular information that is central to an understanding of biological activity. Initially, all predictors are weighted equally; although there are co-dependencies between various predictors, this does not emerge as a problem in the overall application of the predictors. Physiochemical properties, reflecting differential drug solubility in lipid and aqueous phases, were represented by logP and $(logP)^2$, where P is the octanol-water partition coefficient; these values were calculated using the ClogP computer program.[11] It has been appreciated in drug design for many years that drugs with logP values of 1.5-3.0 seem to have optimal abilities to diffuse through biological membranes and other biological lipid barriers. Topological properties, representing molecular branching and complexity, were determined using the Zagreb (M1, M2), Platt, and Randic $c_1$ - $c_4$ (R1-R4) indices;[12] these values were calculated using empirical graph theory calculations. Topological predictors such as the Randic indices (R1-R4) are useful in differentiating between isomeric drug molecule substituents, such as *n*-butyl [$-CH_2CH_2CH_2CH_3$] and *t*-butyl [$-C(CH_3)_3$], which have the same molecular weights and volume, but very different "branching" properties. Electronic properties, representing regional electron distribution and dipoles within the drug molecule were represented by the hydrogen bonding number (HBN); this was calculated by the method of Stein.[13] The HBN is a simple numerical representation of the number of hydrogen bonding donors and acceptors within the drug molecule. Bulk properties, related to molecular size, were represented by molecular weight (MW) and molecular volume (Vol) determinations; Vol was calculated by the method of Motoc and Marshall.[14]

### Training and Test Set Selection

Compilation of training and test sets required a database of compounds with meaningful measures of BBB permeability. The brain/plasma ratio (BPR) and the brain-uptake index (BUI) method of Oldendorf[15] were the measures of BBB permeability used in this study. These are time-honoured indices that have an extensive history of use in the study of BBB permeability. An extensive literature search identified compounds with BPR and/or BUI values measured reproducibly in mammals using comparable experimental methods. A total of 34 compounds with reported BPRs were used as the training set (BS1) to derive an equation relating BPR to the predictors; 44 compounds with reported BUIs were used as the training set (BS2) to derive the equation for BUI. These compounds are listed in Table 1. (The values of BPR for the drugs listed in BS1 were converted to the percentage of molecule in the brain: BPP=amount in brain/(amount in brain+plasma), to give a proportion).

The test sets were composed of molecules not present in the training set. TS1, the test set for the equations derived for BPP from BS1, thus consisted of those molecules in BS2 not common to BS1. TS2, the test set for BS2, consisted of those molecules in BS1 not common to BS2. In addition, a further 17 molecules, 10 of which were qualitatively known to cross the BBB and seven of which were known not to cross the BBB, with neither BPPnor BUI reported, were added to both test sets (these molecules are listed in Table 2).

Empirically determined prediction cut-off values were calculated for both BPP and BUI to convert the estimated response from the equation to a qualitative "does" or "does not" cross. A predicted value $\geq P_c$ meant a prediction of "crossing"; a value $\leq P_d$ meant "not-crossing". A value between $P_c$ and $P_d$ indicated that equation could not accurately resolve if the molecule crossed. $P_c$ and $P_d$ were determined from an examination of the range of values (BPP/BUI) of the molecules that crossed and of those that did not in BS1 and BS2.

### Statistical Methods

Regression analyses were used to find the best equations expressing BPPor BUI as a function of structural predictors. The BPP data in BS1 was fit with a general linearized model, utilizing a logit transformation[16] on the response (BPP) and quasi-likelihood function;[17] where logit(BPP)=ln(BPP)/(1-BPP). (*Note:* logit(BPP) is henceforth referred to as the "BS1 response"). The BUI data in BS2 was analyzed using multiple linear regression. For the BS2 data, Variance Inflation Factors (VIF) (Montgomery & Peck 1992) and $r^2_{prediction}$ were calculated for each of the fits.[16] The Box-Cox transformation[16] was performed on the BUI values to confirm the correct transformation. (*Note:* the transformed BUI is now referred to as the "BS2 response".) The regression analysis was done on an IBM RS/6000 320 RISC Workstation. Splus was used to compute glm for BS1; Minitab was used for BS2. The methodology of analysis included the following six step strategy:

***Step I:*** The model was fit to the initial factors logP, HBN, MW, Vol, M1, M2, Platt Index and R1-R4 indices, and the significance was analyzed via t-values and $r^2$. Residuals versus fitted values were plotted and outliers noted. Plots were made of response versus factor, as were all the partial residual plots. These were analyzed (first and second order regressions were



https://doi.org/10.1017/S0317167100003759 Published online by Cambridge University Press

**Table 1: Test set molecules based on brain/plasma concentration ratios**

| Compounds in Data Set BS1 | | Compounds in Data Set BS2 | |
|---|---|---|---|
| **Compounds that cross the BBB** | **Compounds that do not cross the BBB** | **Compounds that cross the BBB** | **Compounds that do not cross the BBB** |
| Antipyrine[23] | Acebutolol[35] | Antipyrine[38] | Aldosterone[40] |
| Bromperidol[24,25] | Atenolol[25,32] | Beta-Phenethylamine[22] | Ascorbic Acid[38] |
| Bupropion[26] | Atropine[28] | Caffeine[38] | Chloramphenicol[10] |
| Carbamazepine[27] | 3,4-Dimethyloxynorepinephrine[30] | Carbisocaine[10,39] | Cortisol[40] |
| Chlorpromazine[24,25,28] | Dopamine[28] | Codeine[38] | Cytosine Arabinoside[38] |
| Diazepam[29] | Epinephrine[28] | Corticosterone[40] | Dopamine[22] |
| 3,4-DiMethoxyl-N-MethylEpinephrine[30] | Mesoridazone[24] | Diazepam[10] | Epinephrine[22] |
| 3,4-DiMethyloxyepinephrine[30] | Norepinephrine[28,37] | Ethanol[38] | Glutamine[22] |
| Fluphenazine[24] | Paraephrine[30] | Felbamate[20] | Histamine[22] |
| Haloperidol[24,28] | Reserpine[28] | Heroin[38] | 5-Hydroxytryptamine[22] |
| Imipramine[25,31] | Sotalol[35] | Heptacaine[10,41] | 5-Iodo-2-Deoxyuridine[38] |
| Metoprolol[32] | Triamterene[23] | Imipramine[38] | Mannitol[43] |
| N-Methylmetaeprine[30] | | Isopropanol[38] | Mescaline[22] |
| Phenobarbital[33] | | Levomethadone[38] | Methotrexate[38] |
| Phenytoin[33] | | Nicotine[38] | Morphine[38] |
| Promazine[24,25] | | Paraldehyde[20] | Norepinephrine[21,22] |
| Propranolol[28,32,34,35] | | Phenobarbital[20,38] | Putrescine[44] |
| Rolipram[36] | | Phenytoin[19] | Spermidine[44] |
| Sulforidazine[24] | | Procaine[38] | Spermine[44] |
| Thioridazine[24] | | Testosterone[40] | Sucrose[45] |
| Thyrotropin Releasing Hormone[5] | | Tryptamine[22] | Tyramine[22] |
| | | Tryptophol[42] | |
| | | Zonisamide[19,20] | |

(Literature references for each drug given in suprascript)

**Table 2: Additional Molecules Added to TS1 and TS2 Data Sets**

| Drugs that cross the BBB | Drugs that do not cross the BBB |
|---|---|
| Bicuculline[46] | Baclofen[51] |
| Clonidine[25,47] | Carebastine[25] |
| Desipramine[25] | Cetirizine[48,52] |
| Hydoxyzine[48] | Loperamide[25] |
| Metoclopramide[49] | Ranitidine[53] |
| Metrizamide[50] | Roxatidine[53] |
| Pentylenetetrazole[46] | Trimelamol[54] |
| Perphenazine[25] | |
| Promethrazine[25] | |
| Tamitinol[25] | |

(Literature references for each drug given in suprascript)

performed) to determine if higher-order terms were needed, which were then implemented.

***Step II:*** Multiple techniques were employed to reduce the number of variables required. Factors whose t-values indicated a lack of significance were removed gradually, the regression was re-run, and the $r^2$ (and VIF, $r^2_{pred}$ for BS2) were compared. Stepwise and best-subsets regression were employed when necessary.

***Step III:*** Any outliers identified in Part I were removed from the data set, and Part II was repeated.

***Step IV:*** Those equations that appeared significant were then tested with their appropriate test set; $P_c$, $P_d$ values defined above were used.

***Step V:*** After noting the apparent lack of predictability by logP, $(logP)^2$ was introduced as a factor, and Parts I-IV were repeated.

***Step VI:*** As two previous studies[9,10] have suggested the relationship log(BUI)*sqrt(MW) = k*logP+ b, this simple linear regression was also executed on BS2, as was log(BUI)*sqrt(MW) regressed on HBN. Furthermore, in accord with these previous studies, $logP*MW^{-05}$ and $HBN*MW^{-05}$ were introduced as additional factors for BS2 and Parts I-IV were repeated.



https://doi.org/10.1017/S0317167100003759 Published online by Cambridge University Press

## Results

The prediction cutoffs were determined to be: $P_c$, BPP= 0.35; $P_d$, BPP = 0.15; $P_c$, BUI = 0.30; $P_d$, BUI = 0.15. The Box-Cox transformation plot for BUI data in BS2 had a considerable minimum at $\lambda = 0$, which means that there is a minimum sum of squares error when BUI is transformed to ln(BUI). The partial residual plots and the plots of response vs. factor indicated no need for any transformation or addition of terms.

Mesoradazine was an outlier for the BS1 data. Sucrose was an outlier for the BS2 data for the original regression (and after $(logP)^2$ was added as a factor). However, no outliers were present once the factors of $logP*MW^{-0.5}$ and $HBN*MW^{-0.5}$ were introduced into the model.

The significant regression results are summarized in Tables 3-5.

## Discussion

This study employs 14 different descriptors (See Table 6). As a descriptor of molecular bulk, MW is the most common predictor of size in examinations of BBB diffusibility;[5,6,9,10] Vol is also a logical estimator. As a physiochemical descriptor, logP is a commonly used predictor of lipophilicity. $(LogP)^2$ was added as an additional lipophilicity predictor because initial results indicated that logPwas not significant. Although this unexpected observation is in apparent contradiction with the literature, Hansch et al[18] have indicated that $(logP)^2$ is an acceptable physiochemical measure of lipophilicity. Electronic descriptors reflect desolvation properties. Since a molecule must break hydrogen bonds within its surrounding water hydration shell prior to crossing the BBB,[13] hydrogen bonding ability should be a good predictor of ability to cross. Accordingly, HBN was provided as another predictor. Finally, descriptors of the molecule's topological complexity should be considered since a highly branched molecule and a linear one with the same molecular weight may diffuse across the BBB differently. A molecule with a linear alkyl chain may insert among the alkyl tails of membrane phospholipids, whereas an analogous, but more branched molecule, will behave differently.

In the drug design literature, BPR and BUI have been used for more than twenty years; they are established and "time-honoured" accepted indices. Although some of the data date back more than 15 years, BPR and BUI were chosen because they have an extensive history of use as indices of cerebral access for drug molecules. In recent years, the permeability coefficient-surface area product (PA) has emerged as a rigorous, quantitative and analytically sound measure of blood-brain transfer. Currently, a large body of experimental data of calculated PA's across mammalian cerebral microvasculature is also appearing in the literature. In future refinements of our mathematical algorithm for crossing the BBB, the PA will be a logical replacement for the BPR and BUI as indices of brain penetration.

Many molecules have had their BPRs reported, thus rendering BPR a useful measure of diffusibility. Statistically, as BPR is a ratio, conversion to a percentage between 0 and 1 allows regression via a general linearized model with the logit transformation. This allowed utilization of both non-simple linear regression and simple linear regression, optimizing the advantages of both regression techniques. Using both techniques prevented the results from being prejudiced by an initial assumption that a certain model would provide the best fit. However, there are problems with using BPR. The most significant problem is that different authors calculate BPR at different times post injection: e.g. if a molecule has a BPR at 40 minutes post injection, there is no certainty that its BPR at three hours post injection will be the same. To address this difficulty, where possible, the average BPR was used. Another problem was that a diverse range of mammals was used to calculate BPR; different species may have different BPR values for the same molecule. It is for these reasons that the BPR results are treated with caution; an equation is considered only if a high correlation was achieved.

The BUI is another widely used experimental method for measuring BBB permeability. The technique is relatively standardized, and thus literature results are consistent and comparable. As $0 \leq BUI \leq$ infinity and regression analysis requires a response between (-)infinity and (+)infinity, a transformation was required. Ln(BUI) fulfills this requirement; the Box-Cox method assured that this was correct.

As with BPR, however, there are problems with using BUI. The BUI was chosen because it is standard technique; virtually all the data comes from experimentation on rats, and measurements are done precisely 15 seconds post injection. The only discrepancy is that a slightly modified version of Oldendorf's original method is also in use.[9,10] However, it has been noted that these methods are essentially the same for molecules that cross: e.g., the BBB permeability for the drug zonisamide has been determined by both methods and the two results agree.[9,19,20] There may be a problem for molecules that do not cross; indeed, norepinephrine has a BUI of 1.20 by one method[21] and 4.5 by another.[22] However, this was not a problem because norepinephrine was the only molecule that did not cross which had BUI calculated by Oldendorf's second method. Since both methods are the same for molecules that cross, then, essentially all of the data used herein were calculated by one method.

The molecules chosen to form the data set were selected to optimize the likelihood that the method of BBB traversal was by passive diffusion. Traditionally in drug design, a number of "classic trans-BBB transport systems" are recognized: D-glucose transporter, large neutral amino acid transporter, carboxylic acid transporter. Accordingly, any molecule structurally resembling D-glucose, L-phenylalanine or other actively transported molecules were rejected. This restriction was applied because no equation can be derived confidently to determine if a molecule will be actively transported; either it has a specific transporter, or it does not. Inclusion of transported molecules would skew the equation because they have a different BUI than similar non-transported molecules. However, it is extremely difficult to correctly identify actively transported molecules. Although the "classic trans-BBB transporters" are recognized, P-glycoprotein and other drug efflux transporters also seem to play a major role in determining blood-brain barrier distribution; unfortunately, these transport systems are too promiscuous to allow their substrates to be categorically identified prospectively. Therefore, it is possible that certain molecules included in our data set may in fact cross the BBB by active transport rather than by passive diffusion. While this is a potential limitation of our study, it is offset by the fact that the number of such actively transported molecules is small in number.



https://doi.org/10.1017/S0317167100003759 Published online by Cambridge University Press

**Table 3: Regression Results for BS1**

| Stage | Equation | $R^2$ |
|---|---|---|
| 1a-S1 | y=0.4884-0.287*HBN-0.117*M2+0.245*PI | 0.318 |
| 1a-S2 | y=7.143+0.520*logP-0.0570*Vol+1.717*HBN+0.153*PI+1.313*R3-2.335*R4+0.00755*Vol*HBN | 0.600 |
| lb-S1 | y=-0.427-0.328*HBN+0.876*R3-0.721*R4 | 0.480 |
| 1b-S2 | y=5.6707+0.671*logP-0.021*MW-1.492*HBN+0.093*M2+0.9296*R3-3.089*R4+0.2097*HBN*R3 | 0.687 |
| 2a-S1 | same as 1a-S1 | |
| 2a-S2 | y=5.902-0.568*HBN+0.434*logP-0.0199*$(logP)^2$-0.852*R3+0.117*HBN*$(logP)^2$ | 0.722 |
| 2b-S1 | same as 1b-S1 | |
| 2b-S2* | y=4.2634+0.8979*logP-0.3524*HBN-0.4218*R1 +0.8153*R2-0.4885*R3-0.0694*logP*HBN-0.0454*logP*R1 | 0.751 |

NOTE: y refers to the response (logit(BPP)); S1 refers to stepwise of order 1; S2 refers to stepwise of order 2.
Stage 1: Before the addition of $(logP)^2$ as a factor: a): before removal of outlier; b): after removal of outlier.
Stage 2: After the addition of $(logP)^2$ as a factor: a): before removal of outlier; b): after removal of outlier.
*Further elimination was done to remove predictors with low t-values which did not contribute greatly to the overall fit.

**Table 4: Stepwise Regression Results for BS2**

| Stage | Equation | $R^2$ | $R^2p$ | # |
|---|---|---|---|---|
| 1a-S1 | y=4.188-0.341*HBN | .535 | .453 | 0 |
| 1a-S2 | y=4.26+0.90*logP-0.352*HBN-0.422*R1 +0.815*$R^2$-0.489*R3-0.0694*logP*HBN-0.0454*logP*R1 | .751 | .662 | 5 |
| 1b-S1 | y=4.01-0.436*HBN+0.379*$R^2$-0.378*R3 | .668 | .614 | 0 |
| 1b-S2 | y=3.45+1.80*logP-0.322*HBN-0.664*M1+0.629*M2-0.895*PI+5.09*$R^2$-0.343*R3+ 0.02*Vol -0.106*logP*HBN-0.577*logP*$R^2$ +0.0371*logP*M1 | .824 | .543 | 9 |
| 2a-S1 | y=3.97+0.042*$(logP)^2$-0.35*HBN+0.0443*PI-0.374*R3 | .618 | .482 | 2 |
| 2a-S2 | y=3.74+1.36*logP+0.233*$(logP)^2$-0.374*HBN-0.768*M1+0.344*M2+1.19*R1+3.32*$R^2$ -0.391*R3 -0.131*logP*R1-0.012*$(logP)^2$*R1 | .796 | .535 | 9 |
| 2b-S1 | y=4.25+0.0259*$(logP)^2$-0.412*HBN +0.0486*PI-0.41*R3 | .662 | .593 | 2 |
| 2b-S2 | y=3.76+1.36*logP+0.23*$(logP)^2$-0.368*HBN-0.765*M1+0.342*M2+1.18*R1+3.31*$R^2$ -0.389*R3 -0.131*logP*R1-0.012*$(logP)^2$*R1 | .793 | .613 | 9 |

NOTE: y refers to the response (ln(BUI))
$R^2_p$ refers to $R^2_{prediction}$
# refers to the number of predictors with Variance Inflation Factor larger than 10.
S1 refers to stepwise of order 1; S2 refers to stepwise of order 2.
Stage 1: Before the addition of $(logP)^2$ as a factor: a): before removal of outlier; b): after removal of outlier.
Stage 2: After the addition of $(logP)^2$ as a factor: a): before removal of outlier; b): after removal of outlier.

**Table 5: Summary of Regression Results with Introduction of HBN*$MW^{-05}$ and logP*$MW^{-05}$ as Predictor Variables**

| Method | Equation | $R^2$ | $R^2p$ | # |
|---|---|---|---|---|
| 1 | log(BUI)*$MW^{0.5}$ = 36.3 + 7.71*logP | .526 | .481 | - |
| 2 | log(BUI)*$MW^{0.5}$ = 60.3 - 4.75*HBN | .417 | .327 | - |
| Be3 | y=4.88-0.119*R3+0.0375*$(logP)^2$-5.84*HBN*$MW^{-0.5}$ | .624 | .536 | 0 |
| Be4 | y=4.11-1.05*logP-0.367*HBN+0.0547*$(logP)^2$ +16.7*logP*$MW^{-0.5}$ | .643 | .547 | 2 |
| Be5 | y=5.01-0.0153*MW+0.641*$R^2$-0.331*R3+0.0501*$(logP)^2$-5.38*HBN*$MW^{-0.5}$ | .678 | .571 | 2 |
| Be6 | y=3.92-1.35*logP-0.412*HBN+0.363*$R^2$-0.377*R3+0.0622*$(logP)^2$+20.2*logP*$MW^{-0.5}$ | .691 | .574 | 4 |

As above, y refers to the response (ln(BUI))
Method 1, 2 : Comparison with equations previously suggested.
Be - Best Subsets Model of given order.



https://doi.org/10.1017/S0317167100003759 Published online by Cambridge University Press

**Table 6: List of Biophysical Descriptors Employed in Study**

**A. Physiochemical Descriptors** (measure of solubility and ability to cross the blood-brain barrier)
1. logP– octanol/water partition coefficient
2. $(logP)^2$ – square of logP

**B. Electronic Descriptor** (measure of electron distribution properties)
3. Hydrogen Bonding Number (HBN) – number of hydrogen bonding donors/acceptors

**C. Topological Descriptors – Graph Theory Indices** (measure of molecular "branching")
4. Zagreb Topological Index M1
5. Zagreb Topological Index M2
6. Platt Topological Index
7. Randic Topological Index R1
8. Randic Topological Index R2
9. Randic Topological Index R3
10. Randic Topological Index R4

**D. Topological Descriptors – MolecularBulk Indices** (measure of molecular volume/size)
11. Molecular Weight
12. Molecular Volume

**E. Composite "Hybrid" Descriptors** (capturing combined properties from Groups A-D)
13. $logP*MW^{-0.5}$
14. $HBN*MW^{-0.5}$

Another restriction applied when assembling the data set was that large numbers of structurally similar molecules were not included. Inclusion of a large number of similar molecules within a given analogue series would cause the regression equation to be weighted to those values contained in the similar set. However, these values may not be an accurate measure for a random molecule. Restricting the number of similar molecules makes the data sets more homologous to a pseudo-random sample.

All BS1 molecules known to cross the BBB had BPP≥ 0.35. While some BPP values of molecules that do not cross the BBB exceeded the $P_d$, BPPof 0.15, 0.15 was set as the maximal upper limit. A value of 0.20, for example, meant that 20% of the drug was present in the brain; this is a considerable amount, and suggests that the drug has an ability to cross. From BS2, there was a distinction between molecules with BUI in single digits, and those over 20%. Rationally, substances with an uptake less tahn 15% of water do not cross. However, once the uptake reaches 30% of that of water, crossing is indicated; these values were therefore selected as the cut-offs.

As inclusion of nonsignificant factors decreases model predictability, stepwise and best-subsets regression were utilized to find the minimum set of descriptors that significantly explained the response. Furthermore, any models with factors that were highly correlated (indicated by VIF greater than 10) were viewed with great caution, as high correlation suggests that the regression coefficients are poorly estimated.[16]

The best BPP model is derived from stepwise regression. However, even though the second order models have somewhat better $r^2$ and satisfy the test sets better, they are not representative models because there is no biological basis to include the presence of mixed second order terms. Hence, the choice is between the two first order models with the best BPP model being:

$$BPP = \frac{\exp(-0.427 - .328HBN + .876R3 - .721R4)}{1 + \exp(-0.427 - .328HBN + .876R3 - .721R4)}$$

However, all BPP models were poor predictors of the ability of test set molecules to cross the BBB. It appears that the BPP models predict erroneously high BPP values.

From best subsets regressions, the most significant model (based on $r^2$, $r^2_{pred}$ and VIF) is the model of order three (after the outlier was removed). When compared with the results from the stepwise regression, the stepwise second order results cannot be considered significant due to a lack of biological basis to include mixed terms. As well, all of these models have serious multicollinearity problems. While some of the stepwise first order models have a better $r^2$, as they all have lower $r^2_{pred}$ and/or a considerable multicollinearity problem, the best subsets model still appears optimal. With regard to the models after introduction of $HBN*MW^{-0.5}$ and $logP*MW^{-0.5}$, these too also have lower $r^2_{pred}$ and/or correlation among the variables.

Nevertheless, the best subsets model is *not* the best predictor model. While it does predict those molecules that cross with the most accuracy, it has the least accuracy for those molecules that do not cross. In particular, the model 2b-S1 is a better overall predictor. It has a similar $r^2$, a slightly lower $r^2_{pred}$; however, there is the presence of some multicollinearity. As well, Be5 (after removal of outlier) also is a better overall predictor, with a higher $r^2$ and $r^2_{pred}$, but multicollinearity problems. It is these multicollinearity problems that cause the choice of the initial best subsets regression as the best overall model:

$$\ln(BUI) = 4.01-0.436*HBN+0.379*R^2-0.378*R3.$$

The final BUI equation is the best model to predict trans-BBB diffusibility. Not only does it have a better $r^2$ than the BPPmodel and predicts the test set much better, but also the problems involved with using the BPP data are avoided. This model is intuitive: the greater the hydrogen-bonding ability, the lower ability to cross. As well, a topological indication of branching is a better predictor of molecular bulk than MW.

The model has good accuracy, with an $r^2$ = 67% indicating significant correlation. Although it is a good predictor of ability to cross for molecules that do so (accurate 77% of the time and only returns an inaccurate answer 8% of the time), it is only a fair predictor of inability to cross (accurate 44%, but inaccurate 33% of the time). A probable reason that prediction (and thus by implication the fit) is worse for molecules that do not cross can be seen from an examination of BS2. There are 21 molecules that do not cross; all have BUI≤10%. Since there are many molecules with different values of the predictors scattered over a relatively small response range (as opposed to 24 molecules with 0.2≤BUI≤130 for those that cross), it will be difficult to ascertain an accurate fit. It is possible that there is no equation, based on



https://doi.org/10.1017/S0317167100003759 Published online by Cambridge University Press

the factors of lipophilicity and size alone, which predicts diffusibility with great accuracy.

The accuracy of our results is similar to those suggested by earlier authors:
$\log P_c = -4.605+0.4115\log[P(MW)^{-0.5}]$, $r^2=0.83$ (Levin[8]) ($P_c$ is permeability coefficient.)
$\log(BUI)*\sqrt{MW} = 6.02\log P+ 14.5$, $r^2 = 0.74$ (Cornford et al[9])
$\log(BUI)*\sqrt{MW} = -3.77HBN + 30.78$, $r^2 = 0.34$ (Cornford et al[9])
$\log(BUI)*\sqrt{MW} = 7.3\log P + 17.7$, $r^2 = 0.74$ (Bezek[10])

Although cross publication comparisons of models based on $r^2$ derived from different data sets are not necessarily reliable, they are nevertheless a good indication that our model is as good (or better) than previously published models.

It is significant to note that the equation derived in this study does not have logPas a predictor, which appears to disagree with literature precedent. Arguably, there may be a physical basis for this observation. LogPmay be too specific and "pure" a measure of lipophilicity for BBB penetrability. It does not adequately encompass the full spectrum of molecular events as a drug molecule is desolvated prior to diffusion into the BBB. From a thermodynamic perspective, the energy associate with water-drug hydrogen bond breaking (as the drug leaves the aqueous serum prior to entering the lipid membrane) may constitute a more significant factor than vaguer hydrophobic interactions and lack thereof between the membrane and molecule as represented by logP. Thus, molecules with different logPbut same HBN may have (by this argument) similar BBB penetrability.

In addition, the discrepancy may also have a statistical explanation. The fact that logP was insignificant in the equation derived herein does not mean that it is not an effective predictor of ability to cross (i.e. log(BUI)*sqrt(MW) on logP has high $r^2$), but that it has low last-in p-values/t-ratio. Last-in t-values are a measure of the significance of a regressor variable after all of the effects of all of the other regressor variables are taken in to account. Hence, it could be that the predictability of logP is simply accounted for by a combination of the other regressor variables of size and lipophilicity.

A method to predict the BBB diffusibility of a molecule is given by the equation: $BUI = \exp(4.01-0.436HBN+0.379R^2-0.378R3)$. A response of BUI ≤ 15% indicates that the molecule does not cross; a response of BUI ≥ 30% indicates that the molecule crosses. A value 15% ≤ BUI ≤ 30% indicates that the equation cannot accurately determine if the molecule will cross. The data used to derive this equation indicated good correlation, with $r^2 = 66.8\%$, and $r^2_{pred} = 61.4\%$. This equation indicates that the HBN and Randic topological indices are important predictors of ability to traverse the BBB. The HBN reflects a variety of properties including sites of hydration on the molecule which must be desolvated prior to crossing the BBB. The Randic indices, as described by Balaban et al,[12] are complex descriptors which reflect the size and branching complexity of the drug molecule.

The development of this algorithm is an important first step in the creation of a computer program with which to predict the ability of any drug molecule to cross the BBB, thereby influencing neurological function. As a result of bioinformatics, chemoinformatics and other informatics-based technologies, the number of small molecules being developed as potential therapeutics is increasing exponentially. A computer-based method to predict BBB penetrability will be a much-needed tool for the evaluation of these molecules.

## Acknowledgements

The authors thank Dr. E. Chow, Department of Mathematics and Statistics, Queen's University, who provided advice and help with the statistical computations. This work was supported in part by an operating grant from the Natural Sciences and Engineering Research Council of Canada. DFW is a Tier 1 Canada Research Chair in Neuroscience.

## References

1. Rapoport SI. Blood-Brain Barrier in Physiology and Medicine. New York: Raven Press. 1976: 43-86, 153-176.
2. Pardridge WM. Recent advances in blood-brain barrier transport. Ann Rev Pharmacol Toxicol 1988; 28:25-39.
3. Pardridge WM. Strategies for drug delivery through the blood-brain barrier. Microbiol Aging 1989; 10:636-637.
4. Pardridge WM. Peptide Drug Delivery to the Brain. New York: Raven Press. 1991: 1-11, 23-51, 63-67, 123-134.
5. Banks WA, Kastin AJ. Peptides and the blood-brain barrier: lipophilicity as a predictor of permeability. Brain Res Bull 1985;15:287-292.
6. Banks WA, Kastin AJ. Peptide transport systems of opiates across the blood-brain barrier. Am J Physiol 1990; 259:E1-E10.
7. Spector R. Fatty acid transport through the blood-brain barrier. J Neurochem 1988; 50:639-643.
8. Levin VA. Relationship of octanol/water partition coefficient and molecular weight to rat brain capillary permeability. J Med Chem 1980; 23:682-684.
9. Cornford EM, Braun LD, Oldendorf WH, Hill MA. Comparison of lipid-mediated blood-brain barrier penetrability in neonates and adults. Am J Physiol 1982; 243:C161-C170.
10. Bezek S, Trnovec T, Scasnar V, et al. Irradiation of the head by 60Co opens the blood-brain barrier for drugs in rats. Experientia 1990; 46:1017-1020.
11. BioByte Corporation (1994) MacLogP™
12. Balaban AT, Motoc I, Bonchev D, Mekenyan O. Topological indices for structure-activity correlations. In: Charton M, Motoc I (Eds). Steric Effects in Drug Design (Topics in Current Chemistry, Vol 114). Berlin: Springer-Verlag, 1983; 21-55.
13. Stein WD. The Movement of Molecules Across Cell Membranes. New York: Academic Press. 1967; 76.
14. Motoc I, Marshall GR. Van der Waals volume fragmental constants. Chem Phys Lett 1985; 116:415-419.
15. Oldendorf WH. Measurement of brain uptake of radiolabeled substances using a tritiated water internal standard. Brain Res 1970; 24:372-376.
16. Montgomery DC, Peck EA. Introduction to Linear Regression Analysis. 2nd ed. New York: John Wiley & Sons. 1992; 103-105, 173-177, 189-192, 265-304, 317-318 .
17. Nelder JA, McCullagh P. Generalized Linear Models. 2nd ed. London: Chapman & Hall. 1991; 323-356.
18. Hansch C, Steward A, Anderson SM, Bentley D. The parabolic dependence of drug action upon lipophilic character as revealed by a study of hypnotics. J Med Chem 1968; 11:1-11.
19. Cornford EM, Landon KP. Blood-brain barrier transport of CI-912: single passage equilibration of erythrocyte-borne drug. Ther Drug Monit 1985; 7:247-252.
20. Cornford EM, Young D, Paxton JW, Sofia RD. Blood-brain barrier penetration of felbamate. Epilepsia 1992; 33:944-954.
21. Bradbury MW, Bloom DS, McDowell M. An inhibitor of cerebral uptake of noradrenaline in jaundiced blood plasma. J Cerebral Blood Flow Metab 1983; 3:516-520.
22. Oldendorf WH. Brain uptake of radiolabeled amino acids, amines, and hexoses after arterial injection. Am J Physiol 1971; 221:1629.
23. Dayton PG, Pruitt AW, McNay JL, Steinhorst J. Studies with triamterene, a substituted pteridine. Unusual brain to plasma ratio in mammals. Neuropharmacol 1972; 11:435-446.

24. Tsuneizumi T, Babb SM, Cohen BM. Drug distribution between blood and brain as a determinant of antipsychotic drug effects. Biol Psychiatry 1992; 32:817-826.
25. Selig A, Gottschlich R, Devant RM. A method to determine the ability of drugs to diffuse through the blood-brain barrier. Proc Natl Acad Sci USA. 1994; 91:68-72.
26. Schroeder DH. Metabolism and kinetics of bupropion. J Clin Psychiatry 1983; 44:79-91.
27. Friis ML, Christiansen J, Hvidberg EF. Brain concentrations of carbamazepine and carbamazepine-10,11-epoxide in epileptic patients. Eur J Clin Pharmacol 1978; 14:47-51.
28. Watanabe T, Matsuhashi K, Takayama S. Placental and blood-brain barrier transfer following prenatal and postnatal exposures to neuroactive drugs: relationship with partition coefficient and behavioural teratogenesis. Toxicol Appl Pharmacol 1990;105:66-77.
29. Dubey RK, McAllister CB, Inoue M, Wilkinson GR. Plasma binding and transport of diazepam across the blood-brain barrier. No evidence for in-vivo enhanced dissociation. J Clin Invest 1989; 84:1155-1159.
30. Evans BD, Vogel WH. Penetration of some O-and/or N-Methylated norepinephrine derivatives through the rat blood-brain barrier. Res Comm Chem Path Pharmacol 1977; 17:61-76.
31. Barkai AI, Suckow RF, Cooper TB. Imipramine and its metabolites: relationship to cerebral catecholamines in rats in-vivo. J Pharmacol Exp Ther 1984; 230:330-335.
32. Neil-Dyer G, Bartlett J, McAinsh J, Cruickshank JM. Beta-adrenoceptor blockers and BBB. Br J Clin Pharm 1981; 11:549-553.
33. Cornford EM, Diep CP, Pardridge WM. Blood-brain barrier transport of valproic acid. J Neurochem 1985; 44:1541-1550.
34. Tocco DJ, Clineschmidt BV, Duncan AEW, deLuna FA, Bear JE. Uptake of the beta-adrenergic blocking agents propranolol and timolol by rodent brain: relationship to central pharmacological action. J Cardiovasc Pharmacol 1980; 2:133-143.
35. Abernethy DR, Arendt RM, Greenblatt DJ. Pharmacologic properties of acebutolol: relationship of hydrophilicity to central nervous system penetration. Am Heart J 1985; 109:1120-1125.
36. Krause W, Kuhn G. Pharmacokinetics of rolipram in the rhesus and cynomolgus monkeys, the rat and the rabbit. Studies on species differences. Xenobiotica 1988; 18:561-571.
37. MacKenzie E, McCulloch J, Harper AM. Cerebral circulation and norepinephrine: relevance of the blood-brain barrier. Am J Physiol 1976; 231:483-488.
38. Oldendorf WH. Lipid solubility and drug penetration of the blood-brain barrier. Proc Soc Exp Invest 1974; 63:1241-1248.
39. Bezek S, Faberova V, Scasnar V, et al. Pharmacokinetics of carbisocaine in rats and mice. Eur J Drug Metab Pharmacokinet 1988; 13:27-34.
40. Pardridge WM. Transport of protein-bound hormones into tissue in-vivo. Endocr Rev 1981; 2:103-123.
41. Scasnar V, Kallay Z, Bezek S, Trnovec T, Durisova M. Pharmacokinetics of the new local anaesthetic N-[2-(2-Heptyloxyphenyl-carbamoyloxy)ethyllpiperidinium chloride in rats and mice. Arzneimittelforschung - Drug Res 1987; 37:783-787.
42. Cornford EM, Bocash WD, Braun LD, et al. Rapid distribution of tryptophol (3-indole ethanol) to the brain and other tissues. J Clin Invest 1979; 63:1241-1248.
43. Zlokovic BV, Segal MB, Begley DJ, Davson H, Rakic L. Permeability of the blood-cerebrospinal fluid and blood-brain barriers to thyrotropin releasing hormone. Brain Res 1985; 358:191-199.
44. Shin WW, Fong WF, Pang SF, Wong PC. Limited blood-brain barrier transport of polyamines. J Neurochem 1985; 44:1056-1059.
45. Zlokovic BV, Begley DJ, Chain-Eliask DG. Blood-brain barrier penetrability to leucine-enkephalin, D-alanine$^{2}$-D-Leu$^{5}$-enkephalin and their N-terminal amino acid (tyrosine). Brain Res 1985; 336:125-132.
46. Engstrom FL, Woodbury DM. Seizure susceptibility in DBA and C57 mice: the effects of various convulsants. Epilepsia 1988; 29:389-395.
47. Nordling J, Meyoff JJ, Hald T. Sympatholytic effect on striated urethral sphincter. A peripheral or central nervous system effect? Scand J Urol Nephrol 1981; 15:173-180.
48. Wasserman S. Histamine and the preclinical pharmacology of cetirizine. Ann Allergy 1987; 57:1-3.
49. Nagahama S, Chen YF, Oparil S. Mechanism of the depressor effect of bromocriptine in the spontaneously hypertensive rat. J Pharmacol Exp Ther 1984; 228:370-375.
50. Kerber CW, Sovak M, Ranganathan RS, Heilman CB. Totrol: a new myelographic agent: 1. Radiography, CT, CSF clearance, and brain penetration. AJNR Am J Neuroradiol 1983; 4:317-318.
51. Feldmen J, Bousquet P. A new procedure for studying central acting cardiovascular drugs: the permeabilization of the blood-brain barrier. Arch Inter Pharmacodynam Ther 1988; 296:7-17.
52. Ryhal BT, Fletcher MP. The second-generation antihistamines. What makes them different? Postgrad Med 1991; 89:87-88.
53. Tryba M, Kurz-Muller K. Penetration of roxatidine into the cerebrospinal fluid. Scand J Gastroenterol 1988; 146:153-158.
54. Judson IR, Rutty CJ, Abel G, Graham MA. Low central nervous system penetration of N2, N4, N6-trihydroxymethyl-N2, N4, N6-trimethylmelamine (Trimelamol): a cytotoxic s-triazine with reduced neurotoxicity. Br J Cancer 1986; 53:601-606.