Prediction of the Brain−Blood Distribution of a Large Set of Drugs from Structurally Derived Descriptors Using Partial Least-Squares (PLS) Modeling†

Juan M. Luco*

Laboratorio de Alimentos, Facultad de Química, Bioquímica y Farmacia, Universidad Nacional de San Luis, Chacabuco y Pedernera, 5700 San Luis, Argentina

Received September 24, 1998

In this study, the multivariate partial least-squares projections to latent structures (PLS) technique was used for modeling the brain−blood concentration ratio (BB) of 61 structurally diverse compounds. The PLS model was based on molecular descriptors that can be calculated for any compound simply from a knowledge of its molecular structure, and the model included several topological and constitutional descriptors. The PLS analysis resulted in a significant three-component model with the following statistics: $r = 0.922$, $Q = 0.867$, $s = 0.318$, $n = 58$, and $F = 102$. The predictive ability of the model was assessed by means of crossvalidation and also by using BB partitioning data, BBB permeability data, and 1 set of qualitative brain penetration data, resulting in BB distribution data for 97 compounds. The results indicate that the PLS model developed is statistically sound and is sufficiently robust for predictive use. Taking into account the great ease of computation and interpretation of the derived model, it may be of general utility in predicting BB ratios for a very wide range of new drugs.

1. INTRODUCTION

Optimization of the physicochemical properties and structural attributes that contribute to the more efficient delivery of therapeutic compounds into the brain is one of the central problems in the design of safer and more effective central nervous system (CNS) active drugs.[1] The main factors that determine the concentration of a compound in the brain include[2] (i) time-dependent plasma concentration profile, (ii) binding to plasma and brain constituents, (iii) cerebral blood flow, (iv) permeability at the blood−brain barrier (BBB), (v) facilitated transport systems, and (vi) metabolic/degradative enzymes at the BBB.

The experimental determination of the brain−blood concentration ratio, defined by eq 1, is not a simple task. The

$$\text{BB} = \text{concentration in brain/concentration in blood} \quad (1)$$

technique is laborious, expensive, and time-consuming and requires a sufficient quantity of the pure compound, often in radiolabeled form, to obtain reliable experimental data. For these reasons, the need for appropriate estimation approaches, either by using physicochemical parameters or by means of theoretical calculations, is a question of great importance.

Various attempts to predict BB transport with several physicochemical parameters, in particular with the octanol−water partition coefficient, as log P_{oct}, have been described in the literature.[3,4] However, in some cases, such as in the brain penetration by H2-receptor histamine antagonists,[5] a poor correlation was found between log BB and log P_{oct}. Several workers[6−9] have shown that the partitioning of compounds between the blood and brain compartments can be described by a combination of the partition coefficient (determined with different solvent pairs such as octanol/water or cyclohexane/water) and a term reflecting the molecular size. A similar conclusion has also been reached by Kaliszan and Markuszewski[10] using four representative sets of brain/blood distribution data. More recently, Salminen *et al.*[11] examined a set of 26 structurally dissimilar drugs using retention factors measured by immobilized artificial membrane (IAM) chromatography. They concluded that the molecular size, the ionization of basic groups, and the lipophilicity reflected either by log P_{oct} or log KIAM are the principal factors contributing to the brain uptake of the compounds examined.

Following a different line of analysis, Abraham *et al.*[12,13] related log BB to the solvation parameters such as the dipolarity/polarizability (π_2^H), hydrogen-bond acceptor basicity (β_2^H), hydrogen-bond donor acidity (α_2^H), excess molar refraction (R_2), and characteristic volume of McGowan (V_x). They obtained a significant five-parameter regression model for a set of 65 structurally diverse compounds. The model, which had a correlation coefficient of $r = 0.95$ for the reduced set of 57 compounds (8 compounds of the original set were excluded as outliers), also showed a very promising predictive quality.[14]

Even if the obtained results using physicochemical approaches appear very encouraging, it should be noted that physicochemical parameters are not easily available for any arbitrary chemical structure. Furthermore, since in drug design one has to deal with a large number of compounds, many of which have not even been synthesized or isolated, the models entirely based on experimental parameters will be of limited value.[15] Recently, some quantitative structure−property relationship (QSPR) studies applying theoretical

* E-mail: jmluco@unsl.edu.ar.

† Dedicated to the memory of my dear friend Dr. Mario J. Blasquez.

10.1021/ci980411n CCC: $18.00
Published on Web 02/27/1999

approaches to predict the BB distribution have been reported.[16−18] In all of these studies, however, the reported models suffer from a lack of generality. This is due to the fact that the correlations were based on a small number of compounds on the one hand or were not sufficiently validated on the other hand.

Among different approaches employing computational chemistry, those based on the chemical graph theory have been useful in establishing quantitative structure−activity and structure−property relationships (QSAR, QSPR).[19−23] There are, however, difficulties in the use of these approaches when the QSARs are derived by means of multiple linear regression (MLR) techniques. Commonly, the topological descriptors are mutually interrelated by simple or multiple correlations, and therefore, fortuitous or artifactual MLR models may be obtained. The disadvantage of the MLR method has recently been overcome through the development of a partial least-squares (PLS) method[24] that circumvents the problems of collinearity among the variables and also offers the advantage of handling data sets where the number of independent variables is greater than the number of observations. Furthermore, because the complexity of a PLS model is determined by crossvalidation, the probability of obtaining chance correlations is kept to a minimum.[25]

In the present study, the PLS technique is used for modeling the BB transport of 61 compounds, which were previously analyzed by Abraham *et al.*[13] The QSPR model was based on molecular descriptors that can be calculated for any compound utilizing only the knowledge of its molecular structure, and the model included several topological and constitutional descriptors. The predictive ability of the model was assessed by means of crossvalidation and also by using a large validation set consisting of BB distribution data for 97 compounds.

The main purpose of this work is to attempt to obtain a QSPR model based entirely on nonempirical parameters that could be used to predict BB ratios for a very wide range of new therapeutic agents.

2. MATERIALS AND METHODS

2.1. Biological Data. The brain/blood concentration ratios for the 56 compounds comprising the training set were taken from the studies reported by Young *et al.*[5] (**1**-**30** in Chart 1) and Abraham *et al.*[13] (**31**-**56** in Table 1). In addition, five acidic compounds were included in the training set (**57**-**61** in Chart 1), and their log BB values were taken from Salminen *et al.*[11] and Greig *et al.*[1] The validation set included 70 compounds, and their log BB values along with permeability data were collected from several literature sources.[1,11,14,17,26−29] In addition, data for a further 27 compounds classified according to CNS availability[30−32] were also included in the validation set.

2.2. Structural Descriptors. The following indexes based on molecular topology were considered: the Wiener index,[33] the valence and connectivity molecular indexes,[34] the kappa shape indexes,[35] the differential molecular connectivity indexes,[36] and the charge and geometrical indexes recently introduced by Galvez *et al.*[37,38] The other group of variables included the molecular volume, molecular weight, and several constitutional descriptors. To quantify the total H-bond capacity of the compounds under study, the parameters HBA and HBD as defined by Magee[39] were used. A list of all structural descriptors considered in this study is given in Table 2.

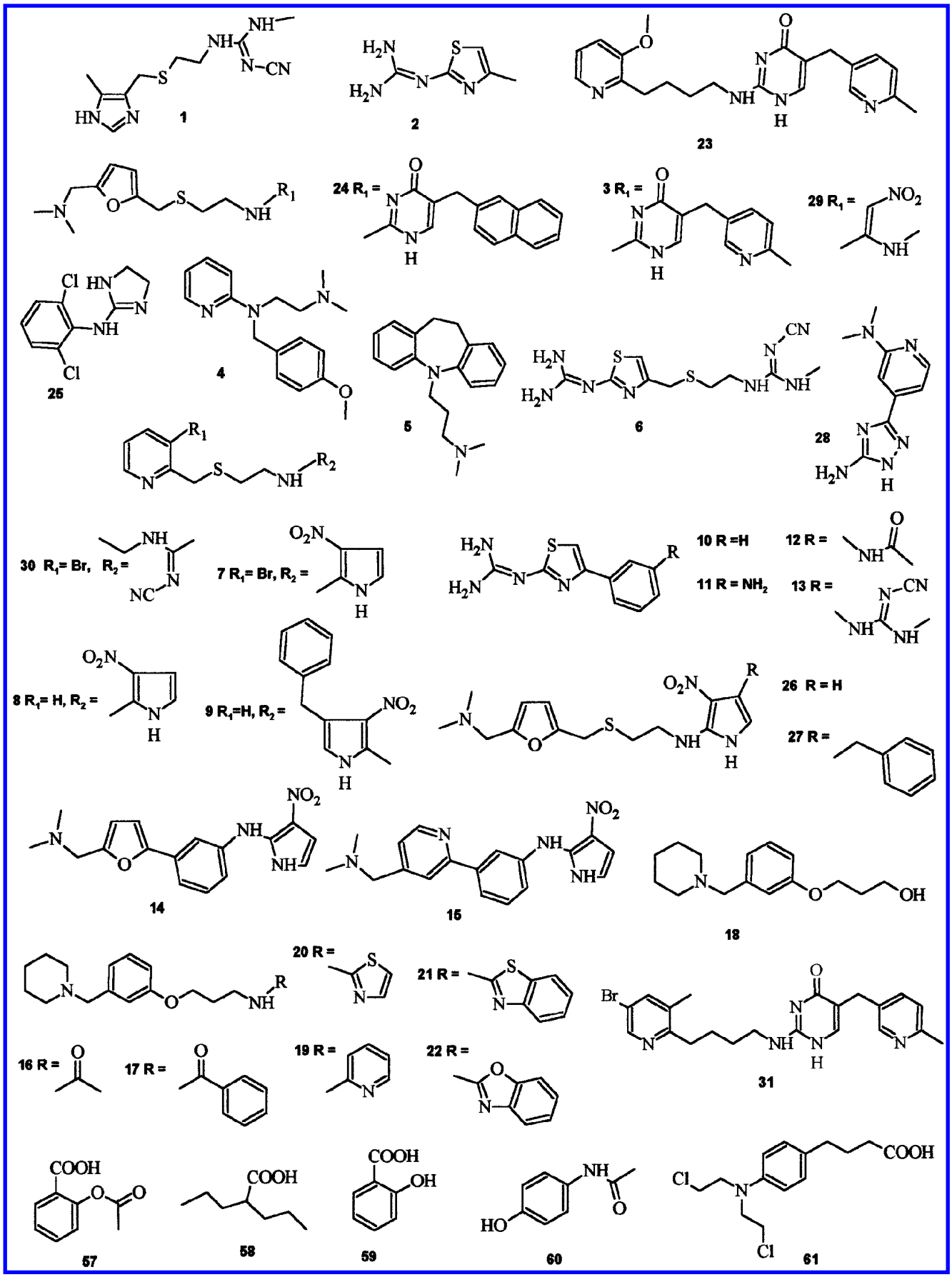
Table 1. Observed and Calculated log BB Values for Compounds Comprising the Training Set

compd	log BB exptl[b]	log BB calcd	log BB residuals
1	−1.42	−1.183	−0.237
2[a]	−0.04	−0.969	0.929
3	−1.06	−1.420	0.360
4	0.49	−0.127	0.617
5	0.83	0.605	0.225
6[a]	−0.82	−1.990	1.170
7	−0.67	−0.515	−0.155
8	−0.66	−0.534	−0.126
9	−0.12	−0.297	0.177
10	−0.18	−0.699	0.519
11	−1.15	−1.364	0.214
12	−1.57	−1.361	−0.209
13	−1.54	−1.811	0.271
14	−0.27	−0.436	0.166
15	−0.28	−0.364	0.084
16	−0.46	−0.082	−0.378
17	−0.24	−0.036	−0.204
18	−0.02	0.138	−0.158
19	0.69	0.008	0.682
20	0.44	0.011	0.429
21	0.14	0.132	0.008
22	0.22	−0.057	0.277
23	−2.00	−1.368	−0.632
24	−1.30	−1.376	0.076
25	0.11	−0.122	0.232
26	−1.12	−0.651	−0.469
27	−0.73	−0.569	−0.161
28	−1.17	−1.123	−0.047
29	−1.23	−0.828	−0.402
30[a]	−2.15	−0.728	−1.422
temelastine (**31**)	−1.88	−1.158	−0.722
butanone (**32**)	−0.08	0.151	−0.231
benzene (**33**)	0.37	0.549	−0.179
3-methylpentane (**34**)	1.01	0.876	0.134
3-methylhexane (**35**)	0.90	0.887	0.013
2-propanol (**36**)	−0.15	−0.240	0.090
2-methylpropanol (**37**)	−0.17	0.023	−0.193
2-methylpentane (**38**)	0.97	0.766	0.204
2,2-dimethylbutane (**39**)	1.04	0.704	0.336
1,1,1-trifluoro-2-chloroethane (**40**)	0.08	0.327	−0.247
1,1,1-trichloroethane (**41**)	0.40	0.342	0.058
diethyl ether (**42**)	0.00	−0.002	0.002
enflurane (**43**)	0.24	0.437	−0.197
ethanol (**44**)	−0.16	−0.233	0.073
fluroxene (**45**)	0.13	0.188	−0.058
halothane (**46**)	0.35	0.557	−0.207
heptane (**47**)	0.81	0.859	−0.049
hexane (**48**)	0.80	0.826	−0.026
isoflurane (**49**)	0.42	0.528	−0.108
methylcyclopentane (**50**)	0.93	0.673	0.257
pentane (**51**)	0.76	0.789	−0.029
propanol (**52**)	−0.16	−0.062	−0.098
propanone (**53**)	−0.15	−0.081	−0.069
teflurane (**54**)	0.27	0.609	−0.339
toluene (**55**)	0.37	0.694	−0.324
trichloroethene (**56**)	0.34	0.582	−0.242
acetylsalicylic acid (**57**)	−0.50	−1.181	0.681
valproic acid (**58**)	−0.22	−0.815	0.595
salicylic acid (**59**)	−1.10	−1.209	0.109
p-acetamidophenol (**60**)	−0.31	−0.458	0.148
chlorambucil (**61**)	−1.70	−1.156	−0.544

[a] These compounds were not used to derive the PLS model. [b] Experimental values taken from refs 1, 5, 11, and 13.

2.3. Statistical Methods. The PLS method was employed to search for relationships between the BB distribution data

Chart 1



and the structural descriptors. This principal component-like method is based on the projection of the original multivariate data matrices down onto smaller matrices (**T**,**U**) with orthogonal columns, which relates the information in the

Table 2. Symbols and Definitions for the Molecular Descriptors Applied in This Study

${}^{m}\chi^{n}_{\mathrm{p}}$	simple and valence ($n = v$) path connectivity index of order $m = 0-10$
${}^{m}\chi^{n}_{\mathrm{C}}$	simple and valence ($n = v$) cluster connectivity index of order $m = 3, 4$
${}^{m}\chi^{n}_{\mathrm{PC}}$	simple and valence ($n = v$) path−cluster connectivity index of order $m = 4$
${}^{m}\chi^{n}_{\mathrm{ch}}$	simple and valence ($n = v$) chain connectivity index of order $m = 5-7$
${}^{4}C^{n}_{\mathrm{pc}}$	simple and valence ($n = v$) connectivity index defined as ${}^{4}\chi^{n}\mathrm{pc}/({}^{3}\chi^{n}_{\mathrm{C}} + 1)$
${}^{m}\kappa$	simple kappa shape index of order $m = 1-3$
${}^{m}\kappa_{\alpha}$	α- kappa shape index of order $m = 0-4$
${}^{m}\Delta\chi$	differential connectivity index of order $m = 0-3$ defined as ${}^{m}\chi - {}^{m}\chi^{v}$
${}^{n}G_{k}{}^{a}$	simple and valence ($n = v$) charge index of order $k = 1-5$
${}^{n}J_{k}{}^{a}$	simple and valence ($n = v$) mean charge index of order $k = 1-5$ defined as $J_k = G_k/(N - 1)$
W^{a}	Wiener index calculated as the half-sum of the off-diagonal elements of the distance matrix of a graph plus the number of vertices pairs separated by three edges
L^{a}	topological molecular length defined as the counted distance in the number of edges between the molecule's two most separate atoms by the shortest means
S^{a}	molecular surface parameter calculated as the sum of the predetermined S values for several molecular fragments
E^{a}	shape index defined as $E = S/L^2$
$\mathrm{Pr}_1{}^{a}$	number of pairs of ramifications separated by one edge
$\mathrm{Pr}_2{}^{a}$	number of pairs of ramifications separated by two edges
$\mathrm{Pr}_3{}^{a}$	number of pairs of ramifications separated by three edges
V_x	McGowan characteristic molecular volume
M_{w}	molecular weight
HBA[b]	count of electron pairs on O and N
HBD	count of O−H and N−H bonds
∑elem	count of elements in the molecule
∑N	sum of nitrogen atoms in the molecule
I_{COOH}	−1 or 0 for the presence or absence of an acidic group
I_{s}	1 or 0 for the presence or absence of sulfur atoms in the molecule

[a] Charge and geometrical indexes were calculated according to Gálvez *et al.*[37,38] [b] According to Magee,[39] only one available acceptor pair is considered for each oxygen in the nitro group.

response matrix **Y** to the systematic variance in the descriptor matrix **X**, as shown below:

$$\mathbf{X} = \bar{\mathbf{X}} + \mathbf{TP}' + \mathbf{E}$$

$$\mathbf{Y} = \bar{\mathbf{Y}} + \mathbf{UC}' + \mathbf{F}$$

$$\mathbf{U} = \mathbf{T} + \mathbf{H} \text{ (the inner relation)}$$

where $\bar{\mathbf{X}}$ and $\bar{\mathbf{Y}}$ are the corresponding mean value matrices, **T** and **U** are the matrices of scores that summarize the *x* and *y* variables, respectively, **P** is the matrix of loadings showing the influence of the *x* variable in each component, **C** is the matrix of weights expressing the correlation between **Y** and **T**(**X**), and **E**, **F**, and **H** are the corresponding residuals matrices. The PLS calculations also give an auxiliary matrix **W** (PLS weights), which expresses the correlation between **U** and **X** and is used to calculate **T**. Determinations of the significant number of model dimensions (*A*) was made by crossvalidation.[40]

The computer software (Molconn-X) used to calculate the molecular connectivity-type topological indexes was obtained from L. H. Hall (Eastern Nazarene College, Quincy, MA). Calculations of the charge and geometrical indexes were performed with the INDIS program kindly offered by Dr. Jorge Gálvez (University of Valencia, Spain). PLS analysis was carried out using the SIMCA-S 5.1a software package obtained from Umetri AB (Umea, Sweden).

3. RESULTS AND DISCUSSION

3.1. PLS Analysis. All variables used in the PLS calculations were initially autoscaled to zero mean and unit variance to give each descriptor equal importance in the PLS analysis. The statistical significance of the screened models was judged by the correlation coefficient (*r*), the standard deviation (*s*), and the *F* statistic. The predictive ability was evaluated by the crossvalidation coefficient (*Q*), which is based on the prediction error sum of squares (PRESS). Although the PLS method offers the advantage of handling data sets where the number of independent variables is greater than the number of observations, it can be seen that considerably worse predictions are obtained if many irrelevant descriptors are included in the PLS model. Because of the large number of structural descriptors considered in this study, the VIP (variable importance for the projection) parameter[24] was used to unravel which descriptor variables were the most relevant to explain log BB. Preliminary analysis of the obtained models showed that different combinations of descriptor variables yielded PLS models of similar statistical relevance. Thus, the selection of the definitive model was carried out on the basis of predictions for the compounds comprising the validation sets.

In the statistical process, 3 strong outliers (**2**, **6**, and **30**) were detected and removed from the original 61 compound set. None of the descriptors used in this study can account for the reported log BB values of these compounds. The reason for this is not evident, especially taking into account that other structurally analogous compounds in the training set (**10**−**13**) were well-predicted by the selected model. Thus, this may be due to inaccuracies in the measurement of the BB ratio on the one hand or may result from the intervention of metabolic factors or active transport systems on the other.[13]

The PLS analysis for the remaining 58 compounds resulted in a significant 3-component model ($\mathrm{PLS}_{\mathrm{selec}}$) with the following statistics: $r = 0.922$, $Q = 0.867$, $s = 0.318$, and $F = 102$. The model accounted for 85% (49.8%, 26.5%, and 8.7%, respectively) of the variance in log BB. The *Q* value was calculated using three crossvalidation groups and represents the highest obtained value. The residual values showed a reasonable central tendency and the agreement

Table 3. PLS Model Parameters[c]

descriptor	WC[1][a]	WC[2][a]	WC[3][a]	coeff MR[b]
$^3\chi^v$	−0.1928	**0.3652**	0.0479	**0.1303**
$^4C_{pc}$	−0.2301	0.2039	**0.2875**	**0.1079**
$^0\Delta\chi$	**−0.2813**	−0.0321	0.1929	−0.0505
G_3	**−0.2790**	0.1438	0.0660	−0.0001
G_3^v	**−0.2916**	0.0604	−0.2306	**−0.1293**
J_1	0.0045	−0.2353	**0.3897**	−0.0023
J_1^v	−0.1834	−0.2447	0.2041	**−0.1199**
W	−0.2251	0.2324	−0.2207	−0.0213
$(W)^2$	−0.1365	0.0829	**−0.5772**	**−0.1005**
S	−0.1978	**0.3598**	0.1021	**0.1415**
E	0.0508	0.0242	**−0.3391**	−0.0686
Pr_2	−0.2124	**0.2865**	**0.2944**	**0.1558**
Pr_3	−0.2254	0.0090	−0.0007	−0.0678
HBA	**−0.3645**	−0.2199	−0.1222	**−0.2586**
HBD	**−0.3359**	**−0.4228**	−0.1016	**−0.3426**
∑elem	**−0.2627**	−0.0798	0.0407	**−0.1112**
∑N	**−0.3162**	−0.0926	−0.0720	**−0.1667**
I_{COOH}	0.1122	**0.4115**	−0.0262	**0.2292**

[a] PLS weight vectors of the first, second, and third dimension, respectively. [b] Pseudoregression coefficients of autoscaled and mean-centered structural descriptors (log BB is unscaled and uncentered). The constant term is **−0.08**. [c] Descriptors having weights > 0.25 and coefficients > 0.10 have been bold faced.

between the observed and calculated values is very satisfactory as shown in Table 1.

The PLS weights and the corresponding pseudoregression coefficients are shown in Table 3. From these values, it can be seen how much a single variable contributes in each PLS component to the modeling of the BB ratio. The first PLS component is mainly related to the molecular polarity of the compounds, since HBA, HBD, ∑N, G_3^v, G_3, $^0\Delta\chi$, and ∑elem have the largest contribution. The second PLS component is mainly related to the molecular size and the H-donating ability of the compounds, since $^3\chi^v$, S, HBD, and I_{COOH} are among the most influential descriptors. Finally, the third PLS component, which describes a minor part of the variance in log BB, is strongly dominated by the quadratic term in W and also by E, Pr_2, and $^4C_{pc}$, indicating the low but not negligible contribution of the molecular shape.

3.2. Model Validation. It is well-known that the real predictive ability of any QSAR model cannot be judged solely by using internal validation (i.e., crossvalidation); it has to be validated on the basis of predictions for compounds not included in the training set.[24] In order to demonstrate the applicability of the derived PLS_{selec} model, several sets of BB partitioning and BBB permeability data along with one set of qualitative brain penetration data were evaluated. An important point to highlight is that the model has been developed from a given descriptor space (training set) and, therefore, its application is not universal. To evaluate the reliability of the predictions, that is, to identify the compounds to which the model is not applicable, the DModX (observation distances to the PLS model in the descriptor space) parameter[24] was computed. Large DModX values correspond to large extrapolations and indicate probable outliers in the descriptor space.

3.2.1. Test Compounds for Brain−Blood Partitioning. The first validation set included the BB ratios of eight H1-receptor histamine antagonist/agonists[14,26] (**62**−**69**) and six miscellaneous CNS agents[17] (**70**−**75**). The chemical structures along with the observed and predicted log BB data are shown in Table 4. As may be seen from the table, the predictions for compounds **64**−**75** were very good, whereas the log BB values for compounds **62** and **63** were strongly overestimated by the PLS_{selec} model. It should be noted, however, that these compounds were also overestimated in their BB ratios by all the linear free-energy equations reported by Abraham *et al.*[14] and can therefore be considered as outliers. For the remaining 12 nonoutlier compounds, the RMS prediction error is 0.235 and is at the level of the experimental error in the BB determinations.[13] Further, these prediction results are found to be clearly superior to the ones obtained by the models recently reported by Lombardo *et al.*[17] and Norinder *et al.*[18] (RMSE = 1.244 and 0.473, respectively). An interesting point to highlight is the lower precision of the PLS_{selec} model for predicting the BB ratios of compounds **70**−**74**, as compared to the other compounds in this data set. This fact may be attributed to the greater DModX values that these compound data have (see Table 4).

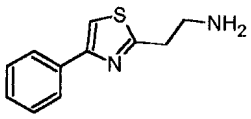
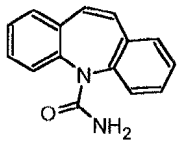
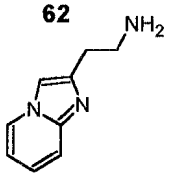
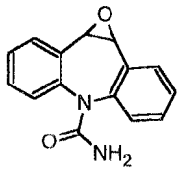
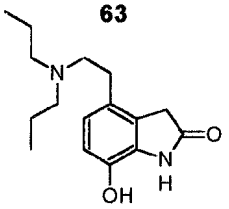
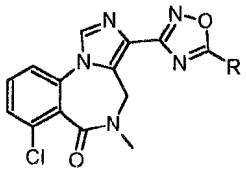
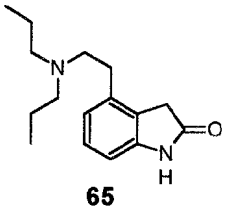
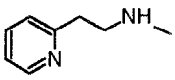
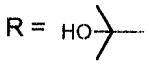
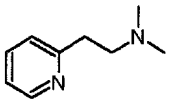
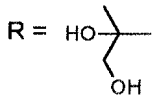
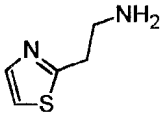
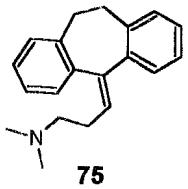
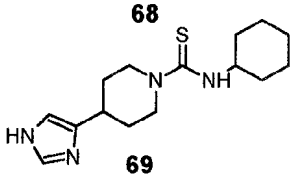
The second validation set included 17 structurally diverse compounds (**76**−**92**), and their log BB values were taken from the Salminen *et al.* compilation.[11] In addition, eight miscellaneous agents (**93**−**100**) were also included, and their log BB values were taken from several literature sources.[1,28,29] The observed and predicted log BB data are shown in Table 5. Inspection of these results shows that the PLS_{selec} model performs reasonably well, especially taking into account the varied experimental conditions under which the BB ratios have been obtained. Only three compounds (**84**, **91**, and **96**) were predicted above three standard deviations and can therefore be considered as outliers. The RMSE value calculated on the 25 validation compounds is 0.541, while the RMSE value for the reduced data set (excluding the 3 outliers) is 0.408.

3.2.2. Test Compounds for Barrier Permeability. Another way of evaluating the PLS_{selec} model consisted in the PLS modeling of 2 representative sets of brain−blood barrier permeability data but using only the 18 selected descriptors shown in Table 3.

The first data set included the rat brain capillary permeabilities for 27 miscellaneous drugs studied by Levin.[27] Six drugs were excluded from the PLS analysis: $^{24}NaCl$, ^3H_2O, bleomycin, adriamycin, vincristine, and epipodophylotoxin. These chemicals (except $^{24}NaCl$) were also omitted from the correlation equation derived by Levin. The PLS analysis resulted in a significant three-component model that described 89.3% (62%, 15.9%, and 11.3%, respectively) of the variance in the permeability data. The other statistical parameters were $r = 0.945$, $Q = 0.773$, $s = 0.356$, and $F = 47.29$. The measured and predicted logarithms of the permeability coefficients (log Pc) are plotted in Figure 1a. As can be seen in this figure, the agreement between the measured and calculated data is very good.

The second data set included the permeabilities of 10 miscellaneous drugs reported by Lombardo and co-workers.[17] The permeability data were measured by using cultured monolayers of endothelial cells from bovine brain microvessels. The PLS analysis resulted in a significant two-component model explaining 95.2% (89% and 6.2%, respectively) of the variance in the permeability data. The other statistical parameters were $r = 0.976$, $Q = 0.90$, $s = 9.30$, and $F = 69.41$. The measured and predicted permeability

Table 4. Experimental[a] and Predicted log BB Values for Compounds Comprising the Test Set 1

Compound	Expl.	Pred.	Compound	Expl.	Pred.
 **62**	-1.30	-0.036 (0.395)[b]	 **70**	0.00	-0.005 (1.141)
 **63**	-1.40	-0.292 (0.599)	 **71**	-0.34	0.005 (2.035)
 **64**	-0.43	-0.263 (0.744)			
 **65**	0.25	0.170 (0.421)	R =  **72**	-0.30	-0.447 (2.434)
 **66**	-0.30	-0.139 (0.497)	R =  **73**	-1.34	-0.931 (2.144)
 **67**	-0.06	0.016 (0.570)	R =  **74**	-1.82	-1.308 (2.091)
 **68**	-0.42	-0.510 (0.542)	 **75**	0.89	0.966 (0.769)
 **69**	-0.16	-0.285 (0.670)			

[a] Logarithm of BB ratios, from refs 14, 17, and 26 (see full text). [b] The figures in parantheses are the DModX values, being the critical distance $D_{crit[3]} = 1.043$ for a significance level of 0.001.

coefficients are plotted in Figure 1b. Again, the agreement between the measured and calculated data is very good.

In summary, taking into account that the PLS_{selec} model was based on measures of the brain−blood distribution at or near the steady state, whereas the brain penetration data used in this validation process refer to a rate process across the blood−brain barrier, the above PLS models clearly show the robustness and usefulness of the selected descriptors for modeling the passage of drugs through the brain−blood barrier.

3.2.3. Qualitative Brain Penetration Data. The predictability of the derived PLS_{selec} model may also be examined

Table 5. Experimental and Predicted log BB Values for Compounds Comprising Test Set 2

compd	log BB exptl[a]	log BB pred	DModX[b]
theophylline (**76**)	−0.290	−0.512	1.228
caffeine (**77**)	−0.055	−0.219	1.452
antipyrine (**78**)	−0.097	0.474	0.725
ibuprofen (**79**)	−0.180	−0.555	1.020
codeine (**80**)	0.550	0.271	2.447
pentobarbital (**81**)	0.120	−0.191	0.893
alprazolam (**82**)	0.044	0.332	1.503
indomethacin (**83**)	−1.260	−1.032	2.158
oxazepam (**84**)	**0.610**	**−0.476**	1.558
hydroxyzine (**85**)	0.390	0.128	0.535
desipramine (**86**)	1.200	0.426	0.769
midazolam (**87**)	0.360	0.400	1.810
verapamil (**88**)	−0.700	−1.111	1.387
promazine (**89**)	1.230	0.832	0.723
chlorpromazine (**90**)	1.060	0.710	1.013
trifluoroperazine (**91**)	**1.440**	**0.459**	0.925
thioridazine (**92**)	0.240	1.062	1.294
BCNU (**93**)	−0.520	−0.570	0.648
phenserine (**94**)	1.000	0.230	0.840
physostigmine (**95**)	0.079	0.007	0.886
terbutylchlorambucil (**96**)	**1.000**	**−0.227**	0.653
didanosine (**97**)	−1.301	−0.816	0.763
zidovudine (**98**)	−0.720	−1.024	0.969
nevirapine (**99**)	0.000	−0.285	1.903
SB-222200 (**100**)	0.300	0.426	1.597

[a] Experimental values taken from refs 1, 11, 28, and 29. [b] $D_{crit[3]} = 1.043$ for a significance level of 0.001.

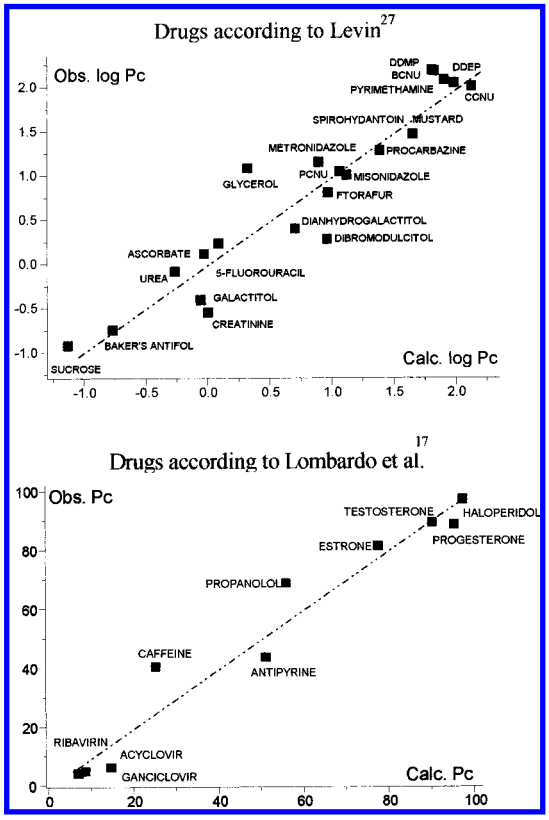


Figure 1. Relationships between the experimental and calculated permeability coefficients: (a) logarithm of permeability coefficient $\times 10^{-6}$ cm/s; (b) permeability coefficient $\times 10^{4}$ cm/s.

in yet another way, that is, to obtain predictions of the BB ratios for compounds to which only discrete and unambiguous data of brain penetration have been established. Recently, Seelig *et al.*[30] studied 28 compounds falling into 2 categories according to CNS availability (high or low CNS penetration). Of these, 22 were selected, omitting 6 which were previously analyzed (compounds **5**, **25**, **86**, **89**, **90**, and **92**). To give a larger selection of compounds, an additional five compounds were selected from the literature.[31,32] Compounds with predicted log BB values of < -1 were classified as poor brain penetrators (CNS^-), whereas compounds with values of > -1 were classified as easy brain penetrators (CNS^+). All predicted values are listed in Table 6. As may be seen from the table, the predictive ability of the PLS_{selec}, model is high enough, having 100% accuracy for classifying CNS^+ compounds and 84.6% accuracy for classifying CNS^- compounds. It should be noted that the CNS^- compounds, domperidone, loperamide, and terfenadine, were considered as correctly classified because these compounds are efficiently transported back to the blood by P-glycoprotein in the blood−brain barrier.[41,42] Thus, only two CNS^- compounds were incorrectly classified, mequitazine (which can cross the BBB when it is applied at higher concentrations) and pirenzepine.

Table 6. Predicted log BB Values for Compounds Comprising the Qualitative Brain Penetration Data Set[f]

compd	CNS activity obsd[a]	CNS activity pred	pred[b] log BB	DModX[c]
(*R*)-apomorphine	+	+	0.447	2.272
doxilamine	+	+	0.341	0.545
cis-flupentixol	+	+	0.061	0.944
haloperidol	+	+	−0.265	0.744
naltrexone	+	+	−0.177	1.904
perphenazine	+	+	0.097	0.975
promethazine	+	+	0.824	1.241
thiopental	+	+	−0.064	0.862
roxinodole	+	+	−0.016	0.565
tamitinol	+	+	−0.353	0.526
N-isopropylpyridine piperazine (U-88703)	+	+	−0.428	0.584
cotinine	+	+	0.342	0.594
nicotine	+	+	0.401	0.701
nornicotine	+	+	0.065	0.600
astemizole	−	−	−1.255	1.578
domperidone	−[d]	+	−0.588	0.708
loperamide	−[d]	+	−0.526	0.918
terfenadine	−[d]	+	−0.982	1.772
atenolol	−	−	−1.359	0.627
mequitazine	−[e]	+	**1.212**	**1.800**
salbutamol	−	−	−1.090	0.542
furosemide	−	−	−2.516	1.101
pirenzepine	−	+	**−0.489**	**1.753**
ebastine	−	−	−1.059	2.326
carebastine	−	−	−3.223	4.653
carmoxirol	−	−	−1.298	1.181
delavirdine	−	−	−2.064	0.859

[a] Reported activity levels taken from refs 30, 31, and 32. [b] Predicted values from the PLS_{selec} model. [c] $D_{crit[3]} = 1.043$ for a significance level of 0.001. [d] P-glycoprotein substrate drugs. [e] CNS activity is dose-limiting. [f] Compounds with easy brain penetration are denoted as CNS^+, and compounds with poor brain penetration are denoted as CNS^-.

It is important to note that the brain penetration data used in this validation process are composite in nature, being related to the affinity of drugs for some CNS receptors on the one hand and to either a rate or an equilibrium transport across the blood−brain barrier on the other hand. However, because these observed CNS activity levels have been modeled from a qualitative point of view, the obtained results are not confusing with respect to their interpretation. Thus,

in the case of the CNS^+ compounds, the results are clear, since, as inferred from their reported CNS activities, these compounds must inevitably cross the blood−brain barrier, either to inducte CNS side effects or occupy the CNS receptors. On the other hand, with respect to the CNS^- compounds, it becomes difficult to reach a definitive conclusion in relation to their permeability properties. However, taking into account that no consistent quantitative data of brain penetration have been established for these compounds, the agreement between the observed and predicted CNS activity levels is very good. In conclusion, this qualitative validation process reinforces the predictive ability of the derived PLS_{SELEC} model.

4. CONCLUSION

This study has shown that the brain penetration of drugs can be properly modeled in terms of structurally-based parameters. The model stability and predictive capacity have been checked in three independent validation procedures that gave quite good results in spite of the diversity of structures considered. Thus, the statistical reliability of the derived PLS_{selec} model suggests that it may be of general utility in predicting BB ratios for a very wide range of new drugs.

ACKNOWLEDGMENT

I thank Dr. Jorge Gálvez for providing the software INDIS used for the topological calculations. I cordially thank Prof. Dr. R. Olsina for careful reading of the manuscript and for his friendly support. I also thank Dr. H. F. Ferretti for his comments and Prof. John A. Luco for linguistic advice. This work has been supported by the University National of San Luis, Argentina.

REFERENCES AND NOTES

(1) Greig, N. H.; Brossi, A.; Pei, X. F.; Ingram, D. K.; Soncrant, T. T. In *New Concepts of a Blood−Brain Barrier*; Greenwood J., *et al.*, Eds.; Plenum: New York, 1995; pp 251−264.
(2) Pardridge, W. M. Brain Drug Delivery and Blood−Brain Barrier Transport. *Drug Del.* **1996**, *3*, 99−115.
(3) Gupta, S. P. QSAR Studies on Drugs Acting at the Central Nervous System. *Chem. Rev.* **1989**, *89*, 1765−1800.
(4) Hansch, C.; Björkroth, J. P.; Leo, A. Hydrophobicity and Central Nervous System Agents: On the Principle of Minimal Hydrophobicity in Drug Design. *J. Pharm. Sci.* **1987**, *76*, 663−687.
(5) Young, R. C.; Mitchell, R. C.; Brown, T. H.; Ganellin, C. R.; Griffiths, R.; Jones, M.; Rana, K. K.; Saunders, D.; Smith, I. R.; Sore, N. E.; Wilks, T. J. Development of a New Physicochemical Model for Brain Penetration and Its Application to the Design of Centrally Acting H_2 Receptor Histamine Antagonist. *J. Med. Chem.* **1988**, *31*, 656−671.
(6) Hansch, C., Leo, A.; Hoekman, D. In *Exploring QSAR-Fundamentals and Applications in Chemistry and Biology*; Heller, S. R., Ed.; American Chemical Society: Washington, DC, 1995; pp 387−390.
(7) van de Waterbeemd, H.; Kansy, M. Hydrogen-Bonding Capacity and Brain Penetration. *Chimia* **1992**, *46*, 299−303.
(8) van Bree, J. B. M. M.; De Boer, A. G.; Danhof, M.; Ginsel, L. A.; Breimer, D. D. Characterization of an "in Vitro" Blood−Brain Barrier: Effects of Molecular Size and Lipophilicity on Cerebrovascular Endothelial Transport Rates of Drugs. *J. Pharmacol. Exp. Ther.* **1988**, *247*, 1233−1239.
(9) Kai, J.; Nakamura, K.; Masuda, T.; Ueda, I.; Fujiwara, H. Thermodynamic Aspects of Hydrophobicity and the Blood−Brain Barrier Permeability Studied with a Gel Filtration Chromatography. *J. Med. Chem.* **1996**, *39*, 2621−2624.
(10) Kaliszan, R.; Markuszewski, M. Brain/Blood Distribution Described by a Combination of Partition Coefficient and Molecular Mass. *Int. J. Pharm.* **1996**, *45*, 9−16.
(11) Salminen, T.; Pulli, A.; Taskinen, J. Relationship Between Immobilised Artificial Membrane Chromatographic Retention and the Brain Penetration of Structurally Diverse Drugs. *J. Pharmaceut. Biomed. Anal.* **1997**, *15*, 469−477.
(12) Kamlet, M. J.; Doherty, R. M.; Bergerova, F. V.; Carr, P. W.; Abraham, M. H.; Taft, R. W. Solubility Properties in Biological Media 9: Prediction of Solubility and Partition of Organic Nonelectrolytes in Blood and Tissues from Solvatochromic Parameters. *J. Pharm. Sci.* **1987**, *76*, 14−17.
(13) Abraham, M. H.; Chada, H. S.; Mitchell, R. C. Hydrogen Bonding. 33. Factors that Influence the Distribution of Solutes Between Blood and Brain. *J. Pharm. Sci.* **1994**, *83*, 1257−1268.
(14) Abraham, M. H.; Chada, H. S.; Mitchell, R. C. Hydrogen-Bonding. Part 36. Determination of Blood Brain Distribution Using Octanol−Water Partition Coefficients. *Drug Des. Discov.* **1995**, *13*, 123−131.
(15) Basak, S. C. Use of Molecular Complexity Indices in Predictive Pharmacology and Toxicology: A QSAR Approach. *Med. Sci. Res.* **1987**, *15*, 605−609.
(16) Basak, S. C.; Gute, B. D.; Drewes, L. R. Predicting Blood−Brain Transport of Drugs: A Computational Approach. *Pharm. Res.* **1996**, *13*, 775−778.
(17) Lombardo, F.; Blake, J. F.; Curatolo, W. J. Computation of Brain−Blood Partitioning of Organic Solutes via Free Energy Calculations. *J. Med. Chem.* **1996**, *39*, 4750−4755.
(18) Norinder, U.; Sjöberg, P.; Österberg, T. Theoretical Calculation and Prediction of Brain−Blood Partitioning of Organic Solutes Using MolSurf Parametrization and PLS Statistics. *J. Pharm. Sci.* **1998**, *87*, 952−959.
(19) Kier, L. B.; Hall, L. H. In *Molecular Connectivity in Chemistry and Drug Research*; Bawden, D., Ed.; Research Studies: Letchworth, Hertfordshire, England, 1986; Chapter 8, pp 225−246.
(20) Luco, J. M.; Sosa, M. E.; Cesco, J. C.; Tonn, C. E.; Giordano, O. S. Molecular Connectivity and Hydrophobicity in the Study of Antifeedant Activity of Clerodane Diterpenoids. *Pestic. Sci.* **1994**, *41*, 1−6.
(21) Luco, J. M.; Yamin, L. J.; Ferretti, H. F. Molecular Topology and Quantum Chemical Descriptors in the Study of Reversed-Phase Liquid Chromatography. Hydrogen-Bonding Behavior of Chalcones and Flavanones. *J. Pharm. Sci.* **1995**, *84*, 903−908.
(22) Luco, J. M.; Ferretti, H. F. QSAR Based on Multiple Linear Regression and PLS Methods for the Anti-HIV Activity of a Large Group of HEPT Derivatives. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 392−401.
(23) Santagata, L. N.; Luco, J. M.; Enriz, R. D. Aproximación Analítica de la Superficie Molecular de Van Der Waals Aplicada en un Estudio de QSAR. *An. Asoc. Quim. Argentina* **1996**, *84*, 573−584.
(24) Wold, S. In *Chemometric Methods in Molecular Design*; van de Waterbeemd, H., Ed.; VCH: Weinheim, 1995; Vol. 2, Chapter 4.4, pp 195−218.
(25) Clark, M.; Cramer, R. D., III. The Probability of Chance Correlation Using Partial Least Squares (PLS). *Quant. Struct.-Act. Relat.* **1993**, *12*, 137−145.
(26) Calder, J. A. D.; Ganellin, R. Predicting the Brain-Penetrating Capability of Histaminergic Compounds. *Drug Des. Discov.* **1994**, *11*, 259−268.
(27) Levin, V. A. Relationship of Octanol/Water Partition Coefficient and Molecular Weight to Rat Brain Capillary Permeability. *J. Med. Chem.* **1980**, *23*, 682−684.
(28) von Sprecher, A.; Gerspacher, M.; Anderson, G. P. Neurokinin Antagonists as Potential Therapies for Inflammation and Rheumatoid Arthritis. *IDrugs* **1998**, *1*, 73−91.
(29) Yazdanian, M.; Glynn, S. L. In Vitro Blood−Brain Barrier Permeability of Nevirapine Compared to Other HIV Antiretroviral Agents. *J. Pharm. Sci.* **1998**, *87*, 306−310.
(30) Seelig, A.; Gottschlich, R.; Devant, R. M. A Method to Determine the Ability of Drugs to Diffuse Through the Blood−Brain Barrier. *Proc. Natl. Acad. Sci.* **1994**, *91*, 68−72.
(31) Chang, M.; Sood, V. K.; Wilson, G. J.; Kloosterman, D. A.; Sanders, P. E.; Hauer, M. J.; Zhang, W.; Branstetter, D. G. Metabolism of the HIV-1 Reverse Trancriptase Inhibitor Delavirdine in Mice. *Drug Metab. Dispos.* **1997**, *25*, 828−839.
(32) Crooks, P. A.; Li, M.; Dwoskin, L. P. Metabolites of Nicotine in Rat Brain After Peripheral Nicotine Administration. *Drug Metab. Dispos.* **1997**, *25*, 47−54.
(33) Kaliszan, R. In *Quantitative Structure−Chromatographic Retention Relationships*; Winefordner, J. D., Ed.; John Wiley & Sons: New York, 1987; Chapter 8, p 138.
(34) Kier, L. B.; Hall, L. H. In *Molecular Connectivity in Chemistry and Drug Research*; de Stevens G., Ed.; Academic: New York, 1976.
(35) Jurs, P. C.; Dixon, S. L.; Egolf L. M. In *Chemometric Methods in Molecular Design*; van de Waterbeemd, H., Ed.; VCH: Weinheim, 1995; Vol. 2, Chapter 2.1, pp 15−38.
(36) Kier, L. B.; Hall, L. H. A Differential Molecular Connectivity Index. *Quant. Struct.-Act. Relat.* **1991**, *10*, 134−140.
(37) Gálvez, J.; García-Domenech, R.; Salabert, M. T.; Soler, R. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 520−525.

(38) Gálvez, J.; García-Domenech, R.; de Julián-Ortiz, V.; Soler, R. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 272−284.
(39) Magee, P. In *Classical and Three-Dimensional QSAR in Agrochemistry*; Hansch, C., Fujita, T., Eds.; American Chemical Society Symposium Series 606; American Chemical Society: Washington, DC, 1995; Chapter 9, pp 120−129.
(40) Wold, S. Cross-Validatory Estimation of the Number of Components in Factor and Principal Components Models. *Technometrics* **1978**, *20*, 397−405.
(41) Schinkel, A. H.; Wagenaar, E.; Mol, C. A. A. M.; van Deemter, L. P-Glycoprotein in the Blood−Brain Barrier of Mice Influences the Brain Penetration and Pharmacological Activity of Many Drugs. *J. Clin. Invest.* **1996**, *97*, 2517−2524.
(42) van Asperen, J.; Mayer, U.; van Tellingen, O.; Beijnen, J. H. The Functional Role of P-Glycoprotein in the Blood−Brain Barrier. *J. Pharm. Sci.* **1997**, 881−884.

CI980411N