

# Prediction of blood–brain partitioning: A model based on ab initio calculated quantum chemical descriptors

Sofie Van Damme<sup>a</sup>, Wilfried Langenaeker<sup>b</sup>, Patrick Bultinck<sup>a,\*</sup>

<sup>a</sup> *Department of Inorganic and Physical Chemistry, Ghent University, Krijgslaan 281 S-3, B-9000 Gent, Belgium*

<sup>b</sup> *Silicos NV, Wetenschapspark 7, B-3590 Diepenbeek, Belgium*

Received 25 July 2007; received in revised form 9 November 2007; accepted 13 November 2007

Available online 21 November 2007

## Abstract

A new model for the prediction of log BB, a penetration measure through the blood–brain barrier, based on a molecular set of 82 diverse molecules is developed. The majority of the descriptors are derived from quantum chemical ab initio calculations, augmented with a number of classical descriptors. The quantum chemical information enables one to compute fundamental properties of the molecules. The best set of descriptors was selected by sequential selection and multiple linear regression was used to develop the QSAR model. The predictive capability of the model was tested using internal and external test procedures and the domain of applicability was determined to identify reliable predictions. The selected set of descriptors shows a significant correlation with the experimental log BB. The proposed model could reproduce the data with an error approaching the experimental uncertainty and satisfies the available validation procedures. The obtained results indicate that the use of quantum chemical information in describing molecules improves the behavior of the model.

© 2007 Elsevier Inc. All rights reserved.

**Keywords:** Blood–brain partition coefficient; Structure–activity relationship; Ab initio quantum chemical calculations; Drug transport; Statistical validation

## 1. Introduction

The brain is a delicate organ, and evolution built very efficient ways to protect it. Unfortunately, the same mechanism that protects against intrusive chemicals can also inhibit therapeutic agents in their work. Many existing pharmaceuticals are rendered ineffective in the treatment of cerebral diseases due to our inability to effectively deliver and sustain them within the brain. The blood–brain barrier (BBB) is a complex cellular system whose purpose is to maintain homeostasis of the central nervous system (CNS) by separating the brain from the systemic blood circulation [1]. The ability of a drug to penetrate the blood–brain barrier is of fundamental importance in drug design. High penetration is needed for CNS-active drugs, while negligible penetration may be desirable in order to minimize CNS-related side effects for drugs with peripheral activity [2–4]. The prediction of this property is important, as the experimental determination (in vivo and in

vitro) [5] of BBB penetration is difficult and costly. Hence, there is a great demand for rapid and efficient methods to assess the penetration of drugs across the BBB, and it would be of great scientific and economic value if a computational method could be used to model blood–brain partitioning at an early stage of the drug discovery program.

A common measure of the degree of BBB penetration is the ratio of the steady-state distribution of molecules between the brain and the plasma usually expressed as  $\log(C_{\text{brain}}/C_{\text{blood}})$  or log BB.

The main purpose of the present research is the development of a new in silico model in which the brain/plasma ratio log BB is predicted of a set of structurally diverse molecules. Furthermore, the final model shall be investigated on the added value of the use of quantum chemical information in the prediction of the compounds.

Early models for the prediction of log BB were based on quantitative structure activity relationship (QSAR) equations using a rather limited number of variables that were directly related to experimentally accessible molecular descriptors. A typical example is the model generated by Young et al. [6] for histamine H<sub>2</sub> receptor antagonists. The best correlation was

\* Corresponding author. Tel.: +32 9 2464423; fax: +32 9 2644983.

E-mail address: [patrick.bultinck@ugent.be](mailto:patrick.bultinck@ugent.be) (P. Bultinck).

found with the experimental quantity  $\Delta \log P$ , the difference between  $\log P$  in octanol/water and  $\log P$  in cyclohexane/water systems. Abraham [7] introduced parameters that express the polarizability and the hydrogen bond donor/acceptor properties of the solute. In a first attempt to predict log BB by computational means, Van de Waterbeemd and Kansy [8] sought correlations with a number of computed descriptors. In particular, a good correlation was found using the polar surface area (PSA) together with the calculated molar volume. The polar surface area is the surface area occupied by N, O, OH and NH. Several other authors [9–11] also recognized the importance of the PSA in predicting the penetration through the BBB. Based on their findings they even concluded that log BB can be predicted by using only PSA related descriptors [11]. In addition to these parameters, descriptors related with molecular weight [12], molecular surface properties [13,14], molecular aromaticity, molecular branching and atomic electronegativity [10,15] have also been found to be important contributors to log BB.

All of the former mentioned models were obtained using multiple linear regression (MLR). Besides MLR, other statistical methods, e.g. partial least squares (PLS) have been applied in the prediction of log BB. Norinder et al. [16] used Molsurf parameterisation to calculate various properties related to the molecular valence region and combined it with PLS to develop a QSAR model. The PLS technique was also used by Luco [15] to develop a model based on topological descriptors and by Crivori et al. [17] to correlate transformed 3D molecule fields with experimental permeation. However, the PLS method generally appears to strip the QSAR from explicit physical insight, and the determination of the principle components of numerous physico-chemical descriptors cannot be easily calculated for an arbitrary compound. Another statistical method used for the prediction of log BB values [18] which gives promising results is neural network modeling.

All the models mentioned above have the same disadvantage: they perform relatively poorly for the set of molecules investigated at present. This poor performance is dual: some models fail in the prediction of the molecular set under investigation. Other models perform well in the prediction of this molecular set but the obtained model lacks a physico-chemical meaning, due to the descriptors or the statistical model building used. This hampers using the model to deduce new, more potent molecules.

Quantum chemical descriptors have proven useful for the prediction of many molecular biological and physico-chemical properties of interest to the pharmaceutical industry [19–24]. Given that it has been shown that certain semi-empirical and ab initio quantum chemical descriptors can be useful as molecular descriptors in the prediction of log BB [25–29], it seems worthwhile to investigate the use of ab initio calculated descriptors for the problem of predicting blood–brain barrier penetration of the molecules under investigation. The models developed will be examined for the added value of quantum chemical descriptors to generate predictive and statistical significant structure–activity relationships in the ADME-environment, where ADME stands for the absorption,

distribution, metabolism and excretion of medicinal molecules through the human body.

## 2. Methodology

### 2.1. Data set

The data set comprises 82 more or less structurally diverse molecules. The data is obtained from a low throughput in vivo test performed on male rats in a screening by Johnson & Johnson Pharmaceutical Research and Development. The actual plasma/brain ratio is determined by calculating the areas under the curve of the concentration of the molecule versus time in the plasma as well as in the brain. In view of the sensitivity of a QSAR model to the quality of the data used, it is important to note that the present experimental data were obtained under strict uniform experimental conditions. Therefore, we chose to employ only this data set of 82 data points.

### 2.2. Calculation of descriptors

Descriptors allow the numerical representation of the molecular structure and properties. The salient constitutional, electronic and steric attributes are represented in a set of descriptors. Two kinds of descriptors are generated: quantum chemical descriptors and classical 2D descriptors. The complete list of the descriptors generated can be found in Table 1.

#### 2.2.1. Quantum chemical descriptors

Methods based on quantum chemistry are used to determine the geometry and properties of the molecules corresponding to the minimum energy conformation. Quantum chemical calculations can, in principle, express all of the electronic and geometric properties of molecules and their interactions. The wave function allows the computation of a large set of

Table 1  
Complete list of descriptors

Quantum chemical descriptors
Dipole moment, polarizability
Equalized molecular electronegativity, molecular hardness, molecular softness and molecular electrophilicity
Most positive and most negative charge on any atom and specific atoms (H, C, O, N, F)
Maximum charge separation between any atom and between atoms of the same element (H, C, O, N, F)
Sum of square of charge of all atoms and atoms of the same element (H, C, O, N, F)
Covalent hydrogen bond acidity and basicity
Molecular electrostatic potential derived properties
Classical descriptors
clog <i>P</i>
Molecular volume, molecular weight
Total charge of the molecule at physiological pH
Flexibility, number of rotational bonds
Constitutional descriptors (number of ...)
TPSA related descriptors
1st, 2nd and 3rd principal component of molecular geometry

molecular descriptors. Numerous types of descriptors have been developed to capture the quantum chemical properties [21–23]. Atomic charges, polarizability, dipole moments and frontier orbital energies are among the most often used descriptors. Energy-related and electrostatic descriptors represent the weak intermolecular interactions, often involved in absorption processes. The electronegativity equalization method [30–33] has been used for the fast geometry-dependent calculation of several reactivity indices. These properties include molecular equalized electronegativity, molecular hardness, molecular softness and molecular electrophilicity.

### 2.2.2. Classical descriptors

Many classical descriptors have proven to be useful in the prediction of the activity through the blood–brain barrier. These descriptors comprise physico-chemical descriptors (clog *P*, molecular weight), counts of simple chemical fragments (atom types, functional group counts and rings), polar surface area related descriptors and a set of descriptors that accounts for the molecular shape. This set of shape-related descriptors is derived from a principal component analysis of the molecular geometry and are computed as the square root of the principal moments of the Cartesian coordinates. The 2D descriptors were chosen on the basis of simplicity, ease of calculation and interpretation.

### 2.2.3. Computational protocol

The computational protocol for obtaining the quantum chemical descriptors can be divided in several parts.

- *Conformational search*: the structures of the investigated compounds are built in Hyperchem [34] and are modeled in their neutral forms. The three-dimensional structures are determined by conformational analysis using the MM+ force field. The lowest energy conformation is kept for further studies.
- *Ab initio calculations*: this conformation is subjected to geometry optimization using the semi-empirical AM1 method in Gaussian 03 [35]. The geometry is then further optimized at B3LYP/6-31G\* level.
- *Calculation of molecular descriptors*: using the ab initio data, quantum chemical descriptors are calculated using in-house developed software.

The shape-related classical descriptors are calculated from the optimized geometry. The calculated clog *P* descriptor is obtained using Daylight [36]. Other classical descriptors like number of rotatable bonds and TPSA [24,37] related descriptors are calculated using Molinspiration [38]. The remaining classical 2D descriptors are calculated using e-Dragon [39,40].

### 2.3. Statistical analysis: data processing and modeling

Prior to model generation with the multiple linear regression technique, data were processed to improve prediction performance. In order to decrease the redundancy in the descriptor data matrix, the correlation among the descriptors

was examined and collinear descriptors were detected and deleted by the unsupervised forward selection algorithm [41,42], a data reduction algorithm that selects from a data matrix a maximal linearly independent set of columns with a minimal amount of multiple correlation. Descriptors lacking any variation (variables with standard deviation lower than 0.0005) are also removed. The remaining descriptors were scaled to set the mean value to zero and the variance to one, allowing all descriptors to have equal impact in the training process. The data is gathered into an  $m \times n$  data matrix *X*, where *m* and *n* are the number of molecules and descriptors, respectively. The activity data are ordered in a data vector. The data set was classified into calibration and external prediction sets. The calibration data set was used in the model development and the external prediction set was used to examine the overall performance of the model. The training/test set splitting is based on the K-means clustering algorithm [43] and from each cluster one compound for the test set is randomly selected. This algorithm guarantees that the members of the test set span the entire range of the experimental measurements. As statistical method multiple linear regression is used with a sequential selection of a chosen number of descriptors out of the total set of generated descriptors. Before performing the multiple linear regression on the data, the data are checked to exhibit normal distribution and the residuals are scattered around zero. The data check and the statistical analysis are performed with the in-house built ARTE-QSAR program. For an overview of the algorithms the reader can consult reference [44]. The validation of the obtained QSAR model is based on the guidelines set up in the Organization for Economic Cooperation and Development (OECD) principles [45]. The OECD guidelines are subdivided in five principles, which form the basis of a conceptual framework for characterizing QSAR models, and of reporting formats for describing the model characteristics in a transparent manner.

## 3. Results

To investigate the importance of the use of quantum chemical descriptors, three kinds of models were generated, all of them based on the training set of 70 molecules. The three models differ in the subset of descriptors used for model building: the set of classical descriptors, the set of purely quantum chemical descriptors and a third set of the combined information of quantum chemical and classical descriptors.

### 3.1. Discussion on the QSAR models for the prediction of log BB penetration

As the training set contains 70 molecules, according to the Topliss ratio [46] the number of descriptors should not exceed 14. An excessive number of descriptors lead to over-correlated equations that are difficult to interpret in terms of interactions and mechanisms. It is well known that increasing the model complexity always increases the multiple correlation coefficient, but if the model complexity is not well supervised then the predictive power of the model

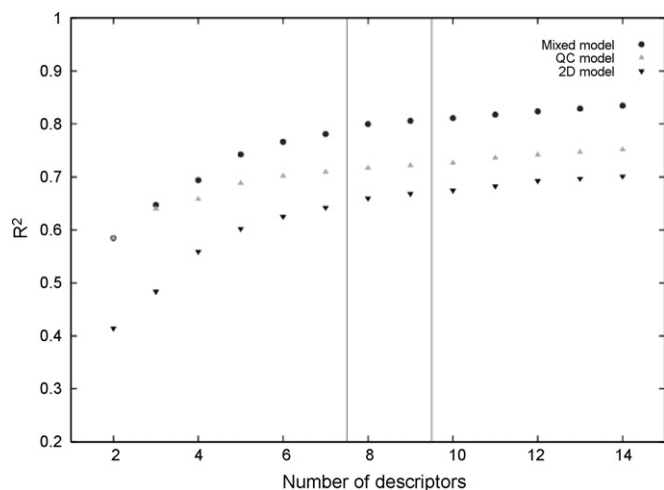


Fig. 1. Evolution of the goodness-of-fit,  $R^2$ , in function of the number of descriptors used in the model.

Table 2

Statistical summary of obtained QSAR models

	$R^2$	$q^2$	S.E.	$F$
Model 1: 2D descriptors	0.6575	0.5068	0.4754	14.6370
Model 2: QC descriptors	0.7142	0.6186	0.4308	19.3682
Model 3: mixed set of descriptors	0.8027	0.7281	0.3636	30.4282

$R^2$  is the coefficient of determination,  $q^2$  the leave-ten-out cross-validated coefficient of determination, S.E. the standard error of the estimate and  $F$  the Fisher statistics.

It is obvious that the model with the mixed set of descriptors is able to best predict the independent variable. From Fig. 1 one can see that independent of the number of descriptors entered, the goodness-of-fit and the statistical significance of the model based on the set of 2D descriptors, never reaches the values for the model based on a mixed set of descriptors or purely quantum chemical descriptors. This confirms the strength of the quantum chemical descriptors in constituting QSAR models. Moreover, it is clear that from the moment some quantum chemical information is entered into the 2D descriptor set, the performance of the model ameliorates a lot. Considering this result and the fact that all of the interpretable classical models from the literature failed on the present molecular set, it can be concluded that quantum chemical information contributes an important ‘added’ value in the prediction of the log BB for this set of molecules. As the model of the mixed set of descriptors is the one which gives the best correlation and validation, it will be tested on the extent of adherence to the OECD guidelines [45].

### 3.2. Diagnostics of the QSAR model following the OECD guidelines

#### 3.2.1. Principle 1: defined endpoint

For QSAR purposes, a well-defined endpoint should ideally be based on experimental data generated by a standardized test protocol. A lot of variation is possible in a test protocol for determining the blood–brain uptake of molecules: there are multiple types of administration, time points of measuring, test animals and techniques for plasma separation and brain homogenizing. All molecules in the present set were characterized experimentally using exactly the same procedure. All data come from an identical low throughput in vivo test in which the brain/plasma ratio is determined of compounds in male rats.

#### 3.2.2. Principle 2: defined algorithm

The need for a QSAR model to be associated with an unambiguous algorithm reflects the need of the end-user to understand how the estimated value was generated, and to be able to reproduce the algorithm and/or the estimates. When the underlying algorithm of a QSAR model is not transparent to the user, the model is sometimes referred as a ‘black box’. The present model is obtained with multiple linear regression. To use this model, one needs to establish certain data characteristics. As can be seen in Fig. 3, the normal probability plot, the data set is normally distributed. This behavior can be quantified by the coefficient of correlation between the ordered residuals

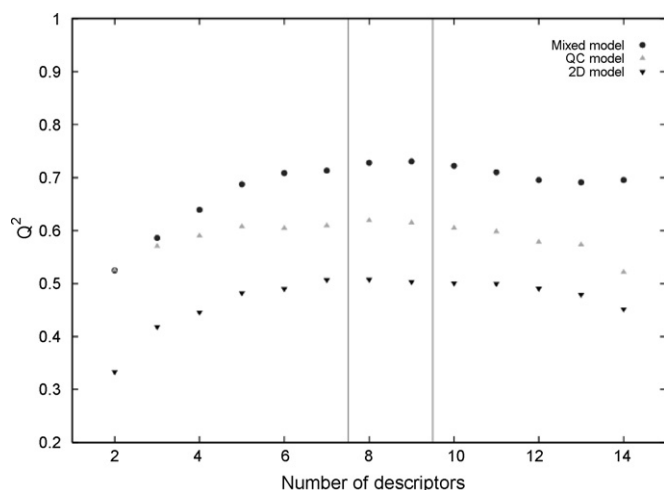


Fig. 2. Evolution of the predictivity,  $q^2$ , in function of the number of descriptors used in the model.

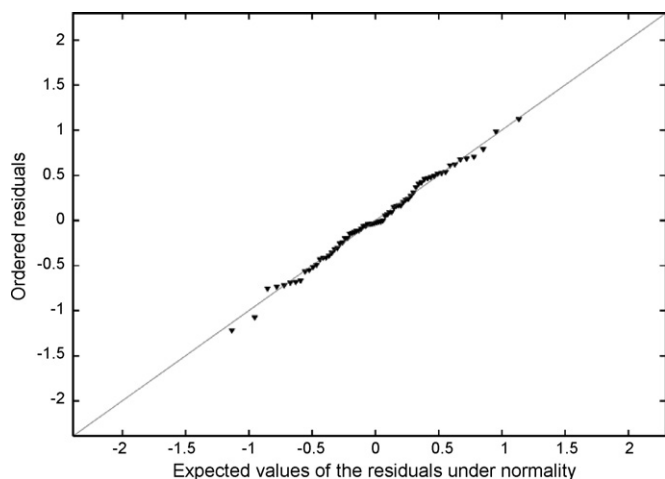


Fig. 3. Normal distribution plot.

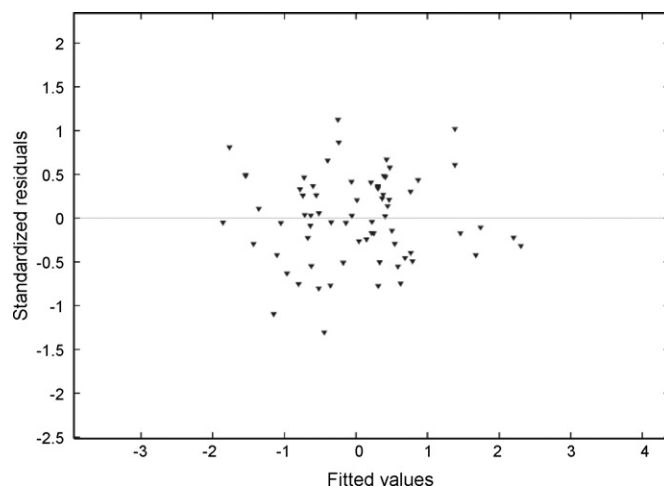


Fig. 4. Residual regression plot.

and their expected values under normality, which is 0.9966. The critical value for normal behavior, for  $\alpha = 0.05$  is 0.983.

Fig. 4 shows the residual plot, the scatter plot of the standardized residuals versus the fitted values. This plot shows whether the assumptions of homoscedasticity and zero mean are satisfied. The residuals have to be scattered around zero and the spread of the residuals has to be nearly the same throughout the graph. Based on the information in the graphs, it is concluded that the data is suitable for multiple linear regression.

The final model is summarized as follows with the values of the descriptors scaled to zero mean and standard deviation one

$$\begin{aligned} \log BB = & 0.3728(\pm 0.0669) \times \text{clog } P - 0.3327(\pm 0.069) \times \mu^2 \\ & - 0.6551(\pm 0.0693) \times q_C^{\max} + 0.4781(\pm 0.0749) \\ & \times \text{TPSA}(\text{NH}) - 0.1850(\pm 0.0610) \times n\text{Cl} \\ & + 0.1478(\pm 0.0680) \times n\text{R11} - 0.1531(\pm 0.0630) \\ & \times q_F^{\max} - 0.1596(\pm 0.0665) \times (q_H^{\max} - q_H^{\min}) \end{aligned} \quad (1)$$

This model becomes, with the unscaled values

$$\begin{aligned} \log BB = & 1.2064(\pm 0.2468) + 0.2013(\pm 0.0364) \times \text{clog } P \\ & - 0.1026(\pm 0.0215) \times \mu^2 - 2.9657(\pm 0.3162) \\ & \times q_C^{\max} + 0.0288(\pm 0.0045) \times \text{TPSA}(\text{NH}) \\ & - 0.2379(\pm 0.0791) \times n\text{Cl} + 0.1880(\pm 0.0873) \\ & \times n\text{R11} - 0.8935(\pm 0.3707) \times q_F^{\max} \\ & - 1.3240(\pm 0.5557) \times (q_H^{\max} - q_H^{\min}) \end{aligned} \quad (2)$$

with  $n = 70$ ,  $R^2 = 0.80$ ,  $R_{\text{adj}}^2 = 0.78$ ,  $s = 0.36$ ,  $Q_{\text{L100}}^2 = 0.7281$  and  $F = 30.43$ .

Here  $n$  is the number of compounds in the training set,  $R^2$  the squared coefficient of determination,  $R_{\text{adj}}^2$  the squared coefficient of determination adjusted for the number of included descriptors,  $s$  the standard error of the estimate,  $Q_{\text{L100}}^2$  the leave-ten-out cross-validated coefficient of determination and  $F$  is the Fisher statistics. An important component of an unambiguous algorithm is the availability of information on the descriptors used to link

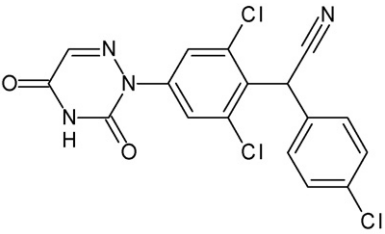
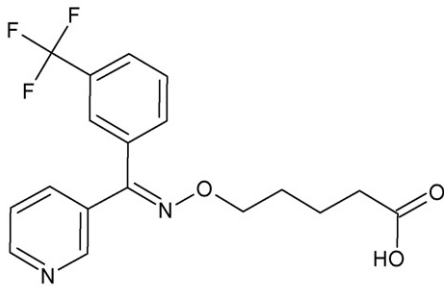
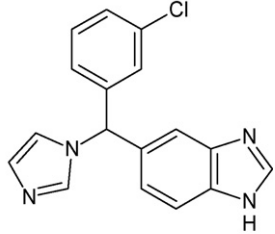
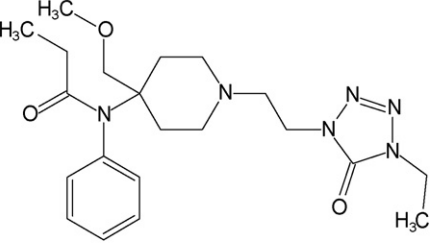
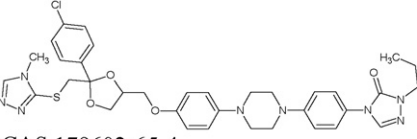
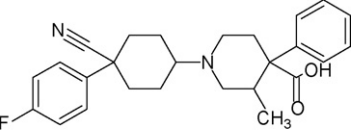
the chemical structures with the predicted endpoint. As important 2D classical descriptors, the number of Cl atoms ( $n\text{Cl}$ ) is found together with the number of the fusion of 6- and 7-membered rings ( $n\text{R11}$ ) in the molecules.  $\text{clog } P$  is the theoretically determined value for the octanol–water partition coefficient. The  $\text{TPSA}(\text{NH})$  descriptor is the topological polar surface area, where each N-containing fragment in the molecule contributes to the total value as described in Ref. [37]. The descriptors based on theoretical quantum chemical calculations are the maximum Mulliken charge-derived descriptor on the carbon atom ( $q_C^{\max}$ ) and on the fluorine ( $q_F^{\max}$ ), the maximal separation in Mulliken charges on the hydrogens in the molecule ( $q_H^{\max} - q_H^{\min}$ ) and the dipole moment ( $\mu^2$ ). All of them are calculated at the B3LYP/6-31G\* level. The values of the experimental and predicted  $\log BB$  for some molecules in the training and test set are mentioned in Table 3, together with the values for the selected descriptors. Because of propriety reasons we are only allowed to show this selection of molecules. Nevertheless, they are a representative set for the complete dataset and by providing the descriptor vectors for each molecule, the reader can be convinced of the validity of the model. To allow the reader to reproduce the obtained QSAR model, we also provide the QSAR model, obtained on the selected set of molecules shown and with the same descriptors used in the main QSAR model. The authors want to stress the fact that the model given in Eq. (3), has no use in the regulatory area at all, as it is only build on a set of 22 molecules. This equation is only given in order to meet the current requirement of reproducibility of QSAR models published.

$$\begin{aligned} \log BB = & 0.8694(\pm 0.2626) + 0.2341(\pm 0.0432) \times \text{clog } P \\ & - 0.1011(\pm 0.0208) \times \mu^2 - 2.5976(\pm 0.3061) \\ & \times q_C^{\max} + 0.0319(\pm 0.0047) \times \text{TPSA}(\text{NH}) \\ & - 0.2427(\pm 0.0640) \times n\text{Cl} + 0.2136(\pm 0.1929) \\ & \times n\text{R11} - 0.9460(\pm 0.5056) \times q_F^{\max} \\ & - 1.4198(\pm 0.6833) \times (q_H^{\max} - q_H^{\min}) \end{aligned} \quad (3)$$

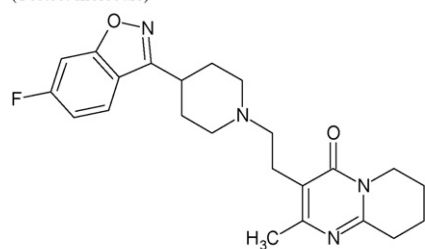
with  $n = 22$ ,  $R^2 = 0.92$ ,  $R_{\text{adj}}^2 = 0.75$ ,  $s = 0.22$ ,  $Q_{\text{L100}}^2 = 0.7501$  and  $F = 18.01$



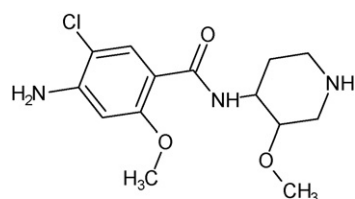
Table 3  
Representation of molecular dataset

Structure <sup>a</sup>	Ref. <sup>b</sup>	Exp. <sup>c</sup>	Pred. <sup>c</sup>	1 <sup>d</sup>	2 <sup>d</sup>	3 <sup>d</sup>	4 <sup>d</sup>	5 <sup>d</sup>	6 <sup>d</sup>	7 <sup>d</sup>	8 <sup>d</sup>
 <p>Cas 101831-37-2</p>	[47]	−1.30	−1.34	5	1.47	0.81	0.00	3.00	0.00	0.00	0.22
 <p>Cas 110140-89-1</p>	[48]	−1.00	−0.69	4.5	3.31	0.79	0.00	0.00	0.00	−0.26	0.27
 <p>CAS 115575-11-6</p>	[49]	−1.00	−0.63	3.2	9.07	0.35	0.00	1.00	0.00	0.00	0.20
 <p>CAS 69049-06-5</p>	[50]	−0.82	−1.01	2	0.58	0.83	0.00	0.00	0.00	0.00	0.09
 <p>CAS 179602-65-4</p>	[51]	−0.70	−1.17	5.6	7.38	0.79	0.00	1.00	0.00	0.00	0.12
 <p>CAS 79547-78-7</p>	[52]	−0.70	−0.56	1.9	3.04	0.58	0.00	0.00	0.00	−0.30	0.30

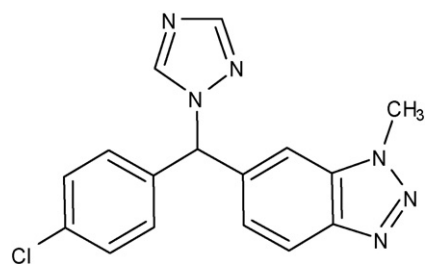
Structure <sup>a</sup>	Ref. <sup>b</sup>	Exp. <sup>c</sup>	Pred. <sup>c</sup>	1 <sup>d</sup>	2 <sup>d</sup>	3 <sup>d</sup>	4 <sup>d</sup>	5 <sup>d</sup>	6 <sup>d</sup>	7 <sup>d</sup>	8 <sup>d</sup>
(Testset molecule)											
 CAS 106266-06-2	[53]	-0.40	-0.44	2.60	6.02	0.58	0.00	0.00	0.00	-0.29	0.08
 CAS 202590-69-0	[54]	-0.22	-0.21	0.6	6.31	0.59	50.0	1.00	0.00	0.00	0.26
 CAS 129731-10-8	[55]	-0.16	-0.08	2.5	1.37	0.37	0.00	1.00	0.00	0.00	0.23
 CAS 179474-81-8	[56]	-0.10	-0.27	1.3	4.72	0.60	38.05	1.00	0.00	0.00	0.24
 CAS 155428-00-5	[57]	-0.01	0.01	1.4	3.26	0.63	36.09	0.00	0.00	0.00	0.24
 CAS unknown	[58]	0.07	0.17	5.5	1.45	0.63	0.00	0.00	0.00	0.00	0.09



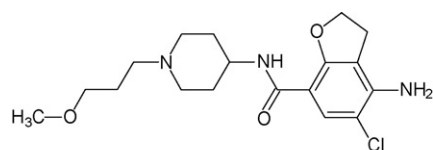
CAS 106266-06-2



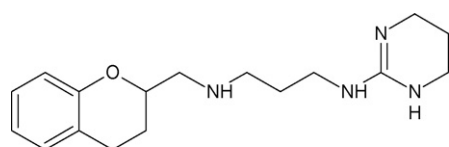
CAS 202590-69-0



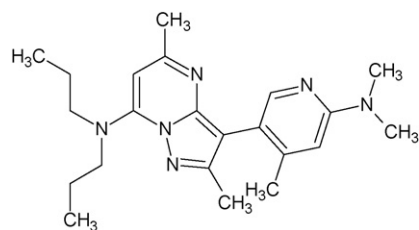
CAS 129731-10-8



CAS 179474-81-8



CAS 155428-00-5



CAS unknown

Table 3 (Continued)

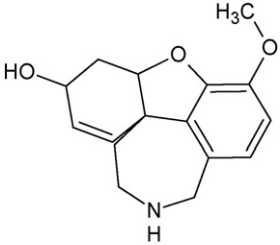
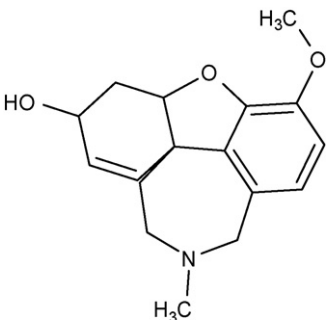
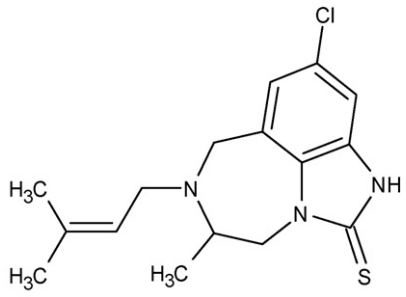
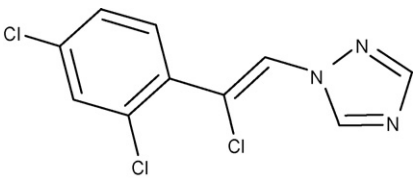
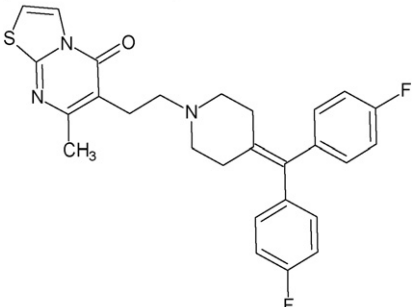
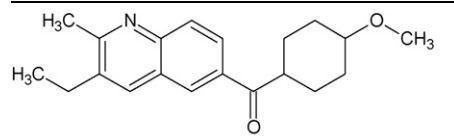
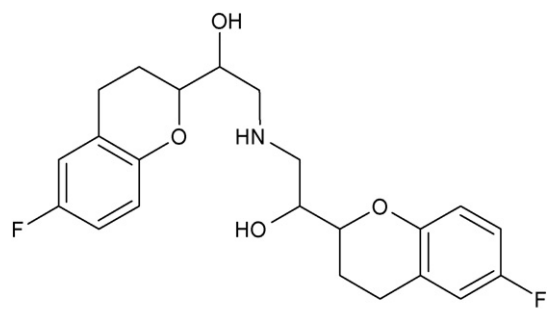
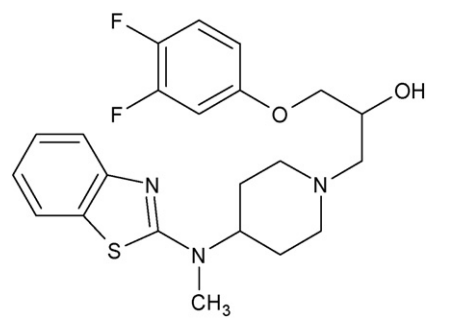
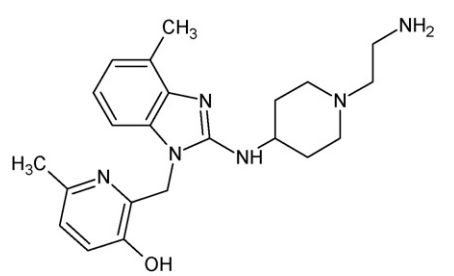
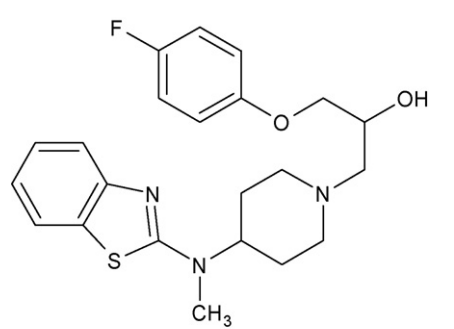
Structure <sup>a</sup>	Ref. <sup>b</sup>	Exp. <sup>c</sup>	Pred. <sup>c</sup>	1 <sup>d</sup>	2 <sup>d</sup>	3 <sup>d</sup>	4 <sup>d</sup>	5 <sup>d</sup>	6 <sup>d</sup>	7 <sup>d</sup>	8 <sup>d</sup>
 <p>CAS unknown</p>	/	0.18	0.53	0.6	0.15	0.32	12.0	0.00	1.00	0.00	0.27
 <p>CAS 357-70-0</p>	[59]	0.20	0.23	1	0.47	0.32	0.00	0.00	1.00	0.00	0.27
 <p>CAS 126347-69-1</p>	[60]	0.28	0.31	4.20	3.49	0.35	0.00	1.00	1.00	0.00	0.22
 <p>CAS 117857-45-1</p>	[61]	0.28	0.46	3.9	0.77	0.23	0.00	3.00	0.00	0.00	0.04
<p>(Testset molecule)</p> 	[62]	0.30	0.35	4.6	1.96	0.58	0.00	0.00	0.00	−0.30	0.10



Table 3 (Continued)

Structure <sup>a</sup>	Ref. <sup>b</sup>	Exp. <sup>c</sup>	Pred. <sup>c</sup>	1 <sup>d</sup>	2 <sup>d</sup>	3 <sup>d</sup>	4 <sup>d</sup>	5 <sup>d</sup>	6 <sup>d</sup>	7 <sup>d</sup>	8 <sup>d</sup>
 CAS unknown	[63]	0.33	0.47	3.8	1.97	0.40	0.00	0.00	0.00	0.00	0.08
 CAS: 99200-09-6	[64]	0.57	0.63	3.5	3.77	0.38	12.03	0.00	0.00	−0.31	0.30
 CAS 144665-07-6	[65]	0.59	0.44	4.3	3.21	0.39	0.00	0.00	0.00	−0.29	0.31
 CAS unknown	[66]	0.63	0.52	3	0.42	0.65	38.05	0.00	0.00	0.00	0.31
 CAS 104383-17-7	[67]	0.78	0.62	4.2	1.85	0.38	0.00	0.00	0.00	−0.31	0.31

<sup>a</sup> Structures with CAS number (when available).<sup>b</sup> References to literature for synthesis and use.<sup>c</sup> Experimental and predicted values for log BB.<sup>d</sup> Values for each of the remaining descriptors 1 = clog P 2 =  $\mu^2$  3 =  $q_C^{\max}$  4 = TPSA(NH) 5 = nCl 6 = nR11 7 =  $q_F^{\max}$  8 =  $q_H^{\max} - q_H^{\min}$ .

Table 4  
Statistical summary of QSAR models in the literature

	$R^2$	$R^2_{adj}$	$n$	M	Topliss ratio	S.E.
[26]	0.87	0.85	90	12 (2)	7.5	0.309
[29]	0.94	0.92	42	12 (0)	3.5	0.357
[28]	0.71	0.70	103	3 (1)	34.3	0.400
This model	0.80	0.78	70	8 (0)	8.75	0.360

$R^2$  is the coefficient of determination,  $R^2_{adj}$  is the squared coefficient of determination adjusted for the number of included descriptors,  $n$  the number of molecules in the training set,  $m$  the number of included descriptors (between brackets the number of non-significant descriptors) and S.E. the standard error of the estimate.

Quantum chemical properties have been used previously [26,28,29] to predict the penetration of the blood–brain barrier for different sets of molecular structures. It is always useful to compare the behavior of the currently obtained model with the performance of former obtained models in the literature, which are constructed based on the same statistical methodology. The statistical results for these models are summarized in Table 4.

The model obtained by Hutter [26] seems at first sight the most promising one. The only criticism which can be raised is that two out of the 12 descriptors are not significantly different from zero at the 5% significance level. Despite from the otherwise promising statistical characteristics, the performance of this model for the current set of molecules is very low, only 33% of the total variation is explained by the regression line. This can point to the fact that the model is molecule-specific and not applicable for the present set of molecules. On the other hand one can test the application of the current model on a selection of molecules present in the molecular set of Hutter [26] which lie within the domain of applicability of the model under investigation. A set of 21 molecules is predicted with 66% of the total variation explained by the regression line, which are encouraging results.

Al-Fahemi et al. [29] constructed a model based on 12 descriptors with a promising statistical outcome. The portion of total variation which is explained by the regression line is 0.94, although the Topliss-ratio [46] of 3.5 is rather low and using 12 descriptors might be somewhat too much. As a lower limit, a Topliss ratio of at least 5 is recommended. A recently developed model by Wichmann et al. [28] is built based on a molecular set of 103 molecules. They obtained a model with only three

descriptors, which provides the analysis with a high value for the Topliss ratio. The rather low value for  $R^2$ , 0.71, can be seen as a demerit for the model, but seems to be within the range of the results from other models [14,68,69] based on the same molecular set. From Table 5 it is clear that the presently developed model for the current molecular set has comparatively good statistical characteristics with a sufficiently high Topliss ratio. Moreover, as reported above, application of the descriptors used by Hutter [26] gives very poor results for the current molecules, whereas the application of the current model on a selection of molecules out of Ref. [26] gives predictive results.

### 3.2.3. Principle 3: defined domain of applicability

This principle can be defined by the so-called row diagnostics. For any of the 70 compounds in the training set the row diagnostics are computed, to reveal influential, leverage and outlier data. Once this is done, the 12 molecules from the test set can be analyzed on outlying cases. The leverage and outlier data for training and test molecules are summarized in the regression diagnostic plot in Fig. 5, the so-called Williams plot. The regression diagnostic plot indicates that three molecules from the training set are outlying with respect to their  $X$  values (high leverage).

The next step in the row diagnostics is to ascertain whether or not these cases are influential. A case is considered influential if its exclusion causes a major change in the fitted regression function. Three measures of influence that are widely used in practice are investigated, each one based on the omission of a single case to measure its influence. DFFITS represents the change in the predicted value that results from the exclusion of a particular case. DFBETAS stands for the change in the regression coefficient that results from the exclusion of a particular case. A value is computed for each term in the model, including the constant. Cook's distance is a measure of how much the residuals of all cases would change if a particular case were excluded from the calculation of the regression coefficients. Large values for each of these three

Table 5  
Row diagnostics for each of the outlying cases

Outlying diagnostic	Compound	
DFFITS < 0.6761	23	0.287
	32	0.038
	35	−0.145
Max(DFBETAS) < 0.239	23	0.008
	32	0.157
	35	0.075
Cook's distance < 1.000	23	0.170
	32	0.002
	35	0.040

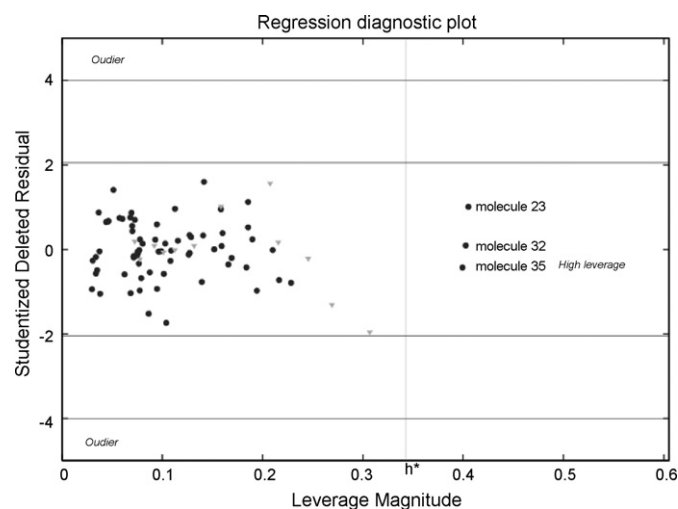


Fig. 5. Regression diagnostic plot: the domain of applicability in William's plot.

measures indicate that excluding a case from computation of the regression statistics changes the coefficients substantially. These diagnostics and their upper limit are summarized for the three outlying cases in Table 5. Based on these values, one can conclude that there are no influential cases in the training set. Any future prediction of unknown molecules should be preceded by the insurance that this chemical fits into the domain of applicability.

#### 3.2.4. Principle 4: statistical validation

A QSAR model should be associated with appropriate measures of goodness-of-fit, robustness and predictivity.

**3.2.4.1. Internal model performance: goodness-of-fit and robustness.** The goodness-of-fit of the model is a measure of how well the model accounts for the variance of the response in the training set and is given by the correlation coefficient and the standard error of the estimate. The present model yields a squared correlation coefficient of 0.80 and a standard error on the prediction of 0.3636. In view of the range of experimental log BB values from approximately  $-1.5$  to  $2$  log units this seems a considerable error. In a review of Clark [70], a variety of log BB models was compared and it was found that the predictive performance of all of them is about the same with approximately  $0.4$  log units. This appears to be close to the experimental error of approximately  $0.3$  log units. A standard error of the estimate smaller than the experimental error of the biological data can be an indication for an over fitted model, which is not the case for the present model. Before considering the model suitable for prediction and interpretation the robustness of the model needs also to be defined. The robustness of the model refers to the stability of its parameters and consequently the stability of its predictions when a perturbation is applied to the training set. The stability of the eight descriptors is summarized in the column diagnostics. The significance level for a descriptor has to be lower than  $0.05$ , the pairwise correlation coefficient lower than  $0.9$ . A pairwise correlation coefficient lower than  $0.9$  can also be expressed by a tolerance level higher than  $0.1$ . The column diagnostics are provided Table 6, together with the descriptors related with these values.

Further internal validation techniques used in this analysis are cross-validation and Y-scrambling. The cross-validation used in this analysis is the Leave-10-out procedure. The data set is divided into a number of blocks containing 10 molecules. At each step, all the compounds belonging to a block are left out from the derivation of the model and the cross-validation measure  $Q_{L100}^2$  amounts to  $0.7281$ . Because this value is high and it does not differ substantially from the  $R^2$  value, we can conclude that this model is not unstable and the obtained model is robust. Y-scrambling or the Y-randomization test is a widely used alternative to ensure the robustness of a QSAR model. In this test, the dependent variable vector is randomly shuffled and a new QSAR model is developed using the original independent-variable matrix. The process is repeated 500 times. It is expected that the resulting QSAR models should generally have low  $R^2$  and low  $q^2$  values. The results of the Y-

Table 6

Column diagnostics for the eight final used descriptors

Significance level	
clog P	0.000
$\mu^2$	0.000
$q_C^{\max}$	0.000
TPSA(NH)	0.000
Number of Cl atoms	0.004
Number of 11-membered rings	0.035
$q_F^{\max}$	0.019
$q_H^{\max} - q_H^{\min}$	0.020
Tolerance	
clog P	0.723
$\mu^2$	0.679
$q_C^{\max}$	0.673
TPSA(NH)	0.576
Number of Cl atoms	0.868
Number of 11-membered rings	0.698
$q_F^{\max}$	0.815
$q_H^{\max} - q_H^{\min}$	0.732

randomization are summarized in Fig. 6. Because all QSAR models obtained in the Y-randomization test have relatively low  $R^2$  and  $q^2$  values, it implies that an acceptable QSAR model can be obtained for the given data set by the current modeling method, without chance correlation.

**3.2.4.2. External model performance: predictivity.** External validation should be regarded as a supplementary procedure to internal validation, rather than as a superior alternative [71]. This is because a model that is externally predictive should also be robust, although a robust model is not necessarily predictive. A high value of the leave many out cross-validated correlation coefficient can be regarded as a necessary, but insufficient, condition for a model to have a high predictive power. The predictivity of a regression model is estimated by comparing the predicted and observed values of a sufficiently large and representative external test set of compounds that were not used in the model development. The quality of these predictions is

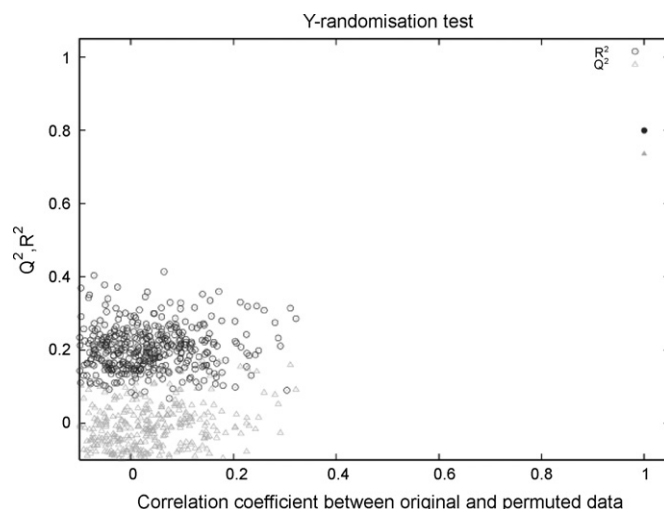


Fig. 6. Y, randomisation plot.

Table 7  
Statistical characteristics of the test set

Conditions	Performance of the model
$q_{\text{ext}}^2 > 0.5$	0.755
$R^2 > 0.6$	0.763
$(R^2 - R_0^2)/R^2 < 0.1$	0.001
$0.85 \leq k \leq 1.15$	0.91

$q_{\text{ext}}^2$  as mentioned in Eq. (3),  $R^2$  is the coefficient of determination of the regression line between the predicted and observed activities of the test set and  $R_0^2$  is the determination coefficient and  $k$  is the slope of the regression lines through the origin.

defined in terms of external explained variance  $q^2$  defined as follows:

$$q_{\text{ext}}^2 = 1 - \frac{\sum_{i=1}^{\text{test}} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{\text{test}} (y_i - \bar{y}_{\text{training}})^2} \quad (4)$$

with  $y_i$  is the known activities and  $\hat{y}_i$  the predicted activities of the molecules in the external test set and  $\bar{y}_{\text{training}}$  is the mean value of the activities of the molecules in the training set. It is the performance accuracy of the QSAR model on this test set that determines the actual predictive power of a QSAR model. A QSAR model is considered predictive if the conditions mentioned in Table 7 are satisfied.

It is demonstrated [71] that if a model meets all of the criteria mentioned, the predictive ability of a QSAR model can be certified. This is the case for the obtained BBB model.

### 3.2.5. Principle 5: mechanistic relevance

Note that the QSAR models do not express any causal relationships. QSAR findings can be useful in analyzing causal relationships, but they do not by themselves establish causal patterns. Any mechanistic interpretation should be treated with caution, because the assessment of the likelihood of that interpretation will depend partially on the state of the knowledge in the field (i.e. knowledge of the features captured by the model descriptors and the relevance of the association between the descriptors and the endpoint). The assessment will also be partially subjective, depending of the training and experience in the field of the endpoint. The selected set of descriptors shows a possible direction in which one have to search to obtain more potent molecules but is not a unique statement about the problem under investigation.

In accordance with a commonly found conclusion [7,26,68] concerning the blood–brain barrier, the results of this analysis indicates that non-polar molecules cross the barrier more easily than polar molecules. This can be seen in the sign of the coefficients in (1) of the atomic charges, the dipole moment and the clog  $P$  value. The presence of the polar surface area due to the nitrogen atoms in the model describes the significance of the polar interaction between the solutes and the blood–brain barrier.

## 4. Conclusion

The multiple linear regression technique was used to obtain predictive QSAR models for log BB of a diverse set of

molecules using quantum chemical descriptors derived from ab initio calculations. An eight-parameter model could reproduce the log BB data, based on the information of clog  $P$ , TPSA(NH), several Mulliken charge related descriptors, the number of Cl atoms and 11-membered rings and the dipole moment. The log BB is a function of the lipophilicity and electronic properties of the molecule. By using the ab initio method for more accurate calculations of these electronic properties, the proposed model could reproduce the log BB data with an error approaching the experimental uncertainty. It is obvious from this analysis that the inclusion of quantum chemical information improves the behavior of the model. This encourages the research on the use of quantum chemical descriptors for the prediction of ADME-Tox properties.

## Acknowledgements

The authors thank Ghent University and the Fund for Scientific Research-Flanders (FWO-Vlaanderen) for their grants to the Quantum Chemistry group at Ghent University. S.V.D. thanks the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen) for the PhD. grant. Johnson&Johnson Pharmaceutical Research and Development is kindly acknowledged for providing the authors with the molecular dataset.

## References

- [1] M.W.B. Bradbury, The blood–brain-barrier—transport across the cerebral endothelium, *Circ. Res.* 57 (1985) 213–222.
- [2] A.G. de Boer, P.J. Gaillard, Drug targeting to the brain, *Annu. Rev. Pharmacol. Toxicol.* 47 (2007) 323–355.
- [3] A. Misra, S. Ganesh, A. Shahiwal, S.P. Shah, Drug delivery to the central nervous system: a review, *J. Pharm. Pharm. Sci.* 6 (2003) 252–273.
- [4] I. Tamai, A. Tsuji, Drug delivery through the blood–brain barrier, *Adv. Drug Deliv. Rev.* 19 (1996) 401–424.
- [5] E.G. Chikhale, K.Y. Ng, P.S. Burton, R.T. Borchardt, Hydrogen-bonding potential as a determinant of the in-vitro and in-situ blood–brain-barrier permeability of peptides, *Pharm. Res.* 11 (1994) 412–419.
- [6] 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, T.J. Wilks, Development of a new physicochemical model for brain penetration and its application to the design of centrally acting H-2-receptor histamine-antagonists, *J. Med. Chem.* 31 (1988) 656–671.
- [7] M.H. Abraham, The factors that influence permeation across the blood–brain barrier, *Eur. J. Med. Chem.* 39 (2004) 235–240.
- [8] H. Vandewaterbeemd, M. Kansy, Hydrogen-bonding capacity and brain penetration, *Chimia* 46 (1992) 299–303.
- [9] H. van de Waterbeemd, G. Camenisch, G. Folkers, J.R. Chretien, O.A. Raevsky, Estimation of blood–brain barrier crossing of drugs using molecular size and shape, and H-bonding descriptors, *J. Drug Target.* 6 (1998) 151–165.
- [10] D.E. Clark, Rapid calculation of polar molecular surface area and its application to the prediction of transport phenomena. 2. Prediction of blood–brain barrier penetration, *J. Pharm. Sci.* 88 (1999) 815–821.
- [11] J. Kelder, P.D.J. Grootenhuys, D.M. Bayada, L.P.C. Delbressine, J.P. Ploemen, Polar molecular surface as a dominating determinant for oral absorption and brain penetration of drugs, *Pharm. Res.* 16 (1999) 1514–1519.
- [12] R. Kaliszan, M. Markuszewski, Brain/blood distribution described by a combination of partition coefficient and molecular mass, *Int. J. Pharm.* 145 (1996) 9–16.

- [13] F. Lombardo, J.F. Blake, W.J. Curatolo, Computation of brain-blood partitioning of organic solutes via free energy calculations, *J. Med. Chem.* 39 (1996) 4750–4755.
- [14] G.M. Keseru, L. Molnar, High-throughput prediction of blood–brain partitioning: a thermodynamic approach, *J. Chem. Inf. Comput. Sci.* 41 (2001) 120–128.
- [15] J.M. Luco, Prediction of the brain–blood distribution of a large set of drugs from structurally derived descriptors using partial least-squares (PLS) modeling, *J. Chem. Inf. Comput. Sci.* 39 (1999) 396–404.
- [16] U. Norinder, P. Sjöberg, T. Osterberg, Theoretical calculation and prediction of brain–blood partitioning of organic solutes using MolSurf parametrization and PLS statistics, *J. Pharm. Sci.* 87 (1998) 952–959.
- [17] P. Crivori, G. Cruciani, P.A. Carrupt, B. Testa, Predicting blood–brain barrier permeation from three-dimensional molecular structure, *J. Med. Chem.* 43 (2000) 2204–2216.
- [18] B. Hemmateenejad, R. Miri, M.A. Safarpour, A.R. Mehdipour, Accurate prediction of the blood–brain partitioning of a large set of solutes using ab initio calculations and genetic neural network modeling, *J. Comput. Chem.* 27 (2006) 1125–1135.
- [19] A. Agatonovic-Kustrin, I.G. Tucker, M. Zecevic, L.J. Zivanovic, Prediction of drug transfer into human milk from theoretically derived descriptors, *Anal. Chim. Acta* 418 (2000) 181–195.
- [20] S. Agatonovic-Kustrin, R. Beresford, A.P.M. Yusof, Theoretically-derived molecular descriptors important in human intestinal absorption, *J. Pharm. Biomed. Anal.* 25 (2001) 227–237.
- [21] G.R. Famini, L.Y. Wilson, Linear free energy relationships using quantum mechanical descriptors, *Rev. Comput. Chem.* 18 (2002) 211–255.
- [22] M. Karelson, V.S. Lobanov, A.R. Katritzky, Quantum-chemical descriptors in QSAR/QSPR studies, *Chem. Rev.* 96 (1996) 1027–1043.
- [23] M. Karelson, in: P. Bultinck, H. De Winter, W. Langenaeker, J.P. Tollenaere (Eds.), *Quantum-Chemical descriptors in QSAR*. In *Computational Medicinal Chemistry for Drug Discovery*, Decker Inc., New York, 2004, pp. 641–667.
- [24] M.J. Soric, R.A. McKinnon, J.O. Miners, D.A. Winkler, P.A. Smith, Rapid prediction of chemical metabolism by human UDP-glucuronosyltransferase isoforms using quantum chemical descriptors derived with the electronegativity equalization method, *J. Med. Chem.* 47 (2004) 5311–5317.
- [25] X.C. Fu, G.P. Wang, W.Q. Liang, Q.S. Yu, Predicting blood–brain barrier penetration of drugs using an artificial neural network, *Pharmazie* 59 (2004) 126–130.
- [26] M.C. Hutter, Prediction of blood–brain barrier permeation using quantum chemically derived information, *J. Comput. Aided Mol. Des.* 17 (2003) 415–433.
- [27] X.L. Ma, C. Chen, J. Yang, Predictive model of blood–brain barrier penetration of organic compounds, *Acta Pharmacol. Sin.* 26 (2005) 500–512.
- [28] K. Wichmann, M. Diedenhofen, A. Klamt, Prediction of blood–brain partitioning and human serum albumin binding based on COSMO-RS sigma-moments, *J. Chem. Inf. Model.* 47 (2007) 228–233.
- [29] J.H.A. Al-Fahemi, D.L. Cooper, N.L. Allan, Investigating the utility of momentum-space descriptors for predicting blood–brain barrier penetration, *J. Mol. Graph. Model.* 26 (2007) 607–612.
- [30] P. Bultinck, W. Langenaeker, P. Lahorte, F. De Proft, P. Geerlings, C. Van Alsenoy, J.P. Tollenaere, The electronegativity equalization method II: applicability of different atomic charge schemes, *J. Phys. Chem. A* 106 (2002) 7895–7901.
- [31] P. Bultinck, R. Carbo-Dorca, Algebraic relationships between conceptual DFT quantities and the electronegativity equalization hardness matrix, *Chem. Phys. Lett.* 364 (2002) 357–362.
- [32] P. Bultinck, W. Langenaeker, P. Lahorte, F. De Proft, P. Geerlings, M. Waroquier, J.P. Tollenaere, The electronegativity equalization method I: parametrization and validation for atomic charge calculations, *J. Phys. Chem. A* 106 (2002) 7887–7894.
- [33] P. Bultinck, W. Langenaeker, R. Carbo-Dorca, J.P. Tollenaere, Fast calculation of quantum chemical molecular descriptors from the electronegativity equalization method, *J. Chem. Inf. Comput. Sci.* 43 (2003) 422–428.
- [34] Hyperchem(TM) [7.0]. Hypercube, Inc., 1115 NW 4th Street, Gainesville, Florida 32601, USA, 2007.
- [35] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, J.A. Montgomery, T. Vreven, K.N. Kudin, J.C. Burant, J.M. Millam, S.S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G.A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J.E. Knox, H.P. Hratchian, J.B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, P.Y. Ayala, K. Morokuma, G.A. Voth, P. Salvador, J.J. Dannenberg, V.G. Zakrzewski, S. Dapprich, A.D. Daniels, M.C. Strain, O. Farkas, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J.V. Ortiz, Q. Cui, A.G. Baboul, S. Clifford, J. Cioslowski, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-laham, C.Y. Peng, A. Nanayakkara, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, M.W. Wong, C. Gonzalez, J.A. Pople. Gaussian 03, Revision B. 06.03.2007, Pittsburgh PA, 2003.
- [36] T. Leo, D. Weiniger, Daylight clogP 4.0. Daylight Chemical Information Systems, Santa Fe, New Mexico, 2007.
- [37] P. Ertl, B. Rohde, P. Selzer, Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties, *J. Med. Chem.* 43 (2000) 3714–3717.
- [38] N.L. Allinger, Conformational-analysis. 130. Mm2—hydrocarbon force-field utilizing V1 and V2 torsional terms, *J. Am. Chem. Soc.* 99 (1977) 8127–8134.
- [39] I.V. Tetko, Computing chemistry on the web, *Drug Discov. Today* 10 (2005) 1497–1500.
- [40] I.V. Tetko, J. Gasteiger, R. Todeschini, A. Mauri, D. Livingstone, P. Ertl, V. Palyulin, E. Radchenko, N.S. Zefirov, A.S. Makarenko, V.Y. Tanchuk, V.V. Prokopenko, Virtual computational chemistry laboratory—design and description, *J. Comput. Aided Mol. Des.* 19 (2005) 453–463.
- [41] D.J. Livingstone, D.W. Salt, Variable selection-spoilt for choice? *Rev. Comput. Chem.* 21 (2006) 287–348.
- [42] D. Whitley, M.G. Ford, D.J. Livingstone, Unsupervised forward selection: a method for eliminating redundant variables, *J. Chem. Inf. Comput. Sci.* 40 (2004) 1160–1168.
- [43] A. Golbraikh, M. Shen, Z.Y. Xiao, Y.D. Xiao, K.H. Lee, A. Tropsha, Rational selection of training and test sets for the development of validated QSAR models, *J. Comput. Aided Mol. Des.* 17 (2003) 241–253.
- [44] S. Van Damme, P. Bultinck, A new computer program for QSAR-analysis: ARTE-QSAR, *J. Comput. Chem.* 28 (2007) 1924–1928.
- [45] A.P. Worth, A. Bassan, A. Gallegos, T.I. Netzeva, G. Patlewicz, M. Pavan, I. Tsakovska, M. Vracko, The Characterisation of (Quantitative) Structure–Activity Relationships: Preliminary Guidance, 2007.
- [46] J.G. Topliss, R.P. Edwards, Chance factors in studies of quantitative structure–activity relationships, *J. Med. Chem.* 22 (1979) 1238–1244.
- [47] J. Dekock, M. Desmet, R. Sneyers, Determination of Diclazuril in animal feed by liquid-chromatography, *J. Chromatogr.* 606 (1992) 141–146.
- [48] J. Vanreempts, B. Vandeuren, M. Borgers, F. Declerck, R68070, a combined Txa2-synthetase Txa2-prostaglandin endoperoxide receptor inhibitor, reduces cerebral infarct size after photochemically initiated thrombosis in spontaneously hypertensive rats, *Thromb. Haemost.* 58 (1987) 182.
- [49] P.E. Goss, A. Oza, R. Goel, J.M. Nabholz, R. De Coster, J. Bruynseels, C. Reid, N. Wadden, M. Crump, L.M. Tye, Liarazole fumarate (R85246): a novel imidazole in the treatment of receptor positive postmenopausal metastatic breast cancer, *Breast Cancer Res. Treat.* 59 (2000) 55–68.
- [50] B. Kay, B. Pleuvry, Human volunteer studies of Alfentanil (R39209), a new short-acting narcotic analgesic, *Anaesthesia* 35 (1980) 952–956.
- [51] L. Lammens, H. Borghys, P. Roevens, J. Vandenberghe, Histological changes in liver and jejunum induced by R103757, a new inhibitor of microsomal triglyceride transfer protein (MTP), *Atherosclerosis* 144 (1999) 39.
- [52] J. Monbaliu, M. Michiels, R. Woestenborghs, I. Geuens, R. Gasparini, J. Heykants, Pharmacokinetics of levocabastine (R50547) in the dog,

- Archives Internationales de Pharmacodynamie et de Therapie 274 (1985) 330–331.
- [53] T.F. Meert, P. Dehaes, P.A.J. Janssen, Risperidone, a potent and complete Lsd-antagonist with antipsychotic properties, *Psychopharmacology* 96 (1988) S37.
- [54] W. Meuldermans, J. Hendrickx, W. Lauwers, R. Hurkmans, E. Mostmans, E. Swysen, J. Bracke, A. Knaeps, J. Heykants, Excretion and biotransformation of cisapride in rats after oral-administration, *Drug Metab. Dispos.* 16 (1988) 410–419.
- [55] W. Wouters, E. Snoeck, R.V. Decoster, A specific nonsteroidal aromatase inhibitor, *Breast Cancer Res. Treat.* 30 (1994) 89–94.
- [56] J.P.R.M. Bosmans, J.A.J. Schuurkes, G.H.P. Vandaele, M.A.J. Decleyn, R93877—the first selective colonkinetic agent, Abstracts of Papers of the American Chemical Society 1997, 214, 266-MEDI.
- [57] V. Limmroth, D. Wermelskirchen, F. Tegtmeier, H.C. Diener, R 91274 (Alniditan) blocks neurogenic inflammation in the rat meninges more effectively than sumatriptan by activation of 5HT1D-receptors, *Neurology* 48 (1997) 1137.
- [58] A.W. Zobel, T. Nickel, H.E. Kunzel, N. Ackl, A. Sonntag, M. Ising, F. Holsboer, Effects of the high-affinity corticotropin-releasing hormone receptor 1 antagonist R121919 in major depression: the first 20 patients treated, *J. Psychiatr. Res.* 34 (2000) 171–181.
- [59] R.B. Rosse, S.I. Deutsch, Adjuvant galantamine administration improves negative symptoms in a patient with treatment-refractory schizophrenia, *Clin. Neuropharmacol.* 25 (2002) 272–275.
- [60] E.L. White, R.W. Buckheit, L.J. Ross, J.M. Germany, K. Andries, R. Pauwels, P.A.J. Janssen, W.M. Shannon, M.A. Chirigos, A tibo derivative, R82913, is a potent inhibitor of Hiv-1 reverse-transcriptase with heteropolymer templates, *Antivir. Res.* 16 (1991) 257–266.
- [61] F. Debeukelaar, L. Tritsmans, Loreclezole, *Epilepsy Res.* (1991) 125–128.
- [62] A. Bosio, S. Marini, E. Verzelletti, A. Ermentini, Preliminary clinical-evaluation of ritanserin (R55667) in anxiety disorders, *Int. J. Neurosci.* 32 (1987) 424.
- [63] Y. Huang, N.N. Guo, Z.H. Zhu, R. Narendran, A. Lesage, F. Bischoff, D.R. Hwang, H. Moore, M. Laruelle, A new radioligand for the metabotropic glutamate 1 receptor: synthesis and in vivo evaluation of [ $^{11}\text{C}$ ]R176898, *Neuroimage* 22 (2004) T33.
- [64] D. Duprez, R. Lefebvre, T. Debacker, P. Desutter, J. Trouerbach, D.L. Clement, Influence of nebivolol on the cardiovascular hemodynamics during postural changes and isometric-exercise, *Cardiovasc. Drugs Ther.* 5 (1991) 709–717.
- [65] G. Uceda, A.G. Garcia, J.M. Guantes, P. Michelena, C. Montiel, Effects of  $\text{Ca}^{2+}$  channel antagonist subtypes on mitochondrial  $\text{Ca}^{2+}$  transport, *Eur. J. Pharm. Mol. Pharmacol. Sect.* 289 (1995) 73–80.
- [66] P.R. Wyde, S.N. Chetty, P. Timmerman, B.E. Gilbert, K. Andries, Short duration aerosols of JNJ 2408068 (R170591) administered prophylactically or therapeutically protect cotton rats from experimental respiratory syncytial virus infection, *Antivir. Res.* 60 (2003) 221–231.
- [67] H. Geerts, R. Nuydens, R. Nuyens, F. Cornelissen, Sabeluzole accelerates neurite outgrowth in different neuronal cell-lines, *Restor. Neurol. Neurosci.* 4 (1992) 21–32.
- [68] K. Rose, L.H. Hall, L.B. Kier, Modeling blood–brain barrier partitioning using the electrotopological state, *J. Chem. Inf. Comput. Sci.* 42 (2002) 651–666.
- [69] F. Ooms, P. Weber, P.A. Carrupt, B. Testa, A simple model to predict blood–brain barrier permeation from 3D molecular fields, *Biochim. Biophys. Acta Mol. Basis Dis.* 1587 (2002) 118–125.
- [70] D.E. Clark, In silico prediction of blood–brain barrier permeation, *Drug Discov. Today* 8 (2003) 927–933.
- [71] A. Golbraikh, A. Tropsha, Beware of  $q(2)!$  *J. Mol. Graph. Model.* 20 (2002) 269–276.