Prediction of Drug Distribution in Rat and Humans Using an Artificial Neural Networks Ensemble and a PBPK Model

ARTICLE *in* PHARMACEUTICAL RESEARCH · MAY 2014

Impact Factor: 3.42 · DOI: 10.1007/s11095-014-1421-4 · Source: PubMed

4 AUTHORS, INCLUDING:

Paulo Paixao
University of Lisbon
11 PUBLICATIONS **151** CITATIONS
SEE PROFILE

Luis F. Gouveia
University of Lisbon
43 PUBLICATIONS **549** CITATIONS
SEE PROFILE

José A G Morais
University of Lisbon
68 PUBLICATIONS **655** CITATIONS
SEE PROFILE

Pharm Res (2014) 31:3313–3322
DOI 10.1007/s11095-014-1421-4

RESEARCH PAPER

Prediction of Drug Distribution in Rat and Humans Using an Artificial Neural Networks Ensemble and a PBPK Model

Paulo Paixão • Natália Aniceto • Luís F. Gouveia • José A. G. Morais

Received: 21 February 2014 / Accepted: 12 May 2014 / Published online: 28 May 2014

ABSTRACT

Purpose To develop a QSAR model, based on calculated molecular descriptors and an Artificial Neural Networks Ensemble (ANNE), for the estimation of rat tissue-to-blood partition coefficients ($K_{t:b}$), as well as the assessment of the applicability domain of the model and its utility in predicting the drug distribution in humans.

Methods A total of 1460 individual $K_{t:b}$ values (75% train and 25% validation), obtained in 13 different rat tissues were collected in the literature. A correlation between simple molecular descriptors for lipophilicity, ionization, size and hydrogen bonding capacity and $K_{t:b}$ data was attempted by using an ANNE.

Results Similar statistics were observed between the train and validation group of data with correlations, between the observed values and the predicted average ANNE values, of 0.909 and 0.896, respectively. A degradation of the correlations was observed for predicted values with high uncertainty, as judged by the standard deviations of the ANNE outputs. This was further observed when using the ANNE $K_{t:b}$ values in a Physiologically based pharmacokinetic (PBPK) model for predicting the Human Volume of distribution of another 532 drugs.

Conclusions This model (available as a MS Excel® workbook in the Supporting material of this article) may be a valuable tool for prediction and simulation in early drug development, allowing the *in silico* estimation of rat $K_{t:b}$ values for PBPK purposes and also indicating its applicability domain.

KEY WORDS Artificial Neural Networks · rat tissue-to-blood partition coefficients · QSAR · PBPK · volume of distribution

Electronic supplementary material The online version of this article (doi:10.1007/s11095-014-1421-4) contains supplementary material, which is available to authorized users.

P. Paixão (✉) • N. Aniceto • L. F. Gouveia • J. A. G. Morais
iMed.UL, Faculdade de Farmácia, Universidade de Lisboa, Av. Prof. Gama Pinto, 1649-003 Lisboa, Portugal
e-mail: ppaixao@ff.ul.pt

INTRODUCTION

PBPK models are becoming increasingly used in the various phases of the drug development process, as these may include data from different origins in an integrated way, allowing the simulation of the pharmacokinetic behaviour for mechanistic interpretation (1). *In silico*, *in vitro* or *in vivo* data characterizing any of the individual ADME single processes, may thus be introduced in the drug related part of the PBPK models and, by changing the physiologic parameters of these models, extrapolations between different species may be obtained (2).

Within this context, tissue-to-blood partition coefficients ($K_{t:b}$) in the different organs are fundamental parameters that must be determined in order to describe the distribution of the drug, and to simulate the concentration profiles for each individual organ (3). Volume of distribution at steady-state (Vss), which is defined as the proportionality factor that relates the concentration of drug in the blood (or plasma) to the total amount of drug in the body at steady state and is related to the extent of drug distribution in the body, may be determined by the sum of the products of $K_{t:b}$ and the corresponding tissue volume in addition to the blood volume (3). However, because it is difficult to determine $K_{t:b}$ values in man they are typically determined in other mammals. To that end both *in vitro* (using animal tissues homogenates) or *in vivo* approaches are available (4), but these protocols are expensive, laborious, ethically questionable and require the physical existence of the drug molecule itself, which hinders its utilization in the early phases of the drug development process.

To overcome some of these limitations, a number of advanced physiological models have been developed, which utilize a combination of human *in vitro* measurements, physicochemical data, and physiological data, such as tissue composition, to define tissue distribution, obviating the need for animal data altogether (5–8). These models are being extensively used with stimulating results but they still require the physical existence of the drug molecule (9). It would be very

appealing if these same parameters could be predicted based only on computational models, and indeed some *in silico* QSAR models are already available (10–12). However, these models are based on a limited number of drugs or were built based only on volatile organic compounds (VOC). Although these models presented some interesting results, it is well established that the performance of a QSAR model is highly dependent on the experimental design and diversity of compounds used (13), making the applicability of these available QSAR models of limited scope in the drug development scenario.

In this paper we present the development of a QSAR model for the *in silico* prediction of $K_{t:b}$ values in 13 different rat tissues obtained from different experimental approaches, based on an artificial neural networks ensemble (ANNE). Simple and easily calculated molecular descriptors were used. Validation was performed both on a group of external $K_{t:b}$ data and also on the ability to predict the Vss in humans in a large amount of drugs. The applicability domain was also characterized based on the variability of the predicted values. The final model is available as an MS Excel® workbook in the Supporting material of this paper.

MATERIALS AND METHODS

Data Base of Rat $K_{t:b}$ Values

The QSAR models were built based on the two datasets described and characterized previously (14). Dataset 1 included *in vitro* data on volatile and non-volatile compounds determined in 13 different rat tissues, and contains 143 compounds (64.3% were VOCs) on a total of 657 $K_{t:b}$ values. Dataset 2 consisted of *in vivo* derived data, also obtained in 13 different rat tissues, and included a total of 196 compounds (8.7% were VOCs) and 817 $K_{t:b}$ values. It was previously reported that *in vitro* and *in vivo* $K_{t:b}$ values, although correlated, are statistically different (14,15). A statistically significant difference was again observed between the different tissues, but good correlations were also observed between several of them (14), indicating that common distribution characteristics are expected in the different tissues for several drugs. As such, in order to increase the applicability domain by using a larger data base, *in vitro* and *in vivo* $K_{t:b}$ values were both included in the model by introducing an additional binary input to the ANN having a value of zero for *in vitro* data and a value of one for *in vivo* data. Four additional binary inputs were also included in order to indicate the different rat tissues. These 4 bits (half a byte, or a nibble) are able to represent 16 different values/tissues. Values, from 1 to 13 were assigned to the different tissues with no particular order, (*e.g.* from 1, 0, 0, 0 to describe the muscle to 1, 0, 1, 1 to describe the pancreas) (Table I). The total data (nitric oxide and cyclosporine were excluded due to impossibility to calculate the required molecular descriptors and only 1460 of the available 1474 individual $K_{t:b}$ values were considered) were randomly split in a 75/25 ratio within each individual tissue between a train (n=1098 $K_{t:b}$ values) and an external validation group of data (n=362 $K_{t:b}$ values), not used in the QSAR model building procedure. For further validation, Vss values were also collected for another 532 drugs (Vss Validation Data) from the work of Obach et al. (16). These drugs were included neither in the train nor in the external validation data. Since these Vss values were determined in plasma, blood-to-plasma ratios were calculated according to Paixão *et al.* (17) and used to transform to blood Vss values the original plasma Vss values. These dataset are available in the Supplementary material as supporting data.

Table I Binary Representation of the Different Rat Tissues as Used in the ANN Model

Tissue	Binary selector			
	4	3	2	1
Muscle	0	0	0	1
Bone	0	0	1	0
Brain	0	0	1	1
Heart	0	1	0	0
Intestine	0	1	0	1
Skin	0	1	1	0
Lungs	0	1	1	1
Spleen	1	0	0	0
Adipose	1	0	0	1
Liver	1	0	1	0
Kidney	1	0	1	1
Stomach	1	1	0	0
Pancreas	1	1	0	1

In Silico Calculation of the Molecular Descriptors

The following methodology was used for the calculation of the *in silico* descriptors: SMILES notation of each molecule was obtained using the on-line PubChem Compound database (http://www.ncbi.nlm.nih.gov). Ionization descriptors (pKa_{acid}; pKa_{base}) were predicted by using ChemAxon (http://www.chemicalize.org). For drugs without an ionisable acid group, a value of 15 was attributed to pKa_{acid}. For drugs without an ionisable basic group, a value of -1 was attributed to pKa_{base}. Lipophilicity (LogP) and intrinsic solubility (LogS) were obtained using the on-line ALOGPS 2.1 program (18). The remaining descriptors, related to size, hydrogen bonding potential, lipophilicity and others, were obtained from the on-line E-Dragon 1.0 software using CORINA to convert the SMILES notation to the 3D representation of the molecule (19). Molecular descriptors were not

possible to calculate for nitric oxide and cyclosporine, and these two molecules were removed from the dataset.

ANNE Optimization

The Artificial Neural Network (ANN) non-linear regression was performed using the backpropagation neural modelling system Qnet for Windows v.2000 build 751 (Vesta Services inc., USA) and an in-house developed Microsoft Excel® VBA procedures for process automation. Both the *in silico* molecular descriptors and the binary descriptors for experimental protocol and tissue characterization were included as model inputs. $K_{t:b}$ values (log transformed) were introduced in the ANN as the model output. In order to allow the calculation of the relative relevance of the used molecular descriptors, all networks were built using normalised variables both at the input and output, and a sigmoid transfer function was used in all connections. Early stopping of training was used in order to prevent over fitting (20). For that, 25% of the train molecules were randomly selected to act as a sub-set for the internal testing of the model, not effectively used in the regression process, and ANN train was performed until degradation on the RMSE for the internal test data was observed. Additionally, each network was started 20 times, with random initial values and different sub-sets of the internal test cases, to minimize training convergence to local minima.

Network optimization was performed in a three-step process. The first step consisted of the reduction of the molecular descriptors space. This was done by removing highly correlated ($r>0.90$) molecular descriptors, allowing the removal of the redundant ones, which contained information already within another descriptor. The second step consisted of the optimisation of the network structure for the most relevant molecular descriptors. Several ANN's structures were tested, varying the number of hidden layers (up to three) and the number of hidden neurons in order to obtain a ratio between the number of patterns (832 $\log K_{t:b}$ values effectively used in the training) to the number of connections above 1. Finally, after the optimization of the selected structures, and based on the RMS of the internal test data, the 10 best ones were selected in order to obtain the final QSAR ANNE model. A MS Excel workbook (Tiblisi - Tissue to Blood drug partition in-silico prediction) was developed and made available in the Supplementary material that allows the calculation of the $K_{t:b}$ values.

ANNE Validation

External validation was done by comparing the values predicted by the ANNE to the observed $K_{t:b}$ values of the drugs in the external validation group of data, not previously used in the training and optimization process. In addition, validation was also tested by assuming that rat and human $K_{t:b}$ values are similar and evaluating the ability to predict the Human blood Vss in the 532 drugs of the Vss Validation Data by including the tissue $K_{t:b}$ values calculated by the ANNE models in the following Physiologically Based equation,

$$V_{ss}^{human} = V_{blood}^{human} + \sum_{i=1}^{n}\left(K_{t:b(i)}^{Rat} \times V_{tissue(i)}^{human}\right) + K_{t:b(average)}^{Rat} \times V_{tissue(remaining)}^{human} \quad (1)$$

where V_{blood}^{human} is the volume of blood in the human (5.53 L), $K_{t:b(i)}^{Rat}$ are the different tissue-to-blood distribution coefficients in the rat calculated for each drug, $V_{tissue(i)}^{human}$ are the corresponding tissue physiological volumes according to Brown at al. (21) and $K_{t:b(average)}^{Rat}$ is the average of the available $K_{t:b(i)}^{Rat}$ values weighted by the corresponding $V_{tissue(i)}^{human}$ values. Finally $V_{tissue(remaining)}^{human}$ is the remaining tissue value not described by the available sum of $V_{tissue(i)}^{human}$, that considering the 13 tissues described by the ANNE model only accounts for 3.8% of the total Human volume of 79.3 L.

Applicability Domain

In order to try to establish the applicability domain, the accuracy and variability of predicted values under the ANNE were calculated. This was performed by predicting each case in both the train and external validation data sets using the 10 best ANN models, one at a time. Average and standard deviation (SD) of each predicted case (molecule) were then calculated. In order to see if the applicability domain could be defined based on the variability of the predictions, drugs in the train and test sets were divided into 8 classes of increasing ANNE prediction SD, and the average squared error of prediction $\left(APE = \sum\left(\overline{Y_i} - y_i\right)^2/n\right)$ was calculated within each such block.

In addition, we also studied the effect of this prediction variability on the ability to correctly predict the Human blood Vss in the 532 drugs of the Vss Validation Data. In this case, and because the final Vss value is the result of multiple tissue predictions with different tissue weights according to Eq. 1), the individual prediction errors were included in the variability of the final determination as follows:. since each tissue $\log K_{t:b}$ can be characterized as $a \pm \sigma_a$, the standard deviation on $K_{t:b}$ was calculated as $\sigma_b = 2.303 \times \sigma_a \times 10^a$. By consequence, since Vss is determined by Eq. 1, the final standard deviation value, due to the error propagation, can be approximated to $\sigma_{Vss} = \sqrt{\sum_{i=1}^{13}\left(V_{tissue(i)}^{Human} \times \sigma_b\right)^2}$. This allows the estimation of the variability coefficient of the Vss prediction as $CV(\%) = 100 \times \sigma_{Vss}/Vss$. Again, data was sorted by ascending values of CV(%), divided into 8 classes, and again the APE was calculated within each such class.

RESULTS

ANNE Optimization

Optimization of the ANN model was done as described under methods. Regarding the reduction of the molecular descriptors space, the identification of correlated descriptors allowed the removal of four molecular descriptors, namely the number of H-bond donors and acceptors, the molar refractivity and total topological polar surface area. The final inputs included in the modelling process are described in Table II, and included a set of 8 molecular descriptors, 1 binary discriminator for the experimental protocol (*in vitro* vs *in vivo*) and 4 binary discriminators for the different tissues.

Based on these descriptors, a network structure optimization was pursued using a brute force approach. In order to maintain a practical computational time, the network architectural space was swept between 1 and 3 hidden layers. The number of hidden neurons by layer was also changed taking into consideration the ratio between the number of patterns to the number of connections. This ratio was maintained above 1 in order to reduce the ability of the network to memorise the data and avoid over-fitting (22,23). Since a large number of structures were possible (>1500) and in order to make the computational time reasonable, a random sub-set of 100 structures was selected for optimization. Each structure was randomly started 20 times, to identify convergence to local minima, and trained until degradation on the RMSE for the internal test data was observed to avoid over-fitting. The overall performance of the optimization procedure can be seen on Fig. 1. The 10 best individual networks in terms of the lowest test RMSE were kept. These were grouped in order to provide the final predictions of $LogK_{t:b}$, (Fig. 2, Table III) allowing also the determination of the prediction variability

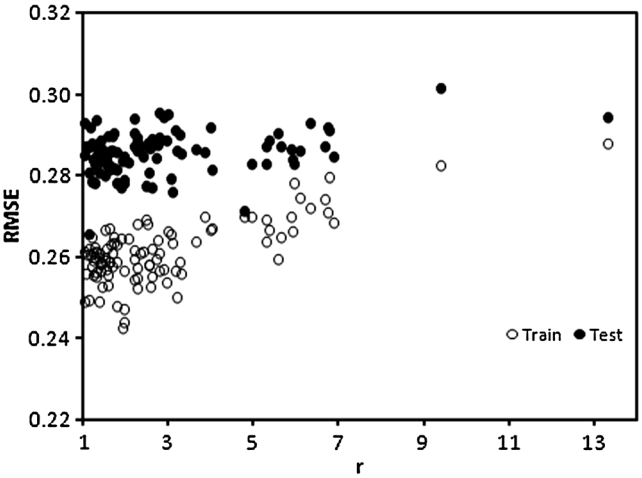

Fig. 1 Performance, measured in terms of the train and internal test RMSE, for the 100 individual ANN during the optimization procedure. Complexity of each individual ANN is described by the ratio between the number of patterns to the number of connections (r).

itself. Similar statistics are observed when data is grouped by experimental protocol or by different tissues, indicating that the overall adjustment was not unbalanced within the different data origins (table S1 and S2 in supplementary material). It is also observed that, parallel to the increase in the prediction variability, an incremental trend is also obtained in the APE (Fig. 3), indicating that the ANNE variability in the predictions could serve as a rationale for the establishment of the applicability domain. Based on this trend, a RMSE of around 0.3 was observed for a SD cut-off value of 0.2 and the overall statistics within these two groups of data is also presented in Table III. Results can be replicated by using the Tiblisi Excel VBA macro, available in the paper Supplementary material. ANNE results and individual ANN

Table II Summary of the 13 Inputs Used in the ANNE Model, Including 8 Molecular Descriptors and 5 Binary Selectors

No.	Inputs		Description
	Type	Name	
1	Molecular descriptors	pKa (basic)	pKa of the strongest basic group
2		pKa (acid)	pKa of the strongest acid group
3		MW	Molecular weight
4		Ui	Unsaturation índex
5		Hy	Hydrophilicity índex
6		TPSA(NO)	Topological Polar Surface Area (including N and O)
7		logP	ALOGPS 2.1 calculated LogP
8		logS	ALOGPS 2.1 calculated intrinsic solubility
9	Binary selectors	IV/IV	Binary selector for *in vitro* (0) or *in vivo* (1) original data
10		1	Binary Selectors for the different rat tissues
11		2	
12		3	
13		4	

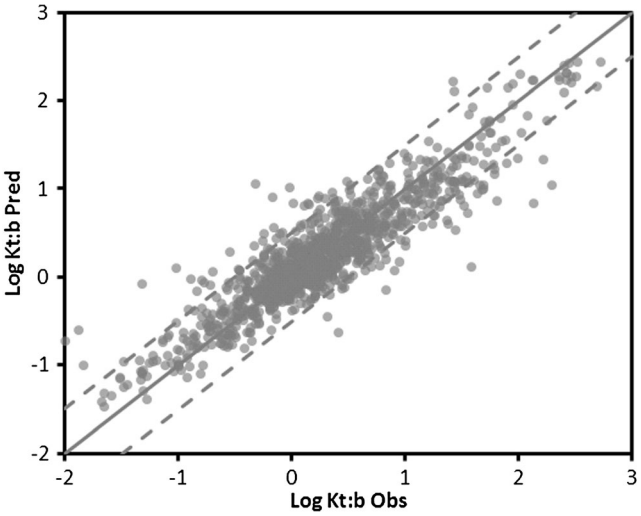

Fig. 2 Plot of the *in vitro/in vivo* observed Log $K_{t:b}$ *vs. in silico* predicted Log $K_{t:b}$ values for the ANNE model on the training dataset. *Solid line* represents the line of unity and the dashed-lines the ±3-fold tolerance value.

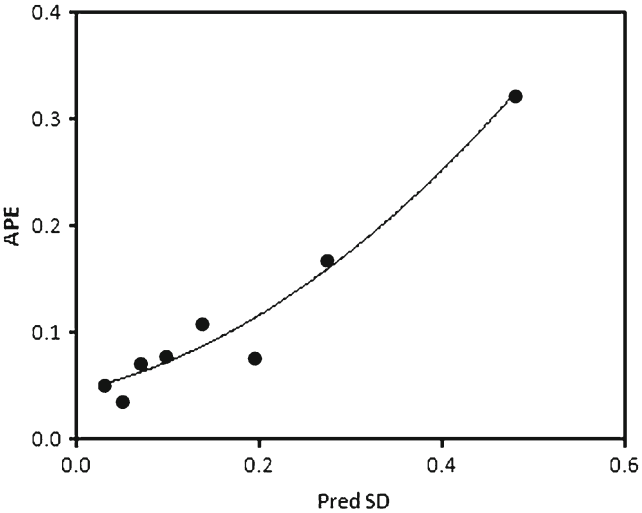

Fig. 3 Plot of the average squared error of prediction (APE) against the average SD on the training dataset divided into 8 classes of increasing ANNE prediction SD.

structures are also made available in the supporting excel data file on the Supplementary material to the paper.

ANNE Validation and Applicability Domain

The ability of the ANNE to predict the Log $K_{t:b}$ for new cases was initially tested in the 369 $K_{t:b}$ values of the external validation data not used in the training process. As can be seen (Table III and Fig. 4) good agreement was observed with the statistics of the train data, indicating that no significant over-fitting was obtained. Again, it was observed an increase in the APE (Fig. 5) following the increase in the ANNE prediction variability, as measured in terms of SD. Splitting the data based on the SD cut-off value of 0.2 resulted in a degradation of the prediction statistics (Table III), indicating that this could be a rationale for the establishment of the applicability domain.

An additional external validation was done by testing the ability to predict the blood Vss for another 532 drugs of the Vss Validation Data. Predicted values for the 13 different rat tissues, assuming an *in vivo* experimental protocol (ANN input #9=1), were obtained by the ANNE as well as the SD for each tissue. These values were used to predict the blood Vss according to Eq. 1) as well as the CV% associated to each Vss prediction. The prediction results, from the total data and by groups of similar CV%, are presented in Table IV and Fig. 6. Figure 7 also presents the variation in APE with the increase in the CV(%) of the predictions itself. As can be seen, more accurate results are obtained when the estimated CV% presents lower values, which again indicates that the variability of the predicted values from the ANNE can be used to evaluate the applicability domain of the QSAR model.

Table III Statistical Evaluation of the Performance of the ANNE Model to Predict Log $K_{t:b}$ Values for the Train and External Validation Data

	Train data			External validation data		
	Total	SD < 0.2	SD > 0.2	Total	SD < 0.2	SD > 0.2
r	0.909	0.911	0.886	0.896	0.899	0.859
ME	0.004	0.004	0.000	−0.023	−0.033	0.021
RMSE	0.304	0.294	0.375	0.322	0.307	0.380
n	1098	978	120	362	291	71

r correlation coefficient; *ME* mean error; *RMSE* root mean square error; *n* number of predicted values

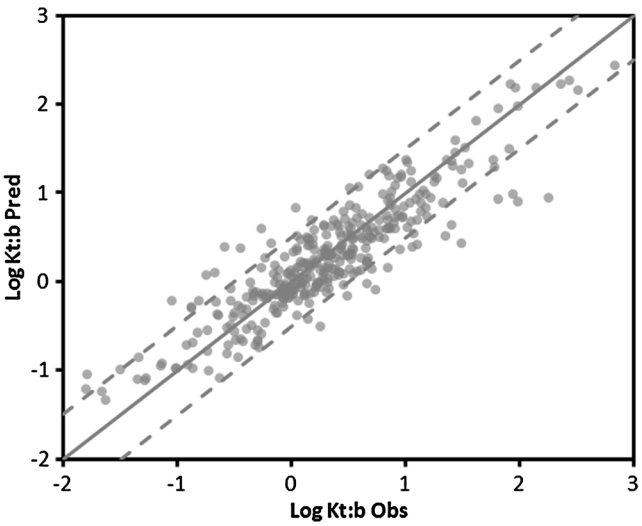

Fig. 4 Plot of the *in vitro/in vivo* observed Log $K_{t:b}$ *vs. in silico* predicted Log $K_{t:b}$ values for the ANNE model on the external validation dataset. *Solid line* represents the line of unity and the *dashed-lines* the ±3-fold tolerance value.

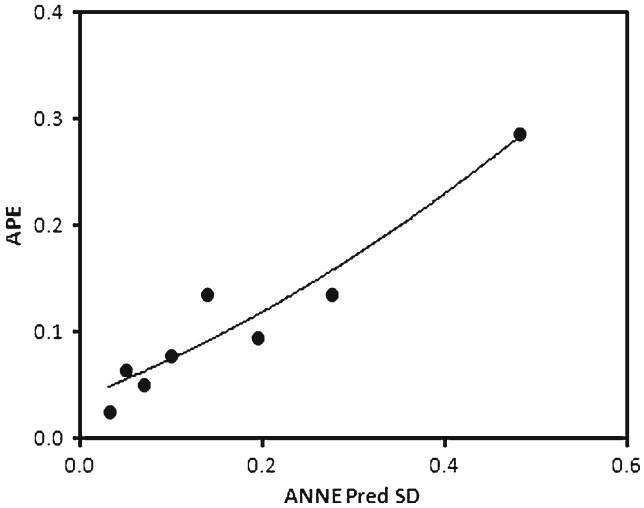

Fig. 5 Plot of the average squared error of prediction (APE) against the average SD on the external validation dataset divided into 8 classes of increasing ANNE prediction SD.

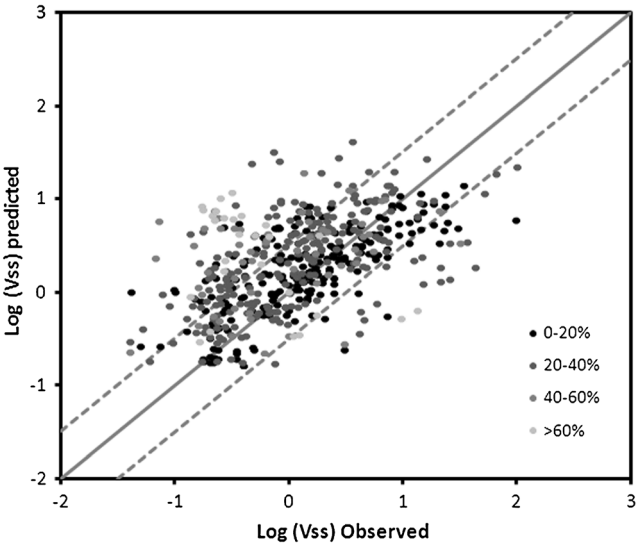

Fig. 6 Plot of the *in vivo* observed *vs. in silico* PBPK prediction of the human Log Vss (L/kg) values for the ANNE model on the Vss validation dataset. Data is grouped in terms of variability (CV%) on the predictions. *Solid line* represents the line of unity and the dashed-lines the ± 3-fold tolerance value (a colour version of figure 6 is provided in the supplementary material as figure S3).

DISCUSSION

There has been a hot debate about the utility of QSAR models during drug development, since sometimes a lack of predictability is observed (13,24–26). Data quality and training sample size are known to greatly influence the reliability of the QSAR models (13). Proper methodologies during model development are also very important and it has been proposed, in the case of ANN, that some basic issues should be addressed with caution, namely (i) Had the network converged to the global minimum or a local minimum? (ii) Was there a quantifiable metric to describe the network's "memory" or data retention? (iii) Was the network making use of the right set of input parameters for the problem domain? (iv) How will the network handle situations when presented data is outside the training set domain or unique from previous training data? (27,28)

We have previously presented a compilation of rat $K_{t:b}$ data obtained both *in vitro* and *in vivo* in 13 different tissues for a total of 309 different drugs (14). The number of individual

Table IV Statistical Evaluation of the Performance of the ANNE Model to Predict the Human Log Vss (L/Kg) on the Vss Validation Dataset, by Using an *in silico* – PBPK Approach. Data is Grouped in Terms of Variability on the Predictions and Compared to the Performance Reported Using a Similar *in vitro/in vivo* – PBPK Approach

	Observed[a]		Predicted[b]				
	In vitro	*In vivo*	Total	CV%			
				0–20	20–40	40–60	> 60
% < 2-fold	56.7	62.1	43.0	55.6	38.4	28.6	11.1
% < 3-fold	70.0	79.3	61.7	74.7	57.6	50.0	14.8
% < 5-fold	80.0	88.5	75.9	87.1	75.4	57.1	29.6
% < 10-fold	93.3	94.3	89.3	96.4	90.2	80.4	51.9
ME	0.088	0.181	0.247	0.111	0.275	0.470	0.692
RMSE	0.484	0.429	0.604	0.458	0.597	0.779	1.106
r	0.648	0.768	0.532	0.687	0.590	0.360	−0.441
n	30	87	532	225	224	56	27

% < (2,3,5,10)-*fold* percentage of drugs well predicted within the defined fold error; *ME* mean error; *RMSE* root mean square error; *r* correlation coefficient; *n* number of predicted values

[a] Vss predictions based on experimental Kt:b as reported on (14)

[b] Vss predictions based on ANNE - predicted Kt:b

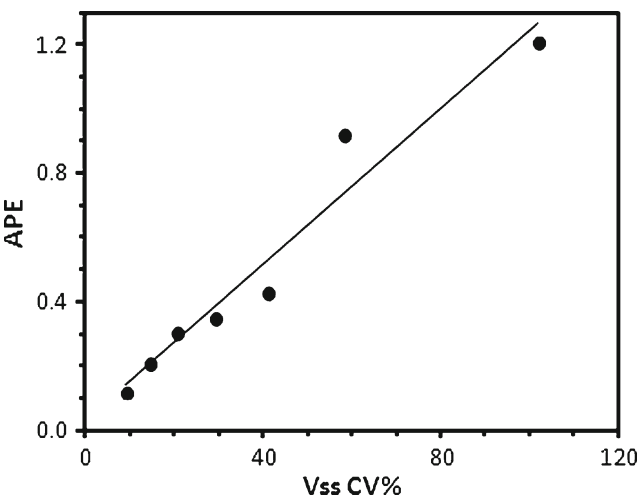

Fig. 7 Plot of the average squared error of prediction (APE) against the average variability (CV%) of the predicted *in silico* PBPK human Log Vss (L/kg) values on the Vss validation dataset divided into 8 classes of increasing PBPK prediction CV%.

$K_{t:b}$ values varied from 253 cases in the muscle to 19 in the stomach, totalling 1474 $K_{t:b}$ determinations in all the 13 tissues. These values were obtained both by using *in vitro* tissue homogenates (657 $K_{t:b}$ values) or *in vivo* PK experiments (817 $K_{t:b}$ values). It was also observed that there were statistically significant differences between tissues and experimental protocols (14). Abraham *et al.* (12) reported that although statistically significant differences between *in vitro* and *in vivo* distribution of VOC from air to muscle were observed in their data, they were able to combine the VOC and drug data by incorporating a new indicator variable in their QSAR models, defined as 1 for the *in vitro* data and 0 for the *in vivo* data. This resulted in a final QSAR model with similar statistics to the individual models using separated datasets, but with increased applicability domain. We had also observed distribution differences between experimental protocols and rat tissues, which could hinder the development of QSAR models for some tissues, where data was limited in number implying that a small applicability domain would be obtained. In order to overcome this issue, since the same "general" distribution characteristic was being quantified in the different protocols/tissues and several similarities were expected to exist between them, we combined the total available data by including additional binary variables to classify the data origin itself. For the tissues, and since these totalled 13 different tissues, 4 binary variables were used (Table I). This approach allowed to model these 26 different combinations using a reduced set of only 5 inputs, with obvious gains in terms of network complexity and computational time.

ANNE Optimization, Validation and Applicability Domain

Optimization of the neural networks was done as reported in methods. Data was initially divided into a train (75%) and an external validation (25%) dataset. Since the creation of the external validation group is known to be of paramount importance in the validation of the QSAR methods (26), the creation of these two groups of data was made by randomly selected cases in a 75/25 ratio within each individual tissue, in order to obtain a balanced distribution of data between the two datasets. During the model optimization procedure, and in each individual training step, the train data was again divided randomly into an actual train dataset and an internal test dataset again in a 75/25 ratio, aiming at every individual $K_{t:b}$ value being used in the training procedure at least once. This was done in order to allow the early-stop of the optimization, by training until a degradation of the RMSE in the internal validation data was observed. It is assumed that the generalization error decreases in an early period of training, reaches a minimum and then increases as training goes on, while the training error monotonically decreases. Therefore, in early stopping, it is considered better to stop training at an adequate time (29). However, the real situation is a lot more complex with generalization curves having almost always more than one local minimum (30). Due to this fact, each network was run for an excessive number of iterations, and the iteration that resulted in the lowest residual mean square error (RMSE) of the testing group was kept. In addition, each of the selected ANN structures was started 20 times with random initial values, in order to sweep the parameters space and avoid convergence to local minima. As can be seen in Fig. 1, RMSE of the train datasets, within each individual structure, after optimization varied between 0.242 and 0.287. The same RMSE of the internal validation datasets varied between 0.265 and 0.301. For the ANNE, a group of 10 structures with the lowest RMSE of the internal validation datasets (ranging from 0.265 to 0.280) were finally selected. Early stop has shown to be an effective procedure to decrease the generalization error when the number of training cases is similar to the number of modified parameters (29). Similarity between the RMSE of the test and internal validation groups seems to confirm this and indicates that indeed overtraining and memorization was avoided (Fig. 1). The small range of optimal RMSE between the different structures also seems to indicate that local minima were avoided by the use of the different initial optimization parameter values.

The performance of the ANNE is presented in Table III and Figs. 2 and 4. As can be seen, statistics of the total train and external test groups of data are very similar, with RMSE around 0.3 and ME around 0. It has previously been estimated that the inter-laboratory standard deviation on brain $K_{t:b}$ would be around 0.20–0.25 log units (15). Based on this, and according to Kramer *et al.* (31), the maximum Pearson correlation coefficient achievable would be around 0.93–0.96, which indicates that our models are not over-fitted and very close to the best possible fit for the error in the data. Similar statistics are observed when data is evaluated separately by experimental protocol or by individual tissues (tables S1 and S2 in supplementary material).

Probably the simplest approach for an initial indication on the applicability domain is looking at the range of the individual descriptors used in the model building procedure. This has, however, several limitations especially when non-linear relationships are involved. ANNs, due to their flexible nature, are able to address complex and non-linear relationships between the input variables in order to relate them to the modelled output, resulting in low bias models. It is also known that an over-trained ANN may be prone to memorization and inability to predict new cases, resulting in high variance models. Although our training procedure consisted in a bias/variance trade-off methodology, it is expected that extrapolations (model predictions for drugs that are significantly different from the ones used in the training) would result in significantly different output results within the optimized individual networks and, as a rule of thumb, the larger the

standard deviation of the predictions, the greater the chance that the molecule is outside the applicability domain, making the prediction less reliable. As such, and in order to establish an improved method for the applicability domain determination, an approach based on the standard deviation of the ensemble model prediction (32) was pursued. As can be seen in Figs. 3 and 5 referring to both the train and external test groups, the increase in the SD of the ANNE output results allowed differentiating compounds with larger prediction errors.

ANNE Interpretation

The input parameters used in the ANN were empirically selected based on previous studies on drug distribution in tissues (6,14,33) and are presented in Table II. The ionization of drugs, characterized by both the pKa of the strongest acid and basic groups, is known to influence the drug's tissue distribution. A different distribution mechanism between strong bases and the remaining types of molecules has been described (6,7). It has also been shown that the extent of ionization is inversely related to the drug distribution in the tissues, by influencing the drug lipophilicity (33). The Molecular weight (MW), has been shown to have a positive contribution when extrapolating V_{ss} values from animal to human (34), but it is also known to have an unfavourable impact, for example, on the distribution of drugs through the blood–brain barrier (35). In our database, MW was highly correlated with the molar refractivity (AMR). The unsaturation index (Ui) represents the presence of multiple bonds in molecules (36). It is also a representation of the flexibility of the molecules, and influences the paracellular passage across membranes (37). The hydrophilicity index (Hy) is a simple empirical index related to hydrophilicity of compounds based on count descriptors (36). In our database it was highly correlated with the number of H-donor bonds (nHdon). Hydrogen-bonding causes an association of molecules forming large aggregates of single molecules. In addition it has consequences on the solubilization of molecules both in water and in lipids. Topological polar surface area (tPSA) is defined as the part of the surface area of the molecule associated with oxygen, nitrogen and the hydrogen bonded to any of these atoms. This surface descriptor is also related to the hydrogen-bonding ability of compounds, and it has been suggested that PSA, in some way, describes the desolvation of a compound as it moves from an aqueous to a lipid environment (38). It is frequently used as a molecular descriptor for permeability in biological membranes (39), which may indicate that for high PSA values, drugs tend to remain solvated with water and not to be able to cross the cell membranes. In our database (14), tPSA was highly correlated with the total topological polar surface area and with the number of H-bond acceptors (nDacc) in the molecules, as expected. LogP is the calculated measure of the partitioning of a compound between a lipid (1-octanol) and an aqueous phase that depends on solute bulk, polar and hydrogen-bonding effects for a molecule in its neutral form (18). The importance of LogP is well described in tissue distribution and is on the basis of several physiological methods (9) being, by itself, a good descriptor for distribution in the adipose tissue (5). Finally LogS, the calculated intrinsic solubility, refers to the ability for a given substance to dissolve in water (40). Since the distribution in tissues depends on equilibrium between aqueous, lipid and other cellular components, it was also included in the calculations.

Due to the "black-box" nature of ANN models, its interpretation is frequently described as difficult, but one possible approach to obviate this drawback is to find how input trends influence the output predictions, by determining the relative importance of each input on that particular output. This was done by varying each input at a time, considering all the others constant with the median value for the corresponding descriptor, and evaluating the change in the output response. As can be seen in Fig. 8, where the relative importance of each input is depicted for each individual tissue, different trends of the molecular descriptors are seen in the distribution for the different rat tissues. This analysis also shows that there is no single descriptor able to characterize the individual tissues distribution, as pointed out previously (14). This seems to confirm the initial considerations that ionization, lipophilicity, hydrogen bonding potential and size are all important determinants of drug distribution.

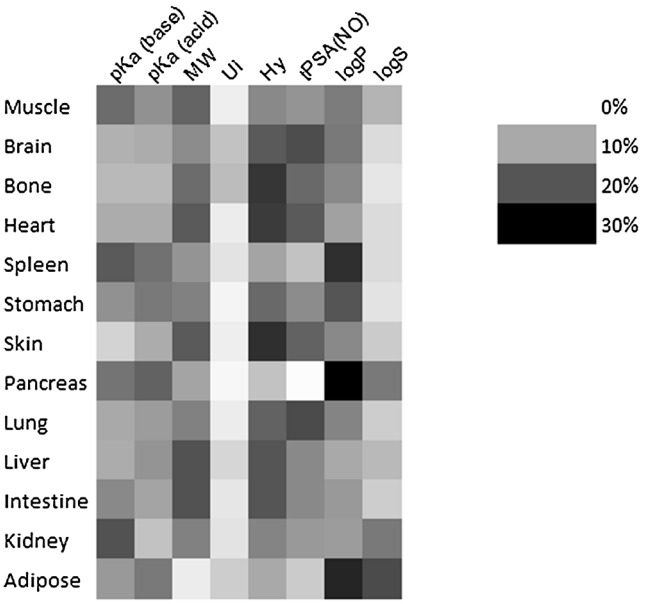

Fig. 8 Relevance of each molecular descriptor in the prediction of log $K_{t:b}$ by the ANNE, obtained by cycling each input for all training patterns and computing the effect on the network's output response at a time, for each individual tissue.

Prediction of Vss in Humans

For an additional evaluation of the predictability of the model, predictions on the various tissues were included in a PBPK model for estimation of the blood volume of distribution in an additional group of 532 drugs not previously used in the model building procedure. Its results are presented in Table IV and Fig. 6. Although different drugs were considered, statistics reported for the same approach but based on *in vitro* or *in vivo* derived $K_{t:b}$ data (14) are also included in Table IV for comparison purposes. As can be seen, QSAR prediction statistics including all available cases are worse than the ones reported for the experimental derived data. However, it is again visible that prediction performance degrades when the model predictions increase their internal variability (Fig. 7). As such, when the CV% of the Vss predictions is below 20%, it is possible to see (Table IV) that the predictive statistics are superior to the ones observed by using $K_{t:b}$ values obtained by *in vitro* experiments with tissue homogenates and almost as good as the prediction on Vss using $K_{t:b}$ values determined by *in vivo* PK rat studies (14). These predictions clearly degrade for CV% larger than 20% reaching a point of absolute lack of predictability for CV% larger than 60%, where the majority of the drugs are large to very large molecules (MW > 500) and/or with high TPSA values (>180A). These were sparse in the training set of data, indicating again that the ANNE predicted variability can be used as an indicator of the applicability domain.

Because the assumptions underlying the QSAR model or PBPK modelling do not take into consideration specific PK features, such as active transport and/or extensive blood distribution, eight drugs presenting higher than 10-fold errors when the CV% of the Vss predictions is below 20% (expectably inside the applicability domain), were further considered on a closer analysis, as follows: (1) for 3 drugs (diazoxide, propylthiouracil and azapropazone), we see that probably a high Rb value was predicted with consequences on the estimation of the *in vivo* blood Volume of distribution. On the contrary, for cladribine, and since this drug is rapidly taken up and phosphorylated by lymphoid and myeloid cells (41), Rb may have been under predicted. (2) Hydroxychloroquine is also a drug with high blood distribution, which can be a confounding factor in the overall model. (3) Two other cases, tamsulosin and 7-hydroxystaurosporine, are drugs that are highly bound to plasma but mainly to α1-acid glycoprotein. Our ANNE model is based on total blood-to-tissue ratios and since α1-acid glycoprotein is only 1 to 3% of the total plasma protein and susceptible to saturation and displacement, its relevance on the drug distribution may not have been well captured by the QSAR model itself. (4) Finally, rosuvastatin is a OATP1B1 substrate and inhibition of its uptake in the liver results in the reduction of the Volume of distribution (42). As stated, active transport processes were not considered, either in the QSAR model or in the PBPK modelling for Vss, as a distribution mechanism. Even so, and although these expected limitations exist, the ANNE model is capable to produce good estimations on the rat tissue distribution when inside its applicability domain.

In a recent report, Jones *et al.* (9) presented an evaluation on the ability for various types of methods to predict the human Vss. Among others, they focused on whole-body physiologically based models. These were all variants of the equations originally presented by Poulin and Theil (5), Rodgers *et al.* (6), and Rodgers and Rowland (7) and dispense with the need for *in vivo* data by estimating the extent of tissue distribution from the physicochemical and *in vitro* plasma protein binding characteristics of the compound. By testing a group of 18 blinded compounds, they concluded that the variation from Berezhkovskiy (8) produced the best results with an r of 0.83, a RMSE of 0.61. Predictions of the observed Vss were 89, 72 and 61% percent of the times correctly estimated within a tenfold, threefold, and twofold error, respectively. These results are very close to the ones obtained with our QSAR model when inside its applicability domain, with the obvious advantage that, since our ANNE relies only on calculated molecular descriptors, it is useable in the early pre-clinical development phase of drug discovery. These results may be further improved if more data on rat $K_{t:b}$ values is made available and used in future training procedures.

REFERENCES

1. Poulin P, Jones RD, Jones HM, Gibson CR, Rowland M, Chien JY, *et al.* PHRMA CPCDC initiative on predictive models of human pharmacokinetics, part 5: Prediction of plasma concentration-time profiles in human by using the physiologically-based pharmacokinetic modeling approach. J Pharm Sci. 2011;100(10):4127–57.
2. Rowland M, Peck C, Tucker G. Physiologically-based pharmacokinetics in drug development and regulatory science. Annu Rev Pharmacol Toxicol. 2011;51:45–73.
3. Poulin P, Theil FP. Prediction of pharmacokinetics prior to in vivo studies. 1. Mechanism-based prediction of volume of distribution. J Pharm Sci. 2002;91(1):129–56.
4. Lin JH, Sugiyama Y, Awazu S, Hanano M. In vitro and in vivo evaluation of the tissue-to-blood partition coefficient for physiological pharmacokinetic models. J Pharmacokinet Biopharm. 1982;10(6):637–47.
5. Poulin P, Schoenlein K, Theil FP. Prediction of adipose tissue: plasma partition coefficients for structurally unrelated drugs. J Pharm Sci. 2001;90(4):436–47.
6. Rodgers T, Leahy D, Rowland M. Physiologically based pharmacokinetic modeling 1: predicting the tissue distribution of moderate-to-strong bases. J Pharm Sci. 2005;94(6):1259–76.
7. Rodgers T, Rowland M. Physiologically based pharmacokinetic modelling 2: predicting the tissue distribution of acids, very weak bases, neutrals and zwitterions. J Pharm Sci. 2006;95(6):1238–57.

8. Berezhkovskiy LM. Volume of distribution at steady state for a linear pharmacokinetic system with peripheral elimination. J Pharm Sci. 2004;93(6):1628–40.
9. Jones RD, Jones HM, Rowland M, Gibson CR, Yates JW, Chien JY, *et al.* PhRMA CPCDC initiative on predictive models of human pharmacokinetics, part 2: comparative assessment of prediction methods of human volume of distribution. J Pharm Sci. 2011. doi: 10.1002/jps.22553.
10. Nestorov IS, Hadjitodorov ST, Petrov I, Rowland M. Empirical versus mechanistic modelling: comparison of an artificial neural network to a mechanistically based model for quantitative structure pharmacokinetic relationships of a homologous series of barbiturates. AAPS PharmSci. 1999;1(4):E17.
11. Abraham MH, Ibrahim A, Acree Jr WE. Air to blood distribution of volatile organic compounds: a linear free energy analysis. Chem Res Toxicol. 2005;18(5):904–11.
12. Abraham MH, Ibrahim A, Acree Jr WE. Air to muscle and blood/plasma to muscle distribution of volatile organic compounds and drugs: linear free energy analyses. Chem Res Toxicol. 2006;19(6):801–8.
13. Stouch TR, Kenyon JR, Johnson SR, Chen XQ, Doweyko A, Li Y. In silico ADME/Tox: why models fail. J Comput Aided Mol Des. 2003;17(2–4):83–92.
14. Paixao P, Aniceto N, Gouveia LF, Morais JA. Tissue-to-blood distribution coefficients in the rat: utility for estimation of the volume of distribution in man. Eur J Pharm Sci. 2013;50(3–4):526–43.
15. Abraham MH, Ibrahim A, Zhao Y, Acree Jr WE. A data base for partition of volatile organic compounds and drugs from blood/plasma/serum to brain, and an LFER analysis of the data. J Pharm Sci. 2006;95(10):2091–100.
16. Obach RS, Lombardo F, Waters NJ. Trend analysis of a database of intravenous pharmacokinetic parameters in humans for 670 drug compounds. Drug Metab Dispos. 2008;36(7):1385–405.
17. Paixao P, Gouveia LF, Morais JA. Prediction of drug distribution within blood. Eur J Pharm Sci. 2009;36(4–5):544–54.
18. Tetko IV, Bruneau P. Application of ALOGPS to predict 1-octanol/water distribution coefficients, logP, and logD, of AstraZeneca in-house database. J Pharm Sci. 2004;93(12):3103–10.
19. Tetko IV, Gasteiger J, Todeschini R, Mauri A, Livingstone D, Ertl P, *et al.* Virtual computational chemistry laboratory–design and description. J Comput Aided Mol Des. 2005;19(6):453–63.
20. Cataltepe Z, Abu-Mostafa YS, Magdon-Ismail M. No free lunch for early stopping. Neural Comput. 1999;11(4):995–1009.
21. Brown RP, Delp MD, Lindstedt SL, Rhomberg LR, Beliles RP. Physiological parameter values for physiologically based pharmacokinetic models. Toxicol Ind Health. 1997;13(4):407–84.
22. So SS, Richards WG. Application of neural networks: quantitative structure-activity relationships of the derivatives of 2,4-diamino-5-(substituted-benzyl)pyrimidines as DHFR inhibitors. J Med Chem. 1992;35(17):3201–7.
23. Turner JV, Maddalena DJ, Cutler DJ. Pharmacokinetic parameter prediction from drug structure using artificial neural networks. Int J Pharm. 2004;270(1–2):209–19.
24. Johnson SR. The trouble with QSAR (or how I learned to stop worrying and embrace fallacy). J Chem Inf Model. 2008;48(1):25–6.
25. Doweyko AM. QSAR: dead or alive? J Comput Aided Mol Des. 2008;22(2):81–9.
26. Huang J, Fan X. Why QSAR fails: an empirical evaluation using conventional computational approach. Mol Pharm. 2011;8(2):600–8.
27. Gramatica P. On the development and validation of QSAR models. Methods Mol Biol. 2013;930:499–526.
28. Taylor BJ, Darrah MA, Pullum LL, Ammar K, Smith JT, Skias ST, et al. Methods and procedures for the verification and validation of artificial neural networks. Taylor BJ, editor. Springer; 2006.
29. Amari S, Murata N, Muller KR, Finke M, Yang HH. Asymptotic statistical theory of overtraining and cross-validation. IEEE Trans Neural Netw. 1997;8(5):985–96.
30. Prechelt L. Automatic early stopping using cross validation: quantifying the criteria. Neural Netw. 1998;11(4):761–7.
31. Kramer C, Kalliokoski T, Gedeck P, Vulpetti A. The experimental uncertainty of heterogeneous public K(i) data. J Med Chem. 2012;55(11):5165–73.
32. Tetko IV, Bruneau P, Mewes HW, Rohrer DC, Poda GI. Can we estimate the accuracy of ADME-Tox predictions? Drug Discov Today. 2006;11(15–16):700–7.
33. Poulin P, Theil FP. A priori prediction of tissue:plasma partition coefficients of drugs to facilitate the use of physiologically-based pharmacokinetic models in drug discovery. J Pharm Sci. 2000;89(1):16–35.
34. Sui X, Sun J, Li H, Pan Y, Wang Y, He Z. Contribution of molecular properties to extrapolation of the volume of distribution in human from preclinical animal species data. Biopharm Drug Dispos. 2010;31(8–9):464–75.
35. Hou TJ, Xu XJ. ADME evaluation in drug discovery. 3. Modeling blood-brain barrier partitioning using simple molecular descriptors. J Chem Inf Comput Sci. 2003;43(6):2137–52.
36. Todeschini R, Consonni V. In: Mannhold R, Kubinyi H, Folkers G, editors. Molecular descriptors for chemoinformatics. 2nd ed. Weinheim: WILEY-VCH; 2009.
37. Paixao P, Gouveia LF, Morais J. Prediction of the human oral bioavailability by using in vitro and in silico drug related parameters in a physiologically based absorption model. Int J Pharm. 2012;429:84–98.
38. van de Waterbeemd H, Camenisch G, Folkers G, Chretien JR, Raevsky OA. Estimation of blood-brain barrier crossing of drugs using molecular size and shape, and H-bonding descriptors. J Drug Target. 1998;6(2):151–65.
39. Paixao P, Gouveia LF, Morais JA. Prediction of the in vitro permeability determined in Caco-2 cells by using Artificial Neural Networks. Eur J Pharm Sci. 2010;41(1):107–17.
40. Tetko IV, Tanchuk VY, Kasheva TN, Villa AE. Estimation of aqueous solubility of chemical compounds using E-state indices. J Chem Inf Comput Sci. 2001;41(6):1488–93.
41. Albertioni F, Lindemalm S, Reichelova V, Pettersson B, Eriksson S, Juliusson G, *et al.* Pharmacokinetics of cladribine in plasma and its 5′-monophosphate and 5′-triphosphate in leukemic cells of patients with chronic lymphocytic leukemia. Clin Cancer Res. 1998;4(3):653–8.
42. Grover A, Benet LZ. Effects of drug transporters on volume of distribution. AAPS J. 2009;11(2):260–1.