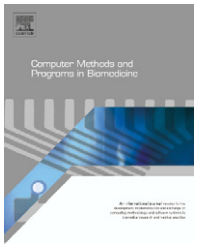
journal homepage: www.intl.elsevierhealth.com/journals/cmpb



A prediction model for blood–brain barrier permeation and analysis on its parameter biologically

Yan Chen[a,b], *Qing-Jun Zhu*[b,*], *Jian Pan*[a], *Yi Yang*[a], *Xiao-Pei Wu*[c]

[a] *Engineering Research Center of Bio-Process, Hefei University of Technology, Ministry of Education, Hefei 230009, China*
[b] *School of Life Science, Anhui University, Hefei 230039, China*
[c] *School of Computer Science and Technology, Anhui University, Hefei 230039, China*

ARTICLE INFO

Article history:
Received 20 August 2008
Received in revised form
20 January 2009
Accepted 13 March 2009

Keywords:
LogBB
Blood–brain barrier permeability
Neural network models
Sensitivity analysis

ABSTRACT

The objective of this paper is to build a reliable model based on the artificial neural network (ANN) for predicting the blood–brain barrier (BBB) permeability and reveal the effects of the molecular descriptor on the BBB permeability. Eight descriptors including high-affinity P-gp substrate probability and plasma protein binding ratio are selected to develop the model. The three layers feedforward neural network (8-5-1) is employed for the prediction of logBB. By analyzing the experimental results, polar surface area (PSA) seems to be the most important factor for BBB permeability. Different from traditional view, the Abraham's hydrogen-bond basicity (HBB) can make a positive contribution to logBB in rational range. The experimental results show that the ANN based model with eight selected descriptors as inputs can achieve good performance for logBB prediction, and the results of sensitivity analysis can be confirmed by the present biological and chemical research.

1. Introduction

The prerequisite to cure neurological disorders is that the drug distribution in central nervous system (CNS) can reach effectively therapeutic concentrations. Blood–brain barrier (BBB) as shield not only maintains the homeostasis of the CNS, but also refuses many potentially important diagnostic and therapeutic agents from entering into the brain. Thus, the ability of drug permeating across BBB becomes critical in the development of new medicines, especially in the design of new drugs which are active in brain tissue. The high BBB penetration is needed for drugs that activate in brain, while low BBB penetration is needed for drugs responsible for peripheral tissues.

Although the experimental analysis of drug permeability is essential, the procedure of experiment is time consuming and complicated.

A theoretical model can always give predictions. There are various methodologies to estimate the distribution of drug in the brain, such as brain uptake index, brain perfusion, blood–brain distribution and so on. The advantages and disadvantages of each technique were thoroughly discussed in Bickel's review [1].

Although some computational models for predicting BBB permeability are based on logPS [2,3], so far most predictive models focus on blood–brain distribution, logBB, which is defined as the logarithm of the brain/blood concentration ratios at steady-statement expressed as below

$$\log BB = \log\left(\frac{C_{brain}}{C_{blood}}\right)$$

In 1988 Young and co-workers found a good correlation between logBB and ΔlogP [4]. van de Waterbeemd and Kansy also found a good correlation with logBB using polar surface

* *Corresponding author*. Tel.: +86 551 5108636.
E-mail address: qingjunzhu@yahoo.com (Q.-J. Zhu).
0169-2607/$ – see front matter © 2009 Published by Elsevier Ireland Ltd.
doi:10.1016/j.cmpb.2009.03.006

area and molecular volume [5]. Afterward, Abraham and coworkers expanded the dataset and introduced the descriptors of the general solvation equation of Abraham into the prediction model [6]. For the expanded dataset, the model they built has a good regression coefficient [7,8]. Lombardo and coworkers used solvation energy (in water) as the only parameter based on Abraham's data set [9]. Many other researchers also made contributions to the model establishments summarized in the review [10].

The development of a reliable predictive model requires more rational and more precise descriptors to reflect molecular properties. Regardless of other atomic and group influences, the number of H-bond acceptors and donors are used to represent the ability to generate hydrogen bond [11,12]. The overall hydrogen-bond acidity and hydrogen-bond basicity suggested by Abraham reflect the ability to generate hydrogen bond more reasonably, but they were usually determined by experiments [13,14]. Thus, the application of these descriptors in the theoretical models was limited.

By the development of chemical software, more and more information about chemical structures can be obtained. Abraham solvation parameters including overall hydrogen-bond acidity and overall hydrogen-bond basicity can be gained easily by ADME Boxes v 4.0 now. The protein binding possibility of the compound can be calculated. For example, the P-gp substrate probability can be calculated by ADME Boxes v 4.0. Bioactivity of drug with GPCR ligands, kinase inhibitors, ion channel modulators and nuclear receptors can be obtained by Molinspiration Drug-Likeness Score V2007.

The factors which influence a drug's distribution in the body are complicated. It was shown that drug disposition was affected by not only its physicochemical properties, but also its biological factors. Feng suggested that four factors including three biological factors determine the steady-statement concentration of drugs in the brain [15]: free concentration of drug in plasma, efflux from brain, metabolic modification by barrier enzymes and permeability of drug through BBB. Previous models were based on the assumption of passive diffusion. Garg and Verma established a new model introducing P-gp substrate probability into the model to reflect the active transport phenomenon [12]. Multiple linear regression (MLR) and partial least square were the common computational methods that researches used. Some models were built on artificial neural network recently [12,16,17].

Different from the logBB predictive model which focus on mathematical analysis, a reliable model based on the artificial neural network (ANN) is developed in this paper, and then how the factors affect BBB permeability is revealed.

2. Methods

2.1. Dataset description and ANN model training

In the paper, dataset is composed of logBB as output of ANN and eight descriptors as inputs of ANN.

The experimental logBB values of 145 molecules have been collected from the published papers [10,12]. The 14 descriptors initially include Abraham's hydrogen-bond acidity (HBA), Abraham's hydrogen-bond basicity (HBB), high affinity P-gp substrate probability (P-gp (H)), LogP, molar refraction, molecular volume, molecular weight, number of rotatable bonds (NRB), plasma protein binding ratio (CSPB), polar surface area (PSA), polarizability, surface area, the energy of the highest occupied molecular orbital (EHOMO), and the energy of the lowest unoccupied molecular orbital (ELUMO). After correlation coefficient analysis among the 14 descriptors, 8 descriptors including HBA, HBB, P-gp (H), LogP, Volume ($Å^3$/100), NRB, CSPB, and PSA ($Å^2$/100) with the corresponding logBB values are selected finally to build dataset.

Then the ANN model (8-5-1) with one hidden layer is built. The eight descriptors and the corresponding logBB value are as inputs and output of ANN model, respectively. For the model training, the dataset are randomly divided into training set ($n = 125$) and testing set ($n = 20$). The transfer function of hidden layer and out layer is selected as tanhAxon.

After inspecting the training results, four compounds including gentisic acid, thioridazine, phenserine, and mesoridazine are removed for the remarkable difference between observed logBB and predicted logBB, which can be seen in Table 1. Here, it is mentioned that all of these compounds are CNS drugs that might have other importing or exporting channel beyond consideration in this paper.

Then, the rest 141 compounds are randomized. 21 compounds are selected as testing set, and 120 compounds are used with sixfold cross-validation. As the number of epochs reaches 1000 or the MSE of the cross-validation set begins to increase, the training will be stopped.

The structure of molecule is drawn using MDL ISIS Draw 2.5. NRB, PSA, P-gp(H) are calculated by using ADME Boxes v 4.0 (Trial Version on www.pharma-algorithms.com). The CSPB is obtained by using ChemSilico Property Prediction Software (free version on www.chemsilico.com). Molecular geometry is optimized based on AM1 methods by using Hyperchem 7.52 Evaluation Version. Then, the rest descriptors are calculated. Smiles notations of molecules are calculated on www.vcclab.org by submitting the molecular structure.

Table 1 – The removed compounds.

Compound	Volume	logP	PSA	P-gp(H)	CSPB	HBA	HBB	NRB	logBB (observed)	logBB (predicted)
Gentisic acid	1.243	1.180	0.778	0.005	0.834	1.200	1.090	1.000	0.080	−0.996
Thioridazine	3.508	4.180	0.065	0.101	0.969	0.000	1.130	4.000	0.240	1.322
Phenserine	3.211	4.190	0.448	0.164	0.839	0.370	1.580	4.000	1.000	−0.144
Mesoridazine	3.557	3.050	0.236	0.266	0.934	0.000	1.950	4.000	−0.360	1.232