ISSN 1607-6729, Doklady Biochemistry and Biophysics, 2016, Vol. 470, pp. 371–374. © Pleiades Publishing, Ltd., 2016.
Original Russian Text © A.S. Dyabina, E.V. Radchenko, V.A. Palyulin, N.S. Zefirov, 2016, published in Doklady Akademii Nauk, 2016, Vol. 470, No. 6, pp. 720–723.

BIOCHEMISTRY, BIOPHYSICS, AND MOLECULAR BIOLOGY

Prediction of Blood-Brain Barrier Permeability of Organic Compounds

A. S. Dyabina[a], E. V. Radchenko[a, b, *], V. A. Palyulin[a, b, **], and Academician of the RAS N. S. Zefirov[a, b]

Received April 29, 2016

Abstract—Using fragmental descriptors and artificial neural networks, a predictive model of the relationship between the structure of organic compounds and their blood-brain barrier permeability was constructed and the structural factors affecting the readiness of this penetration were analyzed. This model ($N = 529$, $Q^2 = 0.82$, $RMSE_{cv} = 0.32$) surpasses the previously published models in terms of the prediction accuracy and the applicability domain and can be used for the optimization of the pharmacokinetic parameters during drug development.

DOI: 10.1134/S1607672916050173

The penetration of drugs and other physiologically active compounds through the blood-brain barrier (BBB) is among the most important pharmacokinetic processes that determine their bioavailability in the central nervous system and have a crucial influence on the efficacy, pharmacological profile, mode of use, and safety of drugs. For the compounds acting on the central nervous system, such penetration must be maximized; however, it should be minimized for the peripherally acting drugs to avoid side effects [1, 2]. Despite the significant advances in the development of efficient and minimally invasive experimental methods for evaluation of penetration of compounds through the BBB, the need for reliable methods for its prediction *in silico* for new structures during the search for and optimization of lead compounds remains relevant. Although numerous studies with the use of chemoinformatics techniques were devoted to this issue [3, 4], the models presented in the literature do not have a sufficiently broad applicability domain and/or high accuracy for predicting the penetration of diverse organic compounds.

In view of above, the aim of this study was to build more reliable predictive models of quantitative relationship between the structure of potential drugs and their ability to penetrate through the blood-brain barrier, as well as to analyze the structural factors affecting the readiness of this penetration. Preliminary results of the study were presented earlier [5].

During this work, on the basis of analysis of numerous publications, we prepared an extensive database characterizing the ability of organic compounds to penetrate the BBB, which was represented as the logarithm of the effective coefficient of distribution of compounds between the brain and blood $LogBB = \lg(C_{brain}/C_{blood})$. The database was built using the Instant JChem software package [6]. In total, the training set comprised 529 compounds of various chemical classes and covered a broad range of *LogBB* values. Apparently, it is the most complete and accurate among the sets containing open quantitative data described in the literature.

To build the models for the relationship between the structure of compounds and the *LogBB* parameter, we used the method of artificial neural networks implemented in the NASAWin 2.0 software [7, 8]. The molecular structure was described using fragmental substructural descriptors that reflect the occurrence number of various types of fragments containing up to 10 atoms (paths, cycles, and branched fragments with a multilevel classification of atoms taking into account their type, valence state, set of bonds, and the number of attached hydrogen atoms) in the structure of the compound [8, 9]. Rare fragments occurring in only one or two compounds were excluded from the analysis. From the resulting set comprising several thousand descriptors, 60 to 160 most significant descriptors were then selected using the stepwise multiple linear regression. When building a neural network model, we considered networks of the three-layer architecture with a different number of neurons in the hidden layer. The 5 × 4-fold double cross-validation procedure was employed, which uses independent subsets of data for

[a] *Department of Chemistry, Lomonosov Moscow State University, Moscow, 119991 Russia*
[b] *Institute of Physiologically Active Compounds, Russian Academy of Sciences, Severnyi proezd 1, Chernogolovka, Moscow oblast, 142432 Russia*
**e-mail: genie@qsar.chem.msu.ru,*
***e-mail: vap@qsar.chem.msu.ru*

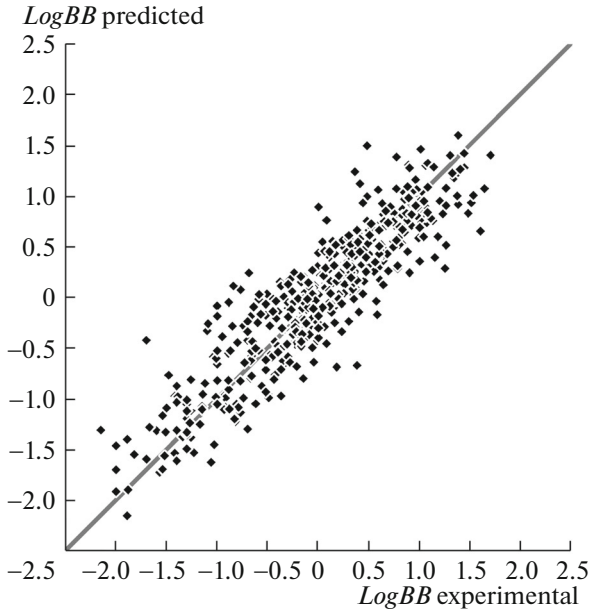

Fig. 1. Correspondence of experimental values of BBB permeability and the values predicted during cross-validation for the compounds of the training set.

neural network training, selection of its optimal configuration, and evaluation of the model predictivity [10]. Since, for the majority of compounds, sufficiently complete and accurate data on the role of various passive and active transport mechanisms in their transport through the BBB and the binding to blood plasma proteins and brain tissues are missing [1], these parameters were not explicitly included in the model; however, the related patterns in the effect of structure on the BBB penetration should be taken into account indirectly by the neural network-based fragmental model.

Table 1. Characteristics of neural network models of the relationship between the structure of organic compounds and their ability to penetrate BBB

Fragment size	N_d	N_h	Q^2	$RMSE_{cv}$
5	120	4	0.78	0.35
6	160	6	0.81	0.33
7	120	6	0.80	0.34
8	160	2	0.80	0.33
9	160	3	0.82	0.32
10	120	2	0.76	0.36

Models were built on the basis of a training set comprising 529 compounds of different classes. Designations: N_d—number of descriptors, N_h—number of neurons in the hidden layer of the neural network architecture, Q^2—cross-validation parameter, $RMSE_{cv}$—root mean square error for cross-validation.

The characteristics of the optimal models built using different subsets of fragmental descriptors are presented in Table 1. It can be seen that the best model is based on the fragments comprising up to nine atoms. In its statistical parameters (Q^2 = 0.82, $RMSE_{cv}$ = 0.32), it is comparable to the most reliable models available in the literature or even surpass them [4, 11], and the average prediction error is close to the error of experimental determination of *LogBB* (0.3 log units [12]). It should be noted that the significantly better representativeness of the training set ensures a broader applicability domain of the model, covering more diverse compounds. The comparison of the experimental *LogBB* values with the values predicted as a result of double cross-validation procedure also confirms the high prediction accuracy for the vast majority of compounds (Fig. 1).

To further validate the model, we used a test set comprising 2053 compounds with qualitative (classification) assessment of the BBB penetration [13]. Two classes of compounds were considered: penetrating (BBB+) and non-penetrating (BBB–) the BBB. Among these compounds, only 247 (12%) were present in our training set. Similarly to [13], the quantitative *LogBB* values predicted by the neural network model were converted to a qualitative scale using the threshold value *LogBB* = –1, corresponding to the penetration of approximately 10% of a compound into the brain. Table 2 shows the classification confusion matrices as well as the parameters calculated using these matrices—overall recognition accuracy, sensitivity (accuracy of detection of penetrating compounds), and specificity (accuracy of detection of non-penetrating compounds) for the full test set and its subset not overlapping with the training set. As can be seen, the constructed model correctly recognized more than 90% of BBB-penetrating compounds in the independent test set. For the BBB– class, the quality of recognition (specificity) was lower. This may be partly due to the fact that the classification of many compounds into this class was based on the indirect data on the absence of observed activity with respect to the central nervous system, which can be explained not only by the absence of BBB permeability but also by other factors. Parameters for the full and non-overlapping test sets were very close, providing additional confirmation of the reliability and broad applicability of our model and allowing its confident use for predicting the penetration of new compounds through the BBB.

For the purposes of qualitative analysis and interpretation of the effect of different descriptors on the endpoint parameter values predicted using the neural network model, we used the numerically determined values of partial derivatives of this parameter with respect to the descriptors averaged over the training set [14]. To take into account the differences in descriptor variation ranges, we calculated the normalized relative response of the model:

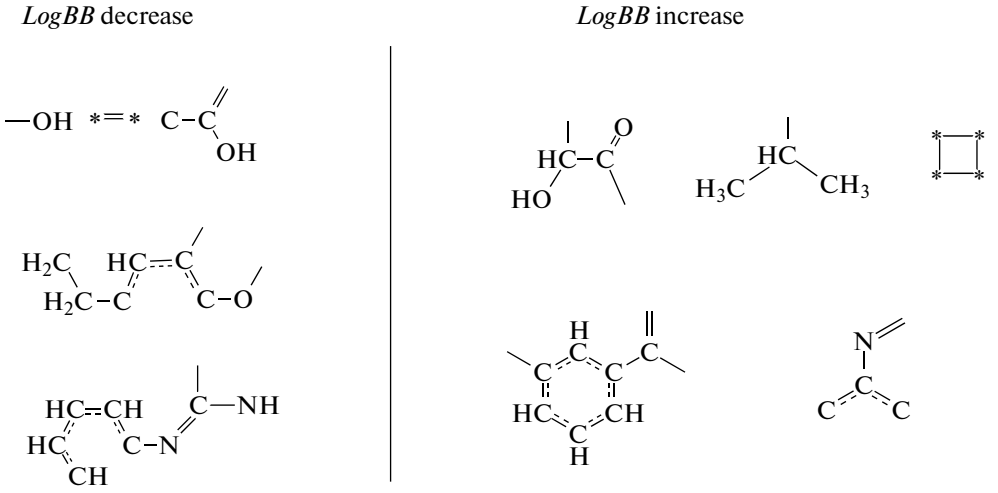

Fig. 2. Fragments having the strongest negative and positive effect on the predicted value of BBB permeability of compounds.

$$RR_i = \text{avg}\left(\frac{\partial y}{\partial x_i}\right) \cdot \frac{\text{range}(x_i)}{\text{range}(y)},$$

where $\text{avg}\left(\frac{\partial y}{\partial x_i}\right)$ is the mean value of the derivative of the endpoint parameter y with respect to descriptor x_i, range(x_i) is the variation range of descriptor x_i, and range(y) is the variation range of parameter y.

The fragments that have the strongest positive and negative effects on the *LogBB* value are shown in Fig. 2. Many of them allow a simple interpretation, which is consistent with the known concepts of the structural characteristics that affect the BBB permeability of compounds [12]. For example, permeability decreases if the structure contains strongly polar groups (hydroxyl, carboxyl, and guanidine), all other conditions being equal. At the same time, the hydrophobic fragments (alkyl or aromatic) increase the permeability of compounds. A significant positive contribution of a fragment containing the hydroxyl group in α-position to the carbonyl group apparently reflects the role of active transport of substituted monocarboxylic acids, which ensures their good permeability, despite the fact that the passive transport of these compounds is hampered by the presence of polar groups in the molecule. Nevertheless, in analyzing the most significant contributions of fragments, it should be borne in mind that the model also includes a large number of other fragmental descriptors affecting the predicted value of BBB permeability. For this reason, to optimize the pharmacokinetic properties of a drug, it is expedient to use a virtual screening-based approach.

Table 2. Confusion matrices and classification quality parameters for the BBB penetration of the test set compounds

Parameter	Prediction, full set		Prediction, non-overlapping set	
Experiment	BBB+	BBB–	BBB+	BBB–
BBB+	1464	105	1255	100
BBB–	281	203	259	192
Total	2053		1806	
Overall accuracy	81%		80%	
Sensitivity	93%		93%	
Specificity	42%		43%	

Our model for predicting the penetration of drug compounds through the blood-brain barrier is included in the integrated web service for predicting ADMET parameters of drugs, which is being developed in the Laboratory of Medicinal Chemistry, Department of Chemistry, Lomonosov Moscow State University (http://qsar.chem.msu.ru/admet/) [15].

Thus, using the fragmental descriptors and artificial neural networks, we have built a predictive model of the relationship between the structure of organic compounds and their ability to penetrate the blood-brain barrier as well as analyzed some structural factors affecting the readiness of penetration. This model is superior in the prediction accuracy and the applicability domain to the models previously published in the literature and can be used for the optimization of the pharmacokinetic parameters during drug development.

ACKNOWLEDGMENTS

We are grateful to the ChemAxon company for providing licenses for the structural data management software. This study was supported by the Ministry of Education and Science of the Russian Federation (project no. RFMEFI60414X0144).

REFERENCES

1. Di, L., Rong, H., and Feng, B., *J. Med. Chem.*, 2013, vol. 56, no. 1, pp. 2–12.
2. Wager, T.T., Liras, J.L., Mente, S., and Trapa, P., *Expert Opin. Drug Metab. Toxicol.*, 2012, vol. 8, no. 5, pp. 531–542.
3. Garg, P., Verma, J., and Roy, N., *In silico* modeling for blood-brain barrier permeability predictions, in *Drug Absorption Studies*, New York: Springer, 2008, pp. 510–556.
4. Lanevskij, K., Japertas, P., and Didziapetris, R., *Expert Opin. Drug Metab. Toxicol.*, 2013, vol. 9, no. 4, pp. 473–486.
5. Palyulin, V.A., Dyabina, A.S., Radchenko, E.V., and Zefirov, N.S., in *ISMC EFMC XXII Intern. Symp. on Medicinal Chemistry, Berlin, September 2–6, 2012*, Berlin, 2012, pp. 374–375.
6. Instant JChem 16.4.11, ChemAxon Ltd., 2016. http://www.chemaxon.com.
7. Baskin I.I., Halberstam N.M., Artemenko N.V., Palyulin V.A., Zefirov N.S., in: *EuroQSAR 2002 Designing Drugs and Crop Protectants: Processes, Problems and Solutions*, Oxford: Blackwell, 2003, pp. 260–263.
8. Artemenko, N.V., Baskin, I.I., Palyulin, V.A., and Zefirov, N.S., *Russ. Chem. Bull.*, 2003, vol. 52, no. 1, pp. 20–29. *Izv. Akad. Nauk, Ser. Khim.*, 2003, vol. 52, pp. 19–28.
9. Zefirov, N.S. and Palyulin, V.A., *J. Chem. Inf. Comput. Sci.*, 2002, vol. 42, no. 5, pp. 1112–1122.
10. Zhokhova, N.I., Baskin, I.I., Palyulin, V.A., Zefirov, A.N., and Zefirov, N.S., *Dokl. Chem.,* 2007, vol. 417, no. 2, pp. 282–284.
11. Muehlbacher, M., Spitzer, G.M., Liedl, K.R., and Kornhuber, J., *J. Comput.-Aided Mol. Des.*, 2011, vol. 25, no. 12, pp. 1095–1106.
12. Clark, D.E., *Drug Discov. Today*, 2003, vol. 8, no. 20, pp. 927–933.
13. Martins, I.F., Teixeira, A.L., Pinheiro, L., and Falcao, A.O., *J. Chem. Inf. Mod.*, 2012, vol. 52, no. 6, pp. 1686–1697.
14. Baskin, I.I., Ait, A.O., Halberstam, N.M., Palyulin, V.A., and Zefirov, N.S., *SAR QSAR Environ. Res.*, 2002, vol. 13, no. 1, pp. 35–41.
15. Radchenko, E.V., Karpov, P.V., Sosnin, S.B., Palyulin, V.A., and Zefirov, N.S., in *XXII Ros. nats. kongress "Chelovek i lekarstvo," Moskva, 6-10 aprelya 2015* (XXII Russ. Natl. Congr. "Man and Drug," Moscow, April 6–10, 2015), Moscow, 2015, p. 250.

Translated by M. Batrukova