MULTI-TARGET NEURAL NETWORK MODEL OF ANXIOLYTIC ACTIVITY OF CHEMICAL COMPOUNDS USING CORRELATION CONVOLUTION OF MULTIPLE DOCKING ENERGY SPECTRA

P.M. Vassiliev*, M.A. Perfilev, A.V. Golubeva, A.N. Kochetkov, D.V. Maltsev

Volgograd State Medical University,
1 Pavshikh bortsov sq., Volgograd, 400131 Russia; *e-mail: pvassiliev@mail.ru

Anxiety disorders are one of the most common mental health pathologies in the world. They require search and development of novel effective pharmacologically active substances. Thus, the development of new approaches to the search for anxiolytic substances by artificial intelligence methods is an important area of modern bioinformatics and pharmacology. In this work, a multi-target model of the dependence of the anxiolytic activity of chemical compounds on their integral affinity to relevant target proteins based on the correlation convolution of multiple docking energy spectra has been constructed using the method of artificial neural networks. The training set of the structure and activity of 537 known anxiolytic substances was formed on the basis of the previously created database, and optimized 3D models of these compounds were built. 22 biotargets presumably relevant to anxiolytic activity were identified and their valid 3D models were found. For each biotarget, 27 multiple docking spaces have been formed throughout its entire volume. Multiple ensemble molecular docking of 537 known anxiolytic compounds into all spaces of relevant target proteins has been performed. The correlation convolution of the calculated energy spectra of multiple docking was carried out. Using seven training options based on artificial multilayer perceptron neural networks, the multi-target model of depending anxiolytic activity chemical compounds on 22 parameters of the correlation convolution of the multiple docking spectra energy was constructed. The predictive ability of the created model was characterized Acc = 91.2% and AUC_{ROC} = 94.4%, with statistical significance of $p < 1\times10^{-15}$. The found model is currently used in the search for new substances with high anxiolytic activity.

Key words: anxiolytic activity; multi-target affinity; multiple molecular docking; docking energy spectra; correlation convolution; artificial neural networks

DOI: 10.18097/PBMC20247006428

INTRODUCTION

According to WHO data for 2023, almost 1 billion people worldwide suffer from mental disorders [1], of these the majority are anxiety disorders. According to Rosstat data for 2023 [2], there are approximately 3.5 million people in Russia suffering from anxiety disorders. Anxiety disorders are characterized by excessive worry, fear, and avoidance behavior that complicate the daily lives of patients and limit their social activity and professional opportunities. All this leads to an intensification of the search for new anxiolytic drugs.

Machine learning methods have long been successfully used in the *in silico* search for compounds with pharmacological activity [3], and the technology of artificial neural networks (collectively called artificial intelligence methods) has proved especially effective in this regard [4]. However, in PubMed we were unable to find a single publication devoted to the *in silico* search for substances with anxiolytic activity using artificial intelligence methods. This is probably due to the fact that anxiolytic effects are of a complex systemic multi-target nature and are caused by the effect of compounds on a set of a sufficiently large number of relevant biotargets.

Thus, the development of new approaches to the search for anxiolytic substances using artificial intelligence methods is one of the current areas of modern bioinformatics and pharmacology.

Our previous work [5] has shown that the spectrum of docking energies of chemical compounds into many spaces of relevant protein is a much more reliable metric of the affinity of ligands to biotargets, in comparison with the unit energy of their docking into the specific site. Taking this into account, we built a based on multiple docking neural network model of the dependence of the anxiolytic activity of chemical compounds on the energy spectrum of their multiple docking in the $GABA_A$ receptor, based on multiple docking and the integral affinity of the predicted compounds for only one biotarget [6].

The aim of this work is to develop a more universal multi-target neural network model of depending anxiolytic activity chemical compounds on the parameters of the correlation convolution of the multiple docking energy spectra.

To achieve this aim, it was necessary to solve the following objectives.

1. To create the training set on the structure and level of anxiolytic activity of known compounds.

2. To identify the optimized 3D models of these compounds.

3. To reveal biotargets putatively relevant to anxiolytic activity.

4. To find the valid 3D models of target proteins relevant to anxiolytic activity.

5. To create the spaces for multiple docking across the entire volume of each validated 3D model of relevant target proteins.

6. To conduct the multiple ensemble molecular docking of known anxiolytic compounds into all spaces of all validated 3D models of relevant target proteins.

7. To perform the correlation convolution of the calculated multiple docking energy spectra.

8. To train the neural networks on the obtained convolutional variables and to create the multi-target neural network model of the dependence of the anxiolytic activity of chemical compounds on the parameters of the correlation convolution of the multiple docking energy spectra.

METHODS

The Training Set on the Structure and Level of Anxiolytic Activity of Known Compounds

The training set was formed on the basis of the original verified database [7] on the structure and anxiolytic activity of 537 known experimentally studied substances with verified structural formulas and a unified level of anxiolytic activity. It included 273 compounds with pronounced anxiolytic activity and 264 low or inactive compounds.

Optimized 3D Models of Compounds

Optimized 3D models of 537 compounds were constructed sequentially by the molecular mechanics method in the MarvinSketch 15.6.15 program [8] and then by the semi-empirical quantum chemical PM7 method in the MOPAC2012 program [9]. First, 10 conformers with the lowest energy were built separately for each compound in MarvinSketch. Then, all the built conformers were optimized in MOPAC2012 and for each compound among its optimized conformers, one conformer with the lowest total energy was selected. In total, 10749 conformers were processed according to this scheme, as a result, 537 optimized 3D models of the compounds of the training set were built.

Biotargets Relevant to Anxiolytic Activity

The Open Targets system [10] yielded a list of 2057 biotargets associated with anxiety disorders based on text mining of existing publications. From the original QSAR database of the Microcosm BioS system 20.6.6 [11] a list of 2697 human biotargets with experimental data on various types of targeted activity was taken. By crossing these two lists, a list of 92 human biotargets that would be putatively relevant to anxiolytic activity confirmed by corresponding experimental data on the activity of compounds was obtained. Using 273 compounds with pronounced anxiolytic activity from the original database [7] using the original IT Microcosm 7.3 [12] and Microcosm BioS 20.6.6 [11] systems, 10 known structurally similar compounds experimentally studied for this type of target activity were found for each selected biotarget by the structural similarity method using the QL-modified Tanimoto coefficient. Based on the data obtained, 92 average indicators of the level of target activity Ind were calculated, the range of changes from Ind = +5 (very high) to Ind = -5 (inactive); Ind = 0 corresponds to average activity. 22 biotargets with Ind values ≥ 1 were selected as the most likely relevant to the pronounced level of anxiolytic activity.

Valid 3D Models of Relevant Biotargets

For 22 possible relevant biotargets, 277 quality 3D models were selected from the PDBe (https://www.ebi.ac.uk/pdbe/) and RCSB PDB (https://www.rcsb.org/) X-ray structural 3D models. The quality criteria were: 1) the maximum length of the simulated amino acid sequence; 2) high resolution; 3) the minimum number of fragments. Among these 277 3D models, according to the methodology described in [13], 22 valid 3D models were identified, one for each biotarget.

Spaces for Multiple Molecular Docking

For each valid 3D model of each relevant biotarget, 27 spaces for multiple molecular docking were constructed using the original MSite 21.04.22 program and the algorithm described in [5], covering the entire volume of a given target protein.

Multiple Ensemble Molecular Docking of Known Anxiolytic Compounds

Ensemble docking was performed according to the method described in [13] and using the AutoDock Vina 1.1.1 program [14]; there were five replicates for each ligand, each time in ten biologically active conformations. The 50 obtained values were used for calculation determination of the five minimum binding energies ΔE. As a result, for each compound, $22 \times 27 \times 5 = 22 \times 135 = 2970$ values of ΔE, reflecting the multi-target multiple affinity of this compound for 22 relevant target proteins were obtained. In total, the full multi-target multiple affinity matrix of 537 compounds of the training set included 1,594,890 docking energy values.

Correlation Convolution of Multiple Docking Energy Spectra

For each biotarget, the 135 values of the minimum binding energies ΔE obtained as a result of multiple

docking can be represented as a fully connected neural network with a symmetric matrix of connections. These 135 parameters act as neurons and are interconnected because they are calculated for the same target protein and the interaction of a ligand with one of the 27 non-overlapping docking regions will lead to conformational changes in the protein and will influence the interaction of this ligand with another docking region. The dependencies between them in a fully connected neural network are assumed to be linear and can be represented by the values of the pair correlation coefficients. In a correlation fully connected neural network, these coefficients are the weights of synapses (interneuronal connections).

Taking into account these conditions, for each biotarget, the convolution parameter of the energy spectrum of multiple docking is the energy W of the fully connected neural correlation network

$$W_l = \frac{1}{2}\sum_{\substack{i,j=1 \\ i \neq j}}^{M} R_{ij} \times \Delta E_{il} \times \Delta E_{jl}\,,\ l=1...N \qquad (1),$$

where R_{ij} – Pearson correlation coefficient between energies ΔE_i and ΔE_j, i≠j;

ΔE_{il} – value of energy i for compound l, $l=1...N$;

ΔE_{jl} – value of energy j for compound l, $l=1...N$;

M – number of energy values for convolution, it is 135;

N – number of compounds (ligands).

The following should be clarified. The values of the minimum binding energies are almost always ΔE << -1, and the Pearson correlation coefficient R varies from -1 to +1. Therefore, the value of W calculated by formula (1) will always be positive.

Thus, as a result of the convolution, the multi-target multiple affinity of each compound for the relevant biotargets was represented by 22 convolution variables.

Training of Neural Networks

The classification training set required for neural network modeling included 31 indicators: 1) compound codes; 2) graded values of anxiolytic activity of compounds; 3) 22 convolutional variables; 4) seven sampling variables. The level of anxiolytic activity of the compounds was indicated by the labels hm (high or moderate) and nhm (low or no activity). Convolutional variables were calculated using formula (1) and reflected the multi-target multiple affinity of each compound for 22 relevant biotargets. Seven sampling variables were specified the options for forming the training, testing, and validation subsets in a ratio of 5:1:1 and were used to build different versions of the neural networks.

The neural networks were trained using the Statistica 7 program [15] according to the scheme described in [13]. According to the Kolmogorov theorem [16], a dependence of any complexity can be approximated using a two-layer artificial neural network. In this case, it is desirable to ensure the convolution of input neuron signals into a smaller number of intermediate patterns. Therefore, a two-layer perceptron MLP k–m–2 with a bottleneck was chosen as the neural network architecture. Here k — the number of input neurons, in this case 22; m — the number of hidden neurons, set by the program from 3 to 20, since 2<m<k. When constructing classification neural networks, it is advisable to use cross entropy as an error function [17]. In this case, the activation function for the output neurons is the multidimensional logistic function, and for hidden neurons, the activation functions can be linear, logistic, hyperbolic tangent, or exponential.

Neural networks were trained using the back propagation error algorithm, by iterating through four different activation functions for the hidden layer of neurons and using seven sampling options.

The training was carried out in two stages. At the first stage, 4000 networks were trained for each sampling option, with automatic selection of the 200 best networks. From these networks, the top 5 were manually selected according to the set of characteristics of the accuracy of training, testing and validation. Of these, the best one was manually selected based on the prediction accuracy indicators on the combined set and the results of ROC analysis. At the second stage, this neural network was subjected to further manual training with a fixed architecture, with the building of 200 neural networks, from which the program selected the 20 best ones. Of these, the best one was manually selected based on the prediction accuracy indicators on the combined set and the results of ROC analysis.

In total, about 30,000 neural networks were trained on 7 sampling variants and 14 neural networks with the best accuracy were found. Of these, the best one was selected based on the prediction accuracy indicators on the combined set and the results of ROC analysis.

As integral indicators of accuracy for all the best neural networks on the combined set, the general prediction accuracy Acc, sensitivity Sens (prediction accuracy of active compounds), specificity Spec (prediction accuracy of inactive compounds) were calculated, and the area under the curve AUC was calculated based on the ROC analysis data.

All calculations were performed on a supercomputers of hybrid architecture with a total peak performance of ~40 Tflops.

RESULTS AND DISCUSSION

As part of the preliminary data preparation, 537 optimized 3D models of compounds from the original verified database [7], containing

information on the structure and anxiolytic activity of known experimentally studied substances, were constructed.

According to the method described above, 22 biotargets relevant to the anxiolytic activity of chemical compounds were found, and 22 valid 3D models of these proteins were identified (Table 1).

It is noteworthy that the number of relevant biotargets identified computationally included such "classic" proteins for the manifestation of anxiolytic activity as the $GABA_A$ receptor, serotonin 5-HT_{1A} and 5-HT_{2A} receptors, α_{2A} and α_{2B} adrenergic receptors, and the glutamate NMDA receptor. This indicates the relevance and validity of the methodology used to identify relevant biotargets.

The summarized results of the correlation convolution of the energy spectra of multiple ensemble molecular docking of 537 known anxiolytic compounds into valid 3D models of 22 relevant biotargets are shown in Table 2.

As can be seen from Table 2, for all biotargets, the mean values of convolutional variables in the class of compounds with pronounced activity with very high statistical significance exceed the mean values of convolutional variables in the class of compounds with low activity. This proves that the correlation convolution of the energy spectra of multiple ensemble molecular docking in the form of the energy of a fully connected neural network calculated using formula (1) is a highly reliable metric of the affinity of ligands to biotargets.

Table 3 shows the architecture and accuracy indicators of the best neural network obtained for pronounced anxiolytic activity during iterative learning in two steps on seven sampling variants.

According to the set of accuracy indicators, the found neural network is statistically very highly significant: for five indicators out of seven, its significance is $p < 1\times10^{-15}$, and for two $p = 0.000170$ and $p = 0.0165$ according to the binomial test [18].

Thus, using the example of anxiolytic activity, the adequacy and high validity of the following newly developed methods have been demonstrated: 1) identification of biotargets potentially relevant to the activity studied; 2) multiple molecular docking, which does not require the presence of specific binding sites in biotargets; 3) correlation convolution of multiple docking energy spectra based on the calculation of energies of a fully connected symmetric neural network; 4) generalization using artificial neural networks for a set of relevant biotargets of a set of obtained convolution variables, with the creation of a high-precision model capable of effectively predicting systemic types of pharmacological activity.

Table 1. Relevant anxiolytic biotargets of chemical compounds and their validated 3D models

Code[1]	Name[2]	3D model PDB code
ADRA1A	Alpha-1A adrenergic receptor	3p0g
ADRA1B	Alpha-1B adrenergic receptor	4amj
ADRA2A	Alpha-2A adrenergic receptor	6kuy
ADRA2B	Alpha-2B adrenergic receptor	3pbl
AGTR1	Type-1 angiotensin II receptor	6os1
CA2	Carbonic anhydrase 2	2weg
CA4	Carbonic anhydrase 4	5jn9
CNR1	Cannabinoid receptor 1	7v3z
GABAR	Gamma-aminobutyric acid A receptor (2α1/2β2/γ2)	6x3x
HTR1A	5-hydroxytryptamine receptor 1A	7e2x
HTR1B	5-hydroxytryptamine receptor 1B	4iar
HTR1D	5-hydroxytryptamine receptor 1D	5d5a
HTR2A	5-hydroxytryptamine receptor 2A	4amj
HTR2B	5-hydroxytryptamine receptor 2B	6j20
HTR2C	5-hydroxytryptamine receptor 2C	4amj
HTR4	5-hydroxytryptamine receptor 4	2rh1
HTR7	5-hydroxytryptamine receptor 5	7e2z
MTNR1A	Melatonin receptor type 1A	7vgz
MTNR1B	Melatonin receptor type 1B	7vh0
NMDAR	N-methyl-D-aspartate receptor (2GRIN1/GRIN2A/GRIN2B)	6irh
SCN11A	Sodium channel protein type 11 subunit alpha	6a90
SLC18A2	Synaptic vesicular amine transporter	3o7q

Note: 1 – standard abbreviations of proteins according to UniProt (https://www.uniprot.org/uniprot/); 2 – nomenclature name of protein, recommended by UniProt (https://www.uniprot.org/uniprot/).

Table 2. Results of correlation convolution of multiple ensemble molecular docking energy spectra of known anxiolytic compounds into valid relevant biotargets 3D models

Code[1]	Convolution mean value, W[2]			p[6]
	W_{hm}[3]	W_{nhm}[4]	$W_{hm} - W_{nhm}$[5]	
ADRA1A	12848	11849	999	3×10^{-6}
ADRA1B	14408	13036	1371	$<5\times10^{-7}$
ADRA2A	12903	11553	1350	$<5\times10^{-7}$
ADRA2B	14366	13098	1268	$<5\times10^{-7}$
AGTR1	13810	12458	1352	$<5\times10^{-7}$
CA2	11453	10552	901	$<5\times10^{-7}$
CA4	10037	9264	773	$<5\times10^{-7}$
CNR1	10225	9363	862	1×10^{-6}
GABAR	15664	14342	1322	$<5\times10^{-7}$
HTR1A	16000	14657	1344	$<5\times10^{-7}$
HTR1B	12182	11125	1057	$<5\times10^{-7}$
HTR1D	13928	12621	1308	$<5\times10^{-7}$
HTR2A	15544	14197	1346	1×10^{-6}
HTR2B	10311	9283	1028	$<5\times10^{-7}$
HTR2C	14822	13558	1265	1×10^{-6}
HTR4	16179	14581	1598	$<5\times10^{-7}$
HTR7	11322	10450	872	1×10^{-6}
MTNR1A	9332	8626	706	1×10^{-6}
MTNR1B	14660	13442	1218	$<5\times10^{-7}$
NMDAR	25090	22544	2546	$<5\times10^{-7}$
SCN11A	17478	15721	1756	$<5\times10^{-7}$
SLC18A2	13293	12300	993	1×10^{-6}

Note: 1 – biotarget designations correspond to Table 1; 2 – average calculated using formula (1) energies of fully connected correlation neural network; 3 – for a class of compounds with pronounced activity; 4 – for a class of compounds with low activity; 5 – difference between mean values W for classes of compounds with pronounced and low activities; 6 – significance of differences in mean values W according to Mann-Whitney criterion [18].

CONCLUSIONS

Using the method of artificial neural networks, the high-precision multi-target model of the dependence of the anxiolytic activity of chemical compounds on their integral affinity to 22 relevant target proteins was built, based on the correlation convolution of multiple docking energy spectra.

The accuracy of predicting anxiolytic activity using the constructed classification multi-target convolutional neural network model is very high and in most tests exceeds 90%. The model is highly statistically significant, in most tests its statistical significance is $p<1\times10^{-15}$.

Table 3. Architecture and accuracy indicators of the multi-target neural network model of depending anxiolytic activity on the parameters of the correlation convolution of the multiple docking spectra energy

General characteristics of the best neural network		
No.[1]	Architecture[2]	
4/2/66	MLP 22-19-2 (Tanh, Softmax)	
Training accuracy[3]		
Data type	**F_0, %[4]**	**p[5]**
Training set	98.4	$<1\times10^{-15}$
Testing set	79.2	1.70×10^{-4}
Validation set	67.5	1.65×10^{-2}
Accuracy testing on combined set		
Rate	**F, %**	**p[5]**
General accuracy *Acc*, %[6]	91.2	$<1\times10^{-15}$
Sensitivity *Sens*, %[7]	91.3	$<1\times10^{-15}$
Specificity *Spec*, %[8]	91.2	$<1\times10^{-15}$
Area under ROC-curve *AUC*, %[9]	94.4	$<1\times10^{-15}$

Note: 1 – the training stage number, sampling number and neural network number are indicated; 2 – multilayer perceptron, 22 input, 19 hidden and 2 output neurons, activation functions of the hidden and output layers are hyperbolic tangent and multidimensional logistic function; 3 – obtained in a two-stage iterative process of building the neural network; 4 – general accuracy Acc; 5 – significance of accuracy indicators according to the binomial criterion [18]; 6 – general prediction accuracy for all compounds; 7 – prediction accuracy of compounds with pronounced activity; 8 – prediction accuracy of compounds with low activity; 9 – for the full set.

Thus, using the example of anxiolytic activity, a new methodology for constructing classification models for predicting systemic types of multi-target activity of chemical compounds has been developed, based on the use of artificial neural network technology and on the correlation convolution of multiple docking energy spectra.

The constructed multi-target neural network model of the dependence of the anxiolytic activity of chemical compounds on the parameters of the correlation convolution of the energy spectra of multiple docking is used in *in silico* search for new highly active compounds of various chemical classes; promising substances have been found.

The created multi-target convolution neural network methodology can be used to search for highly active compounds with other types of pharmacological activity that have a system-wide nature, such as hypoglycemic, anti-inflammatory, and other types of psychotropic activity.

FUNDING

This study was supported in the framework of the state assignment of the Ministry of Health of the Russian Federation No. 23022400009-9 "Development of a methodology for computer searching multi-target pharmacologically active compounds based on multiple docking and technology of the convolutional neural networks of various architectures".

COMPLIANCE WITH ETHICAL STANDARDS

This article does not contain a description of the research conducted by the authors with the participation of humans or the use of animals as objects.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

REFERENCES

1. World Health Organization. Mental Health, Brain Health and Substance Use. Retrieved July 29, 2024, from: https://www.who.int/teams/mental-health-and-substance-use/overview
2. Zdravookhraneniye v Rossii. 2023: Statisticheskiy Sbornik (2023), Rosstat, Moskva, 179 p.
3. *Sarkar C., Das B., Rawat V.S., Wahlang J.B., Nongpiur A., Tiewsoh I., Lyngdoh N.M., Das D., Bidarolli M., Sony H.T.* (2023) Artificial intelligence and machine learning technology driven modern drug discovery and development. Int. J. Mol. Sci., **24**(3), 2026. DOI: 10.3390/ijms24032026
4. *Vora L.K., Gholap A.D., Jetha K., Thakur R.R.S., Solanki H.K., Chavda V.P.* (2023) Artificial intelligence in pharmaceutical technology and drug delivery design. Pharmaceutics, **15**(7), 1916. DOI: 10.3390/pharmaceutics15071916
5. *Vassiliev P.M., Kochetkov A.N., Spasov A.A., Perfilev M.A.* (2021) The energy spectrum of multiple docking as a multi-dimensional metric of the affinity of chemical compounds to pharmacologically relevant bio-targets. Volgogradskiy Nauchno-Meditsinskiy Zhurnal, 3, 57–61.
6. *Vassiliev P.M., Kochetkov A.N., Perfilev M.A.* (2022) Neural network modeling of the dependence of $GABA_A$-agonistic activity of chemical compounds on the energy spectrum of multiple docking. Vestnik VolgGMU, **19**(4), 88–93. DOI: 10.19163/1994-9480-2022-19-4-88-93
7. *Vassiliev P.M., Maltsev D.V., Perfilev M.A., Spasov A.A., Skripka M.O., Kochetkov A.N.* (2022) Soyedineniya s anksioliticheskoy aktivnost'yu (Compounds with anxiolytic activity), Certificate of State Registration of the Database no. 2022621744. (15.07.2022), Offits. Bull. Programmy dlya EVM, baz dannykh i topologiy integral'nykh mikroskhem ETIN (Programs for Computers, Databases and Topologies of Integrated Circuits), **7**, RU 2022621744.
8. ChemAxon: Marvin (2024), Retrieved July 29, 2024, from: https://chemaxon.com/marvin
9. MOPAC (2024), Retrieved July 29, 2024, from: http://openmopac.net/
10. *Ochoa D., Hercules A., Carmona M., Suveges D., Baker J., Malangone C., Lopez I., Miranda A., Cruz-Castillo C., Fumis L., Bernal-Llinares M., Tsukanov K., Cornu H., Tsirigos K., Razuvayevskaya O., Buniello A., Schwartzentruber J., Karim M., Ariano B., Martinez Osorio R.E., Ferrer J., Ge X., Machlitt-Northen S., Gonzalez-Uriarte A., Saha S., Tirunagari S., Mehta C., Roldán-Romero J.M., Horswell S., Young S., Ghoussaini M., Hulcoop D.G., Dunham I., McDonagh E.M.* (2023) The next-generation Open Targets Platform: reimagined, redesigned, rebuilt. Nucleic Acids Res., **51**(D1), D1353–D1359. DOI: 10.1093/nar/gkac1046
11. *Vasilyev P.M., Luzina O.A., Babkov D.A., Appazova D.T., Salakhutdinov N.F., Spasov A.A.* (2019) Studying dependences between the chemotype structure of some natural compounds and the spectrum of their targeted activities correlated with the hypoglycemic effect. J. Struct. Chem., **60**(11), 1827–1832. DOI: 10.1134/S0022476619110179
12. *Vassiliev P.M., Spasov A.A., Kosolapov V.A., Kucheryavenko A.F., Gurova N.A., Anisimova V.A.* (2014) Consensus drug design using it microcosm. In: Application of Computational Techniques in Pharmacy and Medicine (Gorb L., Kuz'min V., Muratov E., eds.). Springer Science + Business Media, Dordrecht (Netherlands), pp. 369–431.
13. *Vassiliev P.M., Spasov A.A., Yanaliyeva L.R., Kochetkov A.N., Vorfolomeyeva V.V., Klochkov V.G., Appazova D.T.* (2019) Neural network modeling of multitarget RAGE inhibitory activity. Biomeditsinskaya Khimiya, **65**(2), 91–98. DOI: 10.1134/S1990750819030107
14. *Trott O., Olson A.J.* (2010) AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J. Comput. Chem., **31**(2), 455–461. DOI: 10.1002/jcc.21334
15. *Hilbe J.M.* (2007) Statistica 7. An overview. The American Statistician, **61**(1), 91–94.
16. *Kolmogorov A.N.* (1957) On the representation of continuous functions of many variables by superposition of continuous functions of one variable and addition. Proceedings of USSR Academy of Sciences, **114**(5), 953–956.
17. *Dubin U.* (2013) Cross-Entropy Method: Theory with Applications. LAMBERT Academic Publishing, Chisinau (Moldova), 148 p.
18. *Glotov N.V., Zhivotovskij L.A., Hovanov N.V., Hromov-Borisov N.N.* (1982) Biometriya, Izd-vo Leningr. Un-ta, L., 264 p.

Received: 01. 10. 2024.
Revised: 25. 10. 2024.
Accepted: 27. 10. 2024.

МУЛЬТИТАРГЕТНАЯ НЕЙРОСЕТЕВАЯ МОДЕЛЬ АНКСИОЛИТИЧЕСКОЙ АКТИВНОСТИ ХИМИЧЕСКИХ СОЕДИНЕНИЙ НА ОСНОВЕ КОРРЕЛЯЦИОННОЙ СВЁРТКИ СПЕКТРОВ ЭНЕРГИЙ МНОЖЕСТВЕННОГО ДОКИНГА

П.М. Васильев*, М.А. Перфильев, А.В. Голубева, А.Н. Кочетков, Д.В. Мальцев

Волгоградский государственный медицинский университет,
400131, Волгоград, пл. Павших борцов, 1; *эл. почта: pvassiliev@mail.ru

Тревожные расстройства являются одной из самых распространённых в мире патологий психического здоровья, которые требуют поиска и создания новых эффективных фармакологически активных веществ. В связи с этим разработка с использованием методов искусственного интеллекта новых подходов к поиску анксиолитических веществ является актуальным направлением современной биоинформатики и фармакологии. В настоящей работе методом искусственных нейронных сетей построена мультитаргетная модель зависимости анксиолитической активности химических соединений от их интегральной аффинности к релевантным белкам-мишеням, основанная на корреляционной свёртке спектров энергии множественного докинга. Для этого на основе ранее созданной базы данных была сформирована обучающая выборка по структуре и активности 537 известных анксиолитических веществ и построены оптимизированные 3D-модели этих соединений. Выявлены 22 биомишени, предположительно релевантные анксиолитической активности, и найдены их валидные 3D-модели. Для каждой такой биомишени по всему её объёму сформированы 27 пространств для множественного докинга. Выполнен множественный ансамблевый молекулярный докинг 537 известных анксиолитических соединений во все пространства релевантных белков-мишеней. Проведена корреляционная свёртка рассчитанных спектров энергий множественного докинга. С использованием семи вариантов обучения на основе искусственных многослойных перцептронных нейронных сетей построена мультитаргетная модель зависимости анксиолитической активности химических соединений от 22 параметров корреляционной свёртки спектров энергий их множественного докинга. Выполнена оценка прогностической способности созданной модели, общая точность которой составила Acc = 91,2% и AUC_{ROC} = 94,4%, при статистической достоверности $p<1\times10^{-15}$. Найденная модель используется в поиске новых веществ с высокой анксиолитической активностью.

Полный текст статьи на русском языке доступен на сайте журнала (http://pbmc.ibmc.msk.ru).

Ключевые слова: анксиолитическая активность; мультитаргетная аффинность; множественный молекулярный докинг; спектр энергий докинга; корреляционная свёртка; искусственные нейронные сети

Финансирование. Работа выполнена в рамках государственного задания Министерства здравоохранения Российской Федерации № 23022400009-9 “Разработка методологии компьютерного поиска мультитаргетных фармакологически активных соединений на основе множественного докинга и технологии свёрточных нейронных сетей различной архитектуры”.

Поступила в редакцию: 01.10.2024; после доработки: 25.10.2024; принята к печати: 27.10.2024.