

ОРИГИНАЛЬНЫЕ ИССЛЕДОВАНИЯ

Научная статья

УДК 615.015.11:544.165:575.112::[544.187.2+004.032.26]

doi: 10.19163/1994-9480-2022-19-4-88-93

## НЕЙРОСЕТЕВОЕ МОДЕЛИРОВАНИЕ ЗАВИСИМОСТИ ГАМК<sub>A</sub>-АГОНИСТИЧЕСКОЙ АКТИВНОСТИ ХИМИЧЕСКИХ СОЕДИНЕНИЙ ОТ СПЕКТРА ЭНЕРГИЙ МНОЖЕСТВЕННОГО ДОКИНГА

**Павел Михайлович Васильев<sup>1✉</sup>, Андрей Николаевич Кочетков<sup>2</sup>,  
Максим Алексеевич Перфильев<sup>3</sup>**

<sup>1,2,3</sup>Волгоградский государственный медицинский университет, Волгоград, Россия

<sup>1</sup>[pvassiliev@mail.ru](mailto:pvassiliev@mail.ru)✉

<sup>2</sup>[akocho@mail.ru](mailto:akocho@mail.ru)

<sup>3</sup>[maxim.firu@yandex.com](mailto:maxim.firu@yandex.com)

**Аннотация.** Выполнен простой и множественный докинг в ГАМК<sub>A</sub>-рецептор его известных агонистов. Для указанных соединений рассчитаны энергии простого докинга в ГАМК-связывающий сайт и спектры энергий множественного докинга по всему объему ГАМК<sub>A</sub>-рецептора. Сформирована модель зависимости уровня ГАМК<sub>A</sub>-агонистической активности химических соединений от энергии их простого докинга в ГАМК-связывающий сайт. Построена нейросетевая модель зависимости уровня ГАМК<sub>A</sub>-агонистической активности от спектра энергий множественного докинга химических соединений. Показано, что прогностическая точность нейросетевой модели на основе множественного докинга существенно превышает прогностическую точность модели на основе простого докинга.

**Ключевые слова:** *in silico*, простой докинг, множественный докинг, спектр энергий, искусственные нейронные сети, зависимости «структура – активность», прогностическая способность

**Финансирование.** Работа выполнена в рамках государственного задания Министерства здравоохранения Российской Федерации № 121060700050-2 «Разработка методологии компьютерного поиска фармакологически активных соединений на основе множественного докинга и технологии искусственных нейронных сетей».

ORIGINAL RESEARCHES

Original article

## NEURAL NETWORK MODELING OF THE DEPENDENCE OF GABA-AGONISTIC ACTIVITY OF CHEMICAL COMPOUNDS ON THE ENERGY SPECTRUM OF MULTIPLE DOCKING

**Pavel M. Vasiliev<sup>1✉</sup>, Andrey N. Kochetkov<sup>2</sup>, Maxim A. Perfiliev<sup>3</sup>**

<sup>1,2,3</sup> Volgograd State Medical University, Volgograd, Russia

<sup>1</sup>[pvassiliev@mail.ru](mailto:pvassiliev@mail.ru)✉

<sup>2</sup>[akocho@mail.ru](mailto:akocho@mail.ru)

<sup>3</sup>[maxim.firu@yandex.com](mailto:maxim.firu@yandex.com)

**Abstract.** Simple and multiple docking into the GABA<sub>A</sub> receptor of its known agonists has been carried out. For these compounds, the energies of simple docking into the GABA binding site and the energy spectra of multiple docking throughout the entire volume of the GABA<sub>A</sub> receptor were calculated. A model of the dependence of the level of GABA<sub>A</sub> agonistic activity of chemical compounds on the energy of their simple docking into the GABA binding site has been formed. A neural network model of the dependence of the level of GABA<sub>A</sub> agonistic activity on the energy spectrum of multiple docking of chemical

© Васильев П.М., Кочетков А.Н., Перфильев М.А., 2022

compounds has been constructed. It is shown that the predictive accuracy of the neural network model based on multiple docking significantly exceeds the predictive accuracy of the model based on simple docking.

**Keywords:** *in silico*, simple docking, multiple docking, energy spectrum, artificial neural networks, structure-activity relationships, predictive ability

**Funding.** The work was carried out within the framework of the state task of the Ministry of Health of the Russian Federation No. 121060700050-2 "Development of a methodology for computer search of pharmacologically active compounds based on multiple docking and artificial neural network technology".

В нашей предыдущей работе [1] было показано, что спектр энергий докинга химических соединений в множество пространств релевантного белка является намного более достоверной метрикой аффинности лигандов к биомешинам, в сравнении с единичной энергией их докинга в специфический сайт. При этом использование множественного докинга для расчета аффинности соединений не требует определения в белке-мишени местоположения специфического сайта связывания.

Известно также, что, в соответствии с теоремой Колмогорова [2], с помощью двухслойной искусственной нейронной сети может быть аппроксимирована зависимость любой сложности.

Таким образом, разработка на основе указанной технологии искусственного интеллекта новой методологии построения *in silico* высокоточных моделей зависимости фармакологической активности химических соединений от спектра их множественной аффинности к релевантным белкам-мишеням, включающая учет взаимодействия лиганда со всей поверхностью белка, является востребованной и весьма актуальной задачей.

## ЦЕЛЬ РАБОТЫ

Построение методом искусственных нейронных сетей модели зависимости уровня ГАМК<sub>A</sub>-агонистической активности от спектра энергий множественного докинга химических соединений и оценка ее прогностической способности, в сравнении с точностью прогноза с использованием энергии простого докинга.

## МЕТОДИКА ИССЛЕДОВАНИЯ

### Исходное допущение

На данном этапе разработки новой методологии прогноза в качестве модельной фармакологической активности была выбрана ГАМК<sub>A</sub>-агонистическая активность, по следующим основаниям: 1) ГАМК-ергические соединения обычно проявляют психотропные свойства и в связи с этим рассматриваются как перспективные кандидаты для создания новых лекарственных средств; 2) имеются достоверные

данные по химической структуре и экспериментальным значениям ГАМК<sub>A</sub>-агонистической активности уже синтезированных известных веществ; 3) доступны валидные экспериментальные рентгеноструктурные 3D-модели ГАМК<sub>A</sub>-рецептора и информация о локализации его специфических сайтов связывания.

### Задачи исследования

Для достижения поставленной цели необходимо было решить следующие задачи.

1. Сформировать верифицированную выборку по структуре и уровню активности известных соединений, испытанных на ГАМК<sub>A</sub>-агонистическую активность.
2. Построить и оптимизировать 3D-модели этих соединений.
3. Выполнить для указанных соединений простой докинг в ГАМК-связывающий сайт ГАМК<sub>A</sub>-рецептора и множественный докинг во все сформированные для этого пространства данного рецептора.
4. Рассчитать энергии простого докинга и спектры энергий множественного докинга указанных известных соединений.
5. Сформировать модель зависимости уровня ГАМК<sub>A</sub>-агонистической активности химических соединений от энергии их простого докинга в ГАМК-связывающий сайт, оценить ее точность.
6. Построить нейросетевую модель зависимости уровня ГАМК<sub>A</sub>-агонистической активности от спектра энергий множественного докинга химических соединений, оценить ее точность.

### Выборка по структуре и активности ГАМК<sub>A</sub>-агонистов

Верифицированные данные по химической структуре и уровню ГАМК<sub>A</sub>-агонистической активности 191 известного вещества в отношении  $\alpha 1$ - $\beta 2$ - $\gamma 2$  подтипа ГАМК<sub>A</sub>-рецептора были взяты из оригинальной QSAR-базы системы Microcosm BioS v20.6.6 [3], содержащей структурированную и обработанную информацию по химической структуре и уровню активности 625888 известных соединений, изученных мировым научным сообществом на 11509 различных видов целевой биологической активности. С помощью программы Statistica 7 [4] методом k-средних был

выполнен кластерный анализ количественных значений ГАМК<sub>A</sub>-агонистической активности с разделением на 2 группы активности: high (24 соединения) и moderate (167 соединений).

#### **Оптимизированные 3D-модели соединений**

В соответствии с методикой [5], для всех 191 соединений методами молекулярной механики с помощью программы MarvinSketch 17.1.23 [6], отдельно для каждого соединения, были построены по 10 конформеров с наименьшей энергией. Построенные конформеры были оптимизированы с помощью программы MORAC2012 [7] с использованием полуэмпирического квантово-химического метода PM7. Среди оптимизированных конформеров были отобраны для каждого соединения по одному конформеру с наименьшей общей энергией.

#### **Простой и множественный докинг**

В расчетах использовалась экспериментальная 3D-модель ГАМК<sub>A</sub>-рецептора 6D6T [8] – наиболее полная и точная из 14 моделей, найденных в банке данных RCSB PDB [9]. Координаты ГАМК-связывающего сайта были определены с помощью программы LigPlot+ 2.2.5 [10]. Кроме того, на указанной модели 6D6T с помощью программы MSite v21.04.22, в соответствии с алгоритмом, подробно описанным в нашей предыдущей работе [1], были построены 27 пространств для множественного докинга.

Ансамблевый докинг проводили с помощью программы AutoDock Vina 1.1.1 [11], каждое соединение в 10 конформерах по 5 раз в каждое пространство докинга, с вычислением по 50 полученным значениям минимальных энергий связывания  $\Delta E$ , как это описано в [5]. Докинг выполняли отдельно в ГАМК-связывающий сайт ГАМК<sub>A</sub>-рецептора и отдельно в каждое из 27 пространств, сформированных для множественного докинга.

Полученная в результате расчетов сводная таблица для последующего моделирования включала в себя 30 колонок: шифры соединений; метка уровня ГАМК<sub>A</sub>-агонистической активности Ind; энергия докинга в специфический ГАМК-связывающий сайт  $\Delta E_0$  и энергии множественного докинга в 27 пространств  $\Delta E_1... \Delta E_{27}$  ГАМК<sub>A</sub>-рецептора.

#### **Модель зависимости активности от энергии простого докинга**

Для построения модели зависимости уровня ГАМК<sub>A</sub>-агонистической активности химических соединений от энергии их простого докинга в ГАМК-связывающий сайт  $\Delta E_0$  все значения этого показателя

для 191 известного ГАМК<sub>A</sub>-агониста с помощью программы Statistica 7 [4] были подвергнуты кластерному анализу методом k-средних, с определением граничного значения  $\Delta \bar{E}_0$ , разделяющего соединения с высокой энергией докинга  $\Delta E_0$  и соединения с умеренной или низкой величиной  $\Delta E_0$ .

Парадигма традиционного докинга постулирует, что высокая расчетная энергия взаимодействия со специфическим сайтом биомишени  $\Delta E_0$  соответствует высокому значению фармакологической активности [12]. Таким образом, основанное на простом докинге решающее правило определяется следующими неравенствами:

1) если  $\Delta E_0 < \Delta \bar{E}_0$ , то прогнозируемое соединение принадлежит к классу соединений с высокой ГАМК<sub>A</sub>-агонистической активностью;

2) если же  $\Delta E_0 \geq \Delta \bar{E}_0$ , то прогнозируемое соединение принадлежит к классу соединений с умеренной или низкой ГАМК<sub>A</sub>-агонистической активностью.

С использованием сформированной модели рассчитаны общая точность прогноза  $F_0$ , точность прогноза высоко активных соединений  $F_a$  (чувствительность) и точность прогноза не высоко активных соединений  $F_n$  (специфичность). С помощью биномиального критерия  $z$  [13] оценена статистическая достоверность  $p$  показателей точности прогноза.

#### **Нейросетевая модель зависимости активности от энергии множественного докинга**

Для построения модели зависимости уровня ГАМК<sub>A</sub>-агонистической активности от спектра энергий множественного докинга химических соединений  $\Delta E_1... \Delta E_{27}$  была использована технология искусственных нейронных сетей [14]. В соответствии с теоремой Колмогорова [2], в настоящем исследовании использовалась архитектура сети в виде двухслойного перцептрона MLP  $k-m-2$  с узким горлом. Здесь  $k$  – число входных нейронов, в данном случае 27 ( $\Delta E_1... \Delta E_{27}$ );  $m$  – число скрытых нейронов, устанавливается программой от 3 до 26, поскольку  $2 < m < k$ . В модуле Statistica Neural Networks [14] программы Statistica [4] при построении перцептронных сетей для скрытого и выходного слоев используются пять наиболее распространенных активационных функций (Identity, Logistic, Tanh, Exponential, Softmax) [14], попарный перебор которых также осуществляется программой.

Обучение нейросетей в программе Statistica проводится с использованием алгоритма обратного распространения ошибок. С целью достижения наилучшего результата обучения, число сетей для обучения было увеличено с 20 до 400, а число наилучших автоматически отбираемых сетей – с 5 до 100. После окончания обучения для заданного уровня активности из 100 лучших отобранных программой нейросетей

по совокупности характеристик точности обучения и тестирования вручную отбирали лучшую.

С использованием построенной модели рассчитаны общая точность прогноза  $F_0$ , точность прогноза высоко активных соединений  $F_a$  (чувствительность) и точность прогноза не высоко активных соединений  $F_n$  (специфичность). С помощью биномиального критерия  $z$  [13] оценена статистическая достоверность  $p$  показателей точности прогноза.

## РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ И ИХ ОБСУЖДЕНИЕ

Для модели зависимости уровня ГАМК<sub>A</sub>-агонистической активности химических соединений от энергии

их простого докинга в ГАМК-связывающий сайт  $\Delta E_0$  получено граничное значение  $\Delta E_0 = 6,0$  Ккал/моль, разделяющее высоко активные и не высоко активные соединения.

В случае нейросетевой модели зависимости уровня ГАМК<sub>A</sub>-агонистической активности от спектра энергий множественного докинга химических соединений  $\Delta E_1... \Delta E_{27}$  получена нейронная сеть следующих характеристик: MLP 27-16-2 (Tanh, Softmax), точность обучения 99,3 %, тестирования 100 %, валидации 67,9 %, также классифицирующая высоко активные и не высоко активные соединения.

В табл. приведены результаты сравнительной оценки на объединенной выборке точности указанных двух моделей.

Точность прогноза уровня ГАМК<sub>A</sub>-агонистической активности  
на основе энергий простого и множественного докинга

Показатель точности прогноза	Значения для модели прогноза с использованием	
	$\Delta E_0$	$\Delta E_1... \Delta E_{27}$
$F_0$ , %	51,3	94,8
$F_a$ , %	66,7	83,3
$F_n$ , %	49,1	96,4
$z_0$	0,58	8,73
$z_a$	1,08	2,23
$z_n$	0,14	8,45
$p_0$	0,281	$<1 \cdot 10^{-15}$
$p_a$	0,139	0,0127
$p_n$	0,555	$<1 \cdot 10^{-15}$

Примечание.

$F_0, F_a, F_n$  – общая точность прогноза, точность прогноза высоко активных соединений, точность прогноза не высоко активных соединений.

$z_0, z_a, z_n$  – биномиальный критерий для  $F_0, F_a, F_n$ .

$p_0, p_a, p_n$  – значимость  $F_0, F_a, F_n$ .

Точность прогноза ГАМК<sub>A</sub>-агонистической активности с использованием модели на основе энергии докинга в специфический сайт  $\Delta E_0$  является статистически незначимой по всем показателям.

В случае прогноза не высоко активных соединений вычисленная оценка даже меньше точности случайного угадывания, составляющей 50 %.

В то же время при использовании нейросетевой модели на основе энергий множественного докинга  $\Delta E_1... \Delta E_{27}$  все оценки точности прогноза являются статистически высоко достоверными. При этом значения всех показателей  $F_0, F_a, F_n$  значительно превышают точность случайного угадывания.

Таким образом, на примере ГАМК<sub>A</sub>-агонистической активности доказано, что нейросетевая модель на основе спектра энергий докинга химических соединений в множество пространств ГАМК<sub>A</sub>-рецептора является намного более точной, чем

модель, полученная на основе единичной энергии их докинга в специфический ГАМК-связывающий сайт.

## ЗАКЛЮЧЕНИЕ

1. С использованием кластерного анализа построена модель зависимости уровня ГАМК<sub>A</sub>-агонистической активности химических соединений от энергии их простого докинга в ГАМК-связывающий сайт.

2. С использованием технологии искусственных нейронных сетей построена модель зависимости уровня ГАМК<sub>A</sub>-агонистической активности от спектра энергий множественного докинга химических соединений по всему пространству ГАМК<sub>A</sub>-рецептора.

3. Статистическая достоверность и точность прогноза уровня ГАМК<sub>A</sub>-агонистической активности химических соединений с использованием нейросетевой модели на основе спектра значений энергий множественного докинга существенно превышает

аналогичные показатели, полученные с применением модели на основе единичного значения энергии докинга в специфический ГАМК-связывающий сайт ГАМК<sub>A</sub>-рецептора.

#### СПИСОК ИСТОЧНИКОВ

1. Васильев П.М., Кочетков А.Н., Спасов А.А., Перфильев М.А. Спектр энергий множественного докинга как многомерная метрика аффинности химических соединений к фармакологически релевантным биомишеням // Волгоградский научно-медицинский журнал. 2021. № 3. С. 57–61.
2. Колмогоров А.Н. О представлении непрерывных функций нескольких переменных в виде суперпозиции непрерывных функций одного переменного // Доклады АН СССР. 1957. Т. 114, № 5. С. 953–956.
3. Studying Dependences Between the Chemotype Structure of Some Natural Compounds and the Spectrum of Their Targeted Activities Correlated with the Hypoglycemic Effect / P.M. Vasilyev, O.A. Luzina, D.A. Babkov [et al.] // Journal of Structural Chemistry. 2019. Vol. 60, iss. 11. P. 1827–1832.
4. Hilbe J.M. Statistica 7: an overview // The American Statistician. 2007. Vol. 61, no. 1. P. 91–94.
5. Neural network modeling of multitarget RAGE inhibitory activity / P.M. Vassiliev, A.A. Spasov, L.R. Yanaliyeva [et al.] // Biochemistry (Moscow) Supplement Series B: Biomedical Chemistry. 2019. Vol. 13, no. 3. P. 256–263.
6. Marvin Pro. URL: <https://chemaxon.com/products/marvin>.
7. MOPAC, Stewart Computational Chemistry. URL: <http://openmopac.net/>.
8. Structure of a human synaptic GABA<sub>A</sub> receptor / S. Zhu, C.M. Noviello, J. Teng [et al.] // Nature. 2018. Vol. 559 (7712). P. 67–72.
9. RCSB PDB, Research Collaboratory for Structural Bioinformatics Protein Data Bank. URL: <https://www.rcsb.org/>.
10. Laskowski R.A., Swindells M.B. LigPlot+: multiple ligand-protein interaction diagrams for drug discovery // Journal of Chemical Information and Modeling. 2011. Vol. 51, iss. 10. P. 2778–2786.
11. Trott O., Olson A.J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading // Journal of Computational Chemistry. 2010. Vol. 31, iss. 2. P. 455–461.
12. Gupta M., Sharma R., Kumar A. Docking techniques in pharmacology: How much promising? // Comput Biol Chem. 2018. Vol. 76. P. 210–217.
13. Глотов Н.В., Животовский Л.А., Хованов Н.В., Хромов-Борисов Н.Н. Биометрия. Л.: Изд-во Ленингр. ун-та, 1982. 264 с.
14. Нейронные сети. Statistica Neural Networks: Методология и технология современного анализа данных. М.: Горячая линия – Телеком, 2008. 392 с.

Авторы заявляют об отсутствии конфликта интересов.

#### Информация об авторах

**П.М. Васильев** – доктор биологических наук, заведующий лабораторией информационных технологий в фармакологии и компьютерного моделирования лекарств, Научный центр инновационных лекарственных

#### REFERENCES

1. Vassiliev P.M., Kochetkov A.N., Spasov A.A., Perfilov M.A. The energy spectrum of multiple docking as a multidimensional metric of the affinity of chemical compounds to pharmacologically relevant biotargets. *Volgogradskiy nauchno-meditsinskiy zhurnal = Volgograd Journal of Medical Research*. 2021;3:57–61. (In Russ.).
2. Kolmogorov A.N. On the representation of continuous functions of several variables as a superposition of continuous functions of one variable. *Doklady AN SSSR = Reports of the Academy of Sciences of the USSR*. 1957;114(5):953–956. (In Russ.).
3. Vasilyev P.M., Luzina O.A., Babkov D.A., Appazova D.T., Salakhutdinov N.F., Spasov A.A. Studying Dependences Between the Chemotype Structure of Some Natural Compounds and the Spectrum of Their Targeted Activities Correlated with the Hypoglycemic Effect. *Journal of Structural Chemistry*. 2019;60(11):1827–1832.
4. Hilbe J.M. Statistica 7: an overview. *The American Statistician*. 2007;61(1):91–94.
5. Vassiliev P.M., Spasov A.A., Yanaliyeva L.R. et al. Neural network modeling of multitarget RAGE inhibitory activity. *Biochemistry (Moscow) Supplement Series B: Biomedical Chemistry*. 2019;13(3):256–263.
6. Marvin Pro. URL: <https://chemaxon.com/products/marvin>.
7. MOPAC, Stewart Computational Chemistry. URL: <http://openmopac.net/>.
8. Zhu S., Noviello C.M., Teng J. et al. Structure of a human synaptic GABA<sub>A</sub> receptor. *Nature*. 2018; 559(7712):67–72.
9. RCSB PDB, Research Collaboratory for Structural Bioinformatics Protein Data Bank. URL: <https://www.rcsb.org/>.
10. Laskowski R.A., Swindells M.B. LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *Journal of Chemical Information and Modeling*. 2011;51(10):2778–2786.
11. Trott O., Olson A.J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *Journal of Computational Chemistry*. 2010;31(2):455–461.
12. Gupta M., Sharma R., Kumar A. Docking techniques in pharmacology: How much promising? *Comput Biol Chem*. 2018;76:210–217.
13. Glotov N.V., Zhivotovskiy L.A., Khovanov N.V., Khromov-Borisov N.N. Biometrics. Leningrad; Leningrad University Publishing House, 1982. 264 p. (In Russ.).
14. Neural networks. Statistica Neural Networks: Methodology and technology for modern data analysis. Moscow: Hot line – Telecom, 2008. 392 p. (In Russ.).

средств с опытно-промышленным производством, профессор кафедры фармакологии и биоинформатики, Волгоградский государственный медицинский университет, Волгоград, Россия;

**А.Н. Кочетков** – системный администратор, инженер-программист лаборатории информационных технологий в фармакологии и компьютерного моделирования лекарств, Научный центр инновационных лекарственных средств с опытно-промышленным производством, Волгоградский государственный медицинский университет, Волгоград, Россия;

**М.А. Перфильев** – старший лаборант лаборатории информационных технологий в фармакологии и компьютерного моделирования лекарств, Научный центр инновационных лекарственных средств с опытно-промышленным производством; ассистент, аспирант кафедры фармакологии и биоинформатики, Волгоградский государственный медицинский университет, Волгоград, Россия.

Статья поступила в редакцию 02.10.2022; одобрена после рецензирования 18.11.2022; принята к публикации 06.12.2022.

**The authors declare no conflicts of interests.**

#### *Information about the authors*

**P.M. Vasiliev** – Doctor of Biological Sciences, Head of the Laboratory of Information Technologies in Pharmacology and Computer Modeling of Drugs, Research Center for Innovative Medicines with Pilot Production, Professor of the Department of Pharmacology and Bioinformatics, Volgograd State Medical University, Volgograd, Russia;

**A.N. Kochetkov** – System Administrator, Software Engineer of the Laboratory of Information Technologies in Pharmacology and Computer Modeling of Medicines, Scientific Center for Innovative Medicines with Pilot Production, Volgograd State Medical University, Volgograd, Russia;

**M.A. Perfiliev** – senior laboratory assistant at the Laboratory of Information Technologies in Pharmacology and Computer Modeling of Medicines, Scientific Center for Innovative Medicines with Pilot Production; Assistant, postgraduate student of the Department of Pharmacology and Bioinformatics, Volgograd State Medical University, Volgograd, Russia.

The article was submitted 02.10.2022; approved after reviewing 18.11.2022; accepted for publication 06.12.2022.