

## Нейросетевое моделирование зависимости RAGE-ингибирующей активности химических соединений от спектра энергий множественного докинга

Павел Михайлович Васильев, Максим Алексеевич Перфильев , Андрей Николаевич Кочетков

Волгоградский государственный медицинский университет, Волгоград, Россия

**Аннотация.** Проведен простой и множественный докинг в RAGE-рецептор его известных ингибиторов. Для этих соединений рассчитаны энергии простого докинга в специфический связывающий сайт и спектры энергий множественного докинга по всему объему RAGE-рецептора. Построена модель зависимости уровня RAGE-ингибирующей активности химических соединений от энергии их простого докинга в специфический связывающий сайт. Сформирована нейросетевая модель зависимости уровня RAGE-ингибирующей активности от спектра энергий множественного докинга химических соединений. Показано, что точность нейросетевой модели на основе множественного докинга значительно превышает точность модели на основе простого докинга. Найденная модель используется в направленном поиске новых соединений с RAGE-ингибирующей активностью, которые станут основой создания инновационных препаратов для лечения осложнений при сахарном диабете.

**Ключевые слова:** искусственные нейронные сети, RAGE-ингибирующая активность, множественный докинг, *in silico*, сахарный диабет

ORIGINAL RESEARCHES

Original article

doi: <https://doi.org/10.19163/1994-9480-2023-20-4-157-161>

## Neural network modeling of the dependence of RAGE inhibitory activity of chemical compounds on the energy spectrum of multiple docking

Pavel M. Vasiliev, Maxim A. Perfiliev , Andrey N. Kochetkov

Volgograd State Medical University, Volgograd, Russia

**Abstract.** Simple and multiple docking into the RAGE receptor of its known inhibitors have been carried out. For these compounds, the energies of simple docking into the specific binding site and the energy spectra of multiple docking throughout the entire volume of the RAGE receptor were calculated. A model of the dependence of the level of RAGE inhibitory activity of chemical compounds on the energy of their simple docking into the specific binding site has been constructed. A neural network model of the dependence of the level of RAGE inhibitory activity on the energy spectrum of multiple docking of chemical compounds has been formed. It is shown that the accuracy of the neural network model based on multiple docking significantly exceeds the accuracy of the model based on simple docking. The found model is used in the directed search for new compounds with RAGE inhibitory activity, which will become the basis for the creation of innovative drugs for the treatment of complications in diabetes mellitus.

**Keywords:** artificial neural networks, RAGE inhibitory activity, multiple docking, *in silico*, diabetes mellitus

Сахарный диабет (СД) характеризуется хронической гипергликемией, что сопровождается значительным усилением неферментативного гликирования белков. В результате образуются конечные продукты гликирования (AGE), которые взаимодействуют со своим специфическим рецептором (RAGE) [1]. Активация RAGE является одним из значимых факторов развития осложнений при СД [1, 2]. По данным ВОЗ, на 05.04.2023 г. в мире в 2014 г. число больных СД составляло 422 млн человек, а в 2019 г. СД был причиной смерти примерно 2 млн человек [3].

В нашей работе [4] было показано, что спектр энергий докинга химических соединений в множество пространств релевантного белка является намного более достоверной метрикой аффинности лигандов

к биомолекулам, 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. В соответствии с теоремой Колмогорова [5], 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 научно востребованной и весьма актуальной задачей.

### ЦЕЛЬ РАБОТЫ

Построение методом искусственных нейронных сетей модели зависимости уровня RAGE-ингибирующей 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

Для достижения цели исследования необходимо было решить следующие задачи.

1. Сформировать верифицированную выборку по структуре и уровню активности известных соединений, испытанных на RAGE-ингибирующую активность.

2. Построить и оптимизировать 3D-модели этих соединений.

3. Выполнить для указанных соединений простой докинг в специфический связывающий сайт RAGE-рецептора и множественный докинг во все сформированные для этого пространства данного рецептора, рассчитать энергии простого докинга и спектры энергий множественного докинга.

4. Сформировать модель зависимости уровня RAGE-ингибирующей активности химических соединений от энергии их простого докинга в специфический связывающий сайт, оценить ее точность.

5. Построить нейросетевую модель зависимости уровня RAGE-ингибирующей активности от спектра энергий множественного докинга химических соединений, оценить ее точность.

**Выборка по структуре и активности RAGE-ингибиторов.** Обучающая выборка по структуре и активности 183 известных веществ, испытанных на RAGE-ингибирующую активность, была сформирована с использованием оригинальной верифицированной базы данных [6], которая включала 99 соединений с высокой активностью и 84 низкоактивных или неактивных соединения.

**Оптимизированные 3D-модели соединений.** Оптимизированные 3D-модели всех 183 соединений были построены в соответствии с методикой [7]. Методами молекулярной механики с помощью программы MarvinSketch 17.1.23 [8], отдельно для каждого соединения, были построены по 10 конформеров с наименьшей энергией. Построенные конформеры были оптимизированы с помощью программы MORAC2012 [9] 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** выполняли с использованием экспериментальной 3D-модели RAGE-рецептора 4LP4 [10] – наиболее точной из 10 моделей, проанализированных в работе [7]. Координаты специфического связывающего сайта были определены с помощью программы LigPlot+ 2.2.5 [11]. Кроме того, на указанной модели 4LP4 с помощью программы MSite v21.04.22, в соответствии с алгоритмом, подробно описанным в нашей работе [4], были построены 27 пространств для множественного докинга.

Ансамблевый докинг проводили с помощью программы AutoDock Vina 1.1.1 [12], 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 работе [7]. Докинг выполняли отдельно в специфический связывающий сайт RAGE-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 RAGE-ингибирующей активности  $Ind$ ; энергия докинга в специфический связывающий сайт  $\Delta E_0$  и энергии множественного докинга в 27 пространств  $\Delta E_1 \dots \Delta E_{27}$  RAGE-рецептора.

**Модель зависимости активности от энергии простого докинга.** Для построения модели зависимости уровня RAGE-ингибирующей активности химических соединений от энергии их простого докинга в специфический связывающий сайт  $\Delta E_0$  все значения этого показателя для 183 известных RAGE-ингибиторов с помощью программы Statistica 7 [13] были подвергнуты кластерному анализу методом k-средних, с определением граничного значения  $\Delta \hat{E}_0$ , разделяющего соединения с выраженной энергией докинга  $\Delta E_0$  и соединения с низкой величиной  $\Delta E_0$ .

Согласно основному постулату традиционного докинга, чем выше расчетная энергия взаимодействия со специфическим сайтом биомишени  $\Delta E_0$ , тем выше значение фармакологической активности [14]. Таким образом, основанное на простом докинге решающее правило определяется следующими неравенствами:

1) если  $\Delta E_0 < \Delta \hat{E}_0$ , то прогнозируемое соединение принадлежит к классу соединений с выраженной RAGE-ингибирующей активностью;

2) если же  $\Delta E_0 \geq \Delta \hat{E}_0$ , то прогнозируемое соединение принадлежит к классу соединений с низкой RAGE-ингибирующей активностью.

С использованием сформированной модели рассчитаны общая точность прогноза  $Acc$ , чувствительность  $Sens$  (точность прогноза высоко активных соединений) и специфичность  $Spec$  (точность прогноза не высоко активных соединений). С помощью биномиального критерия  $z$  [15] оценена статистическая достоверность  $p$  показателей точности прогноза.

Нейросетевая модель зависимости активности от энергии множественного докинга. Для построения модели зависимости уровня RAGE-ингибирующей активности от спектра энергий множественного докинга химических соединений  $\Delta E_1 \dots \Delta E_{27}$  была использована технология искусственных нейронных сетей [16]. В соответствии с теоремой Колмогорова [5], 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 ( $\Delta E_1 \dots \Delta E_{27}$ );  $m$  – число скрытых нейронов, устанавливается программой от 3 до 26, поскольку  $2 < m < k$ . В модуле Statistica Neural Networks [16] программы Statistica [13] при построении перцептронных сетей для скрытого и выходного слоев используются пять наиболее распространенных активационных функций (Identity, Logictic, Tanh, Exponential, Softmax) [16], 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  $Acc$ , чувствительность  $Sens$  и специфичность  $Spec$ . С помощью биномиального критерия  $z$  [15] оценена статистическая достоверность  $p$  показателей точности прогноза.

### РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ И ИХ ОБСУЖДЕНИЕ

Для модели зависимости уровня RAGE-ингибирующей активности химических соединений от энергии их простого докинга в специфический связывающий сайт  $\Delta E_0$  получено граничное значение  $\Delta \dot{E}_0 = -7.0$  Ккал/моль, разделяющее высоко активные и не высоко активные соединения.

В случае нейросетевой модели зависимости уровня RAGE-ингибирующей активности от спектра энергий множественного докинга химических соединений  $\Delta E_1 \dots \Delta E_{27}$  получена нейронная сеть следующих характеристик: MLP 27-12-2 (Exponential, Logistic), точность обучения 93,8 %, тестирования 81,5 %, также классифицирующая высоко активные и не высоко активные соединения.

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

### Точность прогноза уровня RAGE-ингибирующей активности на основе энергий простого и множественного докинга

Показатель точности прогноза	Значения для модели прогноза с использованием	
	$\Delta E_0$	$\Delta E_1 \dots \Delta E_{27}$
$Acc, \%$	17.5	85.2
$Sens, \%$	20.2	93.9
$Spec, \%$	14.3	75.0
$z_0$	3.32	6.72
$z_a$	5.67	8.38
$z_n$	6.81	4.76
$p_0$	1.000	$9.24 \times 10^{-12}$
$p_a$	1.000	$<1 \times 10^{-15}$
$p_n$	1.000	$9.84 \times 10^{-7}$

Примечание:  $Acc, Sens, Spec$  – точность, чувствительность, специфичность прогноза;  $z_0, z_a, z_n$  – биномиальный критерий для  $Acc, Sens, Spec$ ;  $p_0, p_a, p_n$  – значимость  $Acc, Sens, Spec$ .

Точность прогноза RAGE-ингибирующей активности с использованием модели на основе энергии докинга в специфический сайт  $\Delta E_0$  по всем показателям является статистически незначимой и меньше точности случайного угадывания, составляющей 50 %. Таким образом, данная модель является неадекватной.

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

Таким образом, на примере RAGE-ингибирующей активности показано, что нейросетевая модель на основе спектра энергий докинга химических соединений во множество пространств RAGE-рецептора является намного более точной, чем модель, полученная на основе единичной энергии их докинга в специфический связывающий сайт этого же рецептора.

### ЗАКЛЮЧЕНИЕ

С использованием кластерного анализа построена модель зависимости уровня RAGE-ингибирующей активности химических соединений от энергии их простого докинга в специфический сайт. С использованием технологии искусственных нейронных сетей построена модель зависимости уровня RAGE-ингибирующей активности от спектра энергий множественного докинга химических соединений по всему пространству RAGE-рецептора.

Статистическая достоверность и точность прогноза уровня RAGE-ингибирующей 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 сайт RAGE-рецептора.

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

#### СПИСОК ИСТОЧНИКОВ

1. Ansari N.A., Rasheed Z. Non-enzymatic glycation of proteins: from diabetes to cancer. *Biochemistry (Moscow) Supplement Series B: Biomedical Chemistry (Int. Ed.)*. 2009;3(4):335–345.
2. Yan S.F., Ramasamy R., Schmidt A.M. Mechanisms of disease: advanced glycation end-products and their receptor in inflammation and diabetes complications. *Nature Clinical Practice Endocrinology & Metabolism*. 2008;4(5):285–293.
3. World Health Organization: Diabetes, 2019. URL: <https://www.who.int/news-room/fact-sheets/detail/diabetes>.
4. 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;3:57–61.
5. Колмогоров А.Н. О представлении непрерывных функций нескольких переменных в виде суперпозиции непрерывных функций одного переменного. *Доклады АН СССР*. 1957;114(5):953–956.
6. Васильев П.М., Яналиева Л.Р., Спасов А.А. и др. Свидетельство о государственной регистрации базы данных № 2019620160. Ингибиторы рецепторов конечных продуктов гликирования. № 2019620045; заявл. 11.01.2019; зарег. 24.01.2019; опубл. 24.01.2019. Официальный бюллетень «Программы для ЭВМ. БД. ТИМС». 2019;2. URL: <http://www1.fips.ru/ofpstorage/Doc/PrEVM/RUNWDB/000/002/019/620/160/2019620160-00001/DOCUMENT.PDF>.
7. 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.
8. MarvinSketch, ChemAxon Kft. URL: <https://chemaxon.com/products/marvin>.
9. MOPAC, Stewart Computational Chemistry. URL: <http://openmopac.net/>.
10. Yatime L., Andersen G.R. Structural insights into the oligomerization mode of the human receptor for advanced glycation end-products. *The FEBS Journal*. 2013;280(24):6556–6568. doi: 10.1111/febs.12556.
11. 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.

12. 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.

13. Hilbe J.M. Statistica 7: an overview. *The American Statistician*. 2007;61(1):91–94.

14. Gupta M., Sharma R., Kumar A. Docking techniques in pharmacology: How much promising? *Computational Biology and Chemistry*. 2018;76:210–217.

15. Глотов Н.В., Животовский Л.А., Хованов Н.В., Хромов-Борисов Н.Н. Биометрия. Л.: Изд-во Ленингр. ун-та, 1982. 264 с.

16. Нейронные сети. Statistica Neural Networks: Методология и технология современного анализа данных. М.: Горячая линия – Телеком, 2008. 392 с.

#### REFERENCES

1. Ansari N.A., Rasheed Z. Non-enzymatic glycation of proteins: from diabetes to cancer. *Biochemistry (Moscow) Supplement Series B: Biomedical Chemistry (Int. Ed.)*. 2009;3(4):335–345.
2. Yan S.F., Ramasamy R., Schmidt A.M. Mechanisms of disease: advanced glycation end-products and their receptor in inflammation and diabetes complications. *Nature Clinical Practice Endocrinology & Metabolism*. 2008;4(5):285–293.
3. World Health Organization: Diabetes, 2019. URL: <https://www.who.int/news-room/fact-sheets/detail/diabetes>.
4. Vassiliev P.M., Kochetkov A.N., Spasov A.A., Perfiliev M.A. The energy spectrum of multiple docking as a multidimensional metric of the affinity of chemical compounds to pharmacologically relevant biotargets. *Volgogradskii nauchno-meditsinskii zhurnal = Volgograd scientific and medical journal*. 2021;3:57–61 (In Russ.).
5. 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.).
6. Vassiliev P.M., Yanaliyeva L.R., Spasov A.A., et al. Certificate of state registration of the database No. 2019620160. Inhibitors of receptors for advanced glycation end products. No. 2019620045; application 11.01.2019; registered 24.01.2019; publ. 24.01.2019, *Ofitsial'nyy byulleten' "Programmy dlya EVM. BD. TIMS" = Official bulletin "Computer programs. DB. TIMS"*. 2019;2. (In Russ.) URL: <http://www1.fips.ru/ofpstorage/Doc/PrEVM/RUNWDB/000/002/019/620/160/2019620160-00001/DOCUMENT.PDF>.
7. 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.
8. MarvinSketch, ChemAxon Kft. URL: <https://chemaxon.com/products/marvin>.
9. MOPAC, Stewart Computational Chemistry. URL: <http://openmopac.net/>.

10. Yatime L., Andersen G.R. Structural insights into the oligomerization mode of the human receptor for advanced glycation end-products. *The FEBS Journal*. 2013;280(24):6556–6568. doi: 10.1111/febs.12556.

11. 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.

12. 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.

13. Hilbe J.M. *Statistica 7: an overview. The American Statistician*. 2007;61(1):91–94.

14. Gupta M., Sharma R., Kumar A. Docking techniques in pharmacology: How much promising? *Computational Biology and Chemistry*. 2018;76:210–217.

15. Glotov N.V., Zhivotovskiy L.A., Khovanov N.V., Khromov-Borisov N.N. *Biometrics*. Leningrad, Leningrad University Publishing House, 1982, 264 p. (In Russ.).

16. *Neural networks. Statistica Neural Networks: Methodology and technology for modern data analysis*. Moscow, Hot line – Telecom, 2008, 392 p. (In Russ.).

**Конфликт интересов.** Авторы декларируют отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

#### Информация об авторах

*П.М. Васильев* – доктор биологических наук, старший научный сотрудник Высшей аттестационной комиссии, заведующий лабораторией информационных технологий в фармакологии и компьютерного моделирования лекарств, Научный центр инновационных лекарственных средств с опытно-промышленным производством, профессор кафедры фармакологии и биоинформатики, Волгоградский государственный медицинский университет, Волгоград, Россия; pvassiliev@mail.ru

*М.А. Перфильев* – младший научный сотрудник лаборатории информационных технологий в фармакологии и компьютерного моделирования лекарств, Научный центр инновационных лекарственных средств с опытно-промышленным производством, аспирант кафедры фармакологии и биоинформатики, Волгоградский государственный медицинский университет, Волгоград, Россия; <sup>✉</sup> maxim.firu@yandex.com

*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; akocha@mail.ru

Статья поступила в редакцию 12.07.2023; одобрена после рецензирования 30.10.2023; принята к публикации 28.11.2023.

**Competing interests.** The authors declare that they have no competing interests.

#### Information about the authors

*P.M. Vasiliev* – Doctor of Biological Sciences, Senior Researcher at the Higher Attestation Commission, Head of the Laboratory of Information Technologies in Pharmacology and Computer Modeling of Medicines, Scientific Center for Innovative Medicines with Pilot Production, Professor of the Department of Pharmacology and Bioinformatics, Volgograd State Medical University, Volgograd, Russia; pvassiliev@mail.ru

*M.A. Perfiliev* – a junior researcher at the Laboratory of Information Technologies in Pharmacology and Computer Modeling of Medicines, Scientific Center for Innovative Medicines with Pilot Production, postgraduate student of the Department of Pharmacology and Bioinformatics, Volgograd State Medical University, Volgograd, Russia; <sup>✉</sup> maxim.firu@yandex.com

*A.N. Kochetkov* – System administrator, software engineer at 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; akocha@mail.ru

The article was submitted 12.07.2023; approved after reviewing 30.10.2023; accepted for publication 28.11.2023.