

Павел Михайлович Васильев^{1✉}, **Арина Владимировна Голубева**²,
Максим Алексеевич Перфильев³, **Андрей Николаевич Кочетков**⁴

^{1, 2, 3} Волгоградский государственный медицинский университет, Волгоград, Россия

^{1, 2, 3, 4} Научный центр инновационных лекарственных средств, Волгоград, Россия

^{1✉} pvassiliev@mail.ru, <https://orcid.org/0000-0002-8188-5052>

² arina_arina_golubeva@mail.ru, <https://orcid.org/0000-0001-8268-8811>

³ maxim.firu@yandex.com, <https://orcid.org/0000-0002-5326-3299>

⁴ akocha@mail.ru, <https://orcid.org/0000-0003-3077-1837>

МАТРИЧНАЯ СВЕРТКА СПЕКТРА ЭНЕРГИЙ МНОЖЕСТВЕННОГО ДОКИНГА ДЛЯ НЕЙРОСЕТЕВОГО МОДЕЛИРОВАНИЯ МУЛЬТИТАРГЕТНОЙ ФАРМАКОЛОГИЧЕСКОЙ АКТИВНОСТИ ХИМИЧЕСКИХ СОЕДИНЕНИЙ

Аннотация. Изложен алгоритм матричной свертки спектра энергий множественного докинга лигандов в релевантные белки-мишени. На примере анксиолитической активности выполнена матричная свертка спектра энергий мультитаргетного множественного докинга в 22 сверточных переменных. Методами однофакторного дисперсионного анализа, дискриминантного анализа и нейросетевого анализа показана высокая статистическая достоверность использования полученных сверточных переменных для оценки интегральной мультитаргетной аффинности лигандов к совокупности релевантных биомишеней. Как итог, разработан новый метод построения сверточных нейронных сетей для поиска *in silico* фармакологически активных веществ на основе матричной свертки спектров энергий мультитаргетного множественного докинга.

Ключевые слова: *in silico*, фармакологическая активность, релевантные биомишени, мультитаргетный множественный докинг, спектры энергий докинга, матричная свертка, ANOVA, дискриминантный анализ, искусственные нейронные сети

VOLGOGRAD SCIENTIFIC AND MEDICAL JOURNAL. 2024. VOL. 21, NO. 4. P. 54–59.

ORIGINAL ARTICLE

doi: <https://doi.org/10.19163/2658-4514-2024-21-4-54-59>

Pavel M. Vassiliev^{1✉}, **Arina V. Golubeva**², **Maxim A. Perfiliev**³, **Andrey N. Kochetkov**⁴

^{1, 2, 3} Volgograd State Medical University, Volgograd, Russia

^{1, 2, 3, 4} Research Center of Innovative Drugs, Volgograd, Russia

^{1✉} pvassiliev@mail.ru, <https://orcid.org/0000-0002-8188-5052>

² arina_arina_golubeva@mail.ru, <https://orcid.org/0000-0001-8268-8811>

³ maxim.firu@yandex.com, <https://orcid.org/0000-0002-5326-3299>

⁴ akocha@mail.ru, <https://orcid.org/0000-0003-3077-1837>

MATRIX CONVOLUTION OF MULTIPLE DOCKING ENERGY SPECTRUM FOR NEURAL NETWORK MODELING OF MULTI-TARGET PHARMACOLOGICAL ACTIVITY OF CHEMICAL COMPOUNDS

Abstract. An algorithm for matrix convolution of the energy spectrum of multiple docking of ligands into relevant target proteins is presented. Using anxiolytic activity as an example, a matrix convolution of the energy spectrum of multi-target multiple docking into 22 convolutional variables was performed. Using the methods of ANOVA, discriminant analysis and neural network analysis, high statistical significance of using the obtained convolutional variables for assessing the integral multi-target affinity of ligands to a set of relevant biotargets was shown. As a result, a new method for constructing convolutional neural networks for *in silico* search for pharmacologically active substances based on matrix convolution of the energy spectra of multi-target multiple docking was developed.

Keywords: *in silico*, pharmacological activity, relevant biotargets, multi-target multiple docking, docking energy spectra, matrix convolution, ANOVA, discriminant analysis, artificial neural networks

Методы машинного обучения и искусственных нейронных сетей широко применяются мировым научным сообществом в поиске фармакологически активных веществ [1]. При этом обычно используются общепринятые архитектуры нейронных сетей, в частности, сверточные нейронные сети, разработанные для распознавания зрительных образов [2].

Однако при нейросетевом моделировании фармакологических соединений физический смысл и внутренняя структура химико-биологических данных значительно отличаются от таковых для изображений. В связи с этим создание новых методов построения нейросетевых моделей, описывающих зависимости между фармакологической активностью и структурой химических соединений, является актуальной и научно востребованной задачей.

В наших работах [3–5] было показано, что использование метода множественного докинга для построения нейросетевых моделей зависимостей между системными видами фармакологической активности и спектром энергий множественного докинга химических соединений в релевантные биомишени позволяет создавать высокоточные и статистически очень достоверные прогнозные модели.

Следует отметить, что обработка спектров энергий мультитаргетного множественного докинга, вычисленных для совокупности нескольких биомишеней, значительно увеличивает размерность признакового пространства, и поэтому для использования подобных данных в нейросетевом моделировании необходимо разработать валидные методы свертки.

ЦЕЛЬ РАБОТЫ

Доказательство методами многомерной статистики и нейросетевого моделирования валидности использования матричной свертки спектров энергий мультитаргетного множественного докинга как статистически высоко достоверной метрики аффинности химических соединений к совокупности фармакологически релевантных биомишеней.

МЕТОДИКА ИССЛЕДОВАНИЯ

Основная гипотеза. Совокупность значений энергии докинга, рассчитанная для множества пространств нескольких релевантных биомишеней и обработанная с помощью предлагаемой процедуры матричной свертки, поз-

воляет адекватно моделировать воздействие множества молекул лиганда на совокупность нескольких белков-мишеней в целом и с высокой статистической достоверностью отражает системную мультитаргетную фармакологическую активность химических соединений, что позволяет с высокой точностью прогнозировать уровень этой активности.

Задачи исследования. Для достижения поставленной цели необходимо было решить следующие задачи:

1. Разработать алгоритм матричной свертки спектра энергий мультитаргетного множественного докинга.
2. Сформировать верифицированную обучающую выборку по уровню активности и спектрам энергий мультитаргетного множественного докинга известных соединений, испытанных на модельную фармакологическую активность.
3. Провести матричную свертку спектров энергий мультитаргетного множественного докинга.
4. Выполнить однофакторный дисперсионный анализ, устанавливающий статистическую значимость уровня модельной активности известных соединений от параметров матричной свертки спектров энергий мультитаргетного множественного докинга.
5. Провести с помощью дискриминантного анализа оценку точности прогноза активности известных соединений с использованием в качестве независимых переменных параметров матричной свертки спектров энергий мультитаргетного множественного докинга.
6. Выполнить обучение нейронных сетей на указанных сверточных переменных и для лучшей построенной нейросети произвести оценку точности прогноза модельной фармакологической активности.

Алгоритм матричной свертки мультитаргетного спектра энергий множественного докинга

Алгоритм матричной свертки основан на построении матрицы попарных расстояний между значениями переменных и последующем вычислении определителя полученной матрицы. Достоинством алгоритма является отсутствие ограничений на число сворачиваемых переменных, так как он основан на геометрическом подходе. Для одной биомишени и одного соединения k -спектр рассчитанных в результате множественного докинга 135 значений ΔE можно рассматривать как ко-

ординаты точки в многомерном пространстве. В соответствии с этим, матрица расстояний Хэм-

минга между парами значений переменных i и j может быть определена следующим образом:

$$D_k = \{ D_{kij} \} = \{ | \Delta E_{ik} - \Delta E_{jk} | \}, i, j = 1..M, i \neq j, k = 1..N, \quad (1)$$

где D_{kij} – расстояние Хэмминга между энергиями ΔE_i и ΔE_j , $i \neq j$;

ΔE_{ik} – значение энергии i для соединения k , $k=1..N$;

ΔE_{jk} – значение энергии j для соединения k , $k=1..N$;

M – число сворачиваемых значений энергии, равно 135;

N – число соединений.

Для соединения k значение свертки 135 энергий множественного докинга для одной

биомишени вычисляется как определитель матрицы D_k .

$$W_k = \det D_k, k = 1..N. \quad (2)$$

В результате матричной свертки мультитаргетный множественный аффинитет каждого соединения в отношении L -релевантных биомишеней будет представлен L -сверточными переменными.

По смыслу, определитель матрицы равен ориентированному объему многомерного параллелепипеда, заданного векторами в виде столбцов матрицы. В соответствии с формулой (1), элементы матрицы расстояний D_k характеризуют вариативность значений множественного докинга ΔE . Таким образом, если соединение k связывается со всеми пространствами данного белка-мишени с примерно одинаковой интенсивностью, показатель матричной свертки W_k будет близок к нулю. Если же соединение k связывается с пространствами данного белка-мишени с разной интенсивностью, показатель матричной свертки W_k будет принимать тем большие значения, чем больше различия в значениях ΔE . Фактически это означает, что в этом белке имеются «особые» области связывания, с которыми рассматриваемое соединение взаимодействует с гораздо большей интенсивностью, чем с остальными.

Описанный алгоритм был реализован на языке Borland Delphi в виде программы MatrConv v06.11.24.

Формирование обучающей выборки

В качестве модельной фармакологической активности была выбрана анксиолитическая активность, системный характер которой определяется мультитаргетным воздействием соединений на достаточно большое число биомишеней. Исходная верифицированная обучающая

выборка по анксиолитической активности и спектрам энергий мультитаргетного множественного докинга 537 известных соединений в 22 релевантных белка-мишени была сформирована ранее в работе [5] и включала 457 активных и 80 неактивных соединений. Каждое соединение характеризовалось $135 \cdot 22 = 2\,970$ значениями ΔE .

Матричной свертке было подвергнуто $135 \cdot 22 \cdot 537 = 1\,594\,890$ значений ΔE , в результате чего была получена матрица свертчных параметров спектров энергий мультитаргетного множественного докинга размером 22 переменных на 537 наблюдений.

Однофакторный дисперсионный анализ

Для показателя наличия/отсутствия анксиолитической активности с помощью программы Statistica 7 [6] был выполнен однофакторный дисперсионный анализ (ANOVA) [7] зависимости указанного фактора от многомерной матрицы значений свертчных переменных спектров энергий мультитаргетного множественного докинга, полученных в результате матричной свертки. Рассчитаны величины критерия лямбда Уилкса Λ , соответствующего ему критерия Фишера F , и определена статистическая достоверность p используемых свертчных переменных.

Дискриминантный анализ

С целью оценки точности прогноза наличия/отсутствия анксиолитической активности с помощью программы Statistica 7 [6] был выполнен дискриминантный анализ [8], в котором независимыми переменными также служили показатели, полученные в результате матричной свертки спектров энергий мультитаргетного

множественного докинга. По результатам прогноза на объединенной обучающей выборке были рассчитаны общая точность прогноза Acc и, по данным ROC-анализа, площадь под кривой AUC_{ROC} . С помощью биномиального критерия z [9] оценена статистическая достоверность p точности прогноза.

Нейросетевое моделирование

Выполняли с помощью программы Statistica 7 [6]. В качестве архитектуры нейронной сети был выбран двухслойный перцептрон MLP $k-m-2$ с узким горлом. Здесь k – число входных нейронов, в данном случае 22; m – число скрытых нейронов, устанавливается программой от 3 до 21, поскольку $2 < m < k$. При обучении нейросетей для скрытого слоя использовались активационные функции Identity, Logistic, Tanh, Exponential, выбор которых производился программой.

Для выходных нейронов использовалась активационная функция Softmax, которая является наиболее эффективной при обучении классификационных нейронных сетей [10]. При формировании обучающих и тестовых подвыборок использовался метод Монте-Карло.

С целью достижения наилучшего результата обучения, число моделируемых сетей было установлено в 4 000, из которых программой автоматически отбирались 400. После окончания обучения из 400 лучших отобранных программой нейросетей по совокупности характеристик

точности вручную отбирали одну наилучшую. Как и в случае дискриминантного анализа, для лучшей построенной нейросети по результатам прогноза на объединенной обучающей выборке были рассчитаны общая точность прогноза Acc и по данным ROC-анализа – площадь под кривой AUC_{ROC} . С помощью биномиального критерия z [9] оценена статистическая достоверность p точности прогноза.

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ И ИХ ОБСУЖДЕНИЕ

Результаты однофакторного дисперсионного анализа анксиолитической активности показывают, что совокупность параметров матричной свертки спектров энергий мультитаргетного множественного докинга является статистически высоко достоверной интегральной метрикой аффинности химических соединений к релевантным белкам-мишеням.

Вычисленные критерий лямбда Уилкса $\Lambda = 0,914$, соответствующий ему критерий Фишера $F(22,514) = 2,202$ и статистическая достоверность используемых сверточных переменных $p = 1,38 \cdot 10^{-3}$ доказывают, что предлагаемый сверточный метод можно применять в поиске *in silico* фармакологически активных соединений.

В таблице приведено сравнение результатов прогноза анксиолитической активности методом дискриминантного анализа и с помощью сверточной нейросетевой модели.

Точность прогноза анксиолитической активности с помощью дискриминантного анализа и с использованием сверточной нейросетевой модели

Показатель точности прогноза	Значение для метода прогноза	
	Дискриминантный анализ	Нейронная сеть ¹
$Acc, \%$ ²	84,0	87,9
$AUC_{ROC}, \%$ ³	62,9	65,8
z ⁴	4,21	5,16
p ⁵	$1,26 \cdot 10^{-5}$	$1,22 \cdot 10^{-7}$

Примечание.

¹ Архитектура лучшей нейронной сети MLP 22-6-2 (Exponential, Softmax).

² Общая точность прогноза на объединенной обучающей выборке.

³ Площадь под кривой по данным ROC-анализа.

⁴ Биномиальный критерий [9].

⁵ Статистическая достоверность точности прогноза по биномиальному критерию [9].

Показательно, что по результатам как дискриминантного, так и нейросетевого анализа совокупность параметров матричной свертки спектров энергий мультитаргетного множественного докинга

является статистически высоко достоверным способом представления интегральной аффинности химических соединений к релевантным белкам-мишеням при прогнозе фармакологической

активности. В дискриминантном анализе статистическая достоверность точности прогноза составила $p = 1,26 \cdot 10^{-5}$, а при нейросетевом моделировании статистическая достоверность точности прогноза составила $p = 1,22 \cdot 10^{-7}$.

Приведенные в таблице результаты дискриминантного и нейросетевого анализов полностью согласуются с результатами дисперсионного анализа и подтверждают вывод о том, что предлагаемый метод матричной свертки спектров энергий мультитаргетного множественного докинга химических соединений к релевантным белкам-мишеням можно эффективно применять в поиске *in silico* фармакологически активных веществ. Таким образом, на примере анксиолитической активности, с использованием двух методов многомерной статистики – однофакторного дисперсионного анализа и дискриминантного анализа, и технологии искусственных нейронных сетей, доказано, что матричная свертка спектров энергий мультитаргетного множественного докинга является статистически высоко достоверным валидным методом оценки интегральной аффинности химических соединений к совокупности фармакологически релевантных биомишеней.

ЗАКЛЮЧЕНИЕ

1. Разработан алгоритм матричной свертки спектра энергий множественного докинга лигандов в релевантные белки-мишени.

2. Методами многомерной статистики и нейросетевого моделирования доказана валидность использования матричной свертки спектров энергий мультитаргетного множественного докинга как статистически высоко достоверной метрики аффинности химических соединений к совокупности фармакологически релевантных биомишеней.

3. Как итог создан новый метод построения сверточных нейронных сетей для поиска *in silico* фармакологически активных веществ на основе матричной свертки спектров энергий мультитаргетного множественного докинга.

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 сверточных нейронных сетей различной архитектуры».

СПИСОК ИСТОЧНИКОВ

1. Sarkar C., Das B., Rawat V. S. et al. Artificial Intelligence and Machine Learning Technology Driven Modern Drug Discovery and Development. *International Journal Molecular Sciences*. 2023;24(3):2026.
2. Leijnen S., Van Veen F. The Neural Network Zoo. *Proceedings*. 2020;47(4):9.
3. 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.
4. Васильев П. М., Перфильев М. А., Кочетков А. Н. Нейросетевое моделирование зависимости RAGE-ингибирующей активности химических соединений от спектра энергий множественного докинга. *Вестник Волгоградского государственного медицинского университета*. 2023;20(4):157–161.
5. Vassiliev P. M., Perfiliev M. A., Golubeva A. V. et al. Multi-target neural network model of anxiolytic activity of chemical compounds based on correlation convolution of energy spectra of multiple docking. *XXX symposium "Bioinformatics and computer-aided drug discovery": Proceedings book. Moscow: Institute of Biomedical Chemistry*. 2024. 46 p.
6. Hilbe J. M. Statistica 7: an overview. *The American Statistician*. 2007;61(1):91–94.
7. Аренс Х., Лёйтер Ю. Многомерный дисперсионный анализ. Москва: Финансы и статистика, 1985. 230 с.
8. Ким Дж.-О., Мьюллер Ч. У., Клекка У. Р. и др. Факторный, дискриминантный и кластерный анализ. Москва: Финансы и статистика, 1989. 215 с.
9. Глотов Н. В., Животовский Л. А., Хованов Н. В., Хромов-Борисов Н. Н. Биометрия. Ленинград: Изд-во Ленингр. ун-та, 1982. 264 с.
10. Dubin U. Cross-Entropy Method: Theory with Applications. Chisinau (Moldova): LAMBERT Academic Publishing, 2013. 148 p.

REFERENCES

1. Sarkar C., Das B., Rawat V. S. et al. Artificial Intelligence and Machine Learning Technology Driven Modern Drug Discovery and Development. *International Journal Molecular Sciences*. 2023;24(3):2026.
2. Leijnen S., Van Veen F. The Neural Network Zoo. *Proceedings*. 2020;47(4):9.
3. 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. *Volgogradskiy nauchno-meditinskiy zhurnal = Volgograd Journal of Medical Research*. 2021;3:57–61. (In Russ.).

4. Vassiliev P. M., Kochetkov A. N., Perfilev M. A. Neural network modeling of the dependence of GABAA agonistic activity of chemical compounds on the spectrum of multiple docking energies. *Vestnik Volgogradskogo gosudarstvennogo meditsinskogo universiteta = Journal of Volgograd State Medical University*. 2022; 19(4):88–93. (In Russ.).

5. Vassiliev P. M., Perfilev M. A., Golubeva A. V. et al. Multi-target neural network model of anxiolytic activity of chemical compounds based on correlation convolution of energy spectra of multiple docking. XXX symposium "Bioinformatics and computer-aided drug discovery": Proceedings book. Moscow: Institute of Biomedical Chemistry. 2024. 46 p.

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

7. Arens H., Loyter Y. Multivariate Analysis of Variance. Moscow: Finance and Statistics, 1985. 230 p. (In Russ.).

8. Kim J.-O., Mueller C. W., Klekka W. R. et al. Factor, discriminant, and cluster analysis. Moscow: Finance and Statistics, 1989. 215 p. (In Russ.).

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

10. Dubin U. Cross-Entropy Method: Theory with Applications. Chisinau (Moldova): LAMBERT Academic Publishing, 2013. 148 p.

Информация об авторах

П. М. Васильев – доктор биологических наук, с.н.с. ВАК (доцент), заведующий лабораторией информационных технологий в фармакологии и компьютерного моделирования лекарств, профессор кафедры фармакологии и биоинформатики

А. В. Голубева – младший научный сотрудник лаборатории информационных технологий в фармакологии и компьютерного моделирования лекарств, ассистент кафедры фармакологии и биоинформатики

М. А. Перфильев – младший научный сотрудник лаборатории информационных технологий в фармакологии и компьютерного моделирования лекарств, ассистент кафедры фармакологии и биоинформатики

А. Н. Кочетков – системный администратор, инженер-программист лаборатории информационных технологий в фармакологии и компьютерного моделирования лекарств

Авторы заявляют об отсутствии конфликта интересов.

Статья поступила в редакцию 29.09.2024; одобрена после рецензирования 21.10.2024; принята к публикации 12.11.2024.

Information about the authors

P. M. Vasiliev – Doctor of Biological Sciences, Senior Researcher VAK (Associate Professor), Head of the Laboratory of Information Technologies in Pharmacology and Computer Modeling of Drugs, Professor of the Department of Pharmacology and Bioinformatics

A. V. Golubeva – Junior Researcher Laboratory of Information Technologies in Pharmacology and Computer Modeling of Drugs, Assistant of the Department of Pharmacology and Bioinformatics

M. A. Perfil'ev – Junior Researcher of the Laboratory of Information Technologies in Pharmacology and computer modeling of drugs, assistant of the Department of Pharmacology and Bioinformatics

A. N. Kochetkov – system administrator, software engineer at the laboratory of information technologies in pharmacology and computer modeling of drugs

The authors declare no conflicts of interests.

The article was submitted 29.09.2024; approved after reviewing 21.10.2024; accepted for publication 12.11.2024.