

П. М. Васильев, А. Н. Кочетков, А. А. Спасов, М. А. Перфильев

Волгоградский государственный медицинский университет,
Научный центр инновационных лекарственных средств, Россия

СПЕКТР ЭНЕРГИЙ МНОЖЕСТВЕННОГО ДОКИНГА КАК МНОГОМЕРНАЯ МЕТРИКА АФФИННОСТИ ХИМИЧЕСКИХ СОЕДИНЕНИЙ К ФАРМАКОЛОГИЧЕСКИ РЕЛЕВАНТНЫМ БИОМИШЕНЯМ

УДК 615.015.11:544.165:575.112::[544.187.2+519.237.4]

Изложена основная гипотеза множественного докинга и описан алгоритм формирования совокупности пространств для его проведения. Выполнен простой и множественный докинг известных ингибиторов RAGE (рецептор конечных продуктов гликирования). Рассчитаны энергии простого докинга и спектры энергий множественного докинга этих соединений. Методом однофакторного дисперсионного анализа и дискриминантного анализа показано, что матрица энергий множественного докинга является статистически намного более значимой метрикой аффинности лигандов к фармакологически релевантным биомишеням, чем вектор энергий простого докинга.

Ключевые слова: *in silico*, множественный докинг, спектр энергий, аффинность лигандов, биомишень, многомерная метрика, ANOVA, дискриминантный анализ.

P. M. Vasiliev, A. N. Kochetkov, A. A. Spasov, M. A. Perfiliev

THE ENERGY SPECTRUM OF MULTIPLE DOCKING AS A MULTI-DIMENSIONAL METRIC OF THE AFFINITY OF CHEMICAL COMPOUNDS TO PHARMACOLOGICALLY RELEVANT BIO-TARGETS

The main hypothesis of multiple docking is stated and an algorithm for the formation of a set of spaces for its implementation is described. Simple and multiple docking of known RAGE inhibitors was performed. The energies of simple docking and the energy spectra of multiple docking of these compounds are calculated. It has been shown using one-way ANOVA and discriminant analysis that the multiple docking energy matrix is a statistically much more significant metric of the ligand affinity for a pharmacologically relevant biotargets than the simple docking energy vector.

Key words: *in silico*, multiple docking, energy spectrum, ligand affinity, bio-target, multi-dimensional metric, ANOVA, discriminant analysis.

В настоящее время наиболее популярным методом поиска *in silico* новых лекарственных веществ является молекулярный докинг. Общепринятая схема его проведения включает определение в 3D-модели белка-мишени пространства, охватывающего специфический сайт связывания, с последующим расчетом минимальной энергии взаимодействия с этим сайтом 3D-структур докируемых соединений [11]. При этом предполагается, что именно взаимодействие одной молекулы лиганда с локально определенным специфическим сайтом обуславливает конформационные изменения всей биомишени в целом и порождает ее биологический отклик, а единичное значение энергии этого взаимодействия является универсальной метрикой аффинности конкретного химического соединения к данному белку [6]. Часто в рамках этой стандартной схемы рассматриваются дополнительно еще несколько аллостерических сайтов. Но и в этом случае считается, что релевантные

биологическому эффекту общие конформационные изменения белка-мишени детерминированы единичным значением энергии связывания молекулы лиганда с заданным сайтом.

Между тем, очевидно, что в биологически активных концентрациях с биомишенью связывается не одна молекула лиганда, а очень большое их число, причем взаимодействуют они не с изолированным от всей структуры белка сайтом связывания, а со всей поверхностью белка-мишени. Указанное несоответствие является одной из причин весьма низкой прогностической способности простого докинга, что породило множество замечаний, касающихся эффективности его использования [5]. Применение для учета мультikonформационных изменений молекулярной динамики не устраняет вышеописанных противоречий, поскольку и в этом случае рассчитывается энергия взаимодействия одной молекулы с ограниченным пространством заданного сайта [7].

Таким образом, разработка новой методологии оценки *in silico* аффинности химических соединений к фармакологически релевантным белкам-мишеням, основанная на учете взаимодействия лиганда со всей поверхностью белка, является научно востребованной и весьма актуальной задачей.

ЦЕЛЬ РАБОТЫ

Доказательство методами многомерной статистики валидности использования спектра энергий множественного докинга как статистически высоко достоверной метрики аффинности химических соединений к фармакологически релевантным биомишеням.

МЕТОДИКА ИССЛЕДОВАНИЯ

Основная гипотеза. Совокупность значений энергии докинга, рассчитанная для множества пространств, формируемых по всему объему белка-мишени, позволяет адекватно моделировать воздействие множества молекул лиганда на весь белок в целом и статистически более достоверно отражает аффинность докируемых структур к рассматриваемой биомишени, что позволяет с большей точностью прогнозировать уровень фармакологической активности химических соединений.

Адекватность принимаемой к разработке гипотезы иллюстрируют следующие комплексные факты.

1. Белок-мишень окружен огромным числом взаимодействующих с ним молекул лиганда. Количество молекул высоко активных веществ, действующих на целевую биомишень в наномолярной концентрации, составляет, с учетом числа Авогадро, $6,022 \cdot 10^{23} \times 10^{-9} \approx 6 \cdot 10^{14}$.

2. Структуры низкомолекулярных лигандов кратно помещаются в пространство целевого белка. С помощью программы HyperChem 8.0 [8] был рассчитан Ван-дер-Ваальсов объем 3D-модели соединения RAGE-0023 с высокой ингибирующей активностью из верифицированной базы данных [3], получено значение $V = 1578 \text{ \AA}^3$. Для субъединицы экспериментальной 3D-модели RAGE-рецептора 4LP4 [15] аналогичный расчетный показатель составил $V = 34902 \text{ \AA}^3$. По соотношению объемов молекула RAGE-0023 помещается в пространстве RAGE-рецептора не менее 22 раз.

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

1. Разработать алгоритм построения по всему объему белка-мишени пространств для множественного докинга и создать компьютер-

ную программу для расчета координат указанных пространств.

2. С использованием белка-мишени, для которого имеются верифицированные данные по структуре и уровню активности известных соединений, испытанных на активность в отношении выбранного белка, выполнить простой докинг в его специфический сайт связывания и множественный докинг во все сформированные для этого пространства данного белка.

3. Рассчитать энергии простого докинга и спектры энергий множественного докинга указанных известных соединений.

4. Провести однофакторный дисперсионный анализ, устанавливающий статистическую значимость зависимости уровня активности известных соединений: а) от энергии, полученной в простом докинге; б) от спектра энергий, полученного в множественном докинге.

5. Выполнить с помощью дискриминатного анализа оценку точности прогноза уровня активности известных соединений с использованием в качестве метрик аффинности: а) вектора значений энергий, вычисленного в простом докинге; б) матрицы значений энергии, вычисленной в множественном докинге.

Алгоритм построения пространств для множественного докинга

Для выбранной 3D-модели белка определяют минимальные и максимальные значения координат образующих белок N атомов:

$$U_{\min} = \min_{i=1}^N (U_i), \quad U_{\max} = \max_{i=1}^N (U_i), \quad (1)$$

здесь и далее $U = X, Y, Z$.

Размеры параллелепипеда для множественного докинга:

$$D_U = \frac{U_{\max} - U_{\min}}{k - 1}, \quad (2)$$

где k – кратность построения пространств для множественного докинга; минимально $k = 3$, что соответствует 27 пространствам.

Координаты начала и конца пространств для множественного докинга:

$$U_{\text{Init},1} = U_{\min}, \quad U_{\text{End},1} = U_{\min} + D_U, \\ U_{\text{Init},i} = U_{\text{Init},i-1} + \frac{1}{2} D_U, \quad U_{\text{End},i} = U_{\text{Init},i} + D_U, \quad i = 2, \dots, k. \quad (3)$$

Описанный алгоритм был реализован на языке Borland Delphi в виде программы MSite v21.04.22.

Выполнение простого и множественного докинга. В качестве пробной биомишени

для проверки сформулированной выше гипотезы был выбран рецептор конечных продуктов гликирования RAGE. В расчетах использовалась его экспериментальная 3D-модель 4LP4 [15] – наиболее точная из трех валидных моделей, ранее найденных при выполнении работы [14], в которой также были определены координаты специфического сайта этого рецептора. Кроме того, на указанной модели 4LP4 с помощью программы MSite v21.04.22, в соответствии с выше описанным алгоритмом, были построены 27 пространств для множественного докинга.

Верифицированные данные по химической структуре и уровню RAGE-ингибирующей активности 183 известных веществ были взяты из оригинальной базы данных [3]. Градированный уровень активности Ind задавали следующими метками: H – высокая (38 соединений); M – умеренная (61 соединение); L – низкая (39 соединений); I – неактивно (45 соединений). Дополнительно формировали объединенные классы, в которых активность соединения Ind_H , Ind_{HM} , Ind_A обозначалась следующими метками: H / nH – “высокая” / “не высокая”; HM / nHM – “высокая или умеренная” / “не высокая или умеренная”; A / I – “активно” / “не активно”.

Оптимизированные 3D-модели указанных соединений были построены последовательно методами молекулярной механики с помощью программы MarvinSketch 17.1.23 [9] и полуэмпирическим квантово-химическим методом PM7 с помощью программы MOPAC2016 [10], с использованием описанной в работе [14] методики.

Ансамблевый докинг проводили с помощью программы AutoDock Vina 1.1.2 [13], каждое соединение в 10 конформерах по 5 раз в каждое пространство докинга, с вычислением по 50 полученным значениям минимальных энергий связывания ΔE , как это описано в [14]. Докинг выполняли отдельно в специфический сайт RAGE и отдельно в каждое из 27 пространств, сформированных для множественного докинга.

Полученная в результате расчетов сводная таблица для последующего статистического анализа включала в себя 32 колонки: шифры соединений; метки уровней RAGE-ингибирующей активности Ind , Ind_H , Ind_{HM} , Ind_A ; энергия докинга в специфический сайт ΔE_0 и энергии множественного докинга в 27 пространств $\Delta E_1 \dots \Delta E_{27}$ RAGE-рецептора.

Однофакторный дисперсионный анализ. Для показателей уровня активности Ind , Ind_H , Ind_{HM} и Ind_A , каждый из которых выполнял роль группировочной переменной, с помощью программы Statistica 8 [12] был выпол-

нен однофакторный дисперсионный анализ (ANOVA) [1] зависимостей указанных факторов от вектора единичных значений энергии докинга в специфический сайт ΔE_0 и от многомерной матрицы значений энергий множественного докинга $\Delta E_1 \dots \Delta E_{27}$. Для каждого из восьми сравнений рассчитаны величины критерия лямбда Уилкса Λ , соответствующего ему критерия Фишера F и определена статистическая достоверность p используемых метрик аффинности.

Дискриминантный анализ. С целью оценки точности прогноза уровня RAGE-ингибирующей активности по показателям Ind , Ind_H , Ind_{HM} и Ind_A с помощью программы Statistica 8 [12] был выполнен дискриминантный анализ [4], в котором независимыми переменными служили энергия докинга ΔE_0 и энергии множественного докинга $\Delta E_1 \dots \Delta E_{27}$. Для каждой из восьми классификаций рассчитаны общая точность прогноза F_0 , точность прогноза активных соединений F_a (чувствительность) и точность прогноза неактивных соединений F_n (специфичность). С помощью биномиального критерия z [2] оценена статистическая достоверность p общей точности прогноза.

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ И ИХ ОБСУЖДЕНИЕ

В табл. 1 приведены результаты дисперсионного анализа, которые доказывают преимущество использования матрицы энергий множественного докинга для оценки аффинности лигандов к релевантной биомисени, в сравнении с вектором энергий простого докинга.

Энергия докинга в специфический сайт ΔE_0 для всех четырех показателей уровня RAGE-ингибирующей активности не является статистически значимой переменной, отражающей аффинность лигандов к данному рецептору – все вычисленные значения p намного превышают общепринятый порог значимости $p = 0,05$.

Напротив, спектр энергий множественного докинга $\Delta E_1 \dots \Delta E_{27}$ является статистически высоко достоверной метрикой аффинности RAGE-лигандов – все вычисленные значения $p < 0,05$. Очень высоко достоверная зависимость с $p = 0,0007$ выявлена для показателя высокого уровня активности Ind_H . На этом основании можно утверждать, что указанный многомерный параметр $\Delta E_1 \dots \Delta E_{27}$ можно весьма эффективно применять в поиске *in silico* соединений с высокой RAGE-ингибирующей активностью.

Приведенные в табл. 2 результаты дискриминантного анализа полностью согласуются с результатами дисперсионного анализа и подтверждают вывод о том, что матрица энергий

множественного докинга является статистически намного более значимой метрикой аффин-

ности лигандов к релевантной биомишени, в сравнении вектором энергий простого докинга.

Таблица 1

Результаты однофакторного дисперсионного анализа зависимостей уровня RAGE-ингибирующей активности от энергий докинга

Показатель достоверности	Значение для показателя уровня активности			
	<i>Ind</i>	<i>Ind_H</i>	<i>Ind_HM</i>	<i>Ind_A</i>
<i>Энергия докинга в специфический сайт ΔE_0</i>				
Δ Уилкса	0.983	0.992	0.985	0.997
<i>F</i> Фишера	1.013	1.487	2.735	0.640
<i>p</i>	0.3883	0.2243	0.0999	0.4248
<i>Энергии множественного докинга $\Delta E_1... \Delta E_{27}$</i>				
Δ Уилкса	0.480	0.711	0.767	0.784
<i>F</i> Фишера	1.576	2.333	1.745	1.583
<i>p</i>	0.0022	0.0007	0.0191	0.0442

Таблица 2

Результаты прогноза методом дискриминантного анализа уровня RAGE-ингибирующей активности на основе энергий докинга

Показатель точности прогноза	Значение для показателя уровня активности			
	<i>Ind</i>	<i>Ind_H</i>	<i>Ind_HM</i>	<i>Ind_A</i>
<i>Энергия докинга в специфический сайт ΔE_0</i>				
F_0 , %	24.6	54.1	50.3	51.4
F_a , %	6.7	60.5	50.5	55.8
F_n , %	29.0	52.4	50.0	37.8
<i>z</i>	0.00	0.76	0.03	0.24
<i>p</i>	0.500	0.224	0.488	0.405
<i>Энергии множественного докинга $\Delta E_1... \Delta E_{27}$</i>				
F_0	59.6	78.7	70.5	73.2
F_a	54.5	78.9	69.7	70.3
F_n	66.7	78.6	71.4	82.2
<i>z</i>	1.81	5.46	3.90	4.41
<i>p</i>	$3.51 \cdot 10^{-2}$	$2.32 \cdot 10^{-8}$	$4.89 \cdot 10^{-5}$	$5.11 \cdot 10^{-6}$

Точность прогноза F_0 RAGE-ингибирующей активности с использованием в качестве независимой переменной энергии докинга в специфический сайт ΔE_0 является статистически незначимой для всех четырех показателей ее уровня. В ряде случаев вычисленные оценки даже меньше точности случайного угадывания, составляющей 50 %.

В то же время при использовании энергий множественного докинга $\Delta E_1... \Delta E_{27}$ все оценки точности прогноза F_0 являются статистически достоверными – все расчетные величины $p < 0,05$. При этом значения всех показателей F_0 , F_a , F_n превышают точность случайного угадывания.

В дискриминантном анализе подтверждено наличие высоко достоверной зависимости от ($\Delta E_1... \Delta E_{27}$) показателя высокого уровня активности *Ind_H* – в этом случае точность про-

гноза составляет $F_0 = 78,7$ %, что соответствует значимости по биномиальному критерию $p = 2,32 \cdot 10^{-8}$.

Таким образом, на примере RAGE-ингибирующей активности, с использованием двух методов многомерной статистики – однофакторного дисперсионного анализа и дискриминантного анализа, доказано, что спектр энергий докинга химических соединений в множество пространств релевантного белка является намного более достоверной метрикой аффинности лигандов к биомишеням, в сравнении с единичной энергией их докинга в специфический сайт.

Следует особо подчеркнуть, что применение множественного докинга для расчета аффинности соединений не требует определения в белке-мишени местоположения специфического сайта связывания.

ЗАКЛЮЧЕНИЕ

1. Методами многомерной статистики доказана валидность использования спектра энергий множественного докинга как статистически высоко достоверной метрики аффинности химических соединений к фармакологически релевантным биомишеням.

2. Статистическая достоверность и точность прогноза уровня фармакологической активности химических соединений с использованием спектра значений энергий множественного докинга существенно превышает аналогичные показатели, полученные с применением единичного значения энергии докинга в специфический сайт белка-мишени.

3. Для расчета аффинности соединений методом множественного докинга не требуется выявлять в белке специфический сайт связывания.

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».

ЛИТЕРАТУРА

1. Аренс, Х. Многомерный дисперсионный анализ / Х. Аренс, Ю. Лейтер. – Москва: Финансы и статистика, 1985. – 230 с. – Текст : непосредственный.
2. Биометрия / Н. В. Глотов, Л. А. Животовский, Н. В. Хованов, Н. Н. Хромов-Борисов. – Ленинград: Изд-во Ленингр. ун-та, 1982. – 264 с. – Текст : непосредственный.
3. Свидетельство о государственной регистрации базы данных № 2019620160. Ингибиторы рецепторов конечных продуктов гликирования / Васильев П. М., Яналиева Л. Р., Спасов А. А., Кочетков А. Н., Ворфоломеева В. В., Клочкиков В. Г. (Россия). – № 2019620045; заявл. 11.01.2019; зарег. 24.01.2019; опубл. 24.01.2019, Официальный бюллетень «Программы для ЭВМ. БД. ТИМС», № 2, 2019. – 1 с. – Текст : электронный.
4. Факторный, дискриминантный и кластерный анализ / Дж.-О. Ким, Ч. У. Мьюллер, У. Р. Клекка [и др.] / Под ред. И. С. Енюкова. – Москва: Финансы и статистика, 1989. – 215 с. – Текст : непосредственный.
5. Chen, Y.-C. Beware of docking! / Y.-C. Chen. – Text (visual) : unmediated // Trends Pharmacol. Sci. – 2015. – Vol. 36, Iss. 2. – P. 78 – 95.
6. Gupta, M. Docking techniques in pharmacology: How much promising? / M. Gupta, R. Sharma, A. Kumar. – Text (visual) : unmediated // Comput. Biol. Chem. – 2018. – Vol. 76. – P. 210 – 217.
7. Hollingsworth, S. A. Molecular Dynamics Simulation for All / S. A. Hollingsworth, R. O. Dror. – Text (visual) : unmediated // Neuron. – 2018. – Vol. 99, Iss. 6. – P. 1129 – 1143.
8. HyperChem Pro 8.0, HyperCube Inc. – Electronic text. – URL : <http://www.hypercubeusa.com/>.
9. MarvinSketch, ChemAxon Kft. – Electronic text. – URL : <https://chemaxon.com/products/marvin>.
10. MOPAC, Stewart Computational Chemistry. – Electronic text. – URL : <http://openmopac.net/>.
11. Pinzi, L. Molecular Docking: Shifting Paradigms in Drug Discovery / L. Pinzi. – Text (visual) : unmediated // Int. J. Mol. Sci. – 2019. – Vol. 20. – Iss. 18. – Art. 4331.
12. Statistica, StatSoft Inc. – Electronic text. – URL : <http://www.statsoft.com/>.
13. Trott, O. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading / O. Trott, A. J. Olson. – Text (visual) : unmediated // J. Comp. Chem. – 2010. – Vol. 31, Iss. 2. – P. 455 – 461.
14. Neural network modeling of multitarget RAGE inhibitory activity / P. M. Vassiliev, A. A. Spasov, L. R. Yanaliyeva, [et al.]. – Text (visual) : unmediated // Biochemistry (Moscow) Supplement Series B: Biomedical Chemistry. – 2019. – Vol. 13. – № 3. – P. 256 – 263.
15. Yatime, L. Structural insights into the oligomerization mode of the human receptor for advanced glycation end-products / L. Yatime, G. R. Andersen. – Text (visual) : unmediated // (2013) FEBS Journal. – 2013. – Vol. 280. – P. 6556 – 6568.