

УДК 547.787:615.281.873

РАЗРАБОТКА СКРИПТА МНОЖЕСТВЕННОГО МОЛЕКУЛЯРНОГО ДОКИНГА И ЕГО ВАЛИДАЦИЯ ПРИ СРАВНЕНИИ ЭКСПЕРИМЕНТОВ *IN SILICO* ДЛЯ ГЛЮКОКИНАЗЫ

Лахвич Ф. Ф., Прокопеня Я. О.

*Учреждение образования «Белорусский государственный медицинский университет»,
г. Минск, Республика Беларусь*

Реферат. На основании анализа и сравнения принципов молекулярного докинга и алгоритма работы различных программ молекулярного моделирования был проведен скриптинг утилиты на базе AutoDock Vina для виртуального скрининга аффинности лигандов к выбранной биологической мишени. Разработанный алгоритм для создания программы множественного молекулярного докинга демонстрирует высокую эффективность, точность при определении аффинности лигандов к протеинам-мишеням. Тестирование программы на различных наборах данных подтвердило ее способность к корректной обработке и анализу больших объемов данных. Сравнение результатов, полученных с помощью разработанной программы, с результатами, полученными с использованием других программ, в частности в исследовании *in silico* аффинности ряда лигандов к глюкокиназе, подтвердило высокую точность и производительность утилиты, а также соответствие результатов, полученных при работе на базе программного обеспечения AutoDock Vina и AutoDock 4.

Ключевые слова: AutoDock Vina, AutoDock 4, аффинность, глюкокиназа, скриптинг утилиты для множественного докинга.

Введение. Методы моделирования *in silico* являются важным этапом в процессе поиска и разработки новых лекарственных средств (ЛС). Эти методы позволяют предварительно отбирать перспективные химические соединения для их последующего синтеза и исследования *in vitro* и/или *in vivo*, а также дальнейших доклинических исследований. Скрининг *in silico* на начальных этапах разработки ЛС предоставляет возможность проверять большие библиотеки лигандов на соответствие заданным критериям. Одним из ключевых методов скрининга является молекулярный докинг, который оценивает степень комплементарности лиганда и активного сайта связывания рецептора, обеспечивающего необходимый биологический эффект. Результаты докинга включают данные о количественной оценке сродства лиганда к активному сайту связывания рецептора, величине энергии связывания, наборе аминокислотных остатков макромолекулы, участвующих во взаимодействии, и химической природе этих взаимодействий.

В последние десятилетия компьютерное моделирование стало неотъемлемой частью разработки новых ЛС. Методы моделирования

позволяют также изучать механизмы взаимодействия биологически активных низкомолекулярных веществ с протеинами и другими биологическими мишенями. Взаимодействие биологически активных низкомолекулярных лигандов с протеинами лежит в основе многих биохимических процессов, и понимание природы этих взаимодействий позволяет создавать более эффективные и специфичные ЛС. Методы прямого компьютерного моделирования, такие как молекулярный докинг и молекулярная динамика, предполагают прямое моделирование и анализ молекулярных структур, взаимодействий и динамики. Эти методы используют детализированные физические модели для предсказания поведения биологически активных веществ на молекулярном уровне.

Основополагающим принципом является применение физических и химических законов для моделирования молекулярных систем и их взаимодействий. Эти методы позволяют исследовать структуру молекул, их динамические движения и механизмы взаимодействия с биологическими мишенями, обеспечивая детальное понимание процессов на молекулярном уровне [1].

С другой стороны, непрямые методы, включая количественное соотношение структура-активность (SAR) и машинное обучение, используют статистические модели и модели машинного обучения для предсказания поведения лекарств на основе эмпирических данных. Эти методы, как правило, не требуют детальных физических моделей, а основываются на выявлении закономерностей и корреляций в больших массивах данных. Основной принцип заключается в анализе существующих данных о химических структурах и их биологической активности для построения моделей, позволяющих предсказать свойства и поведение новых соединений. Эти методы позволяют быстро и эффективно проводить скрининг больших библиотек соединений, тем самым ускоряя процесс открытия лекарств [1].

Цель работы – скриптинг утилиты для множественного докинга в поиске веществ с высокой активностью к выбранной биологической мишени. В рамках исследования также была проведена валидация разработанного скрипта в эксперименте *in silico* при сравнении результатов взаимодействия экспериментальных активаторов глюкокиназы с аллостерическим центром фермента.

Материалы и методы. Информация о трехмерной структуре фермента глюкокиназа (код белка 4RCH) взята с сайта Protein Data Bank (<https://www.rcsb.org>, структурные формулы лигандов получены с сайта DrugBank (<https://go.drugbank.com>). Для подготовки лигандов и протеинов была использована программа визуализации AutoDock Tools. Для проведения докинга использовали AutoDock 4 и связку программ визуализации UCSF Chimera + AutoDock Vina. Для автоматизации процесса множественного докинга был написан скрипт на языке командной оболочки Windows (.bat файл). Этот скрипт использует функции и команды, предоставленные программным обеспечением AutoDock Vina. С помощью данного скрипта был выполнен скрининг взаимодействия молекул-лигандов с протеинами-мишенями. Разработанный скрипт позволил значительно сократить время и повысить эффективность процесса докинга. Для визуализации комплекса, полученного в AutoDock 4, использовали онлайн-сервер Protein-Ligand Interaction Profiler (PLIP), комплекс, полученный в связке AutoDock 4 и UCSF Chimera отображен в самой программе визуализации. Порядок выполнения исследования:

1. Подготовка протеина: использовали AutoDockTools, были убраны нестандартные структуры (другой лиганд, вода).

2. Подготовка лиганда:

– использовали AutoDockTools, функцию prepare ligand;

– использовали программу, написанную на основе OpenBabel.

3. Заполняли config нашего докинга.

4. Запускали программу множественного докинга лигандов.

Результаты и их обсуждение. Программы молекулярного докинга – важнейшие инструменты в вычислительном поиске ЛС, позволяющие исследователям предсказать топологию связывания малой молекулы с целевым протеином. В данном исследовании были известны пространственная структура лиганда и мишени. Поскольку скриптинг утилиты для программированного скрининга аффинности лигандов к выбранным мишеням и валидизация полученных с его помощью результатов предусматривала использование различных платформ для симуляции, нами было проведено сравнение программных ресурсов, которые будут использоваться для моделирования [2, 3].

Известно, что выделяют как общедоступные (бесплатные), так и полностью коммерческие решения. К первым относятся AutoDock 4, AutoDock Vina, Dock. Ко второй группе относятся Schrödinger's Glide, MOE (Molecular Operating Environment), Gold (Genetic Optimization for Ligand Docking), Sybyl-X Surflex-Dock, LigandScout.

AutoDock 4 является популярным программным продуктом для проведения исследований *in silico*. Однако проведение докинга с AutoDock 4 требует больших временных затрат, что делает его неудобным в сравнении с его прямым конкурентом AutoDock Vina. В этом исследовании проведен скриптинг утилиты для виртуального скрининга взаимодействия малых молекул с протеинами на основе AutoDock Vina. При этом для проверки универсальности разработанной утилиты проведена валидация результатов с использованием AutoDock 4.

Для проверки работоспособности предложенной модели был проведен анализ распространенных программ для докинга лигандов на стандартном наборе данных (рисунки 1, 2).

По результатам сравнения AutoDock 4 и AutoDock Vina можно выделить ряд преимуществ AutoDock Vina:

1. Быстрота проведения эксперимента, что делает программу предпочтительной для виртуального скрининга.

2. Точность симуляции: 78 % результатов докинга имели среднее квадратичное

отклонение меньше 2, в то время, когда у AutoDock 4 только 49 %.

В ходе сравнения также были отмечены удобство и простота использования AutoDock Vina [4, 5].

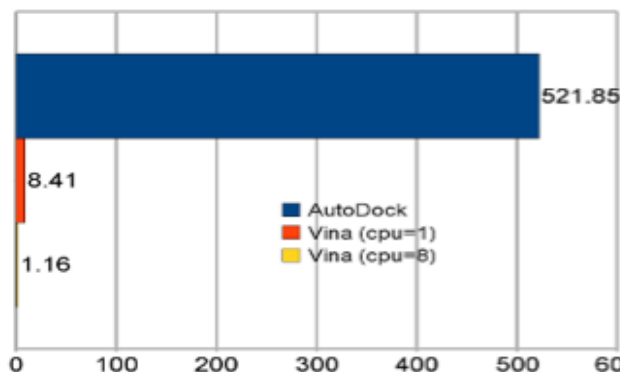


Рисунок 1 – Среднее время на пару рецептор-лиганд в тестовом наборе AutoDock относится к AutoDock 4, а Vina — к AutoDock Vina

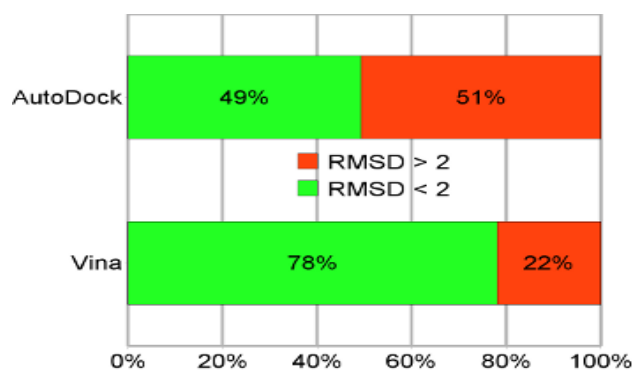


Рисунок 2 – Точность прогнозирования режима привязки на тестовом наборе AutoDock относится к AutoDock 4, а Vina – к AutoDock Vina

Для большинства комплексов пять программ предсказали вполне корректные суперпозиции лигандов в местах связывания. Однако более низкие значения RMSD для поз с наибольшим сродством были расположены следующим образом: Glide 6,6 ($0,99 \pm 0,13$) > AutoDock 3 ($0,90 \pm 0,08$) > AutoDock 4 ($0,84 \pm 0,08$) > DOCK 6,5 ($0,57 \pm 0,06$) > AutoDock Vina ($0,55 \pm 0,05$) [4]. Фактически, значения среднеквадратичного отклонения были довольно низкими (среднее $\text{RMSD} \leq 1,0 \text{ \AA}$) (рисунок 3). Самое низкое значение среднего RMSD было получено с использованием AutoDock Vina (0,55). Исследователи также отмечают высокую скорость работы AutoDock Vina – всего несколько секунд на расчет, тогда как другим программам требовалось около 5 минут на расчет. [6].

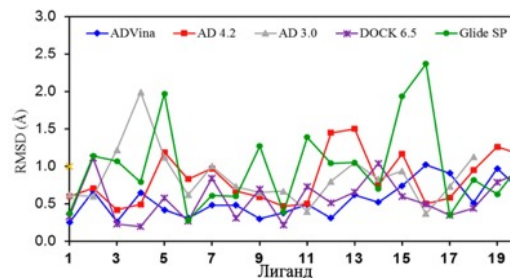


Рисунок 3 – Среднее квадратичное отклонение для самодокинга в программах

Несмотря на то, что многие исследователи проводят эксперимент *in silico* в AutoDock 4, мы провели скриптинг утилиты на базе программной оболочки AutoDock Vina. Это связано с тем, что AutoDock 4 не имеет функции «конфиг» файла, которая облегчает процесс докинга для исследователя. Поэтому перед нами стояла дополнительная задача провести валидацию результатов, полученных на базе AutoDock Vina, сравнивая их с результатами экспериментов на базе AutoDock 4 (рисунок 4). В качестве протеина-мишени была выбрана глюкокиназа, лиганды – ряд веществ, изученных ранее в качестве потенциальных активаторов данного фермента [7]. Нами была разработана утилита для многочисленного молекулярного докинга (рисунок 4).

```
receptor=4RCHT.pdbqt -- подготовленный белок
center_x=30.45
center_y=1.33 -- центр решетки (Grid) для белка
center_z=69.00
size_x=55
size_y=70 -- размер решетки (Grid) для белка
size_z=52
num_modes=20 -- количество пробегов
exhaustiveness=8 -- исчерпаемость (подвижность атомов)
energy_range=3 -- макс. разница в энергии между наилучшим
и наихудшим режимом связывания
```

Рисунок 4 – Конфиг-файл для утилиты докинга

Ниже представлено сравнение результатов докинга с помощью разработанной утилиты для AutoDock Vina в связке с программой визуализации USCF Chimera, а также AutoDock 4 (таблицы 1–3).

Таблица 1 – Результаты докинга AutoDock 4 [5]

№ опыта	Минимальная $E_{\text{связь}}$, ккал/моль	Средняя $E_{\text{связь}}$, ккал/моль
1	–9,41	–8,93
2	–9,41	–8,67
3	–9,28	–8,76
4	–9,26	–8,89

Таблица 2 – Результаты докинга разработки

№ опыта	Минимальная $E_{\text{связь}}$, ккал/моль	Среднеквадратичное отклонение	
		нижняя граница	верхняя граница
1	-7,8	0,000	0,000
2	-7,3	33,583	36,556
3	-6,9	5,569	6,950
4	-6,7	35,890	40,319
5	-6,7	6,726	8,918
6	-6,6	5,560	7,069
7	-6,5	2,448	3,200
8	-6,5	4,381	7,934
9	-6,4	5,802	7,497
10	-6,2	37,189	40,046

Таблица 2 – Результат докинга Chimera+ AD Vina

№ опыта	Минимальная $E_{\text{связь}}$, ккал/моль	Среднеквадратичное отклонение	
		нижняя граница	верхняя граница
1	-8,0	0,000	0,000
2	-8,0	6,620	11,221
3	-8,0	1,903	2,085
4	-7,6	2,242	2,713
5	-7,6	5,575	7,366
6	-7,4	6,548	11,456
7	-7,4	3,982	6,588
8	-7,3	29,541	36,789
9	-7,3	4,751	11,497
10	-7,3	4,923	10,932

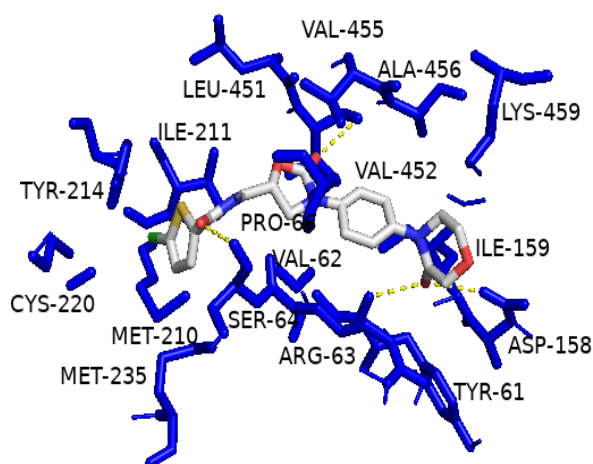


Рисунок 5 – Комплекс протеин-лиганд AutoDock 4 [7]

Для валидации полученных результатов было проведено сравнение топологии комплексов. Было установлено, что лиганды находились в одном кармане и имели сходную конформацию. При этом аминокислотное окружение тоже показало высокую степень совпадения [7].

Сравнение топологии связывания показало (рисунки 5, 6) сходную суперпозицию комплексов лиганд-протеин для экспериментов, проведенных с использованием различных программных ресурсов AutoDock 4 и AutoDock Vina.

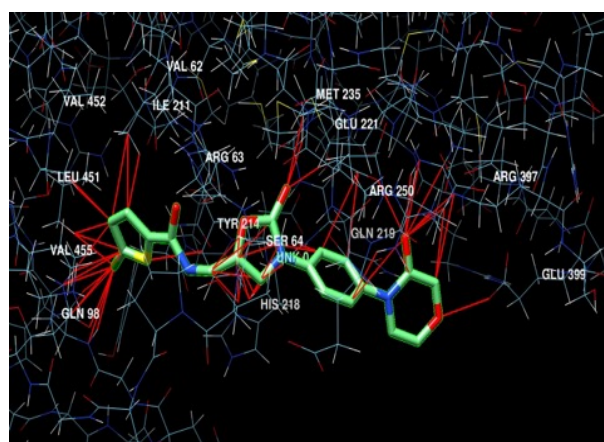


Рисунок 6 – Комплекс протеин-лиганд AutoDock Vina

Вывод. На основании анализа и сравнения принципов молекулярного докинга и алгоритма работы различных программ молекулярного моделирования был проведен скриптинг утилиты на базе AutoDock Vina для виртуального скрининга аффинности лигандов к выбранной биологической мишени.

Разработанный алгоритм для создания программы множественного молекулярного докинга демонстрирует высокую эффективность, точность при определении аффинности лигандов к протеинам-мишеням.

Тестирование программы на различных наборах данных подтвердило ее способность к корректной обработке и анализу больших объемов данных. Сравнение результатов, полученных с помощью разработанной программы, с результатами, полученными с использованием других программ, в частности в исследовании *in silico* аффинности ряда лигандов к глюкокиназе, подтвердило высокую точность и производительность утилиты, а также соответствие результатов, полученных при работе на базе программного обеспечения AutoDock Vina и AutoDock 4.

Список цитированных источников

1. Methods for the Discovery and Identification of Small Molecules Targeting Oxidative Stress-Related Protein-Protein Interactions: An Update / X. Wu [et al.] // *Antioxidants* (Basel). – 2022. – Vol. 11(4). – P. 619.
2. Pagadala, N. S. Software for molecular docking: a review / N. S. Pagadala, K. Syed, J. Tuszynski // *Biophys Rev.* 2017. – Vol. 9(2). – P. 91–102.
3. High-Throughput Virtual Screening of Compounds with Electrophilic Fragments for New Potential Covalent Inhibitors of Bacterial Proteins / P. Yakovets [et al.] // *Proc.* – 2022. – Vol. 12(87).
4. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings / J. Eberhardt [et al.] // *Journal of Chemical Information and Modeling.* – 2021. – Vol. 61(8). – P. 3891–3898.
5. Trott, O. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading / O. Trott, A. Olson // *Journal of computational chemistry.* – 2010. – Vol. 31(2). – P. 455–461.
6. Castro-Alvarez, A. The Performance of Several Docking Programs at Reproducing Protein-Macrolide-Like Crystal Structures / A. Castro-Alvarez, A. M. Costa, J. Vilarrasa // *Molecules.* – 2017. – Vol. 22(1). – P. 136.
7. Лахвич, Ф. Ф. Исследование in silico аффинности Ривароксабана к глюкокиназе / Ф. Ф. Лахвич, О. Н. Ринейская // *Медицинский журнал.* – 2023. – № 2(84). – С. 62–66.

Scripting of a new docking tool for multiple ligands' screening and its validation in silico for glucokinase*Lakhvich T. T., Prakapienia Y. A.**Belarusian State Medical University, Minsk, Republic of Belarus*

When analyzing and comparing different molecular docking tools as well as the operating algorithm of various molecular modeling programs, AutoDock Vina based tool was proposed for virtual screening of the affinity of ligands for a selected biological targets. The developed method for creating a multiple molecular docking script is efficient and accurate in determining the affinity of ligands for protein targets. The utility software was tested on various databases and validated in silico for glucokinase. Comparing the results from utility software developed with the other programs ensures reliability with the high accuracy as well as the consistency of the results obtained when working on the AutoDock Vina and AutoDock 4 software; the data meeting the experimental standards and being compatible with the other programs for screening.

Keywords: AutoDock Vina, AutoDock 4, affinity, glucokinase, scripting utility software for multiple docking.

Поступила 05.06.2024