

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М. В. ЛОМОНОСОВА
ФИЗИЧЕСКИЙ ФАКУЛЬТЕТ

На правах рукописи

Ильин Иван Сергеевич

**Поиск новых ингибиторов для заданных белков-мишеней
методами молекулярного моделирования**

Специальность 03.01.08 – Биоинженерия

ДИССЕРТАЦИЯ
на соискание ученой степени
кандидата физико-математических наук

Научный руководитель:
доктор физико-математических наук, Сулимов Владимир Борисович

ОГЛАВЛЕНИЕ

Введение	4
Актуальность работы	13
Цель исследования	14
Основные задачи	14
Новизна.....	16
Глава 1. Фактор Ха, β -лактамаза TEM-1, белок YB-1 как белки-мишени	19
1.1 Понятие мишени в контексте разработки лекарств.....	19
1.2 Фактор Ха как терапевтическая мишень	20
1.3 Бета-лактамаза TEM-1 как фактор резистентности бактерий	24
1.4 Белок YB-1: онкомаркер и новая мишень для лечения противоопухолевых заболеваний.....	27
Глава 2. Методы виртуального скрининга.	31
2.1 Классификация методов виртуального скрининга.	31
2.2 Структурно-ориентированный виртуальный скрининг	34
2.3 Комбинированный структурно-ориентированный скрининг	38
2.4 Полуэмпирические методы квантовой химии	41
2.5 Расчет энтальпии связывания методом PM7.....	46
Глава 3. Молекулярный докинг	48
3.1 Докинг как основной инструмент виртуального скрининга	48
3.2 Программа докинга SOL	54
Глава 4. Поиск новых ингибиторов фактора Ха	58
4.1 Создание и валидация модели фактора Ха.....	58
4.2 Виртуальный скрининг сфокусированной библиотеки	61

4.3 Идентификация новых ингибиторов фактора Ха на основе 2,2,4-триметил-1,2,3,4-тетрагидрохинолина	67
Глава 5. Поиск новых ингибиторов белка TEM-1	73
5.1 Создание модели аллостерического ингибирования β -лактамазы TEM-1	73
5.2 Виртуальный скрининг библиотеки NCI Diversity set V	74
5.3 Виртуальный скрининг базы ВГУ и библиотеки HitFinder	77
5.4 Применение молекулярной динамики для проверки аллостерического эффекта предсказанных лигандов	83
Глава 6. Поиск аллостерических антагонистов белка YB-1	85
6.1 Создание модели белка на основе ансамбля конформаций	85
6.2 Виртуальный скрининг полной базы ВГУ	88
6.3 Экспериментальное тестирование отобранных кандидатов	91
Заключение	95
СПИСОК ЛИТЕРАТУРЫ	102

Введение

С давних времен человечество ставило перед собой цель поиска биологически активных веществ для борьбы с различными заболеваниями. Самые первые лекарства на основе растительного сырья были найдены эмпирическим путем и составили базис медицины древних государств. В эпоху расцвета алхимии развитие первых методов лабораторной химии позволило более эффективно извлекать лекарственные вещества из природного сырья, что дало возможность снизить количество балластных веществ и использовать улучшенные режимы дозирования. Начиная с эпохи Возрождения, за счет становления научного метода практическая медицина сильно продвинулась в понимании природы заболеваний и разработки эффективных медицинских практик. Развитие органической химии, а именно методов синтеза новых соединений, привело к революции в химии лекарственных средств и дало возможность получения улучшенных аналогов существующих лекарств. В середине прошлого века за счет ускоренного становления современной молекулярной биологии была сформулирована молекулярная гипотеза природы заболеваний, что, в конечном счете, привело к идентификации биомолекул, которые являются важными для патогенеза многих патологических состояний и поэтому могут выступать в качестве мишеней для лекарственных соединений. Расшифровка генома человека дало основу для понимания многих молекулярных аспектов, в том числе ассоциированных с возникновением заболеваний, что привело к нахождению новых белков-мишеней и новых лекарственных молекул. Наконец, развитие методов молекулярного моделирования и хемоинформатики, а также появление больших вычислительных мощностей открыло новый горизонт для поиска лекарственных соединений. Современные методы машинного обучения и технологии искусственного интеллекта уже проявили себя во многих медицинских задачах, например, прогнозировании мишеней для терапии и лиганд-ориентированном виртуальном скрининге[1].

Несмотря на достигнутые результаты, задача разработки лекарственных средств остается очень трудоемкой, ресурсозатратной и длительной. Это связано со сложностью биологии человека, большим объемом существующего химического пространства, а также невозможностью оптимизировать одну переменную (например, растворимость) без одновременного влияния на другие свойства молекулы. Задача идентификации ранних кандидатов, так называемых «хитов», является ключевой для успешного проекта по разработке лекарственных соединений. Для упрощения задачи оптимизации структуры исследователи пытаются идентифицировать несколько десятков химически разнообразных молекул-«хитов» на раннем этапе разработки, что предполагает исследование довольно обширной части химического пространства при поиске активных молекул.

Глобально стратегия поиска новых лекарств начинается с гипотезы о том, что ингибирование определенной биомолекулы в организме может повлиять на патогенез заболевания. Здесь и далее такая биомолекула будет называться мишенью. Чаще всего мишенями являются белковые молекулы. Если мишенью выступает определенный фермент, то его ингибирование может быть достигнуто путем связывания низкомолекулярного соединения либо внутри активного центра, либо в аллостерическом сайте. Такое соединение будет называться ингибитором фермента. Ингибирование других белков-мишеней можно достичь за счет нарушения белок-белковых взаимодействий опять же за счет связывания низкомолекулярного лиганда в сайте, ответственном за данные взаимодействия, либо за счет разобщения ДНК-белкового взаимодействия, если мишенью выступает, например, транскрипционный фактор. В некоторых случаях, когда мишенью выступает рецептор к определенному эндогенному гормону или нейромедиатору, терапевтический эффект может проявиться, наоборот, за счет стимулирования мишени. Другими словами, дизайн таких лекарств будет направлен на выявление агонистов (стимулирующих рецептор лигандов), а не антагонистов (ингибиторов мишени). Однако во всех случаях общим свойством мишень-ориентированной стратегии поиска лекарства

является задача идентификации лигандов, связывающихся с определенным белком, которые влияют на его функциональную активность и на тот сигнальный путь в клетке, в который этот белок вовлечен[2]. Альтернативой мишень-ориентированному подходу является метод на основе фенотипического скрининга, в ходе которого системой оценки биологической активности выступает целая клетка, ткань или даже организм. Несмотря на ряд преимуществ, связанных, прежде всего, с меньшей вероятностью возникновения проблем при переносе результатов скрининга на клинический уровень, именно мишень-ориентированная стратегия является доминирующей в разработке лекарств. Так с 1999 по 2013 год 70% всех лекарств, одобренных управлением по санитарному надзору за качеством пищевых продуктов и медикаментов, были найдены именно с помощью этого подхода[3].

С методологической точки зрения идентификация новых биологически активных соединений может быть достигнута на основе трех обобщенных подходов, которые часто дополняют друг друга: поиск на основе аналога, *de novo* дизайн и метод скрининга[4]. Как следует из названия, поиск на основе аналога предполагает исследование только того химического пространства на предмет идентификации лигандов, которое ограничено аналогами уже ранее обнаруженных лекарств и биологически активных соединений. Очевидно, что исторически в те периоды, когда знание о структурах мишеней было ограничено, именно этот способ был первым среди рациональных подходов для выявления новых лекарственных молекул. Явное ограничение метода на основе аналога – разработанное лекарство будет обладать выраженной гомологией с тем веществом, на основе которого оно было сконструировано. Дальнейшее развитие методов структурной биологии и кристаллографии привели к накоплению знаний о трехмерных структурах белковых комплексов с низкомолекулярными лигандами, что дало возможность конструирования новых лекарств на основе наблюдаемых взаимодействий между белком и лигандом. Такой подход получил название *de novo* дизайн или структурно-

ориентированный *de novo* дизайн. С 1996 по 2014 год 17 одобренных лекарств были разработаны на основе структурной информации о белке-мишени[4].

Метод скрининга отличается от двух предыдущих подходов тем, что подразумевает исследование некоторой библиотеки химических соединений на предмет активности против определенной мишени в рамках определенной системы детекции. Первым ученым, использовавшим данную методологию, считается Пауль Эрлих, который испытал 606 мышьяк-содержащих соединений с целью идентификации лекарства против сифилиса. В ходе его исследования был обнаружен сальварсан или препарат 606, который показал выраженную активность в отношении бледной трепонемы и стал первым синтезированным лекарством направленного действия. Современный экспериментальный скрининг позволяет проводить испытания нескольких тысяч молекул ежедневно[5]. Однако такая производительность требует использования сложного лабораторного оборудования и автоматизированных систем, что является главным недостатком данной методологии. Другими разновидностями экспериментального скрининга являются сфокусированный скрининг, когда исследуются несколько небольших выборок химических соединений, часто имеющих химическое сходство с известными лигандами, и фрагмент-ориентированный скрининг, предполагающий исследование низкомолекулярных соединений малого размера, их кристаллизацию с мишенью и последующее объединение выявленных молекул-«хитов» на основе структурно-ориентированного дизайна. Фенотипический скрининг также направлен на исследование библиотек соединений. Однако, как отмечено выше, данный тип скрининга сфокусирован на идентификации тех эффектов соединений, которые наблюдаются для целых клеточных систем. Краткая характеристика наиболее распространенных методов экспериментального скрининга приведена в таблице 1.

Таблица 1. Методы экспериментального скрининга[6].

Метод	Описание	Комментарий
Высоко-производительный (HTS)	Большое количество соединений тестируется с использованием 384-луночных планшетов	Ввиду высокой стоимости метод доступен по большей части крупным фарм. компаниям
Сфокусированный	Исследуются соединения с уже предполагаемой активностью против определенных классов мишеней (например, ингибиторы протеаз)	Является более дешевой альтернативой HTS, однако выявленные соединения не будут полностью новыми в том смысле, что будут обладать гомологией с уже известными ингибиторами
На основе фрагментов	Кристаллизация соединений с белком-мишенью и поиск фрагментов, которые могут служить структурными блоками для дизайна более крупных и активных лигандов	Объединение фрагментов, для которых выявлена активность, может значительно увеличить активность против мишени
ЯМР скрининг	Использование фрагментов в качестве изучаемых молекул и ЯМР методологии для разрешения структуры белок-лиганд	Фактически является методом скрининга на основе фрагментов, в котором инструментом получения структуры комплекса выступает ЯМР
Фенотипический	Определение эффекта соединений на клетках или тканях	Более низкая пропускная способность

На рисунке 1 показано количество новых одобренных лекарств с 1995 по 2014 год с указанием метода идентификации активной молекулы. Видно, что начиная с 2005 года, методы скрининга становятся доминирующей методологией в поиске лекарств, которая позволяет вполне эффективно идентифицировать инновационные молекулы-«хиты», которые впоследствии становятся полноценными лекарствами. Кроме того, число новых лекарств, выявленных в ходе скрининга, было небольшим до 2009 года, так как эффективность скрининга до какого-то времени сдерживалась отсутствием развитых методов комбинаторного синтеза, помогающих в быстром конструировании репрезентативной химической библиотеки.

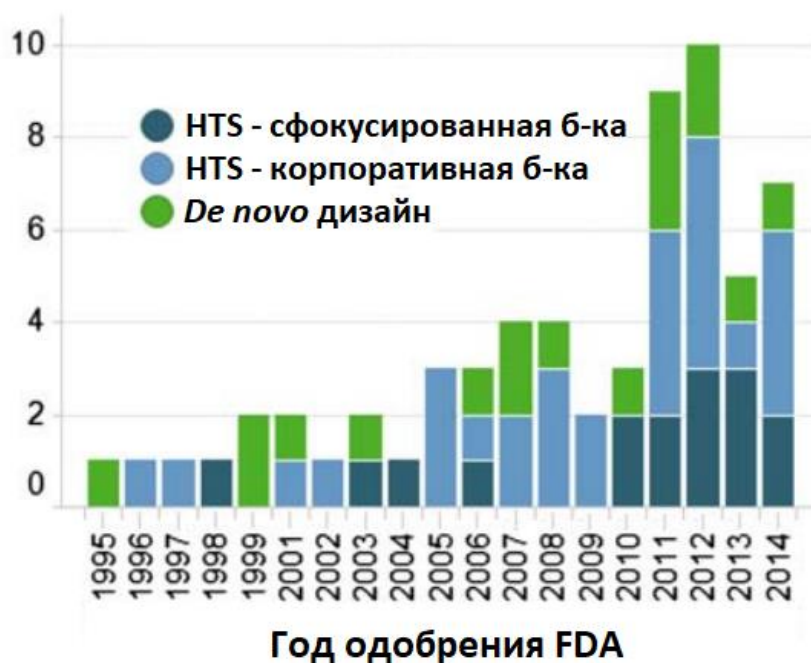


Рисунок 1. Одобренные лекарства с 1995 по 2014 год, которые были разработаны на основе метода экспериментального скрининга (синие и голубые столбики) и метода структурно-ориентированного *de novo* дизайна (зеленые столбики)[4].

Несмотря на свою проявленную эффективность, развитие методов HTS не привело к ожидаемому всплеску новых активных субстанций. Оказалось, что во многих случаях использование экспериментального скрининга является малоэффективным с точки зрения соотношения затрат и числа выявленных

качественных «хитов»[7–9]. Также замечено, что многие проекты по HTS имеют тенденцию к идентификации липофильных плохо растворимых объемных кандидатов, которые сложно оптимизировать в дальнейших стадиях разработки. Сложившаяся ситуация стимулировала развития альтернативных подходов к идентификации биологически активных соединений, среди которых виртуальный скрининг оказался наиболее эффективным с точки зрения экономии времени и материальных затрат.

В отличие от высокопроизводительного экспериментального скрининга, виртуальный скрининг производит отбор активных соединений за счет предсказания их энергии связывания с мишенью с помощью компьютерных программ. При этом можно выполнять предсказания для еще не синтезированных (виртуальных) веществ, которые впоследствии могут быть синтезированы для проверки. Кроме того, такие экспериментальные сложности как образование агрегатов или ограниченная растворимость в целом не нужно учитывать при компьютерном скрининге. Главным требованием этой методологии является знание о пространственных и энергетических аспектах, описывающих процесс связывание лиганда. Другими словами, должна быть известна трехмерная структура комплекса мишени с лигандом, полученная либо рентгеноструктурным анализом, либо ЯМР. Если такой структуры нет в наличии в публичных базах структурных данных, то атомистическая модель мишени может быть получена с помощью моделирования по гомологии, когда на основе структур гомологичных белков удастся воссоздать приблизительную структуру требуемого белка. В случае невозможности построить модель по гомологии знание о структурах активных и неактивных соединений может стать основой для так называемого лиганд-ориентированного компьютерного скрининга. Все это подчеркивает, что *in silico* скрининг во многом является методом на основе знаний.

Многолетний опыт использования виртуального скрининга подтвердил его ценность для процедуры выявления активных молекул в отношении широкого круга мишеней[10]. Одним из ярких примеров применения

компьютерного скрининга является лекарство Aggrastat (Merck & Co) который был обнаружен на основе данной методологии и одобрен для предупреждения инфаркта миокарда. Кроме того, существует множество лекарственных кандидатов, также обнаруженных в ходе виртуального скрининга, которые успешно вошли в стадию клинических испытаний[11]. Есть также примеры, когда *in silico* скрининг выявил молекулы-«хиты», пропущенных первичным запуском HTS оборудования[12]. Лишь повторное тестирование позволило идентифицировать эти активные молекулы. Стратегия расширения списка «хитов» (англ. hit expansion) предполагает объединение экспериментального и компьютерного скрининга, когда на основе информации о структурах соединений, выявленных в ходе первичного эксперимента, происходит генерация серий похожих соединений, их *in silico* оценка активности, отбор самых перспективных молекул и, наконец, проверка активности. Этот и другие примеры подчеркивают комплементарность обоих методов для эффективного поиска лекарств[13].

Долгая история развития позволила умножить первоначальные преимущества *in silico* скрининга и выделить основные ограничения, которые должны быть приняты во внимания для некоторых классов мишеней. Исторически производительность *in silico* отбора лигандов ограничивалась существующими компьютерными мощностями и слабой алгоритмической составляющей. В наши дни виртуальный скрининг является комплексным инструментом, который объединяет различные по точности способы компьютерной абстракции молекулярных систем и позволяет с неплохой точностью (которая зависит как от опыта исследователя, так и от выбранной мишени и упрощений) различать активные и неактивные молекулы для больших библиотек химических соединений. Основным инструментом компьютерного скрининга является молекулярный докинг. Процедура докинга является методом молекулярного моделирования, в ходе которого происходит симуляция взаимодействия двух молекул, чаще всего, белка и низкомолекулярного лиганда. Методологически в ходе докинга достигается

выполнение двух в некоторой степени взаимосвязанных задач: предсказание конформации лиганда в активном центре/выбранном сайте связывания мишени, а также оценка энергии связывания. В рамках скрининга первое позволяет визуально оценить способность предполагаемого лиганда к взаимодействию с белком, а второе дает возможность ранжировать множество соединений по их способности влиять на функциональную активность мишени. Кроме того, предсказанная биоактивная конформация лиганда может быть использована для более точных расчетов, например, оценки сродства лиганда к белку с помощью квантовохимических методов или оценки стабильности комплекса белок-лиганд в симуляции молекулярной динамики. Таким образом, схема виртуального поиска лекарств сводится к следующим шагам. Сначала проводятся некоторые подготовительные процедуры, направленные на создание атомистической модели белка и трехмерных моделей лигандов для всей библиотеки, подлежащей скринингу. После этого проводится молекулярный докинг с целью первичной фильтрации библиотеки химических соединений на основе предсказанной энергии связывания в программе докинга. Главное ограничение этого этапа – невысокая точность оценочных функций, применяемых в докинге, использование которых дает ощутимый процент ложноположительных результатов. Снижение этого процента может быть достигнуто на третьем шаге скрининга, когда для нескольких десятков лучших виртуальных «хитов» выполняется процедура уточнения энергии связывания с привлечением более точных методов. Такой комбинированный подход позволяет различать активные и неактивные молекулы с более высокой точностью.

В рамках данной работы планировалось выполнить процедуру виртуального скрининга с целью идентификации трех классов биологически активных соединений: прямых ингибиторов свертывания Ха, аллостерических ингибиторов β -лактамазы TEM-1 и антагонистов процесса перехода белка YB-1 в ядро клетки.

Актуальность работы

Поиск ингибиторов фактора Ха направлен на выявление новых хемотипов, которые могут стать основой для разработки более безопасных лекарств против тромботических состояний по сравнению с существующими препаратами на основе ингибиторов данного белка. Разработка аллостерических ингибиторов β -лактамаз также является актуальным с точки зрения борьбы с резистентностью бактерий. Воздействие на аллостерический сайт вместо каталитического центра призвано подавить консервативный механизм работы фермента, чтобы предупредить появление мутантных форм, устойчивых к ингибитору. Идентификация ингибиторов работы YB-1 может помочь с разработкой нового класса противоопухолевых агентов, способных излечивать неопластические заболевания, в развитии которых белок YB-1 играет ведущую роль. Центральное место в этих разработках играет суперкомпьютерное молекулярное моделирование, в том числе докинг и квантово-химический метод, которые призваны повысить эффективность разработок.

Степень разработанности избранной темы

На данный момент опубликованы работы, подтверждающие эффективность виртуального скрининга для поиска ингибиторов. Степень разработанности разная для классов соединений, выбранных для разработки в рамках данного исследования. В отечественной литературе нет работ, посвященных идентификации ингибиторов фактора свертывания Ха. На сегодняшний день существуют три лекарства на основе прямых ингибиторов фактора Ха. Однако данные препараты характеризуются большим спектром побочных эффектов, включающим в себя жизнеугрожающие состояния, что подчеркивает необходимость идентификации новых хемотипов с более безопасным профилем. Что касается аллостерических ингибиторов β -лактамаз, то в отечественной литературе есть работа Д.А. Бешновой с соавторами, которая посвящена виртуальному скринингу таких ингибиторов. Однако авторы используют лиганд-ориентированный скрининг, тогда как данная

работа предполагает структурно-ориентированный дизайн, что дает возможность более эффективно исследовать химическое пространство, не ограничиваясь областью применимости модели, построенной на основе структур нескольких известных ингибиторов. Кроме того, пока не существует аллостерических ингибиторов β -лактамаз в виде полноценных лекарств, одобренных к медицинскому применению. Как и в случае поиска ингибиторов фактора Ха, в отечественной литературе нет работ, в которых бы предпринимались попытки идентификации низкомолекулярных лигандов белка YB-1, способных ингибировать его функциональную активность. В мировой литературе есть одно исследование, в ходе которого использование виртуального скрининга позволило идентифицировать такой лиганд. Всего известно три таких лиганда, которые в целом характеризуются небольшим размером и химической простотой. Докинг и методы квантовой химии только начинают широко применяться для разработки ингибиторов заданных белков-мишеней. Особенности этой технологии в значительной степени определяются используемыми методами и программами. В данной работе применяется технология и программы, разработанные в МГУ имени М.В. Ломоносова, а также суперкомпьютерные вычислительные ресурсы, причем последние только начинают применяться в мире для докинга.

Цель исследования

Поиск ингибиторов, влияющих на работу трех терапевтических белков-мишеней: фактор Ха, β -лактамаза TEM-1, белок YB-1 – с использованием виртуального скрининга на основе молекулярного моделирования баз данных органических соединений.

Основные задачи

Основными задачами данного исследования являются:

1. Построение оптимальной атомистической модели фактора свертывания Ха. Отбор кандидатов в ингибиторы фактора Ха с помощью виртуального

- скрининга библиотек химических соединений, сочетающего в себе докинг и квантовохимический постпроцессинг.
2. Оценка влияния постпроцессинга на основе полуэмпирического метода РМ7 на способность выбранного протокола моделирования предсказывать ингибиторы фактора Ха.
 3. Построение атомистической модели β -лактамазы TEM-1. Разработка возможного механизма аллостерического ингибирования этого фермента.
 4. Идентификация потенциальных аллостерических ингибиторов β -лактамазы TEM-1 с использованием виртуального скрининга на основе докинга и квантовохимического постпроцессинга.
 5. Построение обобщенной модели белка YB-1 на основе нескольких белковых конформации для неявного учета подвижности петлевого участка в районе центра связывания в рамках ансамбль докинга. Проведение процедуры виртуального скрининга на основе докинга и квантовохимического постпроцессинга для предсказания антагонистов перехода YB-1 в ядро клетки.
 6. Экспериментальная проверка активности отобранных с помощью моделирования соединений против конкретных белков-мишеней в *in vitro* системе.

Объект и предмет исследований

В данном исследовании объектами выступают три белка в качестве терапевтических мишеней: фактор Ха, β -лактамаза TEM-1 и белок YB-1. Предметом является потенциальные ингибиторы этих белков, выявленные с помощью компьютерного молекулярного моделирования. Таким образом, исследование заключается в проведение процедуры виртуального скрининга с помощью методов докинга и квантовой химии выбранных библиотек химических соединений с целью идентификации молекул, влияющих на активность ряда терапевтически важных белков.

Новизна

Исследована способность разработанного протокола виртуального скрининга, сочетающего докинг с использованием программы SOL и последующее применение квантовохимического полуэмпирического метода, выявлять активные соединения для воздействия на выбранные белки-мишени. С использованием методов молекулярного моделирования впервые найдены новые ингибиторы фактора Ха на основе 1,2,3,4-тетрагидрохинолина. Экспериментально измеренное значение IC_{50} для лучшего найденного ингибитора фактора Ха составило 1.9 мкМ, а один из ингибиторов оказался селективным по отношению к фактору Ха и не оказал влияние на такие гомологически близкие сериновые протеазы, как трипсин, тромбин, фактор XIa и фактор IXa. Впервые разработана теоретическая модель аллостерического ингибирования β -лактамазы TEM-1, подразумевающая воздействие на консервативную область фермента. На основе данной модели идентифицированы потенциальные аллостерические ингибиторы β -лактамазы TEM-1 первые в своем классе, относящиеся к химической группе гидразидов. Впервые идентифицированы потенциальные лиганды YB-1, относящиеся к производным триазолохиназолинона и дигидрохинолина, которые нацелены на подавление процесса перехода данного белка в ядро клетки. Вещества проявили цитостатическое действие на опухолевых клетках и не влияли на рост нормальных клеток, а также ингибировали переход белка YB-1 в ядро *in vitro*.

Теоретическая и практическая значимость работы

Методология приведенного исследования может служить основой при разработке оригинального лекарственного средства. Использование методов молекулярного моделирования может способствовать повышению эффективности поиска молекул-«хитов» и их оптимизации в процессе создания лекарств. Кроме того, использование данных методов позволяет сократить материальные и временные затраты начальной стадии разработки нового лекарственного препарата.

Главная практическая ценность работы – идентификация новых классов ингибиторов таких белков-мишеней как фактор Ха, β -лактамаза TEM-1 и белок YB-1. Найденные структуры ингибиторов могут быть использованы в дальнейшем для получения лидерных соединений в ходе физико-химической оптимизации и проведения дополнительных доклинических испытаний.

Методология диссертационного исследования

Проведенное исследование позволило провести совершенствование методологии виртуального скрининга на основе докинга и квантовохимического постпроцессинга, которая продемонстрировала свою эффективность для выявления экспериментально подтверждённых ингибиторов фактора Ха. Для прогнозирования сродства к белку YB-1 разработана методология применения ансамбль докинга с использованием ансамбля белковых конформаций с их предварительной оптимизацией полуэмпирическим квантовохимическим методом PM7.

Положения, выносимые на защиту

1. Виртуальный скрининг на основе докинга и последующего расчета энтальпии связывания белок-лиганд полуэмпирическим методом PM7 для поиска потенциальных ингибиторов фактора Ха.
2. Влияние постпроцессинга на основе полуэмпирического метода PM7 и модели растворителя COSMO на точность прогнозирования активности в отношении фактора Ха.
3. Модель аллостерического ингибирования β -лактамазы TEM-1, предполагающая изменение положения консервативного петлевого участка белка. Использование разработанной модели аллостерического ингибирования для поиска аллостерических ингибиторов β -лактамазы TEM-1 с помощью виртуального скрининга на основе докинга и квантовохимического постпроцессинга.
4. Модель домена холодового шока белка YB-1 на основе ансамбля белковых конформаций для поиска антагонистов перехода белка YB-1 в ядро клетки.

Степень достоверности и апробации результатов

Степень достоверности следует из корректности постановки задачи и подтверждается тем, что разработанная методика опирается на апробированные процедуры молекулярного моделирования и методы вычислительной химии. Основные результаты диссертации были представлены и обсуждены на 5 международных и всероссийских конференциях.

Личный вклад

Диссертант принимал непосредственное участие в сборе и анализе литературных данных, разработке путей решения поставленных задач, разработке белковых моделей для выбранных терапевтических мишеней, разработке механизма аллостерического ингибирования β -лактамазы TEM-1, выполнении виртуального скрининга в программе SOL для выбранных библиотек соединений, уточнении результатов докинга с помощью полуэмпирического метода PM7 в программе MORAC, отборе соединений на экспериментальное тестирование, процедуре оптимизации первичных кандидатов в аллостерические ингибиторы β -лактамазы TEM-1, подготовке публикаций, и докладов по теме исследования.

Глава 1. Фактор Ха, β -лактамаза TEM-1, белок YB-1 как белки-мишени

1.1 Понятие мишени в контексте разработки лекарств

Ключевым элементом рациональной разработки лекарств является выбор мишени, то есть той молекулы, чью активность предполагается модулировать с помощью лекарства, так как она вовлечена в определенные сигнальные каскады, напрямую связанные с развитием болезни. В большинстве случаев идентификация новых белков-мишеней является трудоемкой задачей, предполагающей проведение многочисленных экспериментов, которые помогают подтвердить связь конкретного белка с патологическим профилем заболевания. Выбрав белок-мишень, должна быть сформулирована концепция того, какое воздействие предполагается достигнуть с помощью низкомолекулярного соединений, то есть лекарства. Чаще всего таким воздействием является ингибирование, например, когда выбранным белком является фермент. Ингибирование фермента может быть достигнуто за счет связывания лиганда либо внутри активного центра, либо внутри аллостерического сайта в случае его наличия. Последнее может быть предпочтительно для тех белков, у которых остатки активного центра способны мутировать, не давая ингибиторам блокировать работу ферментов. Для белков, которые не являются ферментами, влияние на их активность часто может быть достигнуто за счет блокирования тех сайтов на их поверхности, которые распознаются другими белками и опосредуют белок-белковые взаимодействия. Закрытие таких сайтов специфическим лигандом может приводить к нарушению передачи сигнала в интересующем молекулярном каскаде и остановке патологического процесса. Таким образом, помимо понимания того, на какую мишень будет производиться воздействие, проект разработки начинается с формулировки определенного молекулярного механизма, на основе которого можно достичь изменения её активности.

В рамках настоящего исследования были выбраны три белка-мишени, для которых предполагалось разработать соединения, влияющие на их активность. Все три мишени представляют фармакологический интерес, и их связь с

развитием определенного заболевания подтверждена экспериментально. Рассмотрим биохимические аспекты каждого белка-мишени по отдельности, а также те молекулярные механизмы, которые лежат в основе воздействия на их активность.

1.2 Фактор Ха как терапевтическая мишень

Фактор Ха является витамин К-зависимой сериной эндопептидазой, состоящей из двух цепей: легкой, которая состоит из 139 аминокислотных остатков, и тяжелой, включающей в себя 303 аминокислоты. Протеазный домен имеет трипсинподобное строение на основе бета-цилиндра[14]. Каталитическая триада состоит из Ser195, His57 и Asp102. Его главная роль – протеолиз протромбина в тромбин, который, в свою очередь, является финальным участником каскада свертывания и запускает образование фибрина из фибриногена. Одна молекула фактора Ха ведет к образованию около 1000 молекул тромбина[15]. Активация фактора X происходит за счет ограниченного протеолиза его зимогена с участием либо теназы внутреннего пути свертывания (фактор VIIIa в комплексе с фактором IXa), либо теназы внешнего пути свертывания (фактор VII в комплексе с тканевым фактором). Таким образом, оба пути активации свертывания сходятся на этом белке (см. рисунок 2), что подчеркивает факт о том, что фактор Ха играет первостепенную роль для стадии амплификации свертывания.

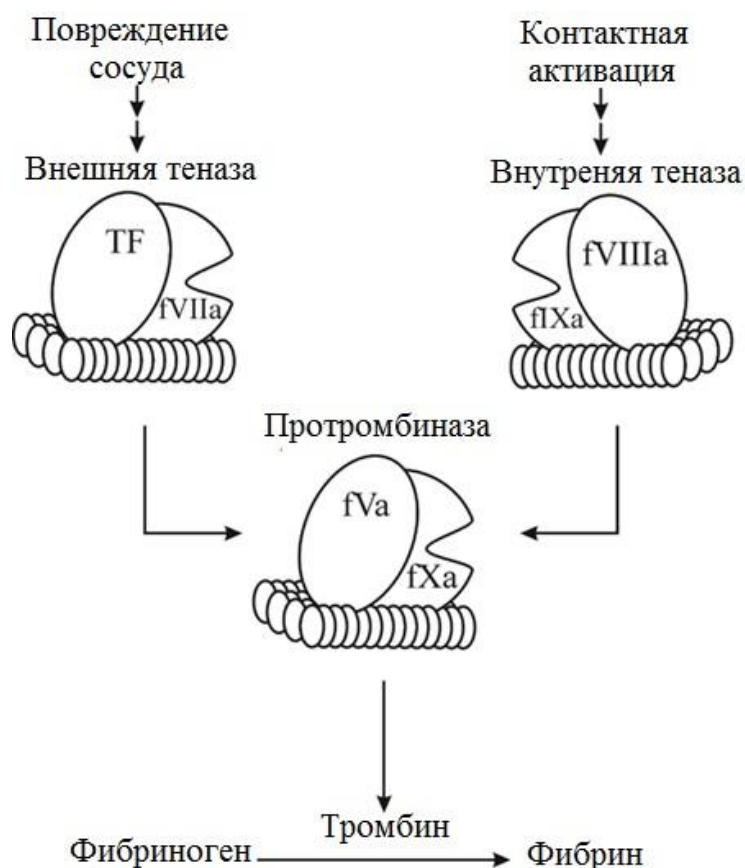


Рисунок 2. Место фактора Ха в системе свертывания крови. TF – тканевой фактор. Картинка взята и адаптирована из[15].

Активный центр фактора Ха состоит из двух выраженных карманов связывания. Карман S1 примыкает к каталитической триаде и является отрицательно заряженной глубокой полостью, образованной такими остатками, как Ser195, Tyr228, Asp189 и Gly216. За счет отрицательного заряда боковой группы Asp189 данный карман имеет сродство к остатку аргинина в составе эндогенного субстрата[16], а также к положительно заряженным группам синтетических ингибиторов этого белка. Вторым карманом связывания, карман S4, образован группой ароматических аминокислот, а именно: Phe174, Tyr99 и Trp215 – и поэтому внутри этого кармана связываются неполярные и ароматические группы лигандов. Другими важными для связывания остатками являются Gly216 и Gly218, которые часто участвуют в формировании водородных связей с центральными ядрами известных ингибиторов фактора Ха. На рисунке 3 показано строение активного центра фактора Ха с указанием остатков наиболее важных с точки зрения медицинской химии.

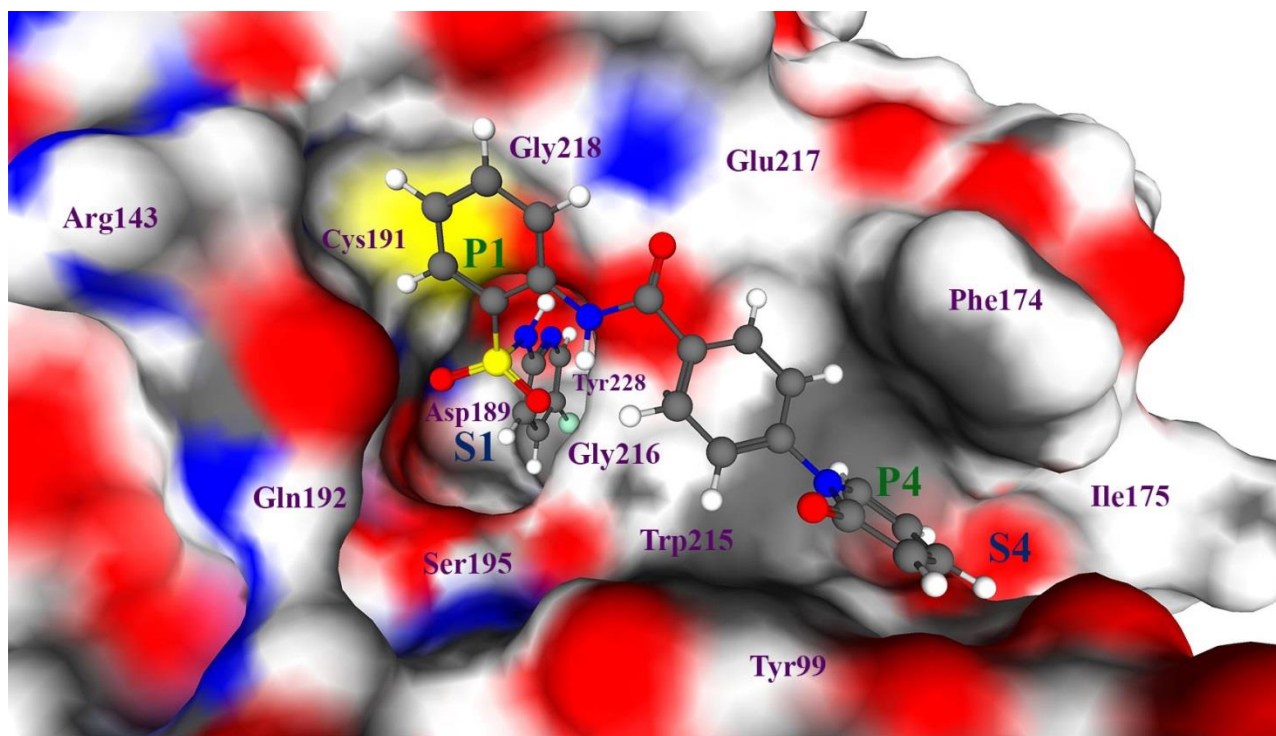


Рисунок 3. Строение активного центра фактора Ха (PDB ID: 3CEN). S1 и S4 указывают на положение соответствующих карманов связывания. P1 и P4 обозначают структурные фрагменты закристаллизованного ингибитора, которые связываются в указанных карманах. Поверхность белка показана как поверхность доступная растворителю, и её цвет соответствует находящимся вблизи нее типам атомов белка: белый цвет соответствует атомам водорода, синий – азота, красный – кислорода, серый – углерода, желтый – атомам серы.

Рисунок сделан с помощью программы молекулярной визуализации MolRed[17,18].

Валидация фактора Ха как терапевтической мишени была проведена в конце прошлого века в ходе многолетних исследований. Первые такие исследования были выполнены на природных ингибиторах этого белка, извлеченных из гематофагов. К ним относится антикоагулянтный пептид клеща *Omithodoros moubata* и антистатин. Исследования *in vitro* и *in vivo* подтвердили, что прямые ингибиторы фактора Ха должны иметь преимущество по сравнению с непрямыми ингибиторами с точки зрения более эффективного влияния на тромбообразование[15,19,20]. Понимание того, что фактор Ха является удобной терапевтической мишенью для лечения тромботических

состояний, привело к разработке многочисленных синтетических ингибиторов фактора Ха. В настоящий момент существует более десяти химических классов прямых ингибиторов фактора Ха[21], а также три соединения, которые были одобрены для медицинского применения: ривароксабан, эдоксабан и аликсабан. Первые прямые низкомолекулярные ингибиторы фактора Ха содержали основные группы, которые заряжаются положительно при физиологическом pH. Это давало им возможность эффективно связываться внутри кармана S1, который содержит заряженный остаток Asp189. Примерами таких соединений служат производные амидина. Главным недостатком данной группы ингибиторов стала их низкая биодоступность, связанная с высокой основностью. Поэтому они были замещены группой слабо основных и неосновных соединений, к которым принадлежит большинство недавно найденных ингибиторов фактора Ха, а также вышеприведенные лекарственные препараты. В отличие от более ранних прямых ингибиторов эти соединения содержат нейтральные при физиологическом значении pH P1-мотивы, которые опосредуют π -взаимодействие ингибитора с остатком Tyr228 кармана S1. В большинстве случаев это взаимодействие достигается за счет использования галоген- или метокси-монозамещенных ароматических систем в качестве P1-мотива. Кроме того, P4-мотивы большинства ингибиторов также состоят из моно- или ди-замещенными ароматических фрагментов. В некоторых случаях ароматические фрагменты соединены с неароматическими гетероциклами. Наряду с большим разнообразием P1- и P4-мотивов существует различные варианты центральных ядер для ингибирования фактора Ха, которые могут быть использованы для сочленения указанных мотивов. Ключевой особенностью любого центрального ядра является способность придавать молекуле ингибитора V- или L-образную форму, что способствует такой пространственной ориентации лиганда во время связывания, при которой ингибитор занимает оба кармана активного центра. Помимо этого центральные ядра подтвержденных ингибиторов фактора Ха часто содержат несколько

доноров и/или акцепторов водородной связи, чтобы взаимодействовать с остатками Gly216 и Gly218.

Несмотря на то, что фактор Ха не является новой мишенью для разработки антикоагулянтов, поиск ингибиторов этого белка остается актуальной задачей, так как существующие лекарства, действующие на фактор Ха, имеют широкий спектр побочных реакций[22]. Поэтому фактор Ха был выбран в качестве одного из белков-мишеней, для которых предполагалось выполнить скрининг потенциальных ингибиторов в рамках данной работы.

1.3 Бета-лактамаза TEM-1 как фактор резистентности бактерий

Появление новых штаммов бактерий, которые устойчивы к существующим антибиотикам, является глобальной проблемой здравоохранения. Самым распространенным типом устойчивости является устойчивость к β -лактамным антибиотикам. Ключевым механизмом развития устойчивости данного типа является выработка бактериями β -лактамаз – ферментов, которые способны разрушать β -лактамное кольцо β -лактамных антибиотиков, что ведет к потере их активности[23]. Так как гены β -лактамаз расположены на мобильных генетических элементах, данный вид устойчивости может быстро распространяться среди бактерий за счет горизонтального переноса генов. Кроме того, многие остатки активных центров β -лактамаз способны мутировать, что ограничивает применение прямых ингибиторов данных ферментов и ведет к появлению новых β -лактамаз, а также β -лактамаз расширенного спектра. Так описано более 2700 видов β -лактамаз, которые делятся на 4 класса: А, В, С и D. Классы А, С и D включают в себя сериновые β -лактамазы – основным каталитическим остатком их активного центра выступает серин. Класс В охватывает β -лактамазы, использующие ионы металлов для гидролиза β -лактамных колец[24].

Одной из признанных терапевтическим мишеней среди β -лактамаз класса А является белок TEM-1. Он был впервые выделен в 60-ых годах прошлого века из крови пациента, и в настоящий момент часто встречается среди

граммотрицательных бактерий[25]. Этот белок массой около 30 кДа имеет компактную структуру, состоящую из одной полипептидной цепи, и не содержит каких-либо остатков углеводов или ионов металлов. Процесс гидролиза β -лактамовых антибиотиков, катализируемого TEM-1, включает в себя два этапа (см. рисунок 4)[26]. Вначале происходит стадия ацилирования β -лактамного кольца – гидроксильный кислород Ser70, активированный за счет аминокислотного окружения активного центра, атакует карбонильный атом углерода внутри β -лактама. Возникает ковалентная связь между Ser70 и молекулой антибиотика, сопровождающаяся разрывом C-N связи в β -лактамном кольце. После этого, за счет стадии деацилирования происходит отщепление связанной с Ser70 молекулы антибиотика с участием активированной молекулы воды, которая атакует вышеупомянутый карбонильный атом углерода, что ведет к разрыву ковалентной связи между этим атомом и атомом кислорода боковой группы Ser70. Активация молекул воды активного сайта β -лактамазы сопряжена, прежде всего, с наличием отрицательно заряженного остатка Glu166 и Asn170, что подчеркивает их ключевую роль в стадии деацилирования[27].

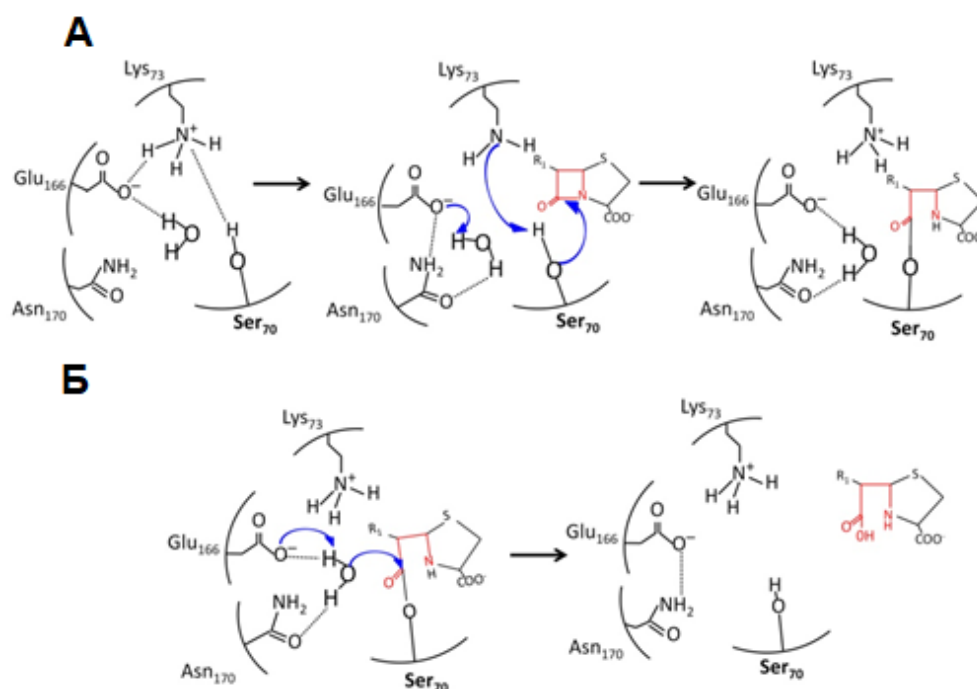


Рисунок 4. Механизм гидролиза β -лактамовых антибиотиков, катализируемый β -лактамазой TEM-1. А – стадия ацилирования, Б – стадия деацилирования.

Частью центра связывания является небольшой петлевой участок, так называемая, омега(Ω)-петля. Данная петля содержит 16 остатков (с 164 по 179 остаток в белке TEM-1) включает в себя вышеупомянутые Glu166 и Asn170, имеет первичное значение для функциональной активности β -лактамазы. N-конец петли и её C-конец располагаются близко в пространстве за счет солевого мостика между Arg164 и Asp179, образуя «бутылочное горлышко» (см. рисунок 5). Исследование мутантных форм TEM-1 показывают, что смещение Ω -петли часто ведет к затруднению процесса деацилирования фрагментов пенициллиновых антибиотиков, обеспечивая инактивацию активного сайта β -лактамазы, так как Ser70 остается ковалентно связан и не способен вступить во взаимодействие с новой молекулой антибиотика[23].

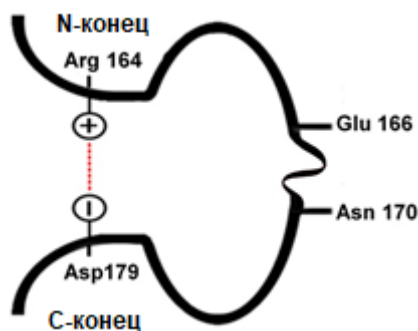


Рисунок 5. Схематичное представление Ω -петли. Остаток Arg164 формирует солевую связь с Asp179, придавая петлевому участку характерную форму.

Симуляция молекулярной динамики для β -лактамаз показывает, что Ω -петля характеризуется некоторой подвижностью, особенно её C-концевой участок[28]. Обнаружено, что подвижностью может возрастать при связывании субстрата, что, по-видимому, можно объяснить участием Ω -петли в доставке молекулы субстрата в активный центр. С учетом вышеизложенных фактов о ключевой роли Ω -петли в катализе и её подвижностью Ω -петля может быть рассматриваться в качестве мишени для аллостерического ингибирования β -лактамазы TEM-1. Аллостерическое воздействие на активность ферментов данного типа является перспективной задачей, так как применение прямых ингибиторов β -лактамаз приводит к появлению их мутантных вариантов устойчивых к данным ингибиторам. В отличие от многих остатков активного

центра белка TEM-1 Ω -петля характеризуется высокой консервативностью, что также подчеркивает её значимость для разработки ингибиторов β -лактамазы TEM-1 с целью борьбы с резистентными штаммами бактерий.

1.4 Белок YB-1: онкомаркер и новая мишень для лечения противоопухолевых заболеваний

Белок YB-1 – это ДНК и РНК-связывающий белок, который получил свое название за счет способности связывать некоторые регуляторные участки ДНК, называемые Y-боксы. Данный белок относится к семейству белков с доменом холодового шока и регулирует различные клеточные процессы, включая пролиферацию и дифференцировку, на транскрипционном и трансляционном уровне. Также YB-1 оказывает влияние на стабильность мРНК. Кроме того, выявлена ведущая роль этого белка в развитии некоторых видов рака, в частности, рака простаты и рака молочной железы, а также выявлена взаимосвязь этого белка с формированием множественной лекарственной устойчивости[29]. Многие исследования показали, что высокая экспрессия YB-1 может вызывать эпителиально-мезенхимальный сдвиг – процесс изменения фенотипа клеток в сторону опухолевой прогрессии. Выявлено, что пациенты с повышенной продукцией белка YB-1 имеют высокий риск метастазирования и рецидива рака. Для клеточных линий рака молочной железы показано, что подавление работы этого белка вызывает остановку клеточного деления, апоптоз и возвращение эпителиального фенотипа[30]. Учитывая все эти данные, YB-1 является не только важным маркером злокачественности, но и удобной терапевтической мишенью для лечения тех видов рака, в которых этот белок играет первичную роль. Ингибиторы этого белка могут стать основой нового класса противоопухолевых лекарств[31,32].

Белок YB-1 содержит 324 аминокислотных остатка и состоит из трех доменов (см. рисунок 6). На N-конце находится небольшой домен, богатый аланином и пролином, за которым идет древний, эволюционно консервативный домен холодового шока. За ним следует протяженный C-концевой домен,

состоящий из чередующихся кластеров положительно и отрицательно заряженных аминокислотных остатков.

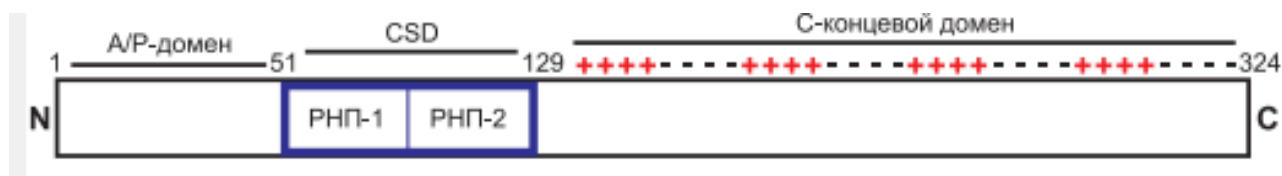
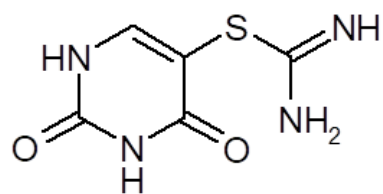


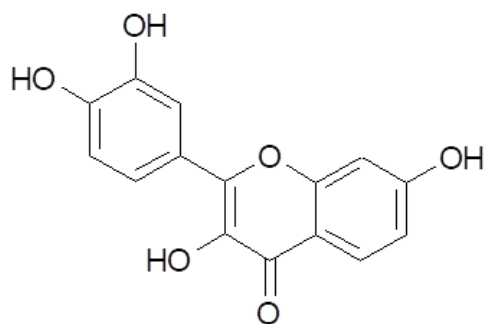
Рисунок 6. Доменная структура белка YB-1. Обозначения: 1) A/P-домен – домен, обогащенный пролином и аланином, 2) CSD – домен холодового шока, 3) РНП-1 и РНП-2 – консенсусные последовательности, участвующие в распознавании ДНК.

Другие важные структурные элементы YB-1 включают в себя ДНК-связывающий сайт, образованный консенсусными мотивами РНП-1 и РНП-2, а также сайт фосфорилирования, который образован петлевым участком домена холодового шока и Ser102 – остатком, на который добавляется фосфатная группа. Примерами киназ, которые распознают данный сайт, являются киназа Akt и белок PKCα. Фосфорилирование Ser102 ведет к активации YB-1, после чего он осуществляет из цитоплазмы в ядро клетки[33]. Ядерная локализация является ключевой для функциональной активности YB-1. Именно после перехода в ядро этот белок способен влиять на профиль экспрессии генов, связанные, в том числе, с малигнизацией и лекарственной устойчивостью[29]. Данные мутагенеза показывают, что замена Ser102 на аланин ведет к потере активности этого белка. Учитывая функциональное значение, место фосфорилирования может служить сайтом связывания антагонистов YB-1, которые будут разобщать белок-белковые взаимодействия между этим белком и клеточными киназами, препятствуя его активации.

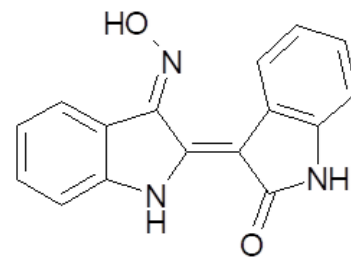
В настоящее время было найдено три антагониста белка YB-1, которые препятствуют переходу этого белка в ядро, снижают уровень его фосфорилирования и ингибируют рост опухолевых клеток, в которых экспрессия YB-1 повышена. Названия антагонистов и их структуры приведены на рисунке 7.



**2,4-дигидрокси-5-пиримидинил
имидотиокарбомат**



Физетин



Индирубин-3'оксим

Рисунок 7. Известные ингибиторы процесса перехода YB-1 в ядро клетки

В литературе имеются работы, в которых для двух антагонистов: 2,4-дигидрокси-пиримидинилимидотиокарбомата и физетина, была выполнена симуляция молекулярного докинга в сайт фосфорилирования[31,32]. После докинга обнаружено высокое теоретическое сродство к данному сайту, что делает эти соединения потенциальными ингибиторами белок-белковых взаимодействий между YB-1 и белками с киназной активностью. На рисунке 8 показан сайт фосфорилирования YB-1 с указанием остатков потенциально важных для связывания внутри этого сайта. Выбор остатков основан на результатах докинга данных ингибиторов, взятых из литературы.

С учетом ведущей роли белка YB-1 в процессе злокачественной трансформации для некоторых видов рака, а также его роли в развитии множественной лекарственной устойчивости опухолевых клеток он был выбран в качестве третьего белка-мишени в рамках данного исследования. В качестве механизма его ингибирования было выбрано нарушение белок-белковых взаимодействий с клеточными киназами за счет связывания низкомолекулярного соединения внутри сайта фосфорилирования.

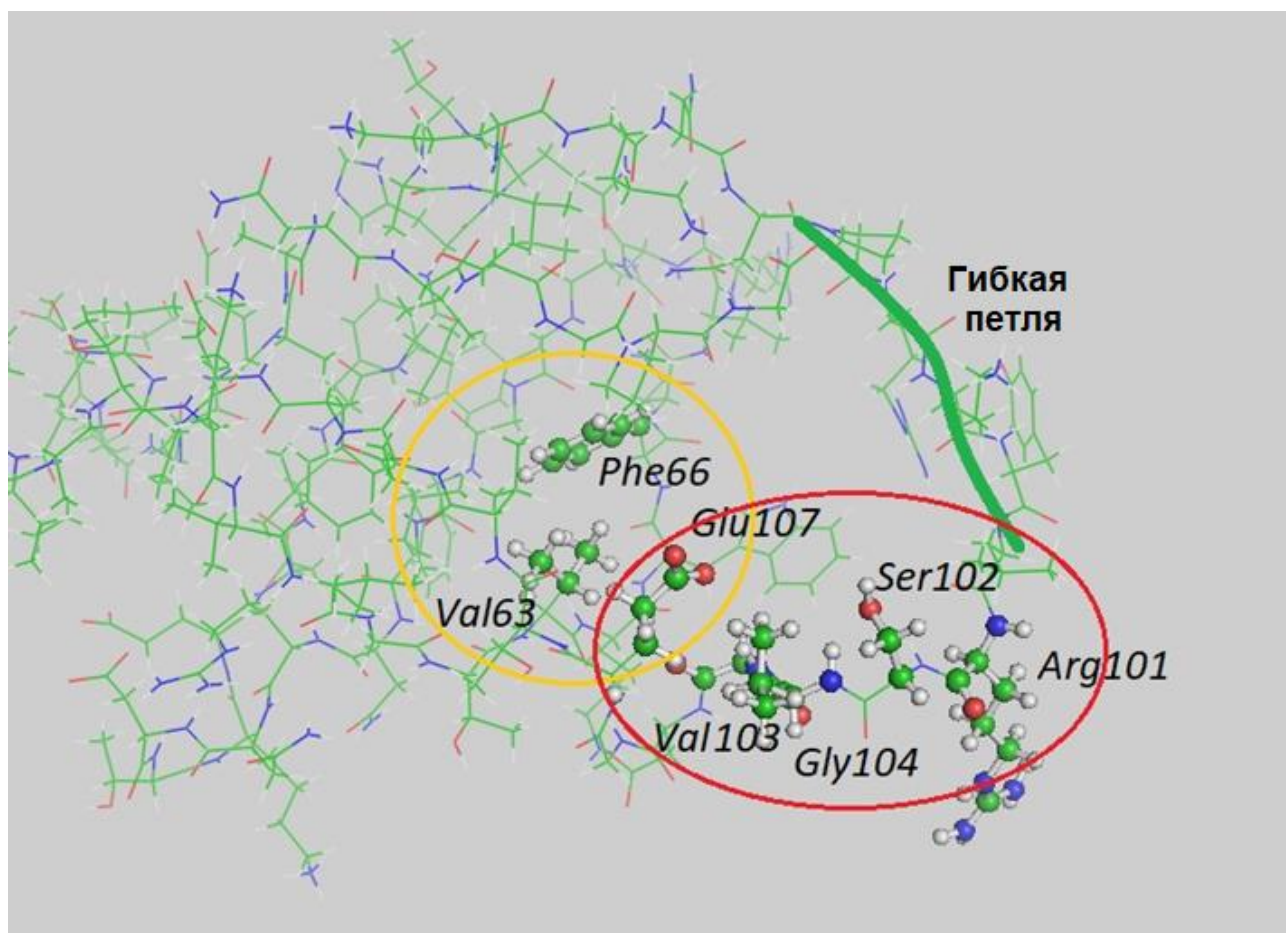


Рисунок 8. Теоретические карманы связывания сайта фосфорилирования белка YB-1. Желтым кругом обведено место физетина, которое он занимает после докинга в работе[32]. Красный круг указывает на теоретическое место связывания 2,4-дигидрокси-пиримидинилимидотиокарбомата, которое также было обнаружено в ходе докинга в рамках исследования Gunasekaran et al.[31]

Глава 2. Методы виртуального скрининга.

2.1 Классификация методов виртуального скрининга.

Виртуальный скрининг представляет собой поиск новых соединений или хемотипов с требуемой биологической активностью внутри химической базы данных[34]. При этом соединения, которые предполагается обнаружить, могут либо служить альтернативой существующим известным лигандам определенного белка-мишени, либо быть первыми в своем роде для тех мишеней, для которых неизвестны экспериментально подтвержденные лиганды, но известно их трехмерное строение. Впервые термин «виртуальный скрининг» был использован в публикации 1997 года[35]. С тех пор данный метод приобрел широкую популярность в качестве важного инструмента начальной разработки лекарств как дополнение к экспериментальным методам.

В общем, виртуальный скрининг можно описать в виде последовательных процедур фильтрации, направленных на сужение исследуемого химического пространства и отбор потенциальных молекул-«хитов», которые имеют активность в отношении выбранного белка-мишени. Важно отметить, что исследуемые низкомолекулярные соединения не обязательно должны быть в готовом виде. Предсказание активности таких виртуальных соединений позволяет значительно экономить материальные ресурсы и синтезировать только те из них, которые показывают наилучшее сродство к белку в ходе компьютерного эксперимента. Эффективность методов виртуального скрининга для выявления активных молекул продемонстрирована для различных классов терапевтических белков-мишеней[36]. В таблице 2 приведена лишь небольшая часть примеров белков-мишеней, для которых виртуальный скрининг был использован для идентификации новых активных молекул.

Таблица 2. Примеры белков-мишеней, для которых были идентифицированы агонисты/ингибиторы с помощью методов виртуального скрининга[36].

Класс белков	Примеры
Рецепторы, сопряженные с G-белками	$\alpha 1A$ адренорецептор, дофаминовый рецептор D3, рецептор к нейрокинину-1, аденозиновый рецептор A2A
Киназы	Bcr-abl киназа, киназа Akt 1, циклин-зависимая киназа 2, киназа гликогенсинтазы, казеин киназа II, киназа рецептора трансформирующего ростового фактора бета
Протеазы	Фактор свертывания Ха, фактор свертывания XIa, тромбин, катепсин D, ВИЧ протеаза, главная протеаза коронавируса SARS
Другие гидролазы	Нейраминидаза вируса гриппа, ацетилхолинэстераза, AmpC β -лактамаза, фосфодиэстераза 4
Оксидазы и редуктазы	Альдоз редуктаза, дигидрофолатредуктаза
Ионные каналы	Kv1.5 калиевый канал, селективный кальциевый канал T-типа
Ядерные рецепторы	Рецептор к ретиноевой кислоте, рецептор тиреотропного гормона
Белковые комплексы	Нарушение белок-белковых взаимодействий для bcl-2, циклофилин A, нарушение белок-белковых взаимодействий для Rac1

Традиционно совокупность методов виртуального скрининга делят на две группы: лиганд-ориентированные методы и структурно-ориентированные методы. Первая группа предполагает использование двухмерных и трехмерных структур известных лигандов (и не-лигандов), чтобы выявить потенциально активные соединения внутри базы данных. В рамках данной методологии

предсказание активности строится либо на некоторой оценке подобия между соединениями, либо на обнаружении общей подструктуры или фармакофора – набора химических групп, электронные и стерические характеристики которых обеспечивают оптимальные взаимодействия с белком-мишенью. Структурно-ориентированный скрининг вместо информации об известных лигандах использует напрямую знание о трехмерной структуре мишени и строится на процедуре докинга – позиционировании исследуемых молекул в функциональный центр белка и предсказании их энергии связывания с данным центром. Классификация методов виртуального скрининга приведена на рисунке 9.

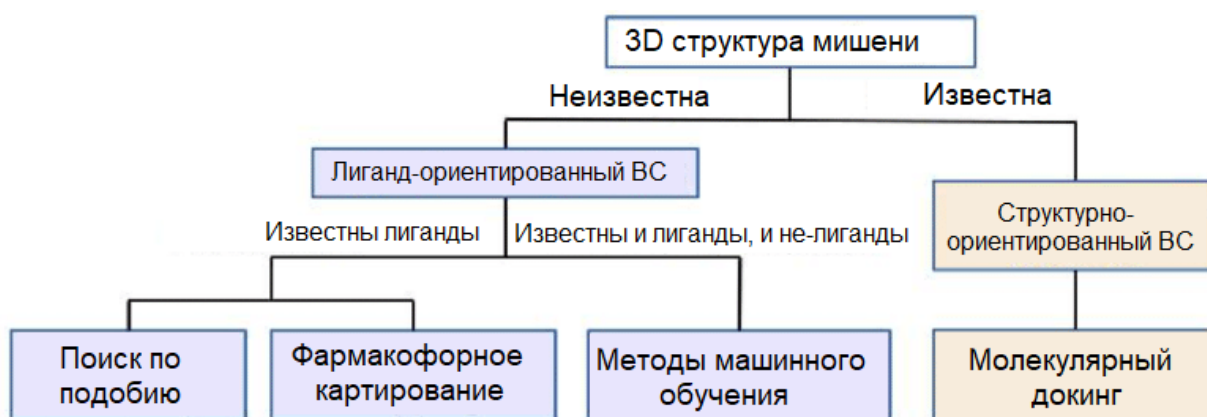


Рисунок 9. Классификация методов виртуального скрининга. Аббревиатура ВС означает виртуальный скрининг.

Выбор метода определяется доступностью экспериментальной информации для конкретной мишени. Наличие данных о трехмерной структуре мишени часто дает однозначный выбор в пользу структурно-ориентированного скрининга и методов докинга, хотя во многих случаях комбинация обоих подходов помогает увеличить вероятность обнаружения активных молекул и сделать финальную выборку кандидатов на эксперимент химически разнообразной[37]. Так как в данной работе поиск активных соединений строился на основе структурно-ориентированного подхода, то рассмотрим эту методологию подробнее.

2.2 Структурно-ориентированный виртуальный скрининг

Структурно-ориентированный скрининг, или по-другому, мишень-ориентированный скрининг, включает в себя три аспекта: молекулярный докинг, трехмерная модель белка, библиотека химических соединений[38]. Докинг является процедурой позиционирования низкомолекулярного соединения в активный или любой другой, заданный сайт белка. При этом генерируется геометрия комплекса двух молекул и рассчитывается энергия связывания на основе некоторой математической (скоринг) функции. Значение скоринг-функции является мерой потенциальной активности условного лиганда. После скрининга соединения в химической библиотеке ранжируются в порядке возрастания значений скоринг-функции и формируется топ предсказанных кандидатов. Подробнее алгоритмическая сторона докинга рассмотрена в главе 3 на примере существующих программ.

Успешность компьютерного скрининга во многом зависит от начальной экспериментальной структуры белка, используемой для подготовки полноатомной модели. При выборе структуры белка в банке белковых структур нужно учитывать такие показатели качества как разрешение, R-фактор и стереохимическая корректность. Кроме того, по возможности нужно выбирать те структуры, в которых вместе с белком закристаллизован низкомолекулярный лиганд. Сайт связывания в таких белках будет уже «подстроен» под взаимодействие с лигандами. Первая задача, которую нужно решить при подготовке правильной белковой модели – добавление атомов водорода, так как метод рентгеноструктурного анализа, основной метод получения трехмерных структур белка, не позволяет идентифицировать положения атомов данного элемента. Кроме того, процедура протонирования должна учитывать, что боковые группы многих остатков, например, лизина или глутаминовой кислоты, ионизируется при физиологическом значении pH. Добавленные атомы водорода должны подвергаться локальной оптимизации, чтобы они сформировали оптимальную сеть водородных связей в белке. В случае наличия в структуре белка пропусков их требуется восстановить, например, на основе

других структур этого же белка. При наличии гибких элементов внутри выбранного сайта связывания модель должна учитывать их подвижность хотя бы в неявном виде за счет применения, так называемого, ансамбля белковых конформаций, отличающихся положениями подвижных участков. При наличии структур известных лигандов подготовленная модель подвергается валидации – оценке того, что она отражает то поведение белка при связывании лигандов, которое наблюдается в реальных экспериментах. Для этого для известных лигандов проводится симуляция докинга. Для нативного лиганда, то есть того, который закристаллизован вместе со структурой белка, на основе которой построена модель, проверяется воспроизведение его связанной конформации. Рассчитывается такой показатель, как среднеквадратичное отклонение (RMSD) между положениями соответствующих атомов в закристаллизованной конформации и конформации после докинга. Считается, что если значение RMSD меньше 2.0 Å, то достигнуто хорошее позиционирование с воспроизведением связанного положения. Наличие кристаллических комплексов белка-мишени с другими известными лигандами позволяет использовать процедуру кросс-докинга для дополнительной валидации модели. В ходе этой процедуры выполняется выравнивание геометрии исходного белка, взятого для построения модели, с геометриями других кристаллических комплексов белка, содержащих закристаллизованные лиганды. После выравнивания эти лиганды становятся «квазинативными» по отношению к исследуемой структуре белка, и качество их позиционирования, а значит, и качество подготовленной модели может быть также оценено с помощью докинга и расчета значений RMSD для каждого известного лиганда. Помимо оценки репрезентативности модели белка, докинг известных ингибиторов позволяет определить тот диапазон значений скоринг-функции для конкретной программы докинга, который имеют подтвержденные активные молекулы. Это дает возможность оценить порог по скоринг-функции, на который стоит ориентироваться при предсказании активности в ходе виртуального скрининга. В том случае, когда в ходе валидации обнаруживается, что не удается получить

хорошее позиционирование для известных ингибиторов белка при использовании построенной модели, модель должна быть подготовлена на основе другой кристаллической структуры, либо должны быть предприняты процедуры улучшения модели. К таким процедурам можно отнести оптимизацию геометрии белка на основе минимизации его энергии с целью релаксации возможных внутренних напряжений, а также неявный учет подвижности активного центра за счет отказа использования одиночной структуры и применения ансамбля конформаций белка.

Последним аспектом скрининга является библиотека химических соединений и процедура её подготовки к докингу. Обычно подготовка начинается с отсеивания тех соединений, что содержат излишне реакционноспособные и токсичные группы. Примерами таких групп являются бензохиноны, ацилгалогенидный фрагмент и нитрогруппа. Для тех молекул, которые представлены солями в библиотеке, удаляются противоионы. Так как взаимодействие белка-мишени и лекарственно-подобной молекулы в подавляющем большинстве случаев происходит в водном окружении при физиологическом значении pH, требуется верное протонирование молекул низкомолекулярных соединений при моделировании этого взаимодействия. В редких случаях (лекарства, работающие в желудочно-кислой среде, или средства, проникающие в лизосомы) моделирование должно предполагать учет величин pH, отличных от физиологического значения. Протонирование/депротонирование функциональных групп органической молекулы происходит согласно уравнению Гендерсона-Гассельбаха (для кислотных групп):

$$pH = pK_a + \log_{10} \frac{[A^-]}{[HA]} \quad (1)$$

где [HA] соответствует молярной концентрации протонированной группы, [A⁻] – молярная концентрация депротонированной формы, pK_a – константа диссоциации (кислоты). Таким образом, для правильного представления зарядовых форм в библиотеке химических соединений требуется знание их значений pK_a . Кроме того, нередко существует несколько зарядовых форм

(протомеров) при физиологическом значении рН, имеющих определенную долю от общего количества вещества. В рамках данного исследования дополнительно исследовались те зарядовые формы лигандов, доля которых была выше 10%.

В большинстве случаев химические библиотеки соединений не содержат данных о значениях pK_a для входящих в их состав молекул. Поэтому требуется привлечение дополнительных программ для предсказания правильных величин pK_a . В рамках данного исследования для протонирования молекул, подлежащих скринингу, использовался программный пакет Marvin[39]. Данная программа имеет несколько плагинов для различных физико-химических расчетов. Плагин pK_a выполняет прогнозирование значений констант диссоциации кислотных/основных групп на основе эмпирического метода, реализованного на принципах количественного соотношения структура-свойство (QSPR). В ходе расчета учитываются три атомных дескриптора: частичный заряд, поляризуемость и внутренняя поправка на водородные связи. На основе этого плагина Marvin дает возможность генерировать различные зарядовые состояния для молекул внутри большой базы химических соединений с учетом заданного значения рН.

Последним этапом подготовки молекул перед процедурой виртуального скрининга является получение трехмерной геометрии молекул из двумерного представления. В данном исследовании для этой задачи использовалась программа Corina[40]. Эта программа представляет собой один из наиболее широко распространённых наборов инструментов для создания качественных 3D моделей органических молекул. Общий алгоритм генерации таутомеров и конформеров базируется на использовании экспериментальных табличных значениях углов и длин связей. При этом для сортировки конформеров и отбора самых низкоэнергетических состояний в программе реализовано с помощью оптимизации в псевдо-силовом поле, содержащем лишь геометрические параметры[41]. Такая упрощенная оптимизация помогает снизить временные затраты на конвертацию отдельных молекул, что позволяет существенно

сократить время подготовки, например, в случае с библиотекой, содержащей несколько десятков тысяч соединений. Программа Corina была протестирована с использованием 25000 структур соединений, полученных с помощью рентгеновской кристаллографии, и показала хорошее воспроизведение экспериментальных данных даже для сложных полимакроциклических систем[42].

По мере развития вычислительной химии и увеличения компьютерных мощностей на смену классической схеме структурно-ориентированного скрининга, описанной выше, пришел комбинированный виртуальный скрининг. Данный тип скрининга состоит в том, что производится уточнение результатов докинга на основе различных вычислительных процедур. Совокупность таких процедур получило название постпроцессинг. Процедура постпроцессинга добавляет дополнительный критерий отбора при прогнозировании активности и сужает число первичных кандидатов, сгенерированных на основе оценочной функции программы докинга. Формально главная задача постпроцессинга – снизить число ложноположительных прогнозов, которые часто дает молекулярный докинг. Поэтому обычно комбинированные протоколы *in silico* скрининга имеют более высокую точность различения активных и неактивных молекул между собой.

2.3 Комбинированный структурно-ориентированный скрининг

Процедуры постпроцессинга, применяемые после докинга для уточнения его результатов и увеличения точности прогнозирования активности, бывают нескольких видов и различаются по сложности проводимых вычислений. Самым простым с точки зрения компьютерных ресурсов, но не производительности является визуальная фильтрация, в ходе которой медицинский химик может проверить, насколько конформация лиганда химически осмысленна, и оценить количество специфических видов взаимодействий между белком и лигандом. Однако визуальный осмотр уже для нескольких сотен конформаций видится затруднительным. Другой

относительно простой группой методов является методы, учитывающие топологические характеристики комплексов. Данные методы могут рассчитывать, например, долю объема лиганда, которая погружена внутрь полости связывания, площадь поверхности лиганда, которая доступна растворителю, и объем липофильных карманов вблизи места связывания. Очевидно, что реальные лиганды максимизируют долю своего объема, погруженного внутрь белка, и минимальную поверхность липофильных частей, которые доступны растворителю. Более сложной группой являются методы консенсус докинга, а самые трудоемкими процедурами являются методы, учитывающие эволюцию системы во времени, и подходы, в которых для оценки энергии связывания используются физические законы с меньшим числом упрощений.

Методы консенсус докинга предполагают объединение результатов нескольких программ докинга в некоторую схему консенсусной оценки аффинности для более точного предсказания активности. Фактически оценка сродства строится с помощью нескольких скоринг-функций вместо одной. Ряд исследований показал, что подходы на основе консенсус докинга, лучше предсказывают геометрию комплексов белок-лиганд и дают лучше выявляют активные молекулы по сравнению с подходом, который основан предсказаниях одной программы и одного скоринг-функции[43]. Кроме того, замечено, что объединение нескольких непохожих по своей природе скоринг-функций дает более высокую точность в сравнении с тем, когда объединяются предсказания похожих скоринг-функций[44]. Например, скоринг-функции, основанные на знаниях (knowledge-based), должны комбинироваться с эмпирическими оценочными функциями, чтобы компенсировать недостатки первых, и наоборот. В большинстве случаев в ходе постпроцессинга на основе консенсус скоринга выполняется пересчет энергии связывания (рескоринг) второй программой докинга для лучших кандидатов, предсказанных с помощью первичной программы, использованной для виртуального скрининга.

Более точным и сложным методом рескоринга является применение метода молекулярной динамики. Метод предполагает интегрирование классических уравнений движения молекулярной системы на отрезке времени. При этом атомы взаимодействуют в приближении молекулярной механики, и энергия рассчитывается с помощью заданного силового поля. Начальные скорости присваиваются атомам случайным образом, после чего рассчитываются силы взаимодействия на основе градиента потенциальной энергии. После этого рассчитываются новые положения атомов и их скорости с учетом временного шага. Этот итеративный процесс, состоящий из этапа расчета сил и этапа сдвига атомов в новые положения, повторяется заданное число шагов, и в конце генерируется траектория молекулы, которая может быть использована для вычисления средних и зависящих от времени свойств системы. Интеграция молекулярной динамики и виртуального скрининга чаще всего предполагает использование этого метода либо для оценки стабильности предсказанного в ходе докинга комплекса белок-лиганд, либо для более точного предсказания сродства и расчета свободной энергии связывания. Первая задача предполагает оценку усредненного значения среднеквадратичного отклонения (RMSD) атомных положений в ходе симуляции, а также проведение анализа среднеквадратичных флуктуаций (RMS fluctuation - RMSF), который позволяет для отдельных остатков позволяет оценить их движение в ходе моделирования. Предполагается, что в стабильном комплексе белка с лигандом наблюдаются только небольшие («тепловые») отклонения в положениях атомов, а остатки активного центра почти не двигаются во время симуляции, так как участвуют в процессе связывания. Вторая задача, направленная на расчет свободной энергии, предполагает использование либо молекулярной механики с методом Пуассона-Больцмана или обобщенной моделью Борна для оценки сольватационных эффектов (MM-PBSA и MM-GBSA), либо термодинамическое интегрирование[45]. Помимо этих задач в контексте компьютерного дизайна лекарств молекулярная динамика часто применяется для оптимизации геометрии белковых молекул

перед построением рабочей модели, а также для генерации ансамбля белковых конформаций для неявного учёта подвижных участков белка во время докинга.

Еще одним важным методом уточнения энергии взаимодействия между белком и лигандом являются полуэмпирические квантовохимические методы, сочетающие в себе баланс между точностью и скоростью. Так как в настоящей работе использовался постпроцессинг на основе полуэмпирического метода RM7, то разумно рассмотреть теоретические аспекты данной группы методов более подробно.

2.4 Полуэмпирические методы квантовой химии

Главным стимулом разработки расчетных методов, базирующихся на квантовой химии, послужил факт заведомо ограниченной точности силовых полей. Прежде всего, это связано с упрощенной формой расчета энергетических вкладов, что делает невозможным учёт некоторых сложных типов взаимодействий, характерных для биологических систем. Например, силовые поля без описания эффектов поляризации завышают энергию парного ионного взаимодействия, имеющего место внутри белкового окружения[46]. Кроме того, многие силовые поля могут быть пригодны только для определенного класса молекул.

Необходимость более точных методов привела к созданию методов квантовой механики на основе линейного масштабирования, которые, однако, оказались непригодными для описания биологических систем в виду высоких вычислительных затрат[47]. Поэтому основой для более точных вычислений для больших молекулярных систем стали полуэмпирические методы. Первая группа методов использует приближения для уравнения Хартри-Фока. Примером таких методов является семейство NDDO методов, которые пренебрегают двухатомным дифференциальным перекрытием. В их основе лежит уравнение Рутаана, являющееся видоизмененным уравнением Хартри-Фока[47]. Другой группой полуэмпирических методов являются те, которые построены на теории функционала плотности в схеме сильной связи, которая, в

свою очередь, является производной метода функционала плотности. Оба набора методов используют минимальный базисный набор и включают в себя упрощенное представление ряда электронных интегралов, что дает значительное снижение вычислительных затрат и, соответственно, времени расчетов[48]. Рассмотрим формализм первой группы методов, чтобы лучше описать тот тип постпроцессинга, который был выбран в рамках данного исследования.

В стандартных методах описания электронной структуры молекул молекулярные орбитали (МО, $\{\psi_i\}$) аппроксимируются линейной комбинацией атомных орбиталей (АО, $\{\chi_\mu\}$):

$$\psi_i(r) = \sum_{\mu} C_{\mu i} \chi_{\mu}(r) \quad (2)$$

В отличие от традиционной теории Хартри-Фока, в рамках которой учитываются все электроны базиса АО, полуэмпирические методы учитывают только валентные электроны, чья плотность описывается минимальным набором орбиталей слэтеровского типа. Орбитальные коэффициенты $C_{\mu i}$, приведенные в формуле (2) вычисляются с помощью уравнения Рутаана:

$$FC = SC\varepsilon, \quad (3)$$

где F , C , S – матрица Фока, матрица орбитальных коэффициентов и матрица перекрывания, соответственно. Матрица ε является диагональной матрицей, содержащей орбитальные энергии. Матрица Фока может быть разделена на одноэлектронную часть (h) и двухэлектронную часть (G), и их элементы рассчитываются следующим образом:

$$h_{\mu\nu} = \langle \mu | -\frac{1}{2} \nabla^2 | \nu \rangle - \sum_A Z_A \langle \mu | -\frac{1}{R_A} \nabla^2 | \nu \rangle \quad (4)$$

$$G_{\mu\nu} = \sum_{\lambda\sigma} P_{\lambda\sigma} \left[\langle \mu\nu | \lambda\sigma \rangle - \frac{1}{2} \langle \mu\lambda | \nu\sigma \rangle \right] \quad (5)$$

Кулоновский и обменный интегралы имеют вид, соответственно:

$$\langle \mu\nu | \lambda\sigma \rangle = \iint \chi_\mu(r_i) \chi_\nu(r_i) \frac{1}{r_j} \chi_\lambda(r_j) \chi_\sigma(r_j) dV_i dV_j \quad (6)$$

$$\langle \mu\lambda | \nu\sigma \rangle = \iint \chi_\mu(r_i) \chi_\lambda(r_i) \frac{1}{r_j} \chi_\nu(r_j) \chi_\sigma(r_j) dV_i dV_j \quad (7)$$

Как уже подчеркивалось выше, главной особенностью полуэмпирических методов является пренебрежение вычислением определенных электронных интегралов. Тип вводимых упрощений разделяет эти методы на несколько семейств. В приближении нулевого дифференциального перекрывания (zero-differential overlap - ZDO) из двухэлектронных интегралов вычисляются только интегралы типа $\langle \mu\mu | \nu\nu \rangle$ [49]. В методе полного пренебрежения дифференциальным перекрыванием (complete neglect of differential overlap - CNDO) используется приближение ZDO, а остальные интегралы параметризуются аппроксимацией Mataga-Nishimoto [49]. Природа этой аппроксимации такова, что p-орбитали описываются тем же самым образом, как и s-орбитали, и это сильно сказывается на точности вычисления относительных энергий, например, энергий вращательных барьеров [47]. Дальнейшей разработкой стал метод частичного пренебрежения дифференциальным перекрыванием (intermediate neglect of differential overlap - INDO) похожий на расчеты в рамках CNDO, но с дополнительной параметризацией интегралов, центрируемых на один и тот же атом [50]. Следующее семейство методов было основано на пренебрежении двухатомным дифференциальным перекрыванием (neglect of diatomic differential overlap - NDDO), которое развило приближение INDO с дополнительным включением описания двухэлектронных интегралов типа $\langle \mu\nu | \lambda\sigma \rangle$, где μ и ν – АО, которые центрированы на одном атоме, а λ и σ – АО, центрированные на другом атоме. Добавленные интегралы заметно повысили точность описания вращательных барьеров и относительных энергий, если сравнивать с методами-предшественниками [51].

Решение уравнения (3) обычно происходит в два шага. Сначала выполняют трансформацию и получают ортогональный базис с последующим решением стандартной задачи собственных чисел. В семействе методов NDDO отсутствует ортогонализация, и матрица Фока *a priori* предполагается ортогональной. Технически это выражается в пренебрежении матрицей перекрывания и заменой уравнения (3) на упрощенную форму:

$$FC = C\varepsilon \quad (8)$$

Приближение, касающееся ортогональности АО базиса, может вносить ощутимую ошибку для одноэлектронной части матрицы Фока. Поэтому было разработано следующее семейство методов, получивших название модифицированное пренебрежение двухатомным перекрыванием (modified neglect of diatomic overlap - MNDO). Методы MNDO улучшили точность метода NDDO за счет замены двуцентровых и двухэлектронных интегралов приближенными интегралами, полученных из мультипольных взаимодействий, которые рассчитываются методом Дьюара-Сабелли-Клопмана. С одной стороны для дальних взаимодействий эти интегралы фактически сходятся к классическому потенциалу, а с другой стороны при близкодействующих взаимодействиях они способны в некоторой степени учитывать электронные корреляции[52].

Основным направлением дальнейшего улучшения полуэмпирических методов стала параметризация энергии взаимодействия между парами ядер, чтобы смягчить неточности, вызванные введенными упрощениями. Так как ядро остается неполяризуемым, что сильно ослабляет эффект ядерного отталкивания и делает длины связей короткими, то в методе MNDO вводятся эмпирические параметры α_A и α_B в уравнение энергии отталкивания:

$$E_{\text{core-core}}^{(\text{MNDO})} = Z_A Z_B \langle s_A s_A | s_B s_B \rangle [1 + \exp(-\alpha_A R_{AB})] + \exp(-\alpha_B R_{AB}) \quad (9)$$

Данные параметры подгоняются, чтобы лучше воспроизвести тестовые данные.

Методология AM1 (Austin model 1) стала дальнейшим развитием семейства MNDO и смягчило неточности в расчетах близкодействующих

взаимодействий, увеличив число гауссоид потенциала энергии парного ядерного взаимодействия до четырех:

$$E_{\text{core-core}}^{(\text{AM1})} = E_{\text{core-core}}^{(\text{MNDO})} + \frac{Z_A Z_B}{R_{AB}} \left[\sum_i (K_{A_i} \exp(L_{A_i} (R_{AB} - M_{A_i})^2)) + \sum_i (K_{B_i} \exp(L_{B_i} (R_{AB} - M_{B_i})^2)) \right] \quad (10)$$

В результате этого число подгоняемых параметров сильно возросло по сравнению с первоначальным методом MNDO. Главным преимуществом методологии AM1 стало улучшенное описание водородных связей.

Метод PM3 (parametric method 3), разработанный Стюартом[53], представлял собой репараметризованный вариант AM1 с двумя гауссоидами для коррекции парного ядерного взаимодействия. Его дальнейшее развитие, метод PM6 (parametric model 6), был улучшен за счет использования более богатого набора молекулярных данных, а также отказа использования гауссоид в потенциале парного взаимодействия и введения разных парных параметров для его расчета. Таким образом, PM6 использует различные потенциалы для описания ядерного отталкивания между такими парами атомов, как N-H, O-H, C-C и Si-O. Другими важным улучшением метода стало добавление *d*-орбиталей в АО базис для некоторых элементов.

Последним на сегодняшний день методом в семействе PMx методов стал метод PM7[54], для увеличения точности которого был использован еще больший набор данных для параметризации, а также были добавлены поправки на образование водородных и галогенных связей и поправки, улучшающие описание дисперсионных взаимодействий. Показано, что данные поправки позволили достичь полуэмпирическим методам такой точности, которая сопоставима с точностью DFT-D, но, очевидно, достигается за более короткое расчетное время[55]. Сравнение точности позиционирования в ходе

суперкомпьютерной валидации методов, основанных на силовых полях, и полуэмпирических методов показало, что метод PM7 лучше описывает энергию комплекса белок-лиганд с точки зрения определения глобального минимума по сравнению с методом PM6-D3H4X при использовании одинаковой модели для описания эффектов растворителя[56]. С учетом его преимуществ именно полуэмпирический метод PM7 был выбран для постпроцессинга в рамках данной работы. Кроме того, для более точного описания взаимодействия белка и лиганда метод PM7 использовался в комбинации с континуальной моделью растворителя COSMO.

2.5 Расчет энтальпии связывания методом PM7

В настоящем исследовании процедура постпроцессинга как завершающий элемент виртуального скрининга для каждой мишени была направлена на оптимизацию самой низкоэнергетической позы после докинга и расчет энтальпии связывания. Для выполнения задач постпроцессинга использовалась программа MORAC[57]. Расчет энтальпии связывания с помощью квантовой химии производился согласно уравнению:

$$\Delta H_{bind} = \Delta H_{PL} - (\Delta H_P + \Delta H_L), \quad (11)$$

где ΔH_P , ΔH_L и ΔH_{PL} представляют собой энтальпию образования свободного белка, энтальпию образования свободного лиганда и энтальпию образования комплекса, соответственно.

Для расчета энтальпии образования комплекса на первом шаге выполнялась локальная оптимизация энергии комплекса при варьировании положений всех атомов лиганда для лучшей позы после докинга внутри активного центра мишени методом PM7. Затем, на втором шаге для этого комплекса производился единичный расчет самосогласованного поля (1SCF) также методом PM7 с включением континуальной модели COSMO[58] для описания эффектов растворителя. Для расчета энтальпии образования свободного лиганда использовалась конформация до докинга – та, что была сгенерирована в программе Corina[40]. Для этой конформации также

выполнялась оптимизация методом PM7. После оптимизации применялся 1SCF расчёт методом PM7 с учетом растворителя в континуальной модели COSMO. Вычисление энтальпии образования свободного белка производилось в ходе 1SCF расчета с помощью метода PM7 с учетом растворителя в модели COSMO без предварительной оптимизации. Выполнение квантовохимических расчетов для большой системы, включающей белок и лиганд, обеспечивалось использованием метода локализованных молекулярных орбиталей в рамках модуля MOZYME[59], пакета MOPAC. Рассчитанная энтальпия связывания служила одним из критериев предсказания активности низкомолекулярных соединений в ходе скрининга наряду со значением скоринг-функции в программе SOL и геометрией комплекса белок-лиганд, наблюдаемой для оптимизированной методом PM7 лучшей позы лиганда после докинга.

Глава 3. Молекулярный докинг

3.1 Докинг как основной инструмент виртуального скрининга

Молекулярный докинг – это метод структурно-ориентированного дизайна лекарств, который выполняет симуляцию молекулярного взаимодействия между белком и лиганда и предсказывает геометрию их комплекса и степень сродства[60]. Докинг остается одним из самых распространенных инструментов разработки лекарств, который используется для таких глобальных задач, как виртуальный поиск активных соединений и предсказание потенциальных мишеней конкретного соединения с помощью методологии обратного докинга[61].

Первые попытки автоматического докинга строились на концепции «ключ-замок», когда жесткий лиганд позиционировался в жесткий белок и геометрическая комплементарность была главным фактором при предсказании сродства. Однако настоящий процесс связывания лиганда предполагает, что лиганд и белок могут изменять свои конформации, чтобы эффективно взаимодействовать друг с другом. Такая модель заменила классическую концепцию «ключ-замок» и получила название «индуцированное соответствие». Как геометрическая, так и энергетическая комплементарность обеспечивает процесс формирования стабильного комплекса между белком и низкомолекулярным лигандом таким образом, чтобы достигался глобальный минимум свободной энергии. Исходя из этого, ключевая парадигма докинга может быть сформулирована следующим образом: лиганд связывается с белком вблизи глобального энергетического минимума, и задача докинга предполагает нахождение данного минимума[62].

Фактически процесс симуляции докинга сводится к трем аспектам. Первый – использование кристаллической структуры белка и выбор сайта докинга. Во многих программах докинга атомная модель белка используется косвенно. Например, методология сеточного докинга предполагает построение сетки заранее вычисленных потенциалов для определенной области белка, подлежащей докингу. Задание области докинга ограничивает пространство,

внутри которого происходит поиск связанной конформации лиганда. В ходе построения сетки в её узлы помещаются пробные атомы, и для них происходит вычисление энергии взаимодействия с белком[63]. Таким образом, каждый узел несет информацию обо всех типах взаимодействия для всех типов пробных атомов. После этого для конкретной конформации лиганда может быть вычислена энергия связывания как сумма энергий его отдельных атомов в поле белка. Непрямой учёт белка помогает значительно ускорить процесс докинга.

Вторым аспектом процесса позиционирования лиганда является процедура конформационного поиска. Эта процедура направлена на нахождение возможных конформаций лиганда, которые могут подходить для связывания в конкретном активном центре белка. С учетом 6 степеней свободы для трансляции и вращения как целого, а также степеней свободы, связанных с вращением вокруг каждой одинарной связи лиганд имеет огромное конформационное пространство, полный перебор которого видится слишком затратным с точки зрения вычислительного времени. Поэтому было разработано большое число различных алгоритмов, использующих некоторые вычислительные упрощения или особые процедуры для эффективного, но в то же время быстрого исследования конформационного пространства лиганда. При этом существующие алгоритмы можно классифицировать на систематические, то есть те, которые систематически исследуют пространство конформаций и дают воспроизводимые результаты, и стохастические, которые содержат элемент случайности. Главной особенностью последних является необходимость проведения нескольких независимых запусков алгоритма и расчет ряда метрик после докинга, которые позволяют удостовериться в полноте проведенного конформационного анализа.

Рассмотрим несколько примеров классических алгоритмов конформационного поиска. Одним из первых алгоритмов автоматического докинга с систематическим исследованием конформационного пространства был алгоритм соответствия формы программы DOCK[64]. При этом форма активного центра представляется в виде минимального набора сфер. Лиганд

«кладется» в этот центр таким образом, чтобы образовывалось соответствие между атомами лиганда и центрами сфер. Более продвинутой версией алгоритма стала методология соответствия фармакофора, когда помимо учёта формы лиганда происходит проверка соответствия фармакофорных признаков между белком и лигандом[65]. С учетом упрощенной оценки активного центра алгоритмы соответствия работают очень быстро.

Другим распространенным методом для систематического конформационного поиска является алгоритм постепенного конструирования. При инициализации лиганд «режется» на отдельные фрагменты за счет разрыва вращательных связей. После этого один из фрагментов (обычно самый большой) выбирается в качестве якорного и для него осуществляется докинг как жесткого тела. Оставшиеся фрагменты добавляются постепенно с дополнительной подгонкой их ориентации внутри активного центра[66].

Генетический алгоритм является хорошо известным стохастическим методом, который используется для решения многих сложных вычислительных задач[67]. Идея генетического алгоритма основана на эволюционной теории Дарвина. Степени свободы лиганда кодируется в виде бинарных строк, называемых генами. Совокупность генов образует хромосому, которая на самом деле представляет собой одну позу лиганда. Для исследования конформационного пространства алгоритм использует два генетических оператора: «кроссовер» (от англ. crossover) и мутация. Мутация вносит случайные изменения в гены, а «кроссовер» вызывает обмен генами между двумя хромосомами. Сгенерированные новые хромосомы проверяются с помощью определенной оценочной функции, и выжившие особи (имеющие хорошие показатели оценочной функции для своих хромосом) могут быть использованы для формирования следующего поколения. Таким образом, алгоритм использует свойство кумулятивности, что дает возможность двигаться к все более и более оптимальным с точки зрения оценочной функции позам лиганда.

Метод Монте-Карло также является распространенным стохастическим вычислительным методом для решения сложных задач[68]. Метод выполняет генерацию исследуемых поз лиганда напрямую – через трансляцию и ротацию как целого и за счет вращения вокруг одинарных связей. Каждая новая конформация проверяется на основе энергетического критерия отбора и, в случае его прохождения, сохраняется для генерации новой конформации. Кроме того, существует критерий уникальности, который отбрасывает те конформации, которые очень близки к тем, что уже исследованы. Данная итерация повторяется до тех пор, пока не будет собрано заданное число конформаций.

Последним аспектом докинга является скоринг-функция. Её целью является приближенная оценка силы взаимодействия между белком и лигандом. Фактически скоринг-функция используется для двух задач. Сначала скоринг-функция в ходе конформационного поиска помогает выделить наиболее вероятную позу лиганда с наименьшей энергией, которая соответствует связанной конформации лиганда. Второй задачей является предсказание активности конкретного соединения, что дает возможность ранжирования химических соединений по их способности связываться с мишенью в ходе виртуального скрининга. Иногда применяются разные скоринг-функции для решения этих двух задач.

Существующие скоринг-функции можно разделить на три вида: основанные на силовых полях, эмпирические и оценочные функции, основанные на знаниях[69]. Классические скоринг-функции, основанные на силовых полях, оценивают энергию связывания путем вычисления суммы несвязевых вкладов таких, как электростатическое взаимодействие и взаимодействие Ван-дер-Ваальса. Первый вклад описывается законом Кулона:

$$E_{\text{Coulomb}} = \frac{1}{4\pi\epsilon} \frac{q_i q_j}{r_{ij}}, \quad (12)$$

где q_i , q_j – атомные заряды i -того и j -того атома, r_{ij} – расстояние между ними, ϵ – диэлектрическая проницаемость, чаще всего зависящая от расстояния. Сила

отталкивания и сближения, входящие в состав взаимодействия Ван-дер-Ваальса, описывается потенциалом Леннарда-Джонса:

$$E_{\text{vdW}} = \varepsilon \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right], \quad (13)$$

где A_{ij} и B_{ij} – силовые константы, r_{ij} – расстояние между двумя атомами, ε – определяет минимум потенциала. Одними из самых частых приближений для скоринг-функций этого типа является «смягчение» потенциала Ван-дер-Ваальса для задания границы возможных близких контактов между белком и лигандом, а также применение граничного радиуса, за пределами которого несвязевые вклады не рассчитываются. Для повышения точности расчета энергии связывания скоринг-функции, основанные на силовых полях, могут содержать дополнительные поправки, например, для грубой оценки водородных связей, эффектов, связанных с влиянием растворителя, и энтропии.

В эмпирических скоринг-функциях энергия связывания также представляет собой линейную комбинацию отдельных вкладов таких, как энергия образования водородных связей, ионное взаимодействие, гидрофобный вклад и энтропия связывания [70]. Каждый компонент умножается на коэффициент, и сумма всех таких компонентов дает окончательное значение скор. Коэффициенты определяются с помощью регрессионного анализа таким образом, чтобы вычисленные энергии соответствовали известным энергиям связывания для комплексов белок-лиганд из тестового набора.

Третьим типом скоринг-функции являются функции, основанные на знаниях. В их основе лежит статистический анализ комплексов белок-лиганд, чтобы получить частоты и расстояния межатомных контактов между белком и лигандом. Главная идея этого подхода – это утверждение о том, что более выгодные взаимодействия встречаются чаще в экспериментальных структурах комплексов. Найденные распределения частот затем конвертируются в парный межатомный потенциал, на основе которого вычисляется конечный скор.

На сегодняшний день существует большое число программ и несколько десятков интернет-сервисов для выполнения докинга (таблица 3). Несмотря на

свою эффективность и выраженную популярность, методология докинга характеризуется рядом ограничений[71]. Основным слабым местом этого метода моделирования является невысокая точность предсказания активности, если опираться только на значение скоринг-функции. Поэтому часто докинг неверно предсказывает активность не только для химически близких молекул, но и для структурно разных соединений. Это, прежде всего, связано с использованием приближенных физических законов при описании энергии системы, что с одной стороны позволяет ускорить расчеты и повысить производительность виртуального скрининга, но с другой ухудшает предсказательную силу метода. Другим ограничением многих программ докинга является невозможность учета подвижности белка и полной реализации концепции «индуцированного соответствия» при связывании. Однако, как уже обсуждалось в главе 2, точность докинга можно повысить с помощью некоторой процедуры рескоринга, использующей более точную методологию, а подвижность участков активного центра может быть учтена в неявном виде за счет использования ансамбля белковых конформаций.

Таблица 3. Некоторые примеры существующих программ докинга.

Название программы	Алгоритм конформационного поиска	Тип скоринг-функции
AutoDock Vina	Глобальная оптимизация на основе локального поиска	Эмпирическая
AutoDock	Ламарковский генетический алгоритм	Эмпирическая
FlexX	Постепенное конструирование	Эмпирическая
GOLD	Генетический алгоритм	На основе силового поля и основанная на знании
DOCK	Постепенное конструирование	На основе силового поля
SOL	Генетический алгоритм	На основе силового поля
FLM	Монте-Карло	Использование силового поля без упрощений

3.2 Программа докинга SOL

Программа SOL была использована в качестве инструмента докинга в рамках данного исследования. Эта программа докинга характеризуется следующими особенностями. Используется жесткая модель, в рамках которой активный центр представляет собой сетку заранее вычисленных потенциалов между пробными атомами и атомами белка. Данные межатомные потенциалы включают в себя электростатический вклад и силы Ван-дер-Ваальса, которые вычисляются на основе силового поля MMFF94[72]. Кроме того, силовое поле MMFF94 используется для вычисления энергии внутренних напряжений лиганда. В отличие от многих программ докинга SOL содержит описание эффектов сольватации-десольватации в рамках приближенной модели обобщенного принципа Борна[73]. Потенциалы этого взаимодействия также вычисляются для каждого узла сетки потенциалов. Сама сетка строится с помощью отдельной программы SOLGRID. Для конформационного поиска SOL использует популярный для многих программ докинга генетический алгоритм. При этом, как и большинстве методов конформационного поиска, генерация конформаций лиганда выполняется в рамках приближения «жесткого ротора» (rigid rotor approximation), то есть изменению подвергаются именно степени свободы, связанные с вращением вокруг одинарных связей, а длины связей и значения валентных углов зафиксированы. Позы лиганда генерируются внутри куба докинга со стандартным размером ребра равным 11 Å. Вышеупомянутые потенциалы электростатического и Ван-дер-Ваальсова взаимодействия, а также сольватации и десольватации вычисляются для сетки размера 101x101x101, которая ограничена данным кубом. При этом используется упрощенная типизация атомов: вместо 99 типов, реализованных для силового поля MMFF94, вычисления происходят лишь для 27 типов пробных атомов. Энергия каждой сгенерированной позы лиганда вычисляется как сумма энергий всех его атомов, а компоненты этих энергий рассчитываются с помощью линейной интерполяции соответствующих энергий пробных атомов для 8 ближайших узлов сетки.

В рамках генетического алгоритма, реализованного в программе SOL для анализа конформационного пространства лиганда, происходит имитация процесса эволюции для решения задачи глобальной минимизации. Каждая отдельная поза лиганда закодирована в виде хромосомы, которая состоит из отдельных генов. Гены несут информацию о координатах геометрического центра конкретной позы, а также указывают на вращение лиганда как целого и значений торсионных углов. Другими словами, «генотип» особи задает конкретное пространственное расположение лиганда. Оценочная функция задает соответствие между набором геном определенной позы и её «фенотипом» – энергией взаимодействия белок-лиганд. Отбор направлен в сторону минимизации значений этой энергии, поэтому предполагается, что спустя определенное число поколений будет найдено оптимальное решение, соответствующее геометрии глобального энергетического минимума системы белок-лиганд.

Алгоритм работает следующим образом. Вначале случайным образом генерируется первая популяция. Её размер, а также размер каждой новой популяции после смены поколения определяется параметром *population size*, который по умолчанию равен 30000. Для каждой особи с её набором генов вычисляется значение оценочной функции и выполняется ранжирование всех особей по значению этой функции. Далее формируется так называемый пул родителей, который используется для генерации следующего поколения и размер которого задается параметром *mating pool*. Он формируется за счет переноса особей с лучшими значениями оценочной функции с учётом «нишинга», а также случайным образом на основе колеса рулетки. В первом случае лучшая особь после ранжирования первого поколения с таким набором генов, который отвечает самой низкой энергии взаимодействия с белком, напрямую переносится в пул родителей. «Нисинг» предполагает назначение штрафа энергии для тех особей, которые имеют близкий набор генов к этой лучшей особи. С учетом этого штрафа происходит новое ранжирование особей и процедура повторяется. Такая процедура предотвращает раннее схождение

алгоритма к одному из локальных минимумов системы белок-лиганд. Во втором случае отбор в пул родителей происходит случайно на основе колеса рулетки, число запусков которой равно значению MATING POOL. При этом чем ниже энергия взаимодействия с белком для определенной особи, тем выше вероятность, что рулетка выберет эту особь для переноса в пул родителей. Сформированный пул родителей ранжируется, и лучшие особи напрямую переносятся в следующее поколение. Для оставшихся особей используются операторы мутации и «кроссовера», которые вносят случайные изменения в гены особей пула родителей, что ведет к образованию нового поколения и цикл повторяется. Процедура смены поколений происходит заданное число раз, определяемое параметром NUMBER OF GENERATION. Результатом одного запуска генетического алгоритма является одна поза лиганда.

С учётом стохастической природы генетического алгоритма требуется производить несколько независимых запусков для подтверждения того, что конформационное пространство конкретного лиганда исследовано репрезентативно. Число независимых запусков генетического алгоритма в программе SOL определяется параметром NUMBER OF RUNS. Проверка охвата конформационного пространства производится с помощью кластерного анализа найденных решений, собранных после нескольких запусков работы алгоритма (обычно пятидесяти). Самая низкая по энергии поза формирует первый кластер и обозначается как лучшая поза после докинга. Позы близкие к ней, то есть те, для которых величина среднеквадратичного отклонения от лучшей позы получается менее 2 \AA , присваиваются первому кластеру. Оставшиеся позы также формируют отдельные кластеры, заполняющиеся близкими решениями. Сходимость результатов, полученных в разных запусках генетического алгоритма, а значит, и полнота охвата конформационного пространства подтверждается тем, что наблюдается высокая населенность первого кластера и небольшое число других кластеров. В тех случаях, когда наблюдается противоположная картина, докинг повторяется с более высокими параметрами генетического алгоритма.

Скоринг-функция программы SOL представляет собой линейную комбинацию отдельных энергетических членов, умноженных на определенные коэффициенты:

$$\Delta G = \alpha E_{Vdw} + \beta E_{Coulomb} + \beta E_{GB-complex} + \gamma E_{GB-protein} + \delta E_{GB-ligand} + \mu N_{torsions}, \quad (14)$$

где E_{Vdw} и $E_{Coulomb}$ – ван-дер-ваальсовое и электростатическое взаимодействие между белком и лигандом, рассчитываемое, соответственно, согласно уравнению (13) и уравнению (12) в рамках силового поля MMFF94; $E_{GB-complex}$, $E_{GB-protein}$, $E_{GB-ligand}$ – отдельные вклады энергии сольватации, рассчитанные, соответственно, для комплекса, белка и лиганда на основе упрощенного метода для обобщенной Борновской модели; $N_{torsions}$ – число торсионных лиганда. Этот вклад добавлен для учёта потери конформационной подвижности лиганда при связывания в качестве энтропийного члена энергии связывания. Коэффициенты α , β , γ , δ , μ подобраны таким образом, чтобы лучше соответствовать экспериментальным значениям свободной энергии связывания для известных лигандов и правильно воспроизводить закристаллизованные положения лигандов с их рецепторами.

В ходе валидации программы SOL были получены хорошие результаты позиционирования для тестового набора кристаллических комплексов белок-лиганд. Так для 50 комплексов из 80 удалось воспроизвести положение закристаллизованного лиганда. Для данного тестового набора точность позиционирования для SOL оказалась лучше, чем для программы AutoDock 3.05[74]. Кроме того, программа докинга SOL была успешно использована для разработки ингибиторов тромбина[75] и урокиназы[76].

Глава 4. Поиск новых ингибиторов фактора Ха

4.1 Создание и валидация модели фактора Ха

Для построения модели белка фактора Ха предполагалось использовать кристаллическую структуру из банка белковых структур Protein Data Bank (PDB). Всего было найдено 47 структур фактора Ха хорошего разрешения ($< 2.0 \text{ \AA}$), среди которых было взято 11 комплексов фактора Ха с известными ингибиторами для дальнейшего анализа. Структура фактора Ха была взята из комплекса 3CEN, так как этот комплекс не содержит каких-либо пропущенных атомов и имеет хорошие показатели качества (разрешение: 1.6 \AA , R-фактор рабочий: 0.249). Комплекс содержит наномолярный ингибитор фактора Ха с идентификатором в базе PDB FXA, который показан на рисунке 10. Его положение в активном центре фактора Ха показано на рисунке 3. При подготовке белка лиганд был извлечен из файла комплекса в отдельный файл, а также были удалены молекулы воды.

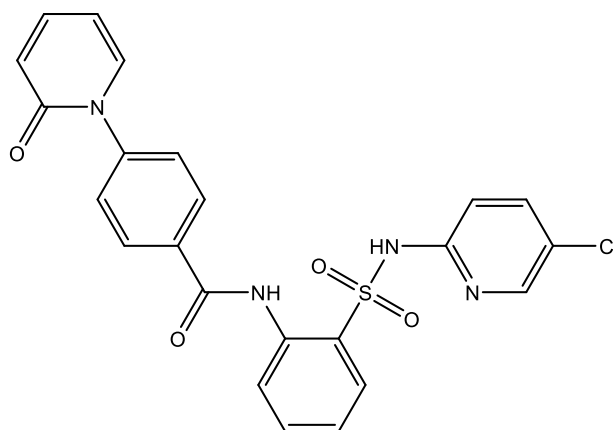


Рисунок 10. Лиганд, закристаллизованный вместе с фактором Ха в комплексе 3CEN. Значение K_i для фактора Ха составляет 5.5 нМ .

Очищенная от других молекул структура белка была протонирована с учётом нейтрального значения pH с помощью программы APLITE. Наряду с протонированием эта программа также выполняет типизацию атомов белка в силовом поле MMFF94, что необходимо для вычисления сетки потенциалов в программе SOLGRID. В качестве центра куба докинга был использован геометрический центр закристаллизованного положения нативного для

комплекса 3CEN ингибитора, для которого также было выполнена процедура протонирования при pH 7.4 программой Avogadro[77]. Размер куба докинга, содержащего сетку потенциалов, которая была рассчитана в программе SOLGRID, был выбран стандартным – 22x22x22 Å. Было проверено, что данный куб охватывает весь активный центр фактора Ха.

Для проверки того, что построенная модель правильно отражает поведение фактора Ха в экспериментах по связыванию подтвержденных ингибиторов, была выполнена валидация на основе докинга. Сначала был выполнен докинг с помощью программы SOL лиганда из комплекса 3CEN, то есть нативного по отношению к структуре фактора Ха, взятой для построения модели. В ходе данного докинга получилось воспроизвести закристаллизованного положение данного лиганда: среднеквадратичное отклонение (RMSD) по всем атомам между лучшей позой после докинга и позой в кристалле составило 1.38 Å. Значение скоринг-функции оказалось равным -6.44 ккал/моль. После этого было выполнено 10 процедур кросс-докинга с использованием других комплексов фактора Ха с известными ингибиторами. Одна процедура кросс-докинга выполнялась следующим образом. Проводилось наложение координат другого комплекса на координаты 3CEN с использованием программы *superimpose_proteins*. Такой «повернутый» комплекс теперь содержал лиганд, являющийся «квазинативным» по отношению к структуре фактора Ха из 3CEN. Преимущество наложения координат двух комплексов в том, что точность докинга лиганда в «квазинативном» положении может быть удобно оценена с помощью расчета RMSD, также как это происходит при докинге лиганда нативного по отношению к структуре белка. Результаты процедур кросс-докинга приведены в таблице 4. Помимо оценки репрезентативности модели кросс-докинг помогает оценить порог отбора активных молекул по скору, который впоследствии используется при виртуальном скрининге.

Таблица 4. Результаты кросс-докинга для молекулярной модели фактора Ха, построенной на основе комплекса 3CEN.

Название комплекса, использованного при кросс-докинге	Заряд лиганда	RMSD между лучшим положением после докинга и «квазинативным» положением	Скоринг-функция в SOL, ккал/моль	$\Delta H_{\text{связывания}}$ PM7, ккал/моль
2P3T	0	0.47	-6.89	-85.12
2P94	0	0.77	-7.39	-91.01
2VH6	0	1.78	-6.78	-73.89
2VVC	+1	1.3	-6.42	-84.28
2W26	0	0.6	-6.87	-73.58
2Y7X	+1	2.35	-7.17	-66.46
3PT	+1	0.95	-6.64	-62.48
3TK6	+1	0.55	-6.95	-71.91
4ZH8	0	1.83	-7.09	-72.48
4ZHA	0	1.49	-5.77	-60.04

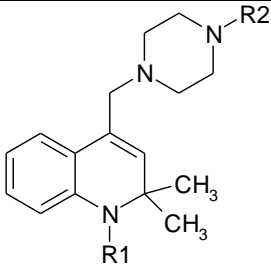
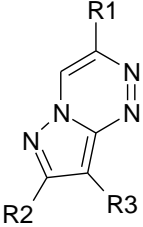
Согласно методу, описанному в пункте 2.5, для известных ингибиторов фактора Ха была рассчитана энтальпия связывания методом PM7 и учетом эффектов растворителя континуальной моделью COSMO (см. таблица 4). Как и расчет значений скоринг-функции в программе докинга SOL, вычисление энтальпии связывания для известных лигандов дает представление о тех значениях рассчитываемых энергий, которые показывают реальные ингибиторы мишени. Это также помогает выбрать порог, на основе которого можно впоследствии разделить активные и неактивные молекулы между собой после проведения виртуального скрининга. С учетом результатов докинга 10 подтвержденных ингибиторов фактора Ха среднее значение сора в программе SOL для известных лигандов этой мишени составило -6.76 ± 0.44 ккал/моль, а среднее значение энтальпии связывания, рассчитанной для лучшей позы после

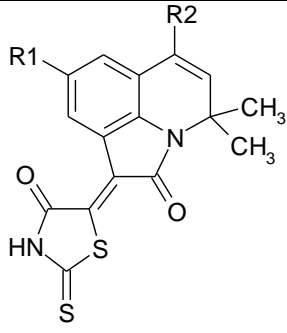
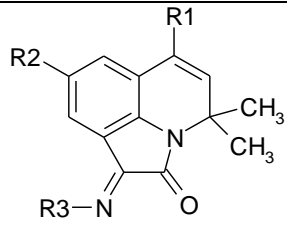
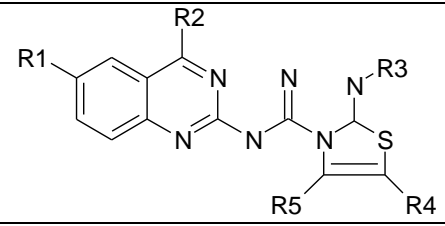
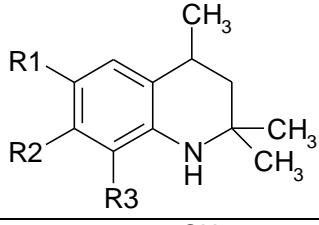
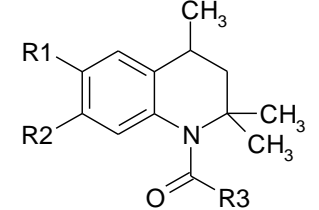
докинга с помощью полуэмпирического метода, получилось равным -74.12 ± 10.05 ккал/моль. Именно такие лиганды, которые показывают расчетные характеристики близкие к данным диапазонам энергий, предполагалось считать потенциально активными при прогнозировании активности против фактора Ха.

4.2 Виртуальный скрининг сфокусированной библиотеки

Первичной библиотекой для скрининга стала небольшая сфокусированная библиотека, которая была разработана сотрудниками кафедры органической химии Воронежского государственного университета (ВГУ). Библиотека содержала 244 соединения, отобранных на основе сходства с ингибиторами фактора Ха, найденными на предыдущих этапах работы. Фактически предсказание активности для аналогов найденных ингибиторов и дальнейшее тестирование лучших кандидатов выполнялось в рамках стратегии расширения списка молекул-«хитов» (hit expansion). Соединения этой библиотеки были сгруппированы на 7 подгрупп согласно строению центрального ядра. Структурные формулы этих центральных ядер приведены в таблице 5.

Таблица 5. Центральные ядра соединений сфокусированной библиотеки.

№	Химическое название центрального ядра	Структурная формула с обозначением мест прикрепления заместителей
1	2,2-диметил-4-пиперазинилметил-1,2-дигидрохиолин	
2	Пиразоло-1,2,4-триазин	

3	Диметилпиррологидрохинолинон-2-тиоксотиазолидин-4-он	
4	Диметилпиррологидрохинолин-3-карбоксиминамид	
5	Амино-N-хинализонил-1,3-дигидротиазол	
6	2,2,4-триметил-1,2,3,4-тетрагидрохинолин	
7	N-ацетил-2,2,4-триметил-1,2,3,4-тетрагидрохинолин	

Подготовка библиотеки к скринингу была основана на методике, описанной в разделе 2.2, которая предполагает использование программы Marvin для протонирования низкомолекулярных лигандов и использование программы Corina для их конвертации в трехмерный формат. С учетом нескольких зарядовых форм, сгенерированных при подготовке соединений, докингу с помощью программы SOL в подготовленную модель фактора Ха подлежало 270 уникальных молекул. По результатам докинга был сформирован список лучших кандидатов по скору после ранжирования молекул по данному показателю. В данном списке было, главным образом, два химических класса молекул, для которых были получены самые отрицательные значения скоры

программы SOL, а именно производные пиррологидрохинолинона (молекулы с центральным ядром №3 и №4 таблицы 5) и производные тетрагидрохинолина (молекулы с центральным ядром №6 и №7 таблицы 5). Позы после докинга производных пиррологидрохинолинона и 2,2,4-триметил-1,2,3,4-тетрагидрохинолина были схожи с конформациями известных ингибиторов, а также наблюдались схожие типы межмолекулярных контактов, такие как водородные связи с аминокислотными остатками Gly216 и/или Gly218, π -стэкинг с остатками Tyr228, Tyr99 и Phe174 и анион- π взаимодействие с Asp189 (см. рисунок 11).

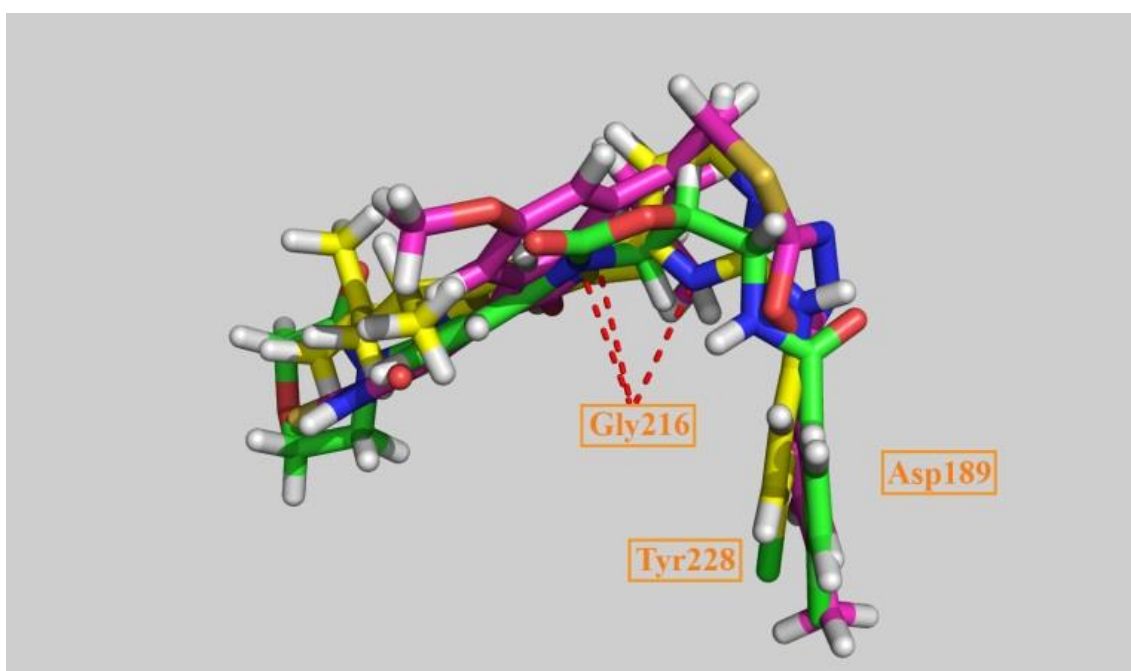


Рисунок 11. Наложение положений после докинга производного тетрагидрохинолина (желтые атомы углерода), производного пиррологидрохинолинона (розовые атомы углерода) и ривароксабана (зеленые атомы углерода). Видно, что все три молекулы образуют схожие контакты с белком с формированием водородной связи с Gly216, анион- π взаимодействия с заряженным остатком Asp189 и π -стэкинга с остатком Tyr228.

В тоже время, несмотря на хорошие значения скоринг-функции, производные N-ацетил-2,2,4-триметил-1,2,3,4-тетрагидрохинолина оказались неспособными занять оба кармана связывания. При этом наблюдались «инвертированные» положения после докинга, когда предполагаемые P1-

мотивы лигандов располагались в кармане S4 вместо кармана S1 (см. рисунок 12). Поэтому данная группа молекул была исключена из списка лучших соединений – процедура постпроцессинга для производных N-ацетил-2,2,4-триметил-1,2,3,4-тетрагидрохинолина не проводилась.

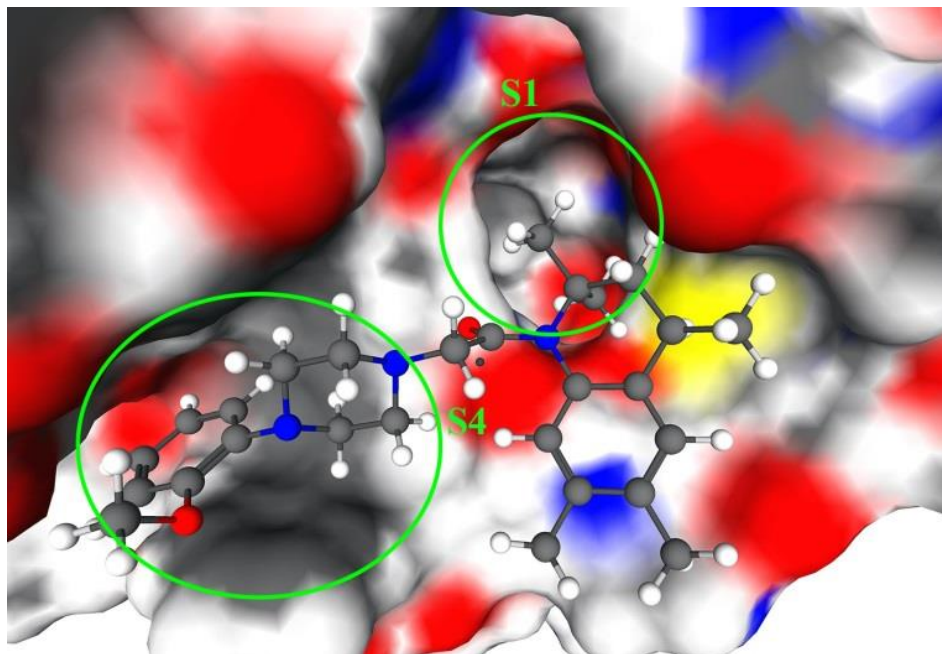


Рисунок 12. «Инвертированное» положение после докинга производного N-ацетил-2,2,4-триметил-1,2,3,4-тетрагидрохинолина.

С учетом результатов визуального анализа для 40 лучших соединений был проведен постпроцессинг методом PM7 для оптимизации положения после докинга и расчета энтальпии связывания. На основе всех расчетных величин и положений в активном центре фактора Ха после оптимизации было отобрано 20 кандидатов на экспериментальное тестирование. Тестирование было основано на амидолитической пробе, в ходе которой оценивается влияние потенциальных ингибиторов на скорость расщепления хромогенного субстрата фактором Ха, детектируемое с помощью спектрофотометра. Эксперименты по проверке активности кандидатов в ингибиторы фактора Ха были выполнены Еленой Николаевной Липец в лаборатории биофизики НМИЦ ДГОИ имени Дмитрия Рогачева. Выявленные активности, классы соединений и их расчетные характеристики приведены в таблице 6. Структуры активных молекул показаны на рисунке 13.

Таблица 6. Результаты тестирования 20 кандидатов в ингибиторы фактора Ха, отобранных на основе виртуального скрининга. Процент ингибирования рассчитан как отношение разницы между активностями фактора Ха без ингибитора и в его присутствии и активностью без ингибитора.

Название соединения	Химический класс центрального ядра *	Скор в SOL, ккал/моль	$\Delta H_{\text{связывания}}$ PM7, ккал/моль	Концентрация в эксперименте, мкМ	Процент ингибирования фактора Ха (\pm SD), %
VGY 0224950	ТГХ	-7.26	-77.2	30	99±0
VGY 0224908	ТГХ	-7.23	-57.22	30	11±6
VGY 0224909	ТГХ	-7.03	-69.5	30	13±4
VGY 0224586	АХД	-6.4	-50.32	30	6±6
VGY 0224867	ТГХ	-7.02	-57.95	30	6±8
VGY 0224933	ПГКИ	-6.2	-45.99	30	51±5
VGY 0225108	ПГКИ	-6.78	-47.39	30	57±3
VGY 0224963	ПГКИ	-6.21	-44.83	30	1±9
VGY 0224865	ТГХ	-7.08	-51.72	30	18±7
VGY 0224918	ТГХ	-6.78	-42.3	30	33±9
VGY 0224945	ТГХ	-7.01	-58.65	30	65±3
VGY 0224887	ТГХ	-7.13	-47.63	30	37±8
VGY 000060	ПГТТ	-8.67	-50.36	1	37±5
VGY 000049	ПГТТ	-8.08	-40.23	1	16±10
VGY 000010	ПГТТ	-7.99	-33.51	1	14±2
VGY 000073	ПГТТ	-7.83	-41.95	1	12±1
VGY 000014	ПГТТ	-6.84	-34.68	1	10±4
VGY 000058	ПГТТ	-6.79	-35.63	1	0±2
VGY 000063	ПГТТ	-6.84	-34.98	1	13±9
VGY 000046	ПГТТ	-6.98	-38.8	1	10±7

* Обозначения классов для центрального ядра: ТГХ – тетрагидрохиолин, АХД - амино-N-хинализонил-1,3-дигидротиазол, ПГКИ - диметилпиррологидрохиолин-3-карбоксиминамид, ПГТТ – диметилпиррологидрохиолинон-2-тиоксотиазолидин-4-он.

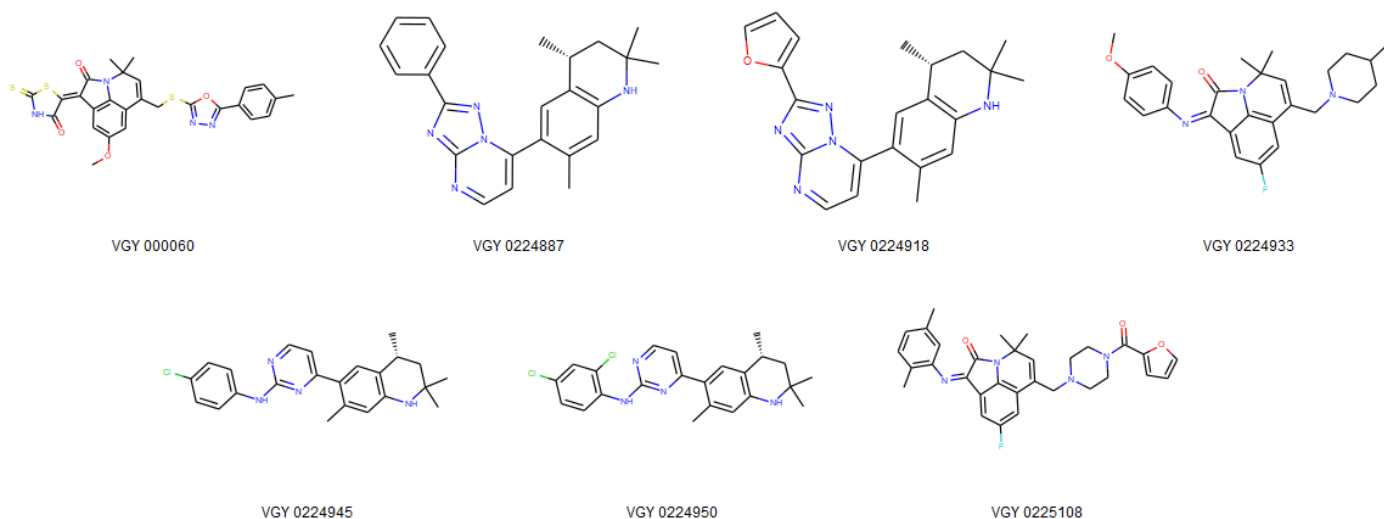


Рисунок 13. Структуры ингибиторов фактора Ха, выявленных в ходе виртуального скрининга сфокусированной библиотеки.

Как видно из таблицы 6, семь из двадцати отобранных кандидатов проявили заметную активность в отношении фактора Ха. Лучшим ингибитором стало соединение VGY 0224950, полностью подавившее работу белка в концентрации 30 мкМ. Для этого вещества были сделаны дополнительные измерения при варьировании его концентрации с целью подтверждения конкурентного механизма ингибирования и определения значения полумаксимальной ингибирующей концентрации (IC_{50}). На графике, приведенном на рисунке 14, видно, что зависимость между изменением концентрации ингибитора и активностью фактора Ха отвечает механизму конкурентного ингибирования. Найденное значение IC_{50} составило 1.9 мкМ.

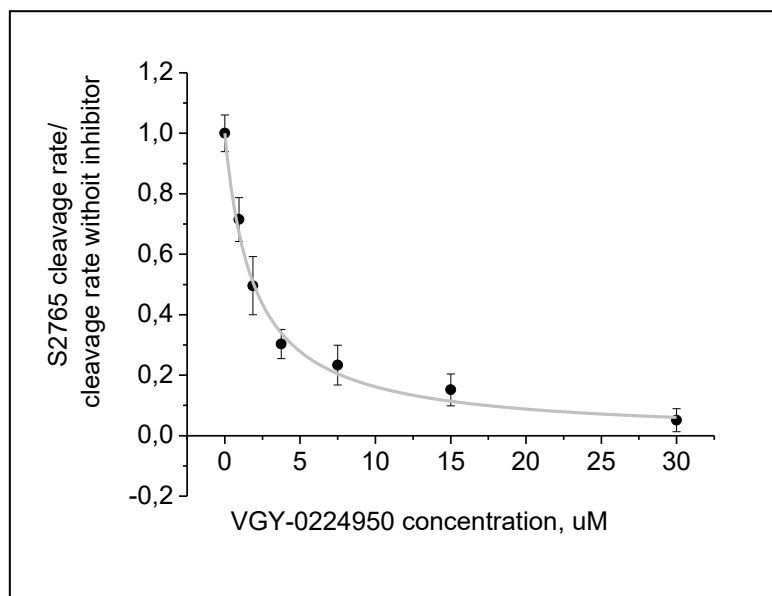
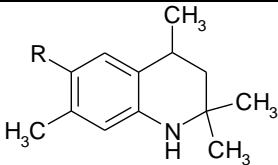
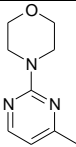
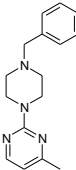
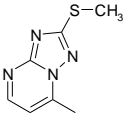
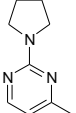

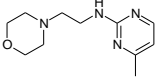
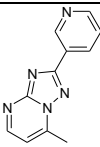
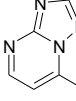
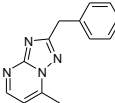


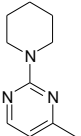
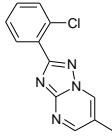
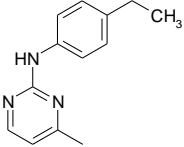
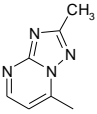
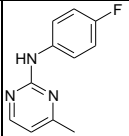
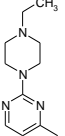
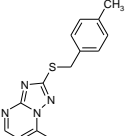
Рисунок 14. Соотношение активности фактора Ха в присутствии и отсутствии VGY 0224950. Точки показывают среднее значение и стандартное отклонение для трех измерений. Серая линия отражает зависимость активности фактора Ха и концентрации VGY 0224950 исходя из допущения о том, что данный ингибитор конкурентно блокирует работу белка.

4.3 Идентификация новых ингибиторов фактора Ха на основе 2,2,4-триметил-1,2,3,4-тетрагидрохинолина

Так как именно соединение на основе тетрагидрохинолина показало максимальную активность в ходе скрининга, была предложена стратегия дополнительного изучения закономерностей структура-активность для производных тетрагидрохинолина. Поэтому эксперимент был проведен еще для 16 соединений, которые были отобраны на основе докинга из подгруппы производных 2,2,4-триметил-1,2,3,4-тетрагидрохинолина. Четыре соединения проявили активность против фактора Ха в концентрации 5 мкМ (таблица 7).

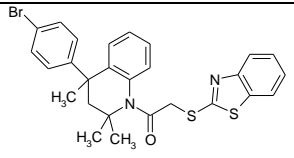
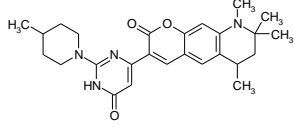
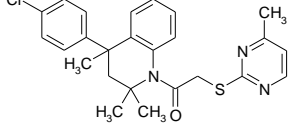
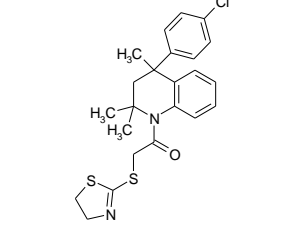
Таблица 7. Дополнительное изучение производных тетрагидрохинолина в качестве ингибиторов фактора Ха. Концентрация всех ингибиторов в эксперименте составила 5 мкМ.

			
Название соединения	R =	Скор в программе SOL, ккал/моль	Процент ингибирования, %
VGY 0224863		-5.46	3±2
VGY 0224864		-5.89	55±4
VGY 0224880		-5.6	5±2
VGY 0224883		-5.1	2±4
VGY 0224893		-5.83	3±6
VGY 0224897		-6.03	19±3
VGY 0224907		-5.62	6±1
VGY 0224912		-5.74	10±4
VGY 0224921		-5.55	11±9

VGY 0224922		-5.24	0±3
VGY 0224948		-5.5	14±2
VGY 0225189		-5.79	35±10
VGY 0225193		-5.23	4±1
VGY 0225195		-6.12	21±4
VGY 0225196		-5.64	12±5
VGY 0225197		-5.82	49±8

Для дополнительного исследования химического пространства соединений данного класса на предмет ингибирующей активности в отношении фактора Ха виртуальному скринингу была подвергнута выборка молекул на основе тетрагидрохинолина из полной базы ВГУ, содержащей около 15 000 соединений[78]. В ходе этого скрининга 4 кандидата в ингибиторы были отобраны на основе расчетных характеристик. Два соединения заметно ингибировали фактор Ха в концентрации 5 мкМ (таблица 8). Одним из ингибиторов оказалось соединение VGY 0224989, которое содержит тетрагидрохинолиновый фрагмент конденсированный с пираноном. Дополнительные измерения активности этого соединения против фактора Ха в различных концентрациях позволили обнаружить, что его IC₅₀ составляет 3 мкМ.

Таблица 8. Производные тетрагидрохинолина, отобранные из полной базы ВГУ и предсказанные в качестве ингибиторов фактора Ха.

Название	Структура	Скор в программе SOL, ккал/моль	$\Delta H_{\text{связывания}}$ PM7, ккал/моль	Процент ингибирования, %
VGY 0006201		-7.82	-48.25	0,08±0,03
VGY 0224989		7.15	-45.8	0,16±0,11
VGY 0006212		-6.95	-56.3	0,85±0,06
VGY 0006208		-6.42	-50.65	0,76±0,09

Для двух лучших ингибиторов, VGY 0224950 и VGY 0224989, были выполнены дополнительные эксперименты с другими сериновыми протеазами, являющимися физиологически близкими к фактору Ха, чтобы оценить их селективность. Оказалось, что VGY 0224950 помимо фактора Ха также ингибирует трипсин и фактор XIa, но не влияет на активность тромбина и фактора IXa. При этом VGY 0224989 – селективный ингибитор фактора Ха, не влияющий на активность других протестированных протеаз.

Таким образом, в ходе трех итераций виртуального скрининга на основе комбинированного подхода, сочетающего докинг в программе SOL и постпроцессинг методом PM7, удалось идентифицировать 12 микромолярных ингибиторов фактора Ха из 40 отобранных кандидатов. Такой результат подчеркивает эффективность используемого подхода для выявления активных

молекул, что является критическим для начальной разработки лекарств. Выявленные ингибитора фактора Ха представляют интерес в качестве «стартовых точек», которые могут быть использованы для дизайна более активных и селективных антагонистов этого белка.

4.4 Влияние постпроцессинга на предсказание активности

Для оценки влияния постпроцессинга на прогнозирование биологической активности были проведены дополнительные расчеты для 129 веществ, среди которых было 28 подтвержденных ингибиторов фактора Ха с активностью на уровне нескольких микромоль и 101 соединение с подтвержденным отсутствием активности в отношении фактора Ха в эксперименте. Фактически требовалось сравнить два вида предсказания активности: предсказание активности только на основе скоринг-функции программы SOL и предсказание активности по двум критериям – и по скоринг-функции, и по энтальпии связывания, рассчитанной методом PM7. В ходе первого метода предсказания использовался только критерий по скору на уровне -5.3 ккал/моль (соединения со скором ниже -5.3 ккал/моль предполагаются активными). Второй метод подразумевал соединение активным, если оно показывало скоринг-функцию ниже -5.3 ккал/моль и энтальпию связывания ниже -35 ккал/моль. После выполнения соответствующих расчетов сравнение двух методов проводилось в терминах точности, специфичности, чувствительности и F1-меры:

$$\text{Точность} = \frac{TP+TN}{N} * 100\% \quad (14)$$

$$\text{Специфичность} = \frac{TN}{TN+FP} \quad (15)$$

$$\text{Чувствительность} = \frac{TP}{TP+FN} \quad (16)$$

$$\text{F1-мера} = \frac{2 TP}{2TP+FP+FN} \quad (17)$$

где TP – истинно положительное предсказание (активность предсказана и она есть в эксперименте), TN – истинно отрицательное предсказание (активность не предсказана и её нет в эксперименте), FP – ложноположительная ошибка

предсказания (активность предсказана, но её нет в эксперименте), FN – ложноотрицательная ошибка предсказания (активность не предсказана, но она есть в эксперименте). Точность – общая доля правильных предсказаний. Чувствительность – это доля правильно предсказанных активных молекул среди всех активных молекул, а специфичность – это доля правильно выявленных неактивных молекул среди пула всех неактивных соединений. Наконец, F1-мера – это общая оценка способности классификатора различать активные и неактивные молекулы. Результаты сравнения двух подходов показаны в таблице 9.

Таблица 9. Сравнения двух подходов по способности прогнозировать активность в отношении фактора Ха.

Подход	TP	TN	FP	FN	Точность	Чувствительность	Специфичность	F1-мера
SOL	22	34	67	6	43,4%	0,78	0,33	0,376
SOL + энтальпия связывания	16	73	28	12	69%	0,57	0,72	0,444

Из приведенной таблицы видно, что учет энтальпии связывания позволяет сильно снижать долю ложноположительных предсказаний, заметно повышая специфичность отбора, хотя и ценой нескольких дополнительных ошибок II рода. В целом, прогнозирование активности на основе и скоринг-функции программы SOL, и энтальпии связывания, рассчитанной полуэмпирическим методом PM7, оказывается более точным с точки зрения метрики точности и F1-меры. Таким образом, применение постпроцессинга на основе квантовой химии в ходе виртуального скрининга оправдано за счет снижения числа ложноположительных ошибок почти в 2,5 раза.

Глава 5. Поиск новых ингибиторов белка TEM-1

5.1 Создание модели аллостерического ингибирования β -лактамазы TEM-1

Теоретическая аллостерическая модель ингибирования β -лактамазы TEM-1, разработанная в рамках данного исследования, строилась на следующих аспектах. Главной структурной частью белка TEM-1, на которую предполагалось воздействовать для блокирования работы фермента, стала Ω -петля – петлевой участок белка, расположенный вблизи активного центра и участвующий во второй стадии гидролиза молекулы антибиотика. Помимо учета её функциональной роли выбор Ω -петли в качестве молекулярной мишени оправдан двумя другими причинами. Во-первых, согласно многим исследованиям этот участок лактамазы отличается высокой консервативностью, что может препятствовать развитию резистентности к разработанным ингибиторам. Во-вторых, анализ поверхности белка, где располагается Ω -петли, показывает наличие полости, которая может служить карманом связывания низкомолекулярных соединений (рис. 15).

Молекулярный механизм ингибирования белка TEM-1 заключался в предположении о том, что связывание в С-концевой части Ω -петли низкомолекулярного лиганда может вызвать смещение положения ключевых остатков в N-концевой части (прежде всего, Glu166 и Asn170), ответственных за стадию деацилирования – вторую стадию процесса гидролиза антибиотика. Пространственное отклонение этих остатков может привести к сильному замедлению этой стадии, а значит, катализ будет остановлен на первой стадии, и β -лактамаза будет ингибирована антибиотиком, ковалентно связанным внутри активного центра. Методологически такая модель ингибирования предполагала предсказание лигандов, которые бы смогли связаться внутри полости, образуемой С-концевой частью Ω -петли белка TEM-1. В качестве инструмента прогнозирования средства был выбран докинг в программе SOL и постпроцессинг полуэмпирическим методом PM7.

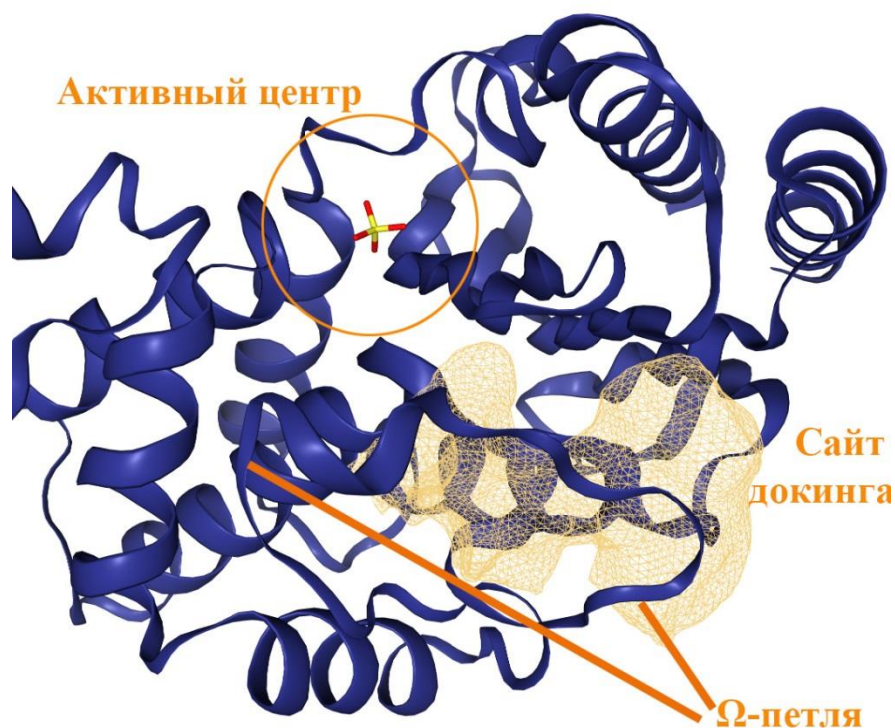


Рисунок 15. Вид аллостерического сайта, образованного С-концевой частью Ω-петли. Объем сайта показан бежевой поверхностью.

Построение атомистической модели β-лактамазы TEM-1 было выполнено на основе структуры, взятой из комплекса 1XPВ. Этот комплекс имеет хорошее разрешение (1.9 Å) и показатель R-фактора (рабочее значение: 0.156) и не содержит каких-либо пропущенных остатков и атомов. Добавление атомов водорода было проведено в программе APLITE. В качестве центра куба докинга был выбран геометрический центр остатка Pro67, располагающегося на внутренней стенке аллостерического сайта вблизи его центральной части. Расчет решетки потенциалов был сделан в программе SOLGRID.

5.2 Виртуальный скрининг библиотеки NCI Diversity set V

Библиотека NCI Diversity set V содержит 1593 полочных соединений, отобранных из нескольких десятков тысяч соединений полной химической базы NCI с учетом химического разнообразия[79]. Разнообразие между молекулами оценивалось на основе фармакофорного профиля. При добавлении в базу фармакофоры для соединения-кандидата сравниваются с набором всех фармакофоров, обнаруженных в структурах, уже принятых в набор. Если

структура соединения-кандидата имеет новый фармакофорный профиль, то она добавляется в набор.

Подготовка библиотеки к скринингу была основана на методике, описанной в разделе 2.2, то есть её протонирование выполнялось расчетным модулем pKa программы Marvin, а генерации трехмерных геометрий – в программе Corina. Конечное число молекул в подготовленном варианте базы NCI Diversity set V составило 3660.

Для подготовленных соединений был произведен докинг в программе SOL на стандартных параметрах. Для повышения производительности расчетов был выполнен параллельный запуск нескольких сотен процедур докинга с использованием суперкомпьютера «Ломоносов-2»[80]. Для 8 лучших соединений, отобранных на основе визуального анализа и величины скоринг-функции программы SOL, был проведен постпроцессинг, в ходе которого производился расчет уточненной энергии связывания на основе метода PM7. Так как для изучаемого сайта связывания не существует известных лигандов, выбор поровых значений энергий, рассчитываемых в ходе виртуального скрининга и которые используются для прогнозирования сродства, строился на основе общих закономерностей, обнаруженных в ходе использования данного протокола для других мишеней и их ингибиторов. Чаще всего известные высокоактивные (значение K_i на уровне наномоль) ингибиторы различных белков показывают энтальпию связывания, в диапазоне от -50 ккал/моль до -80 ккал/моль.

Результаты моделирования приведены в таблице 10. Соединение 639174 показало лучшее значение энтальпии связывания равное -52.6 ккал/моль. Предсказанное положение данного ингибитора внутри аллостерического сайта характеризуется позиционированием положительно заряженной аминогруппы данного соединения вглубь основания Ω -петли вблизи отрицательно заряженного остатка Asp176, что и дает значимый прирост в энергии связывания. Кроме того, наблюдается формирование водородной связи между амидной группой соединения 639174 и атомом кислорода остова остатка Ile173.

Таблица 10. Результаты виртуального скрининга для 8 соединений с лучшими показателями скоринг-функции программы SOL.

№	NSC номер в базе NCI	Скоринг-функция в SOL, ккал/ моль	Населенность 1-го кластера	Число кластеров	$\Delta H_{\text{связывания}} \text{ PM7, ккал/моль}$
1	170955	-4.98	43	6	-27.56
2	311153	-5.13	47	2	-34.25
3	327444	-4.47	9	5	-17.9
4	345845	-5.4	21	7	-40.37
5	372275	-5.32	28	5	-19.09
6	639174	-5.86	4	10	-52.6
7	81703	-5.15	37	6	-35.31
8	85433	-4.75	8	11	-23.55

Таким образом, одно соединение из базы NCI Diversity set V оказалось потенциальным лигандом аллостерического сайта белка TEM-1 согласно результатам докинга. С учетом предсказанного с помощью докинга положения соединения 639174 и его структуры был сформулирован молекулярный профиль возможных аллостерических ингибиторов β -лактамазы TEM-1. По-видимому, для успешного связывания внутри выбранного сайта потенциальные лиганды должны иметь якорный фрагмент, который способен связываться в (см. рисунок 16). Структурным мотивом, который служит якорным фрагментом в молекуле 639174, является 2-аминоацетамидная группа. Исходя из строения аллостерического сайта, потенциальные якорные группы должны иметь ациклическое и линейное строение, чтобы располагаться в ограниченном пространстве внутреннего кармана, и содержать несколько доноров/акцепторов водородной связи для обеспечения взаимодействия с остатками кармана. Такой химический профиль стал основой для дальнейшего поиска возможных аллостерических ингибиторов белка TEM-1.

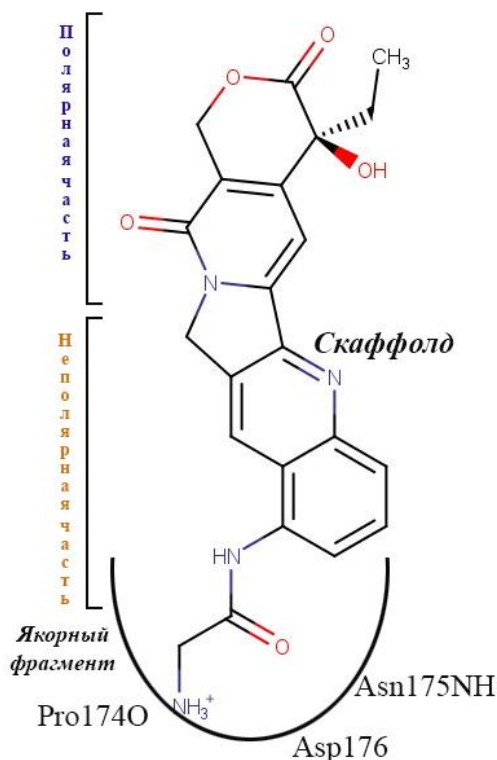


Рисунок 16. Возможное устройство аллостерических ингибиторов β -лактамазы на примере 639174.

5.3 Виртуальный скрининг базы ВГУ и библиотеки HitFinder

Для поиска дополнительных соединений, способных связываться с аллостерическим сайтом белка TEM-1, был выполнен скрининг базы Воронежского государственного университета (ВГУ), содержащей около 15000 различных молекул. Для проведения докинга эта база была протонирована в программе Marvin с генерацией нескольких зарядовых форм, в связи с чем, конечное число молекул, подлежащих молекулярному моделированию, составило 18262. Все молекулы подготовленной базы были конвертированы в трехмерное представление с помощью программы Corina.

Подготовленные молекулы базы ВГУ были задочены в модель бета-лактамазы TEM-1, подготовленную на предыдущем этапе разработки, с использованием программы SOL. По результатам докинга были отобраны 10 лучших соединений, показавших значение скоринг-функции программы SOL в диапазоне от -6.33 и 6.02 ккал/моль. Якорным фрагментом среди лучших соединений оказалась метоксигруппа, связанная с полностью насыщенным

ароматическим кольцом скаффолда (см. для примера рисунок 17). В положениях лигандов внутри аллостерического сайта, предсказанных с помощью докинга, эта группа способна формировать водородную связь с атомом Asn175NH.

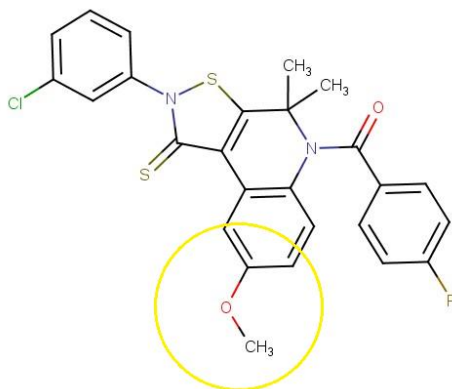


Рисунок 17. Структура соединения VGY 0006417, отобранного на экспериментальное тестирование в рамках скрининга базы ВГУ. Значение скоринг-функции: -6.09 ккал/моль. Желтым кругом выделена метоксигруппа, выполняющая роль якорного фрагмента после симуляции докинга.

Проверка активности отобранных веществ была выполнена на кафедре химической энзимологии химического факультета МГУ имени М.В. Ломоносова Виталием Георгиевичем Григоренко. Экспериментальное тестирование для всех 10 веществ не выявило соединений, имеющих какое-либо влияние на функциональность белка TEM-1 в растворе. Более того, многие вещества не смогли раствориться в условиях эксперимента. Для нескольких растворившихся веществ была проведена общая оценка их способности связываться с β -лактамазой на основе УФ-спектрометрии. Как оказалось, эти соединения, предсказанные как лиганды этого белка процедурой докинга, никак не влияют на спектры поглощения β -лактамазы, а значит, неспособны формировать с ней устойчивые комплексы. Полученные результаты говорят о том, что, во-первых, требуются дополнительные методы оценки влияния лигандов на положение Ω -петли (в идеале – методы молекулярной динамики), и, во-вторых, выбор якорных фрагментов должен фокусироваться на более сложных группах, имеющих повышенное сродство к

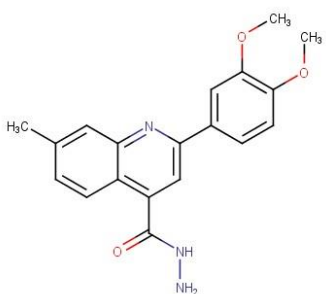
остаткам внутри полости выбранного сайта связывания. Кроме того, важным следствием экспериментального тестирования является понимание того, какие группы могут быть связаны с хорошим показателем скоринг-функции, однако не влиять на реальное связывание. Другими словами, можно выделить ряд групп, которые ассоциированы с ложноположительной ошибкой предсказания активности. Обобщая результаты, к таким группам относятся те, которые располагаются в центре лиганда и заряжены отрицательно, например, аминосульфоновая группа, заряжающаяся отрицательно при рН 7.4. Ложноположительный характер предсказания важности этих групп связан с тем, что в ходе докинга такие группы сильно взаимодействуют с положительно заряженным остатком Arg44 в центре аллостерического сайта. По-видимому, в реальности боковая группа этого остатка вовлечена в солевой мостик с карбоксильной группой Glu65, и на самом деле связывание с Arg44 отрицательно заряженных групп не так выгодно с точки зрения реального взаимодействия с сайтом белка, как это происходит при моделировании и выражается в хорошем score.

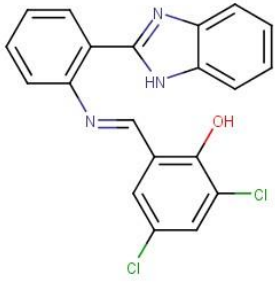
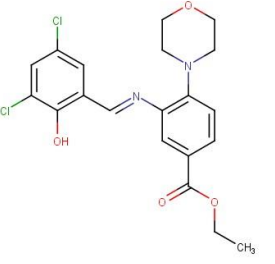
С учётом результатов *in silico* скрининга базы ВГУ и результатов проверки активности появилась задача исследовать якорные фрагменты другой природы. Для этой задачи и задачи идентификации новых потенциальных активных соединений предполагалось выполнить виртуальный скрининг химической библиотеки HitFinder. Данная библиотека отобрана по принципу химического разнообразия из базы Maybridge Screening Collection. Для проведения виртуального скрининга база HitFinder была скачана с сайта maybridge.com и подготовлена согласно принципам, описанным в пункте 2.2. Протонирование базы с генерацией дополнительных зарядовых форм было сделано в программе Marvin. Второй этап подготовки включал конвертацию двумерных молекул базы в 3D модели с помощью программы Corina. Конвертация предполагала формирование дополнительных конформеров макроциклов, так как в программе SOL, основном инструменте докинга в этом исследовании, нет учета гибкости данных химических структур. С учетом

представленного протокола подготовки, конечное количество молекул в базе HitFinder увеличилось до 35500.

Подготовленная база HitFinder была задочена в аллостерический сайт белка TEM-1 с использованием модели этого белка, подготовленной на предыдущем этапе разработке на основе структуры 1XPВ. Далее для молекул, имеющих лучшие показатели скоринг-функции, проводился визуальный анализ их положений в аллостерическом сайте. Из топ 40 ранжированных по убыванию скоринг-функции соединений было отобрано 10 лучших молекул для постпроцессинга на основе критериев: визуально приемлемая поза, выраженность якорной группы, отсутствие фрагментов, ассоциированных с ложноположительной активностью, которая была выявлена на стадии экспериментального тестирования (прежде всего, отсутствие аминосульфоновой группы и вторичной ароматической аминогруппы в центре лиганда). С учетом результатов квантовохимического постпроцессинга методом PM7, направленного на уточнение энергии связывания, было отобрано 3 лучших соединения, структуры и расчетные характеристики которых приведены в таблице 11.

Таблица 11. Три лучших соединения с потенциальной активностью против β -лактамазы, отобранных в ходе виртуального скрининга базы HitFinder.

HitFinder ID	Структура	Скоринг-функция SOL, ккал/моль	$\Delta H_{\text{связывания}}$ PM7, ккал/моль	Якорная группа
5942		-5.61	-62.05	-CONH ₂

9217		-6.33	-53.3472	-C ₆ H ₅ (частично)
10517		-6.04	-47.7693	-COOCH ₃

Самым перспективным с точки зрения скоринг-функции, энтальпии связывания и функциональности якорной группы стало соединение с HitFinder ID: 5942. Если сравнивать с веществами, отобранными из базы ВГУ на эксперимент, а также соединением 639174 из базы NCI Diversity set V, молекула 5942 обладает более отрицательной энтальпией связывания и более сложной функциональной группой. По сравнению с метоксигруппой, характерной для экспериментально измеренных соединений из базы ВГУ, гидразидная группа в составе соединения 5942 обладает возможностью быть и донором, и акцептором для множественных водородных связей внутри выбранного сайта связывания. Согласно положению, предсказанному в ходе докинга, эта группа формирует две водородные связи с атомами кислорода остова остатков Asp177 и Arg178 (см. рисунок 18).

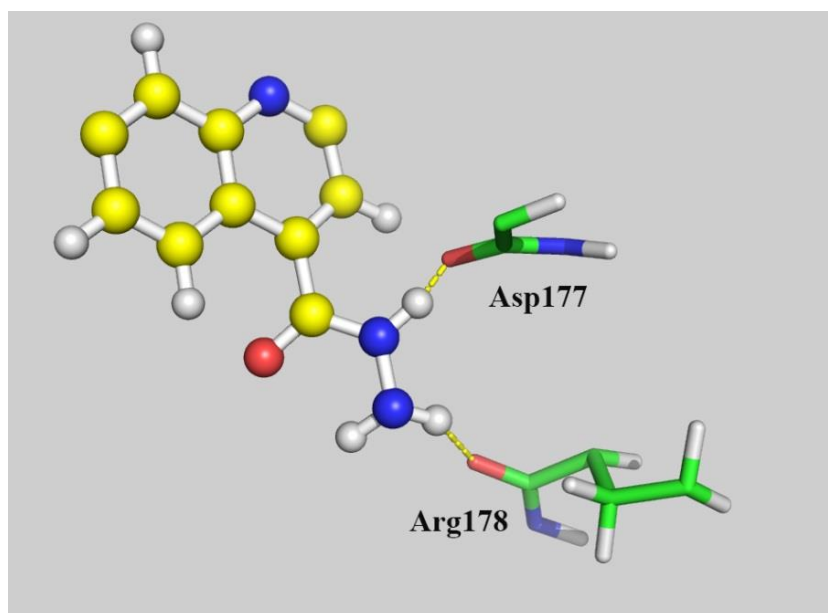


Рисунок 18. Гидразидный фрагмент в качестве якорной группы в составе молекулы соединения 5942 после докинга. Желтыми пунктирными линиями показаны водородные связи. Для краткости часть лиганда не показана.

Учитывая вышеописанную функциональность гидразидного фрагмента с точки зрения потенциального вклада в связывание с белком TEM-1, было решено провести поиск молекул, содержащих данную группу в библиотеке HitDiscover (53000 соединения – полная коллекция соединений Maybridge). При этом предполагался поиск соединений только с концевым расположением гидразидного фрагмента, чтобы он мог выполнять функции якорной группы. Поиск подструктуры производился с использованием программы OpenBabel[81]. В итоге, было обнаружено 239 соединений в базе HitDiscover, которые содержат концевую гидразидную группу. Найденные соединения подлежали докингу для оценки их способности связываться с аллостерическим сайтом в той же мере, в которой связывается соединение 5942 *in silico*.

Гидразид-содержащие соединения, отобранные из библиотеки HitDiscover, были подготовлены для проведения докинга согласно протоколу, который использовался для подготовки базы HitFinder с генерацией дополнительных конформеров макроциклов. Конечное число подготовленных молекул для докинга составило 284 лиганда. Все эти молекулы были задочены в модель белка TEM-1. По результатам докинга было отобрано 8 лучших

соединений, чьи позы после докинга были оптимальны с точки зрения размещения гидразидной группы внутри сайта связывания. Для этих 8 соединений был проведен постпроцессинг методом РМ7 для уточнения энергии связывания. Результаты докинга и постпроцессинга приведены в таблице 12.

Таблица 12. Результаты докинга и постпроцессинга для 8 гидразид-содержащих соединений, отобранных из базы HitDiscover.

№	Код из базы HitDiscover	Скоринг-функция в SOL, ккал/моль	$\Delta H_{\text{связывания}}$ РМ7, ккал/моль
1	HTS09992	-4.77	-47.31
2	HTS12609	-4.89	-41.24
3	HTS12663	-4.83	-48.18
4	HTS12664	-4.63	-44.75
5	HTS12665	-4.95	-50.09
7	KM04727	-4.55	-38.67
8	SEW01689	-4.73	-36.74
9	SPB02296	-4.77	-50.92

Согласно результатам моделирования два соединения, HTS12665 и SPB02296, теоретически способны связываться внутри аллостерического сайта. Их структуры приведены на рисунке 19. Таким образом, соединения 5942, HTS12665 и SPB02296 являются виртуальными молекулами-«хитами», которые потенциально способны ингибировать β -лактамазу TEM-1 аллостерически.

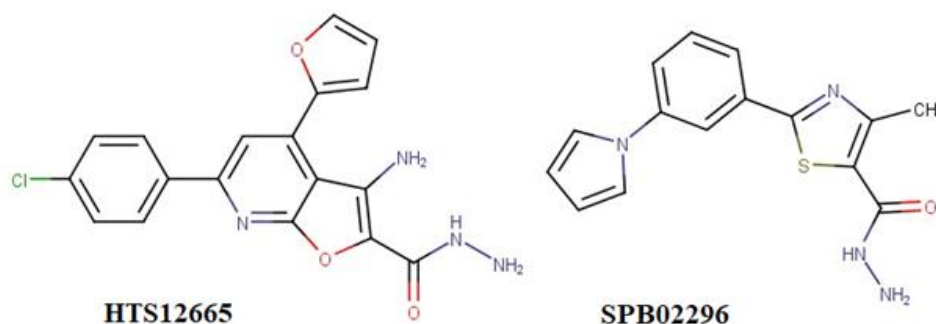


Рисунок 19. Лучшие потенциальные ингибиторы β -лактамазы TEM-1 из базы HitDiscover, имеющих гидразидный фрагмент в качестве якорной группы.

5.4 Применение молекулярной динамики для проверки аллостерического эффекта предсказанных лигандов

Для проверки потенциальной способности трёх отобранных гидразид-содержащих соединений оказывать аллостерический эффект на белок TEM-1

были выполнены расчеты с помощью молекулярной динамики Игорем Владимировичем Упоровым, сотрудником группы Егорова А.М. кафедры химической энзимологии химического факультета МГУ имени М.В. Ломоносова. Расчеты выполнялись с использованием программы AMBER16[82]. В качестве начальных координат были взяты положения лигандов после докинга. Заряды на соединениях рассчитывались программой RESP, которая использовала для их вычисления ограниченный метод Хартри-Фока в базисе 6-31*. Моделирование выполнялось в периодических граничных условиях при постоянной температуре и давлении (NPT-ансамбль). Длина траектории составляла 20 нс.

К концу симуляции ни одно из соединений не покинуло аллостерического сайта. Рассчитанные методом MM-GBSA свободные энергии связывания получились равными -17.2 ккал/моль, -25.8 ккал/моль, -18.2 ккал/моль для соединения 5942, HTS12665 и SPB02296, соответственно. Соединение SPB02296 оказалось способным смещать С-концевую часть Ω -петли белка TEM-1 в растворитель, однако не влияя на положение ключевых для катализа остатков Glu166/Asn170.

Таким образом, с помощью виртуального скрининга на основе докинга были найдены три потенциальных лиганда аллостерического сайта β -лактамазы TEM-1, показавших высокое сродство к этому сайта *in silico*, предсказанное с помощью докинга, полуэмпирических расчетов и симуляции молекулярной динамики. Одно соединение показало способность сдвигать положение Ω -петли, и его структура может быть использована в качестве «стартовой точки» для генерации аналогов – потенциальных аллостерических ингибиторов выбранной мишени. Все три соединения имеют приоритет в экспериментальном тестировании.

Глава 6. Поиск аллостерических антагонистов белка YB-1

6.1 Создание модели белка на основе ансамбля конформаций

В рамках поиска антагонистов белка YB-1 модель ингибирования строилась на том факте, что клеточные эффекты, вызываемые этим белком, напрямую зависят от его ядерной локализации. Процесс перехода YB-1 в ядро происходит после его фосфорилирования по остатку Ser102. Этот остаток располагается на гибкой петле домена холодового шока в составе небольшого сайта, который можно с точки зрения медицинской химии трактовать как сайт аллостерической регуляции активности YB-1. Связывание низкомолекулярного соединения внутри данного сайта должно нарушить переход белка в ядра и остановить клеточный рост и малигнизацию, индуцированную влиянием белка YB-1. Именно сайт фосфорилирования домена холодового шока был выбран в качестве мишени для поиска антагонистов YB-1.

Главным аспектом, который требуется учитывать при построении молекулярной модели белка YB-1 и применении докинга для такой модели, является наличие гибкой петли условного сайта связывания. Длина этой петли составляет 15 остатков (92-107), и явный учет её подвижности ассоциирован с высокими вычислительными затратами, что не совместимо с методологией виртуального скрининга. Поэтому предполагался неявный учет подвижности гибкой петли с помощью использования нескольких или, другими словами, ансамбля белковых конформаций с разным положением петли.

Построение модели было выполнено на основе структуры 1H95, скачанной с сайта Protein Data Bank. Структура соответствует домену холодового шока YB-1. Она была получена методом ЯМР спектроскопии и представляет собой набор 10 белковых конформаций самых низких по энергии. На рисунке 20 показано наложение этих структур друг на друга и видно, что петлевой участок обладает высокой степенью подвижности. Все 10 конформаций были протонированы с помощью программы APLITE, после чего программой SOLGRID был выполнен расчет 10 сеток потенциалов для дальнейшего докинга.



Рисунок 20. Конформации домена холодого шока белка YB-1, полученные методом ЯМР (PDB ID: 1H95). Желтая дуга обводит условный сайт связывания, содержащий гибкую петлю.

Построенный ансамбль из 10 атомных моделей домена холодого шока был валидирован на основе докинга в программе SOL известных антагонистов процесса перехода YB-1, нарушающих фосфорилирование этого белка *in vitro*. Так как для данных антагонистов не существует экспериментально известной геометрии комплекса с белком YB-1, валидация строилась только на энергетическом критерии. Другими словами, предполагалось выбрать в качестве рабочих конформаций те из построенных моделей, для которых хотя бы один из известных лигандов показывает хорошее значение скоринг-функции после докинга. Такие рабочие конформации приблизительно отражают реальное поведение YB-1, и при их использовании в ходе виртуального скрининга в качестве активных кандидатов будут отбираться те соединения, которые потенциально связываются с той же конформацией сайта связываний, с которой связываются известные антагонисты. Так как в ходе валидации все три известных лиганда показали плохие значения скоринг-функции программы докинга SOL, было решено выполнить оптимизацию геометрии всех подготовленных моделей домена холодого шока YB-1 и повторить докинг данных соединений. Оптимизация строилась на полуэмпирическом методе PM7 и включала в себя следующие шаги. В программе MORAC были добавлены

атомы водорода к каждой конформации домена холодого шока из структуры 1H95 с нейтрализацией всех ионизируемых групп (боковые группы Lys, Arg, Asp, Glu, а также концевые аминогруппа и карбоксильная группа белка). Такая нейтрализация позволяет избежать искажений при начальной оптимизации полуэмпирическим методом. Второй шаг подразумевал выполнение оптимизации положений добавленных атомов водорода с потенциальным образованием солевых мостиков, сопровождающимся переносом протона от кислотной группы к основной. Конформации с оптимизированными атомами водорода также вошли в тестовую выборку моделей домена холодого шока. Наконец, в ходе третьего шага выполнялась ограниченная оптимизация всех атомов. Для финальной структуры выполнялось протонирование с ионизацией групп, заряжающихся при рН 7.4. Методологические детали такой подготовки белковых молекул описаны в работе[83]. Таким образом, были сформированы 20 дополнительных моделей: 10 моделей с оптимизированными положениями атомов водорода и 10 моделей с полной ограниченной оптимизацией.

Для оптимизированных конформаций домена холодого шока был выполнен повторный докинг трех известных антагонистов переноса YB-1 в ядро клетки. С учетом результатов докинга полностью оптимизированных и оптимизированных по положению атомов водорода моделей были отобраны 4 рабочие структуры, которые в дальнейшем предполагалось использовать для скрининга. Результаты докинга и описание выбранных моделей приведено в таблице 13.

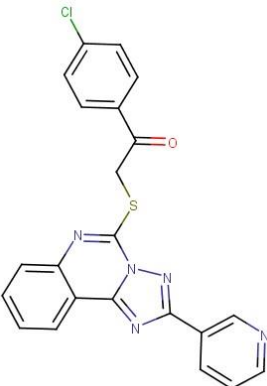
Таблица 13. Рабочие модели домена холодого шока белка YB-1.

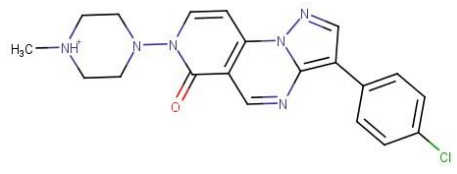
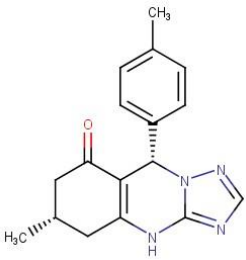
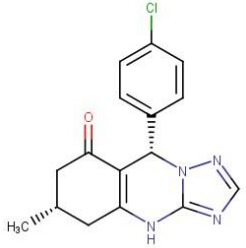
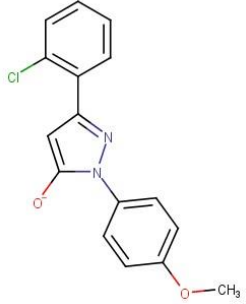
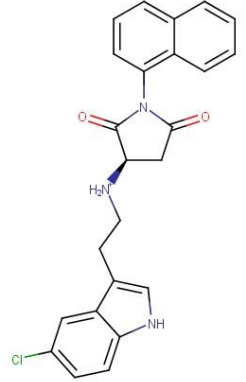
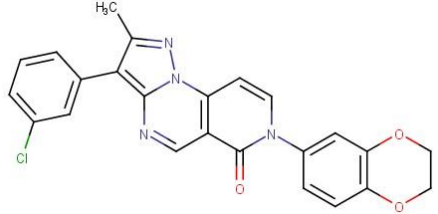
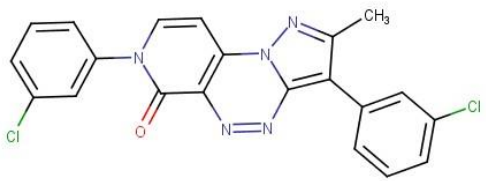
Номер конформера 1H95	Оптимизация	Скор SOL для ИТК, ккал/моль	Скор SOL для физетина, ккал/моль	Скор SOL для индирубина, ккал/моль
1	Только Н	-4.16	-4.08	-3.34
2	Все атомы	-2.27	-5.21	-4.26
5	Все атомы	-3.3	-4.7	-3.78
7	Все атомы	-3.25	-3.46	-4.48

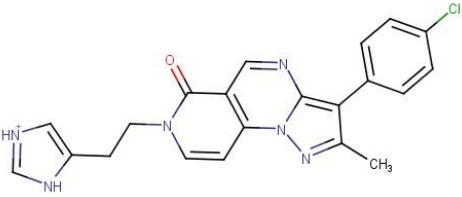
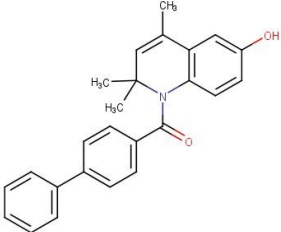
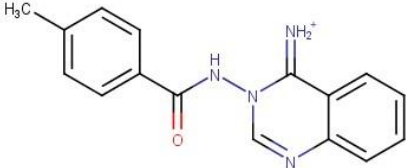
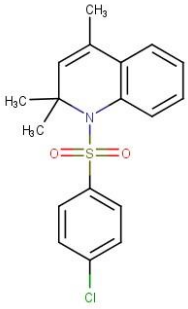
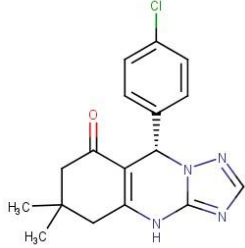
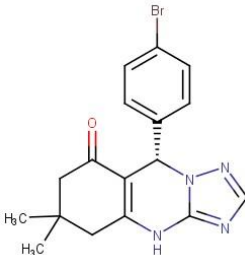
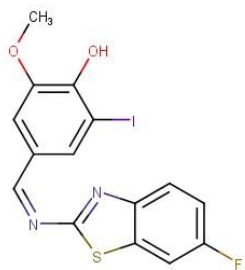
6.2 Виртуальный скрининг полной базы ВГУ

Ранее подготовленная база ВГУ, содержащая 18262 молекул, была использована для виртуального скрининга. Докинг всех молекул этой базы был выполнен с помощью программы SOL для всех 4 моделей холодового шока YB-1. С учетом полученных значений скоринг-функции, а также предсказанных положений внутри сайта связывания были отобраны 14 лучших кандидатов. Для них был выполнен квантовохимический постпроцессинг на основе полуэмпирического метода PM7 согласно ранее описанной методике (см. раздел 2.5) для выявления потенциальных ложноположительных соединений. Кроме того, расчет энтальпии связывания был проведен для известного антагониста YB-1 – физетина. Её значение, используемое в качестве критерия отсеивания потенциально неактивных кандидатов, составило -37 ккал/моль. Все 14 соединений показали предсказанные энтальпии связывания ниже, чем -40 ккал/моль. Дополнительно был выполнен скрининг для базы из 295 химических соединений, предоставленной Athina Geronikaki (Emeritus professor, Aristotle University of Thessaloniki). Одно соединение показало хорошее связывание *in silico*. Все 15 кандидатов (14 соединений из базы ВГУ и одно вещество из базы Athina Geronikaki) были переданы на экспериментальное тестирование. Результаты моделирования для этих соединений, а также их структуры приведены в таблице 14.

Таблица 14. Кандидаты в антагонисты процесса перехода белка YB-1 в ядро.

№	Название – номер в базе ВГУ	Структура	Скоринг-функция SOL, ккал/моль	$\Delta H_{\text{связывания}}$, PM7, ккал/ моль
1	VGY 0121548		-5.24	-41.48

2	VGY 0201463		-5.23	-45.72
3	VGY 0033750		-5.21	-42.3
4	VGY 0049465		-5.28	-46.6
5	VGY 0120533		-5.34	-40.56
6	VGY 0039544		-5.1	-52.27
7	VGY 0134304		-4.9	-42.69
8	VGY 0135674		-4.86	-42.07

9	VGY 0201630		-4.94	-45.92
10	VGY 0223265		-4.87	-42.88
11	VGY 0224984		-5.51	-41.92
12	VGY 0002274		-5.04	-44.68
13	VGY 0018645		-5.05	-49.43
14	VGY 0018615		-5.13	-44.93
15	AthinaYB1		-5.2	-43.57

6.3 Экспериментальное тестирование отобранных кандидатов

Проверка активности предсказанных антагонистов YB-1 на клеточной линии карциномы кишечника (HCT-166) было выполнено Моисеевой Натальей Ивановной, сотрудницей лаборатории генетики опухолевых клеток НМИЦ имени Н.Н. Блохина. В ходе МТТ-теста было показано, что четыре соединения: два на основе триазолохиназолинона (VGY 0018645 и VGY 0018615), одно соединение на основе дигидрохинолина (VGY 0002274) и одно вещество на основе 2-аминобензотиазола (AthinaYB1), имеют цитотоксический эффект в отношении опухолевых клеток. Выявленное значение IC_{50} составило, соответственно, 44, 5, 16 и 51 мкМ. Для более активного соединения, VGY 0018615, была проверена цитотоксичность в отношении нормальных клеток. Оказалось, что даже в высокой концентрации 50 мкМ это соединение не проявляет токсический эффект на псевдонормальные клетки фибробластов эмбриона человека. Таким образом, вещество VGY 0018615 обладает избирательной токсичностью против опухолевых клеток.

Дополнительные эксперименты, направленные на проверку механизма действия для производных триазолохиназолинона, были проведены Мордовкиной Дарьей Алексеевной и Григорьевой Екатериной Михайловной в группе Овчинникова Л.П. лаборатории регуляции биосинтеза белка Института белка РАН. На клеточной линии HeLa было выполнена проверка антагонистического влияния этих веществ на процесс перехода YB-1 в ядро. Внутриклеточную визуализацию YB-1 выполняли методом флуоресцентной микроскопии. Также физетин и кверцетин использовались в качестве контрольных веществ, для которых в литературе имеются данные по связыванию с доменом холодового шока YB-1, и их предположительный механизм действия – ингибирование фосфорилирования Ser102.

Первоначально было подтверждено, что изучаемые вещества не являются индукторами переноса белка YB-1 в ядро. Это же наблюдалось для физетина и кверцетина. После этого, к клеточным культурам был добавлен известный индуктор перехода белка YB-1 в ядро – актиномицин Д, чтобы проверить

способность этих веществ ингибировать переход белка в ядро. Распределение YB-1 в клетках в зависимости от обработки показано на рисунке 21.

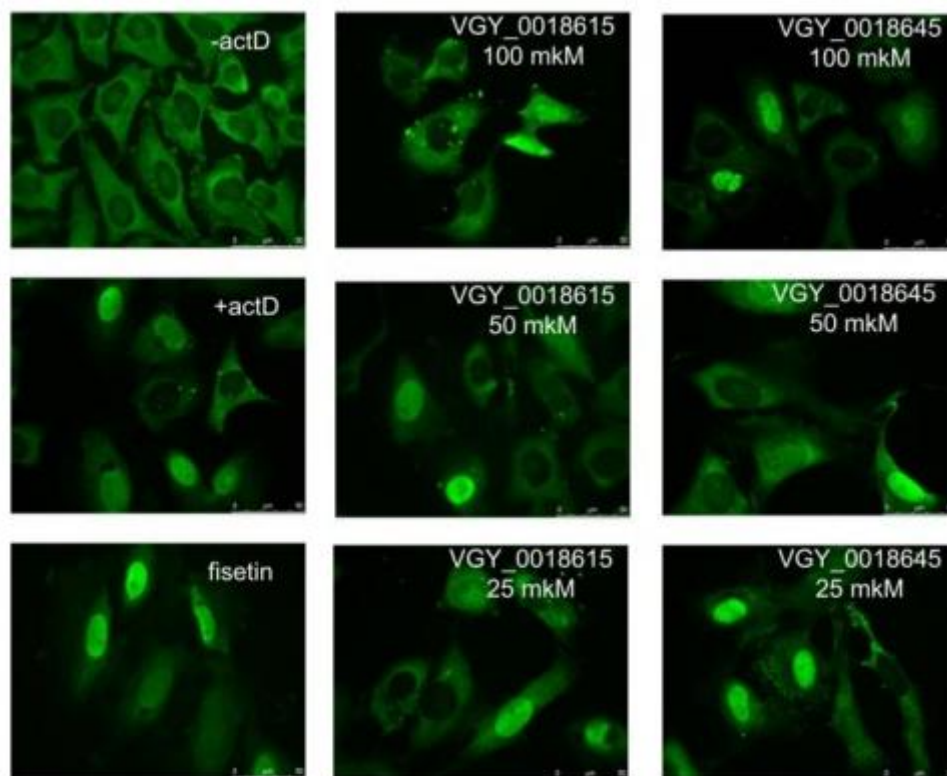


Рисунок 21. Клетки линии HeLa, инкубация в среде с актиномицином Д, изучаемыми соединениями и физетином. Окрашивание против эндогенного YB-1 (флуоресцентный краситель Alexa 488).

Из рисунка 21 видно, что при инкубации клеток в среде, как с изучаемыми веществами, так и с контрольными соединениями белок YB-1 детектируется как в ядре, так и в цитоплазме. Для того чтобы сделать какой-то вывод о способности веществ ингибировать транслокацию YB-1 в ядро, был выполнен подсчет количества клеток с разной локализацией белка YB-1. В зависимости от локализации YB-1 обработанные клетки распределились по трем группам: цитоплазматическая локализация, ядерная локализация и смешанная – и в ядре, и в цитоплазме. Результаты подсчетов показаны на рисунке 22. Из рисунка можно сделать вывод о том, что вещества VGY 0018615 и VGY_0018645 препятствуют переходу белка YB-1 в ядро на уровне физетина и кверцетина, а в некоторых случаях даже лучше.

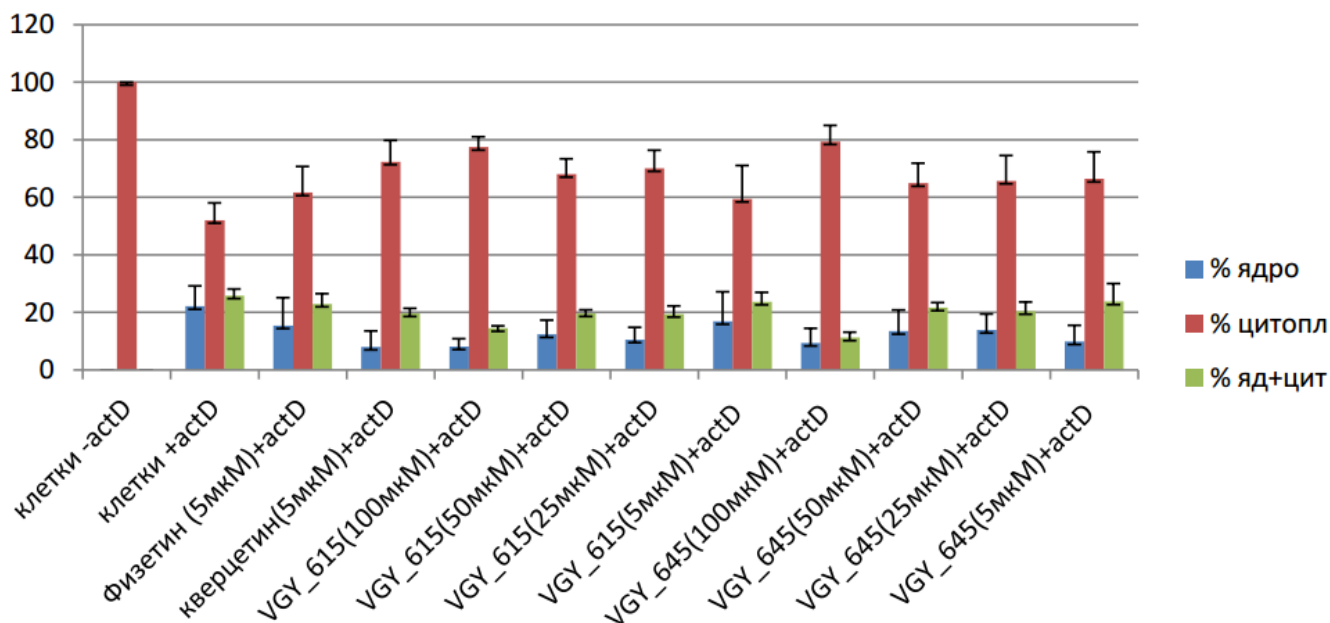


Рисунок 22. Распределение белка YB-1 в клеточных компартаментах.

Процентное содержание в ядре, цитоплазме и смешанная локализация. Планки погрешности показывают стандартную ошибку, рассчитанную по результатам 3-х повторов.

Помимо оценки распределения YB-1 внутри клеток в зависимости от присутствия изучаемых веществ, было оценено их влияние на уровень фосфорилирования YB-1 методом Вестерн-блот. Результаты этого эксперимента показаны на рисунке 23. Видно, что в контрольной точке без добавления актиномицина Д YB-1 практически не фосфорилирован. При добавлении актиномицина Д YB-1 происходит фосфорилирование YB-1 по Ser102. Контрольные вещества физетин и кверцетин препятствуют этому процессу. Вещества VGY 0018615 и VGY 0018645 не только не препятствуют фосфорилированию, а скорее усиливают его. Получается, что обнаруженные антагонисты процесса ядерного перехода YB-1 не влияют на степень фосфорилирования этого белка и, по-видимому, действуют на другие клеточные мишени, вовлеченные в каскад переноса YB-1 в ядро. Дополнительные исследования помогут выявить, как именно такое обходное ингибирование ядерного перехода реализуется для данных веществ.

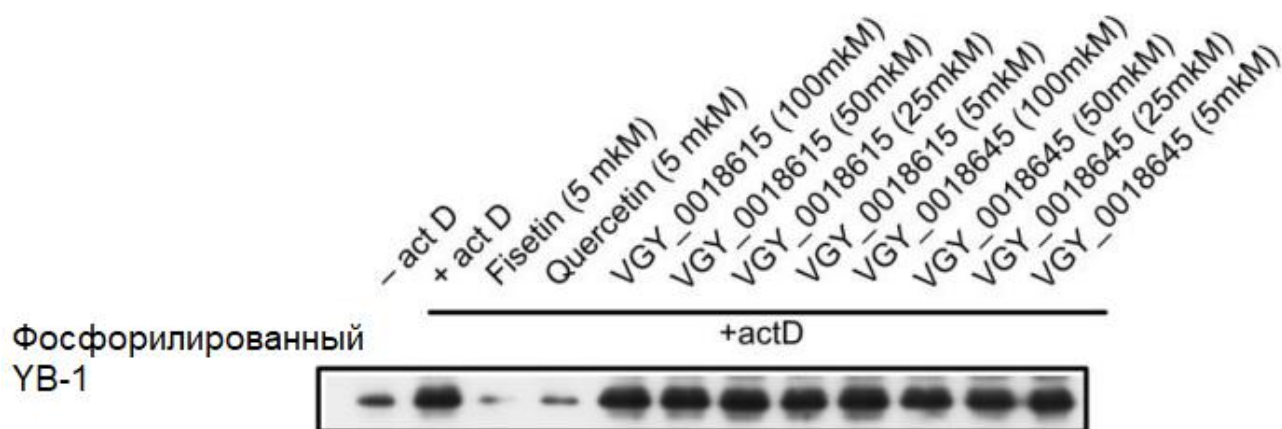


Рисунок 23. Результаты эксперимента Вестрн-блот. Более темное пятно указывает на более высокое содержание фосфорилированного YB-1.

Таким образом, в ходе настоящей работе по поиску антагонистов процесса перехода YB-1 в ядро на основе виртуального скрининга были достигнуты следующие результаты. С помощью докинга на основе ансамбля белковых конформаций и расчетов на основе полуэмпирического метода PM7 было предсказано сродство для 15 соединений к сайту фосфорилирования домена холодового шока YB-1. Четыре из них проявили цитотоксическое действие в отношении опухолевых клеток на уровне десятков микромоляр. Лучшее вещество не показало токсичности против нормальных клеток. Эксперименты, направленные на проверку молекулярного механизма действия двух производных триазолохиназолинона, с одной стороны подтвердили, что они способны препятствовать искусственно индуцированному переходу белка YB-1 в ядро клеток. С другой стороны проверка уровня фосфорилирования не выявило у соединений способности препятствовать этому процессу. По-видимому, соединения влияют на более поздние этапы сигнального пути, в ходе которого YB-1 транспортируется в ядро.

Заключение

В данной работе виртуальный скрининг на основе докинга и квантовохимического постпроцессинга был применен для поиска соединений активных в отношении трех фармакологически значимых белков-мишеней. Продемонстрирована эффективность разработанного метода прогнозирования активности для выявления начальных кандидатов в лекарственные средства. Для одной из мишеней, β -лактамазы TEM-1, предположен механизм аллостерического ингибирования, который показал потенциальную работоспособность в молекулярных симуляциях. В рамках исследования все подтвержденные ингибиторы принадлежат к новым химическим классам, которые ранее не были выявлены в ходе разработок по данным мишеням. Все три группы найденных ингибиторов представляют практический интерес для дальнейшей разработки лекарственных молекул, действующих на представленные белки-мишени.

Результаты и выводы

1. С помощью докинга и последующего постпроцессинга с использованием квантовохимического метода отобран новый класс ингибиторов фактора свертываемости крови Ха на основе 2,2,4-триметил-1,2,3,4-тетрагидрохинолина.
2. Применение постпроцессинга на основе полуэмпирического метода PM7 и модели растворителя COSMO более чем на 40% снижает число ложноположительных предсказаний при прогнозировании активности в отношении фактора Ха.
3. Разработана теоретическая модель аллостерического ингибирования β -лактамазы TEM-1. Модель подразумевает воздействие на консервативную часть фермента – Ω -петлю, которая играет ключевую роль для каталитической активности данного фермента.

4. На основе виртуального скрининга, сочетающего докинг и последующий квантовохимический постпроцессинг, идентифицированы три гидразид-содержащих кандидата в аллостерические ингибиторы β -лактамазы TEM-1.
5. На основе докинга с использованием ансамбля белковых конформаций отобраны молекулы, связывающиеся с доменом холодового шока белка YB-1 *in silico*.
6. Выполнено лабораторное подтверждение активности. Для 12 из 40 предсказанных ингибиторов фактора Ха наблюдалось ингибирование данного фермента в концентрации 5 и 30 мкМ. Значение IC_{50} для лучшего ингибитора фактора Ха составило 3 мкМ. Это соединение оказалось селективным ингибитором фактора Ха и не влияло на активность трипсина, тромбина, фактора IXa и фактора XIa. Четыре соединения из 15 предсказанных антагонистов процесса перехода YB-1 в ядро клетки проявили противоопухолевое действие на клеточных культурах. Значение IC_{50} для лучшего из выявленных веществ составило 5 мкМ. Дополнительные эксперименты для двух соединений подтвердили их способность затруднять переход белка YB-1 в ядро.

В рамках диссертационного исследования сделано 15 публикаций в научных печатных изданиях, включая 10 публикаций в журналах, которые индексируются в системе Web of Science, Scopus, RSCI, а также 5 публикации тезисов докладов на конференциях. Также было принято участие в 5 конференциях.

**ПУБЛИКАЦИИ АВТОРА ПО ТЕМЕ ДИССЕРТАЦИИ в журналах,
индексируемых в базах данных Web of Science, Scopus RSCI:**

1. **Ilin Ivan**, Lipets Elena, Sulimov Alexey, Kutov Danil et al. New factor Xa inhibitors based on 1,2,3,4-tetrahydroquinoline developed by molecular modelling // Journal of Molecular Graphics and Modelling. 2019. Vol. 89. N. June 2019. pp. 215-224.

Импакт-фактор по WoS 1,863

2. Sulimov Alexey, Kutov Danil, Katkova Ekaterina, **Ilin Ivan**, Sulimov Vladimir. New generation of docking programs: Supercomputer validation of force fields and quantum-chemical methods for docking // *Journal of Molecular Graphics and Modelling*. 2017. Vol.78 N. November 2017. pp. 139-147.

Импакт-фактор по WoS 1,863

3. Novichikhina N., **Ilin I.**, Tashchilova A., Sulimov A. Synthesis, Docking, and In Vitro Anticoagulant Activity Assay of Hybrid Derivatives of Pyrrolo[3,2,1-ij]Quinolin-2(1H)-one as New Inhibitors of Factor Xa and Factor XIa // *Molecules*. 2020. Vol. 25. N. 8. pp. 1889-1889.

Импакт-фактор по WoS 3,06

4. Sulimov A.V., Kutov D., **Ilin I.**, Zheltkov D. et al. Supercomputer docking with a large number of degrees of freedom // *SAR and QSAR in Environmental Research*. 2019. Vol. 30. N. 10. pp. 733-749.

Импакт-фактор по WoS 2,287

5. Sulimov A.V., Kutov D.K., **Ilin I.S.**, Sulimov V.B. Docking with combined use of a force field and a quantum-chemical method // *Biomeditsinskaya Khimiya*. 2019. Vol. 65. N. 2, pp. 80-85.

Импакт-фактор по Scopus (SJR) 0,24

6. Sulimov V.B., **Ilin I.S.**, Kutov D.C., Sulimov A.V. Development of docking programs for Lomonosov supercomputer // *Journal of the Turkish Chemical Society Section A: Chemistry*. 2020. Vol. 7, N. 1, pp. 259-276.

Импакт-фактор по Scopus (SJR) 0,22

7. Alexey Sulimov, Danil Kutov, Anna Gribkova, **Ivan Ilin** et al. Search for Approaches to Supercomputer Quantum-Chemical Docking // *Russian Supercomputing Days*. 2019. Vol. 1129. pp. 363-378.

Импакт-фактор по Scopus (SJR) 0,19

8. Сулимов А.В., Кутов Д.К., Тащилова А.С., **Ильин И.С.** и др. Современные методы разработки новых лекарственных средств,

влияющих на систему гемостаза, журнал - Вопросы гематологии/онкологии и иммунопатологии в педиатрии, 2019 г.

Импакт-фактор (РИНЦ) = 0,26

9. Подоплелова Н.А., Сулимов В.Б., Тащилова А.С., **Ильин И.С.** и др. Свертывание крови в XXI-м веке: новые знания, методы и перспективы для терапии, журнал - Вопросы гематологии/онкологии и иммунопатологии в педиатрии, 2019 г.

Импакт-фактор (РИНЦ) = 0,26

10. Сулимов А.В., Кутов Д.К., **Ильин И.С.**, Шихалиев Х.С. и др. Докинг олигопептидов, журнал - *Известия Академии наук. Серия химическая*, 2019 г.

Импакт-фактор по WoS 0,423

**ТЕЗИСЫ ДОКЛАДОВ НА МЕЖДУНАРОДНЫХ И
ВСЕРОССИЙСКИХ НАУЧНЫХ КОНФЕРЕНЦИЯХ,**

опубликованные в журналах:

1. Ilin I., Moiseeva N., Shikhaliev K., Tarasov A. et al. «Search for the new class of anticancer drugs on the basis of yb-1 cold shock domain inhibitors» в сборнике *XXI Mendeleev Congress on General and Applied Chemistry. Book of abstracts in 6 volumes, серия Book 6. Satellite symposia*, том 6, тезисы, с. 68-68. 2019 г.
2. Ильин И.С., Липец Е.Н., Сулимов А.В., Кутов Д.К. и др. «Разработка новых антикоагулянтов с использованием методов молекулярного моделирования ингибиторов белков системы свертывания крови» в сборнике *Сборник материалов XXV Российского национального конгресса «Человек и лекарство». Тезисы докладов/ Главный редактор Чучалин А.Г, серия Секция "Разработка и создание новых лекарственных препаратов"*, с. 69-70. 2018 г.
3. Vladimir Sulimov, Alexey Sulimov, Danil Kutov, Ivan Ilin. «Development of docking programs for Iomonosov supercomputer» в сборнике *Asian*

Federation of Medicinal Chemistry 12th International Medicinal Chemistry Symposium 2019

4. Sulimov A.V., Kutov D.K., Ilin I.S., Zheltkov D. «Supercomputer docking with a large number of degrees of freedom» в сборнике *10th International Symposium on Computational Methods in Toxicology and Pharmacology Integrating Internet Resources СМТПИ-2019*, тезисы, с. 24-24, 2019 г.
5. Сулимов А.В., Кутов Д.К., Ильин И.С., Сулимов В.Б. «Суперкомпьютерный докинг: численные эксперименты» в сборнике *Сборник материалов XXVI Российский национальный конгресс "Человек и Лекарство"*. Тезисы докладов, с. 88-88, 2019 г.

**СПИСОК КОНФЕРЕНЦИЙ, НА КОТОРЫХ БЫЛИ ДОЛОЖЕНЫ
РЕЗУЛЬТАТЫ ДАННЫХ ИССЛЕДОВАНИЙ:**

1. XXI Менделеевский съезд по общей и прикладной химии, г. Санкт-Петербург, Россия, 9-13 сентября 2019. Доклад «Search for the new class of anticancer drugs on the basis of yb-1 cold shock domain inhibitors»
2. Ломоносовские чтения – 2019. Секция «НИВЦ», Москва, Россия, 15-25 апреля 2019. Доклад «Разработка ингибиторов белка YB-1 – нового класса противоопухолевых лекарств»
3. Ломоносовские чтения – 2018. Секция «Вычислительная математика и кибернетика», Москва, Россия, 16-27 апреля 2018. Доклад «Новые синтетические антикоагулянты: молекулярное моделирование и экспериментальная проверка»
4. XXV Российский национальный конгресс "Человек и лекарство", Москва, Россия, 9-12 апреля 2018. Доклад «Разработка новых антикоагулянтов с использованием методов молекулярного моделирования ингибиторов белков системы свертывания крови»
5. XXVII Российский национальный конгресс "Человек и лекарство", Москва, Россия, 6-9 апреля 2020 (онлайн). Доклад «Компьютерное моделирование аллостерических ингибиторов бета-лактамазы TEM-1»

Благодарности

Автор выражает искреннюю признательность и благодарность своему научному руководителю доктору физико-математических наук Сулимову Владимиру Борисовичу за помощь на всех этапах выполнения диссертации. Автор благодарит научную группу лаборатории Вычислительных систем и прикладных технологий программирования НИВЦ МГУ имени М.В. Ломоносова, сотрудников Алексея Владимировича Сулимова, Данила Константиновича Кутова, Ольгу Анатольевну Кондакову и Федора Васильевича Григорьева за ценные замечания и помощь в работе. Автор также благодарит Елену Николаевну Липец лаборатории биофизики НМИЦ ДГОИ имени Дмитрия Рогачева за проверку предсказанных ингибиторов фактора Ха и Надежду Александровну Подоплелову лаборатории клеточного гемостаза НМИЦ ДГОИ имени Дмитрия Рогачева, за проверку селективности лучших выявленных ингибиторов, а также Виталия Георгиевича Григоренко из группы академика Егорова А.М. кафедры химической энзимологии химического факультета МГУ за *in vitro* тестирование отобранных из базы ВГУ потенциальных ингибиторов β -лактамазы TEM-1 и Игоря Владимировича Упорова из этого же подразделения за проведение симуляции молекулярной динамики для проверки аллостерического эффекта предсказанных ингибиторов. Автор выражает благодарность Наталье Ивановне Моисеевой, лаборатория генетики опухолевых клеток НИИ канцерогенеза ФГБУ "НМИЦ онкологии им. Н.Н. Блохина", за проверку противоопухолевой активности предсказанных антагонистов перехода белка YB-1 в ядро, академику Льву Павловичу Овчинникову, группа регуляции биосинтеза белка, Института белка РАН, г. Пущино, и профессору Гелене Петровне Генс, зав. кафедрой Московского Государственного Медико-Стоматологического Университета им. А.И. Евдокимова, за постановку проблемы поиска антагонистов перехода белка YB-1 в ядро, а также сотрудницам Института Белка РАН г. Пущино Мордовкиной Дарье Алексеевне и Григорьевой Екатерине Михайловне за

эксперименты, направленные на проверку механизма действия выявленных антагонистов белка УВ-1. Автор выражает признательность сотрудникам и преподавателям кафедры фармацевтической химии, фармакогнозии и организации фармацевтического дела факультета фундаментальной медицины МГУ за поддержку данной научно-исследовательской работы.

СПИСОК ЛИТЕРАТУРЫ

1. Noe M.C., Peakman M.-C. Drug Discovery Technologies: Current and Future Trends // Reference Module in Chemistry, Molecular Sciences and Chemical Engineering, 2016.
2. Drug Design: Structure and Ligand-Based Approaches / ed. Merz K.M., Ringe D., Reynolds C.H., Cambridge: Cambridge University Press, 2010.
3. Croston G.E. The utility of target-based discovery. // Expert opinion on drug discovery, 2017, Vol. 12, No. 5, P. 427–429.
4. Ettmayer P., Schnitzer R., Bergner A., Nar H. Hit and Lead Generation Strategies // Reference Module in Chemistry, Molecular Sciences and Chemical Engineering, 2016.
5. Caraus I., Alsuwailem A.A., Nadon R., Makarenkov V. Detecting and overcoming systematic bias in high-throughput screening technologies: a comprehensive review of practical issues and methodological solutions // Brief. Bioinform., 2015, Vol. 16, No. 6, P. 974–986.
6. Hughes J.P., Rees S., Kalindjian S.B., Philpott K.L. Principles of early drug discovery // Br. J. Pharmacol., 2011, Vol. 162, No. 6, P. 1239–1249.
7. Klebe G. Virtual ligand screening: strategies, perspectives and limitations // Drug Discov. Today, 2006, Vol. 11, No. 13–14, P. 580–594.
8. Bolten B.M., DeGregorio T. Trends in development cycles // Nat. Rev. Drug Disc., 2002, Vol. 1.
9. Fishman M.C., Porter J.A. Pharmaceuticals: a new grammar for drug discovery. // Nature, 2005, Vol. 437, No. 7058, P. 491–493.
10. Matter H., Sottriffer C. Applications and Success Stories in Virtual Screening // Methods and Principles in Medicinal Chemistry, Virtual Screening, 2011, P. 319–358.
11. Clark D.E. What has virtual screening ever done for drug discovery? // Expert Opin. Drug Discov., 2008, Vol. 3, No. 8, P. 841–851.
12. Brenk R., Irwin J.J., Shoichet B.K. Here be dragons: docking and screening in

- an uncharted region of chemical space // *J. Biomol. Screen.* 2005/09/16, 2005, Vol. 10, No. 7, P. 667–674.
13. Doman T.N., McGovern S.L., Witherbee B.J., Kasten T.P., Kurumbail R., Stallings W.C., Connolly D.T., Shoichet B.K. Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B. // *J. Med. Chem.*, 2002, Vol. 45, No. 11, P. 2213–2221.
 14. Padmanabhan K., Padmanabhan K.P., Tulinsky A., Park C.H., Bode W., Huber R., Blankenship D.T., Cardin A.D., Kisiel W. Structure of human des(1-45) factor Xa at 2.2 Å resolution. // *J. Mol. Biol.*, 1993, Vol. 232, No. 3, P. 947–966.
 15. H. Y.C., C. F.J., I. W.J. Oral Direct Factor Xa Inhibitors // *Circ. Res.*, 2012, Vol. 111, No. 8, P. 1069–1078.
 16. Hsu H.-J., Tsai K.-C., Sun Y.-K., Chang H.-J., Huang Y.-J., Yu H.-M., Lin C.-H., Mao S.-S., Yang A.-S. Factor Xa active site substrate specificity with substrate phage display and computational molecular modeling // *J. Biol. Chem.* 2008/02/22, 2008, Vol. 283, No. 18, P. 12343–12353.
 17. Zhabin S.N., Sulimov V.B. Visualization and editing of molecules: MolRed application // *Sci. Vis.*, 2009, Vol. 1, P. 108–114.
 18. Жабин С.Н., Сулимов В.Б. Реализация интерактивности в молекулярном редакторе MOLRED // *Научная визуализация*, 2010, Vol. 2, No. 1, P. 59–81.
 19. Prasa D., Svendsen L., Stürzebecher J. Inhibition of thrombin generation in plasma by inhibitors of factor Xa. // *Thromb. Haemost.*, 1997, Vol. 78, No. 4, P. 1215–1220.
 20. Nicolini F.A., Lee P., Malucky J.L., Lefkovits J., Kottke-Marchant K., Plow E.F., Topol E.J. Selective inhibition of factor Xa during thrombolytic therapy markedly improves coronary artery patency in a canine model of coronary thrombosis. // *Blood Coagul. fibrinolysis an Int. J. Haemost. Thromb.*, 1996, Vol. 7, No. 1, P. 39–48.
 21. Patel N.R., Patel D. V., Murumkar P.R., Yadav M.R. Contemporary developments in the discovery of selective factor Xa inhibitors: A review // *Eur*

- J Med Chem. 2016/06/21, 2016, Vol. 121, P. 671–698.
22. Ilin I.S., Lipets E.N., Sulimov A. V., Kutov D.C., Shikhaliev K.S., Potapov A.Y., Krysin M.Y., Zubkov F.I., Saprionova L. V., Ataulakhanov F.I., Sulimov V.B. New factor Xa inhibitors based on 1,2,3,4-tetrahydroquinoline developed by molecular modelling // *J. Mol. Graph. Model.*, 2019, Vol. 89, P. 215–224.
 23. Egorov A., Rubtsova M., Grigorenko V., Uporov I., Veselovsky A. The Role of the Ω -Loop in Regulation of the Catalytic Activity of TEM-Type β -Lactamases // *Biomolecules*, 2019, Vol. 9, No. 12, P. 854.
 24. Bonomo R.A. beta-Lactamases: A Focus on Current Challenges // *Cold Spring Harb Perspect Med.* 2016/10/16, 2017, Vol. 7, No. 1.
 25. Pimenta A.C., Fernandes R., Moreira I.S. Evolution of drug resistance: insight on TEM beta-lactamases structure and activity and beta-lactam antibiotics. // *Mini Rev. Med. Chem.*, 2014, Vol. 14, No. 2, P. 111–122.
 26. Pan X., He Y., Lei J., Huang X., Zhao Y. Crystallographic Snapshots of Class A beta-Lactamase Catalysis Reveal Structural Changes That Facilitate beta-Lactam Hydrolysis. // *J. Biol. Chem.*, 2017, Vol. 292, No. 10, P. 4022–4033.
 27. Roccatano D., Sbardella G., Aschi M., Amicosante G., Bossa C., Nola A., Mazza F. Dynamical Aspects of TEM-1 β -Lactamase Probed by Molecular Dynamics // *J. Comput. Aided. Mol. Des.*, 2005, Vol. 19, P. 329–340.
 28. Fisette O., Gagné S., Lagüe P. Molecular dynamics of class A β -lactamases-effects of substrate binding // *Biophys. J.* 2012/10/16, 2012, Vol. 103, No. 8, P. 1790–1801.
 29. Gunasekaran V., Nishi K., Ravikumar V., Mathan G. Nuclear shuttling of Y Box binding protein-1, its clinical relevance in cancer and as a therapeutic target // *Bangladesh J. Pharmacol.*, 2016, Vol. 11, 2 SE-Mini Review.
 30. Bergmann S., Royer-Pokora B., Fietze E., Jürchott K., Hildebrandt B., Trost D., Leenders F., Claude J.-C., Theuring F., Bargou R., Dietel M., Royer H.-D. YB-1 Provokes Breast Cancer through the Induction of Chromosomal Instability That Emerges from Mitotic Failure and Centrosome Amplification // *Cancer Res.*, 2005, Vol. 65, No. 10, P. 4078 LP – 4087.

31. Gunasekaran V.P., Nishi K., Sivakumar D., Sivaraman T., Mathan G. Identification of 2,4-dihydroxy-5-pyrimidinyl imidothiocarbamate as a novel inhibitor to Y box binding protein-1 (YB-1) and its therapeutic actions against breast cancer. // *Eur. J. Pharm. Sci. Off. J. Eur. Fed. Pharm. Sci.*, 2018, Vol. 116, P. 2–14.
32. Khan M.I., Adhami V.M., Lall R.K., Sechi M., Joshi D.C., Haidar O.M., Syed D.N., Siddiqui I.A., Chiu S.-Y., Mukhtar H. YB-1 expression promotes epithelial-to-mesenchymal transition in prostate cancer that is inhibited by a small molecule fisetin. // *Oncotarget*, 2014, Vol. 5, No. 9, P. 2462–2474.
33. Evdokimova V., Ruzanov P., Anglesio M.S., Sorokin A. V., Ovchinnikov L.P., Buckley J., Triche T.J., Sonenberg N., Sorensen P.H.B. Akt-mediated YB-1 phosphorylation activates translation of silent mRNA species. // *Mol. Cell. Biol.*, 2006, Vol. 26, No. 1, P. 277–292.
34. Lavecchia A., Di Giovanni C. Virtual screening strategies in drug discovery: a critical review. // *Curr. Med. Chem.*, 2013, Vol. 20, No. 23, P. 2839–2860.
35. Horvath D. A Virtual Screening Approach Applied to the Search for Trypanothione Reductase Inhibitors // *J. Med. Chem.*, 1997, Vol. 40, No. 15, P. 2412–2423.
36. Kubinyi H. Success Stories of Computer-Aided Design // Wiley Online Books, Computer Applications in Pharmaceutical Research and Development, 2006, P. 377–424.
37. Gimeno A., Ojeda-Montes M.J., Tomás-Hernández S., Cereto-Massagué A., Beltrán-Debón R., Mulero M., Pujadas G., Garcia-Vallvé S. The Light and Dark Sides of Virtual Screening: What Is There to Know? // *Int. J. Mol. Sci.*, 2019, Vol. 20, No. 6, P. 1375.
38. Glaab E. Building a virtual ligand screening pipeline using free software: a survey // *Brief. Bioinform.*, 2015, Vol. 17, No. 2, P. 352–366.
39. Marvin, P. A full featured chemical editor for making science.
40. CORINA Classic [Electronic resource], P. High-Quality 3D Molecular Models, URL: <https://www.mn-am.com/products/corina>.

41. Sadowski J. 3D Structure Generation // Wiley Online Books, Handbook of Chemoinformatics, 2003, P. 231–261.
42. Mannhold R., Kubinyi H., Folkers G. Molecular Drug Properties: Measurement and Prediction // Methods and Principles in Medicinal Chemistry, , Wiley, 2008.
43. Houston D.R., Walkinshaw M.D. Consensus Docking: Improving the Reliability of Docking in a Virtual Screening Context // J. Chem. Inf. Model., 2013, Vol. 53, No. 2, P. 384–390.
44. Cheng T., Li X., Li Y., Liu Z., Wang R. Comparative assessment of scoring functions on a diverse test set. // J. Chem. Inf. Model., 2009, Vol. 49, No. 4, P. 1079–1093.
45. Genheden S., Ryde U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities // Expert Opin. Drug Discov. 2015/04/02, 2015, Vol. 10, No. 5, P. 449–461.
46. Leontyev I., Stuchebrukhov A. Accounting for electronic polarization in non-polarizable force fields // Phys. Chem. Chem. Phys., 2011, Vol. 13, No. 7, P. 2613–2626.
47. Christensen A.S., Kubař T., Cui Q., Elstner M. Semiempirical Quantum Mechanical Methods for Noncovalent Interactions for Chemical and Biochemical Applications // Chem. Rev., 2016, Vol. 116, No. 9, P. 5301–5337.
48. Cui Q., Elstner M. Density functional tight binding: values of semi-empirical methods in an ab initio era // Phys. Chem. Chem. Phys., 2014, Vol. 16, No. 28, P. 14368–14377.
49. Pople J.A., Segal G.A. Approximate Self-Consistent Molecular Orbital Theory. II. Calculations with Complete Neglect of Differential Overlap // J. Chem. Phys., 1965, Vol. 43, No. 10, P. S136–S151.
50. Voityuk A. Intermediate neglect of differential overlap for spectroscopy // Wiley Interdiscip. Rev. Comput. Mol. Sci., 2013, Vol. 3.
51. Sustmann R., Williams J.E., Dewar M.J.S., Allen L.C., Schleyer P. von R. Molecular orbital calculations on carbonium ions. II. Methyl, ethyl, and vinyl

- cations. The series $C_3H_7^+$ // *J. Am. Chem. Soc.*, 1969, Vol. 91, No. 19, P. 5350–5357.
52. Dewar M.J.S., Thiel W. A semiempirical model for the two-center repulsion integrals in the NDDO approximation // *Theor. Chim. Acta*, 1977, Vol. 46, No. 2, P. 89–104.
53. Stewart J.J.P. Optimization of parameters for semiempirical methods II. Applications // *J. Comput. Chem.*, 1989, Vol. 10, No. 2, P. 221–264.
54. Stewart J.J. Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters // *J. Mol. Model.* 11/29, 2013, Vol. 19, No. 1, P. 1–32.
55. Korth M. Third-Generation Hydrogen-Bonding Corrections for Semiempirical QM Methods and Force Fields. // *J. Chem. Theory Comput.*, 2010, Vol. 6, P. 3808–3816.
56. Sulimov A. V., Kutov D.C., Katkova E. V., Ilin I.S., Sulimov V.B. New generation of docking programs: Supercomputer validation of force fields and quantum-chemical methods for docking // *J. Mol. Graph. Model.*, 2017, Vol. 78, P. 139–147.
57. Stewart J.J.P. Stewart Computational Chemistry. MOPAC2016 [Electronic resource], Colorado Springs, CO, USA, 2016, URL: <http://openmopac.net/MOPAC2016.html> (accessed: 30.07.2020).
58. Klamt A., Schuurmann G. COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient // *J. Chem. Soc. Perkin Trans. 2*, 1993, No. 5, P. 799–805.
59. Stewart J.J.P. Application of localized molecular orbitals to the solution of semiempirical self-consistent field equations // *International Journal of Quantum Chemistry*, 1996, Vol. 58, No. 2, P. 133–146.
60. Sousa S.F., Fernandes P.A., Ramos M.J. Protein-ligand docking: current status and future challenges // *Proteins*. 2006/07/25, 2006, Vol. 65, No. 1, P. 15–26.
61. Fan J., Fu A., Zhang L. Progress in molecular docking // *Quant. Biol.*, 2019, Vol. 7, No. 2, P. 83–89.

62. Sulimov V.B., Kutov D.C., Sulimov A. V. Advances in docking // *Curr. Med. Chem.*, 2019, Vol. 26, No. 42, P. 7555–7580.
63. Sulimov A. V., Kutov D.C., Katkova E. V., Sulimov V.B. Combined docking with classical force field and quantum chemical semiempirical method PM7 // *Adv. Bioinformatics*, 2017, Vol. 2017, Article ID 7167691, 6 pages.
64. Chen R., Li L., Weng Z. ZDOCK: An initial-stage protein-docking algorithm // *Proteins Struct. Funct. Bioinforma.*, 2003, Vol. 52, No. 1, P. 80–87.
65. Meng X.Y., Zhang H.X., Mezei M., Cui M. Molecular docking: a powerful approach for structure-based drug discovery // *Curr Comput Aided Drug Des.* 2011/05/04, 2011, Vol. 7, No. 2, P. 146–157.
66. Rarey M., Kramer B., Lengauer T., Klebe G. A Fast Flexible Docking Method using an Incremental Construction Algorithm // *J. Mol. Biol.*, 1996, Vol. 261, No. 3, P. 470–489.
67. Goldberg D.E. Genetic algorithms in search, optimization, and machine learning // *Read. Addison-Wesley*, 1989.
68. Pagadala N.S., Syed K., Tuszynski J. Software for molecular docking: a review // *Biophys. Rev.* 2017/05/17, 2017, Vol. 9, No. 2, P. 91–102.
69. Huang S.-Y., Grinter S.Z., Zou X. Scoring functions and their evaluation methods for protein–ligand docking: recent advances and future directions // *Phys. Chem. Chem. Phys.*, 2010, Vol. 12, No. 40, P. 12899–12908.
70. Guedes I.A., Pereira F.S.S., Dardenne L.E. Empirical Scoring Functions for Structure-Based Virtual Screening: Applications, Critical Aspects, and Challenges // *Frontiers in Pharmacology* , 2018, Vol. 9, P. 1089.
71. Chen Y.C. Beware of docking! // *Trends Pharmacol Sci.* 2014/12/30, 2015, Vol. 36, No. 2, P. 78–95.
72. Romanov A.N., Kondakova O.A., Grigoriev F. V., Sulimov A. V., Lushekina S. V., Martynov Y.B., Sulimov V.B. The SOL docking package for computer-aided drug design // *Numerical methods and programming*, 2008, Vol. 9, No. 2, P. 213-233 (in Russian).
73. Mongan J., Simmerling C., McCammon J.A., Case D.A., Onufriev A.

Generalized Born Model with a Simple, Robust Molecular Volume Correction // *J. Chem. Theory Comput.*, 2007, Vol. 3, No. 1, P. 156–169.

74. Sulimov A. V., Kutov D.K., Katkova E. V., Sulimov V.B. Application of the docking program SOL for CSAR benchmark // 244th ACS National Meeting, Philadelphia, Pennsylvania, USA, 2012, Vol. 244, P. 1.
75. Sinauridze E.I., Romanov A.N., Gribkova I. V., Kondakova O.A., Surov S.S., Gorbatenko A.S., Butylin A.A., Monakov M.Y., Bogolyubov A.A., Kuznetsov Y. V., Sulimov V.B., Ataulakhanov F.I. New synthetic thrombin inhibitors: Molecular design and experimental verification // *PLoS One*. 2011/05/24, 2011, Vol. 6, No. 5, P. e19969.
76. Sulimov V.B., Katkova E. V., Oferkin I. V., Sulimov A. V., Romanov A.N., Roschin A.I., Beloglazova I.B., Plekhanova O.S., Tkachuk V.A., Sadovnichiy V.A. Application of molecular modeling to urokinase inhibitors development // *Biomed Res. Int.* 2014/06/27, 2014, Vol. 2014, P. 625176.
77. Hanwell M.D., Curtis D.E., Lonie D.C., Vandermeersch T., Zurek E., Hutchison G.R. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform // *J. Cheminform.*, 2012, Vol. 4, No. 1, P. 17.
78. Voronezh State University database [Electronic resource].
79. Nicklaus M.C. National Cancer Institute. NCI Database [Electronic resource], URL: <https://cactus.nci.nih.gov/download/nci/> (accessed: 30.07.2020).
80. Voevodin V. V., Antonov A.S., Nikitenko D.A., Shvets P.A., Sobolev S.I., Sidorov I.Y., Stefanov K.S., Voevodin V. V., Zhumatiy S.A. Supercomputer Lomonosov-2: Large scale, deep monitoring and fine analytics for the user community // *Supercomput. Front. Innov.*, 2019, Vol. 6, No. 2, P. 4–11.
81. O'Boyle N.M., Banck M., James C.A., Morley C., Vandermeersch T., Hutchison G.R. Open Babel: An open chemical toolbox // *J Cheminform.* 2011/10/11, 2011, Vol. 3, P. 33.
82. Case D., Betz R., Cerutti D.S., Cheatham T., Darden T., Duke R., Giese T.J., Gohlke H., Götz A., Homeyer N., Izadi S., Janowski P., Kaus J., Kovalenko A., Lee T.-S., LeGrand S., Li P., Lin C., Luchko T., et al. Amber 16, University of

California, San Francisco., 2016.

83. Brandon C.J., Martin B.P., McGee K.J., Stewart J.J., Braun-Sand S.B. An approach to creating a more realistic working model from a protein data bank entry // J Mol Model. 2015/01/22, 2015, Vol. 21, No. 1, P. 3.