ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«КАЗАНСКИЙ (ПРИВОЛЖСКИЙ) ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»
ИНСТИТУТ ФУНДАМЕНТАЛЬНОЙ МЕДИЦИНЫ И БИОЛОГИИ

На правах рукописи

Андрианов Григорий Васильевич

Разработка новых биоинформатических подходов для подбора ингибиторов киназной активности

1.5.8. Математическая биология, биоинформатика

ДИССЕРТАЦИЯ
на соискание ученой степени
кандидата биологических наук

Научный руководитель:
кандидат биологических наук
Илья Генрихович Серебрийский

Москва – 2025

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ 4
1 ОБЗОР ЛИТЕРАТУРЫ 12
1.1 Киназы и их функциональная роль 12
1.2 Киназы как терапевтические мишени 13
1.3 Структурная характеристика каталитического киназного домена 15
1.4 Типы киназных ингибиторов 19
1.5 Виртуальный скрининг в разработке лекарственных средств 25
1.6 Фрагментно-ориентированная разработка лекарственных средств 27
1.7 Модели машинного обучения в разработке лекарственных средств 31
2 МАТЕРИАЛЫ И МЕТОДЫ 33
2.1 Разработка новых киназных ингибиторов с использованием фрагментно-ориентированного дизайна 33
2.2 Моделирование взаимодействия мишень-PROTAC-лигаза E3 36
2.3 Поиск киназных ингибиторов традиционными методами виртуального скрининга 40
2.4 Синтез киназных ингибиторов и определение их активности 42
2.5 Подготовка тренировочного набора и обучение vScreenML 45
3 РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ 49
3.1 Фрагментно-ориентированная разработка ингибиторов ПК 49
3.1.1 Создание библиотек фрагментов 50
3.1.2 Скрининг фрагментов и последующее слияние 53
3.1.3 Попарная аддитивность энергий взаимодействия 55
3.1.4 Селективность из аддитивной энергии фрагментов 57
3.1.5 Скрининг масштабных химических библиотек 59
3.2 Моделирование структуры тройного комплекса 64
3.2.1 Влияние длины линкера PROTAC на эффективность деградации 64
3.2.2 Селективность PROTAC в деградации киназных мишеней 66

3.3 Поиск новых ингибиторов с помощью традиционных методов виртуального скрининга 70
3.4 Классификация предсказанных моделей при помощи vScreenML 72
ЗАКЛЮЧЕНИЕ 79
ВЫВОДЫ 84
СПИСОК СОКРАЩЕНИЙ И ОБОЗНАЧЕНИЙ 85
СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ 87
ПРИЛОЖЕНИЕ А 105
ПРИЛОЖЕНИЕ Б 107
ПРИЛОЖЕНИЕ В 108
ПРИЛОЖЕНИЕ Г 110
ПРИЛОЖЕНИЕ Д 113

ВВЕДЕНИЕ

Актуальность темы исследования и степень ее разработанности

Протеинкиназы (ПК) представляют собой одно из самых больших семейств ферментов (538 киназ были идентифицированы у человека) и играют важную роль в передаче внутриклеточных сигналов [1,2]. ПК катализируют перенос γ-фосфатной группы АТФ на определенные аминокислотные остатки (тирозин, серин или треонин) белка-мишени, что приводит к определенным конформационным трансформациям и последующему изменению активности [3]. Этот механизм играет важную роль в регуляции множества биологических процессов, таких как транскрипция, клеточный цикл, апоптоз, дифференциация, метаболизм и межклеточная коммуникация [4–9].

Нарушения в активности ПК часто ассоциируются с развитием различных видов заболеваний [10]. Поэтому ПК являются важными терапевтическими мишенями при разработке новых стратегий лечения. К 2024 году был одобрен 81 низкомолекулярный ингибитор ПК для клинического применения, причем значительная часть из них была одобрена в последние годы [11]. Более 180 ингибиторов сейчас находятся на этапе клинических испытаний [12]. Тем не менее, процесс разработки новых ингибиторов ПК сталкивается с рядом проблем: 1) высокая консервативность АТФ-связывающего сайта относительно всего кинома (т.е. совокупности всех ПК) приводит к низкой селективности ингибиторов, что может вызывать побочные эффекты через взаимодействие с нежелательными мишенями [13,14]) существование киназ с перекрывающимися функциям усложняет селективное подавление патологических сигналов, поскольку часто требуется ингибирование нескольких ПК одновременно для достижения терапевтического эффекта [15].

Таким образом, разработка новых низкомолекулярных ингибиторов ПК является сложной задачей и требует всестороннего исследования влияния потенциальных лекарственных кандидатов на кином, соблюдая баланс между безопасностью и эффективностью. Из-за высокой стоимости и продолжительности

экспериментов по проверке ингибиторов ПК активно применяются методы виртуального скрининга, которые позволяют с помощью вычислительных ресурсов предварительно оценить перспективность соединений [16–18]. Эти методы позволяют уже на ранних этапах отсечь значительное количество неработающих соединений и определить перспективных кандидатов для дальнейшего тестирования *in vitro*, что влияет на сокращение временных и трудозатрат в процессе разработки.

Тем не менее, несмотря на эффективность традиционных методов виртуального скрининга, остается несколько нерешенных проблем: 1) химическое пространство потенциальных лекарственных соединений огромно и оценивается в диапазоне от 10^{30} - 10^{60} соединений [19,20], что делает прямой перебор невозможным, 2) покрытие исследуемого химического пространства неоднородно, и зачастую сосредоточено вокруг уже известных ингибиторов [21].

Для решения этих проблем разрабатываются новые стратегии виртуального скрининга. Одним из подходов является фрагментно-ориентированная разработка лекарств, где оцениваются небольшие молекулярные фрагменты (≤ 300 дальтон), что снижает структурное разнообразие по сравнению с полноразмерными лекарственными соединениями [22,23]. Это позволяет уменьшить количество исследуемых соединений на несколько порядков и ускоряет выделение перспективных кандидатов.

Кроме того, в последнее время исследуются возможности применения молекулярных клеев, способствующих контактам между белками, которые обычно взаимодействуют редко [24,25]. Одним из производных данного типа соединений являются протеолиз-таргетированные химеры (англ. proteolysis targeting chimera, PROTAC), которые представляют собой бифункциональные молекулы, в которых один функциональный фрагмент взаимодействует с мишенью, а другой - с убиквитинлигазой E3. После образования тройного комплекса мишень-PROTAC-лигаза E3, происходит перенос убиквитина с лигазы E3 на мишень, что приводит к деградации целевого белка через протеасому. В отличие от традиционных ингибиторов ПК, такие соединения обуславливают деградацию мишени через

взаимодействие с неконсервативной внешней поверхностью каталитического домена ПК, что потенциально может решить проблему низкой селективности [26]. Тем не менее разработка вычислительных методов для предсказания эффективности PROTAC (и того, как образуются тройные комплексы мишень-PROTAC-лигаза E3) является сложной и до сих пор нерешенной задачей.

Традиционные методы виртуального скрининга, такие, как молекулярный докинг или поиск по молекулярному подобию, по-прежнему широко применяются при разработке новых соединений, однако они часто страдают низкой точностью предсказаний [27]. Для преодоления этих ограничений активно внедряются модели машинного обучения, корректирующие ошибки, возникающие при использовании традиционных методов [28–30]. Например, в данном контексте традиционные методы виртуального скрининга выступает в роли "генератора гипотез", тогда как модели машинного обучения играют роль "дискриминатора", отбирая наиболее правдоподобные предсказания [31].

Таким образом, существующие проблемы низкой селективности ингибиторов, сложности в покрытии обширного химического пространства и ограниченная точность традиционных методов виртуального скрининга подчеркивают необходимость разработки новых подходов, которые позволят преодолеть эти ограничения и повысить эффективность поиска перспективных соединений для ингибирования ПК.

Цели и задачи исследования

Целью данной диссертационной работы является разработка новых вычислительных подходов для подбора регуляторов киназной активности с помощью различных методов виртуального скрининга.

Для достижения цели работы нами были поставлены следующие задачи:

1. Создать новый высокоэффективный фрагментно-ориентированный метод для поиска потенциальных ингибиторов протеинкиназ в масштабных библиотеках, полученных путем виртуального комбинаторного синтеза.
2. Разработать метод предсказания взаимного расположения членов тройного комплекса, образованного для протеолитической деградации

протеинкиназ при помощи молекулы PROTAC, и проанализировать его эффективность и селективность.

3. Создать научно-методическую базу для оценки эффективности и сравнительного анализа предсказательной способности новых разработанных вычислительных методов по сравнению с традиционными методами виртуального скрининга.

Научная новизна работы

Разработан новый вычислительно-эффективный метод поиска новых ингибиторов ПК, который совмещает фрагментно-ориентированную разработку лекарств и комбинаторную химию, что позволяет эффективно анализировать многомиллионные библиотеки низкомолекулярных соединений в короткое время. С использованием разработанного метода были предсказаны новые химические соединения, которые могут обладать активностью против CDK2, CDK9, $EGFR^{T790M}$, CHK1 и ACK1. Также был представлен новый метод для моделирования трехмерной структуры тройного комплекса мишень-PROTAC-лигаза Е3 и был разработан метод, позволяющий оценить совместимость PROTAC со структурами мишени и лигазы Е3. Усовершенствован вычислительный метод vScreenML, который позволяет сократить количество ложно-положительных предсказаний, полученных с помощью традиционных методов виртуального скрининга.

Теоретическая и практическая значимость

Проведенное исследование способствует ускорению разработки новых эффективных и селективных ингибиторов киназ. Фрагментно-ориентированный скрининг совместно с методами виртуальной комбинаторной химии позволяет проанализировать объемные специализированные библиотеки низкомолекулярных соединений и отобрать наиболее эффективные химические соединения в короткое время.

Метод предсказания возможных моделей тройного комплекса мишень-PROTAC-лигаза Е3 позволяет оценить возможную эффективность деградации мишени уже на ранних этапах разработки новых соединений и, следовательно,

облегчить оптимизацию структуры перспективных молекул-кандидатов. В итоге данная комбинация методов способна повысить шансы найти новые эффективные и селективные соединения для терапий, основанных на ингибировании аномальной активности киназ.

Проведение виртуального скрининга коммерчески доступной библиотеки химических соединений Enamine, ранжирование при помощи vScreenML и проверка активности отобранных соединений *in vitro* позволило создать базу для объективного сравнения эффективности работы разрабатываемых вычислительных методов.

Методология исследования

Для решения поставленных задач использовались методы молекулярного моделирования, виртуального скрининга и машинного обучения. Виртуальный синтез соединений осуществлялся с использованием предопределенных шаблонов SMARTS и программного обеспечения RDKit. Скрининг соединений проводился с применением программного обеспечения OMEGA для генерации конформеров, ROCS для трехмерного выравнивания относительно известных ингибиторов, PyRosetta для белок-белкового докинга и минимизации энергии комплексов. Тренировка и оценка модели машинного обучения осуществлялись при помощи программных пакетов scikit-learn и mlxtend.

Основные положения, выносимые на защиту

1. Вычислительный метод поиска новых ингибиторов протеинкиназ при помощи совмещения подходов фрагментно-ориентированной разработки лекарств и комбинаторной химии позволяет анализировать многомиллионные библиотеки низкомолекулярных соединений в тысячи раз быстрее по сравнению с традиционными методами виртуального скрининга.
2. Метод моделирования взаимного положения компонентов тройного комплекса, использующий разработанную нами новую метрику доли совместимых комплексов позволяет предсказать избирательность взаимодействия PROTAC с мишенью и убиквитинлигазой E3.

3. Последовательное применение традиционных методов виртуального скрининга, модели машинного обучения vScreenML, и определение ингибирующей способности соединений позволило создать научно-методическую базу для сравнения эффективности различных вычислительных методов для разработки новых ингибиторов протеинкиназ.

Достоверность результатов

Представленные результаты получены при помощи современных экспериментальных подходов и проанализированы с использованием соответствующих статистических методов. Результаты являются воспроизводимыми, опубликованы в российских и международных журналах, индексируемых WoS, Scopus и РИНЦ и представлены на научных конференциях и семинарах.

Место выполнения работы и личный вклад соискателя

Основные результаты, представленные в диссертационной работе, получены лично автором. Его вклад включает анализ научной литературы, разработку новых вычислительных методов, планирование и проведение вычислительных экспериментов, обработку и интерпретацию данных, подготовку публикаций и участие в научных конференциях. В работе G.V. Andrianov et al. 2021 непосредственно автором был разработан метод дизайна новых ингибиторов ПК из отдельных фрагментов, а также проанализирована его эффективность. В исследовании N. Bai et al. 2021 им были созданы алгоритм и код для моделирования структуры PROTAC и построения тройного комплекса. В работе Г.В Андрианов и др. 2021 автор осуществил отбор новых потенциальных ингибиторов CDK2 и CDK9, синтез которых выполнялся компанией Enamine, а тестирование ингибирующей активности - компанией Reaction Biology. В работе G.V. Andrianov et al. 2024 автором был разработан код для вычисления дескрипторов, подготовлен новый тренировочный набор, обучена модель машинного обучения и оценена точность ее предсказаний.

Публикация результатов исследования

По теме диссертации опубликовано пять научных работ, из которых четыре в международных и отечественных рецензируемых журналах, индексируемых в базах данных Web of Science, Scopus и RSCI:

1. **Andrianov G.V.**, Gabriel Ong W.J., Serebriiskii I., Karanicolas J. Efficient Hit-to-Lead Searching of Kinase Inhibitor Chemical Space via Computational Fragment Merging // *Journal of Chemical Information and Modeling*. - 2021. - Vol. 61, № 12. - P. 5967-5987. doi: 10.1021/acs.jcim.1c00630. JIF (для WoS) = **5,6**, (1,31 / 0,33) [1].
2. Bai N., Miller S.A., **Andrianov G.V.**, Yates M., Kirubakaran P., Karanicolas J. Rationalizing PROTAC-Mediated Ternary Complex Formation Using Rosetta // *Journal of Chemical Information and Modeling*. - 2021. - Vol. 61, № 3. - P. 1368-1382. doi: 10.1021/acs.jcim.0c01451. JIF (для WoS) = **5,6**, (0,93 / 0,16)[1].
3. **Андрианов Г.В.**, Серебрийский И.Г. Идентификация новых ингибиторов киназной активности CDK2 и CDK9 методами молекулярного моделирования и высокоэффективного скрининга // *Ученые записки Казанского университета. Серия Естественные науки / Uchenye Zapiski Kazanskogo Universiteta. Seriya Estestvennye Nauki.* - 2021. - Т. 163, кн. 4. - С. 543–556. doi: 10.26907/2542-064X.2021.4.543-556. JIF (для WoS) = **0,4**, (0,86 / 0,43) [1].
4. **Andrianov G.V.**, Haroldsen E., Karanicolas J. vScreenML v2.0: Improved Machine Learning Classification for Reducing False Positives in Structure-Based Virtual Screening // *International Journal of Molecular Sciences.* – 2024. – Vol. 25, №22 – P. 12350. doi: 10.3390/ijms252212350. JIF (для WoS) = **4,9**, (0,74 / 0,25) [1].
5. Miller S.A., **Andrianov G.V.**, Mischley V., Wharton K.A., Chen J.J., Karanicolas J. Computational Modeling of PROTAC Ternary Complexes and Linker Design. In Inducing Targeted Protein Degradation // Inducing Targeted Protein Degradation: From Chemical Biology to Drug Discovery and Clinical Applications / Ed. by P. Cromm.

[1] В скобках приведён объём публикации в печатных листах и вклад автора в печатных листах.

Weinheim: Wiley-VCH, 2023. P. 151-176. doi: 10.1002/9783527836208.ch5, (1,63 / 0,27) [1].

Апробация результатов

Результаты исследования докладывались на двух научно-практических конференциях SummerRosettaCon (Сиэттл, 2019 и 2020, США), а также на конференции для разработчиков программного обеспечения по моделированию биологических систем WinterRosettaCon (Нью-Йорк, 2020, США) и на симпозиуме по компьютерному дизайну лекарств (Балтимор, США, 2023). Кроме того, результаты докладывались на 25-й и 26-й ежегодной конференции онкологического института Фокс Чейз (Филадельфия, США, 2020 и 2023).

Структура и объем диссертационной работы

Диссертационная работа состоит из введения, обзора литературы, материалов и методов исследования, результатов, их обсуждения, заключения, выводов, списка цитированной литературы и пяти приложений. Работа изложена на 116 страницах машинописного текста, включает 23 рисунка, 5 таблиц и 5 приложений. Библиография включает 152 библиографических источника.

[1] В скобках приведён объём публикации в печатных листах и вклад автора в печатных листах.

1 ОБЗОР ЛИТЕРАТУРЫ

1.1 Киназы и их функциональная роль

Киназы представляют собой большое семейство ферментов, которое относится к классу фосфотрансфераз, которые катализируют перенос фосфатной группы с АТФ (иногда ГТФ [32]) на целевой субстрат. Приблизительно треть всех белков в эукариотических клетках может подвергаться фосфорилированию в определенных условиях, что вызывает определенные конформационные изменения, влияющие на активность мишени [1–3]. Геном человека содержит 538 киназ (что составляет примерно 2% всей совокупности генов), 518 из которых являются протеинкиназами (ПК), осуществляющими перенос фосфатной группы на определенный аминокислотный остаток белкового субстрата, а оставшиеся 20 являются липидными киназами [33], участвующими в фосфорилировании жирных кислот. Около 90% всех ПК имеют консервативный каталитический домен, который часто называют “эукариотическим”, в то время как оставшиеся 10% являются “атипичными”, так как их каталитический домен имеет низкое сходство с “эукариотическим” доменом. Представителями данного семейства, например, являются бромодоменные киназы [34].

Фосфорилирование играет ключевую роль во многих регуляторных сетях в клетке [35]. Благодаря этому, киназы участвуют во многих аспектах работы клетки: начиная от регуляции апоптоза, клеточного цикла, транскрипции, дифференциации и метаболизма, и заканчивая регуляцией клеточной подвижности и развитием цитоскелета, а также межклеточной коммуникацией [4–9].

Мутации в каталитическом домене могут приводить к их спонтанной активации, вызывая неадекватный клеточный ответ и способствуя развитию различных заболеваний [36,37]. Повышенная активность киназ часто связана с различными онкологическими, нейрологическими и кардиоваскулярными заболеваниями, а также с отклонениями иммунной системы. Более 85% всех киназ человека связаны хотя бы с одним заболеванием [36]. Например, мутация B-Raf

V600E в сигнальном каскаде Ras/Raf/MEK/ERK ассоциируется с неконтролируемой клеточной пролиферацией и выявлена в 7% всех видов рака, а также в 66% случаев злокачественной меланомы [38,39]. Анализ показал, что замена гидрофобного аминокислотного остатка на отрицательно заряженный в каталитическом домене дестабилизирует структуру, что ведет к активации киназы [40]. Часто ПК имеют дублирующие функции в регуляторных каскадах, обеспечивая распространение сигнала через альтернативные пути, если основная киназа не функционирует по каким-то причинам. Это повышает надежность биологической системы, но также может приводить к неконтролируемой активации множества регуляторных путей, что в конечном итоге способствует развитию заболеваний [15].

Лечение заболеваний, вызванных нарушениями в активности киназ, обычно основывается на использовании антагонистического подхода, где чрезмерная активность киназ блокируется с помощью специально подобранных ингибиторов, конкурирующих с АТФ за связывание с мишенью [41].

1.2 Киназы как терапевтические мишени

В последние три десятилетия ПК активно исследуются как потенциальные терапевтические мишени для лечения множества заболеваний [10,42]. Из-за их функциональной роли они являются привлекательными мишенями для терапевтического вмешательства с целью подавления неадекватной активности нарушенных биологических процессов. Помимо измененной функции киназ в онкологических заболеваниях было показано, что они играют важную роль в развитии иммунологических, воспалительных, дегенеративных, метаболических, сердечно-сосудистых и инфекционных заболеваниях (рисунок 1) [41,43,44].

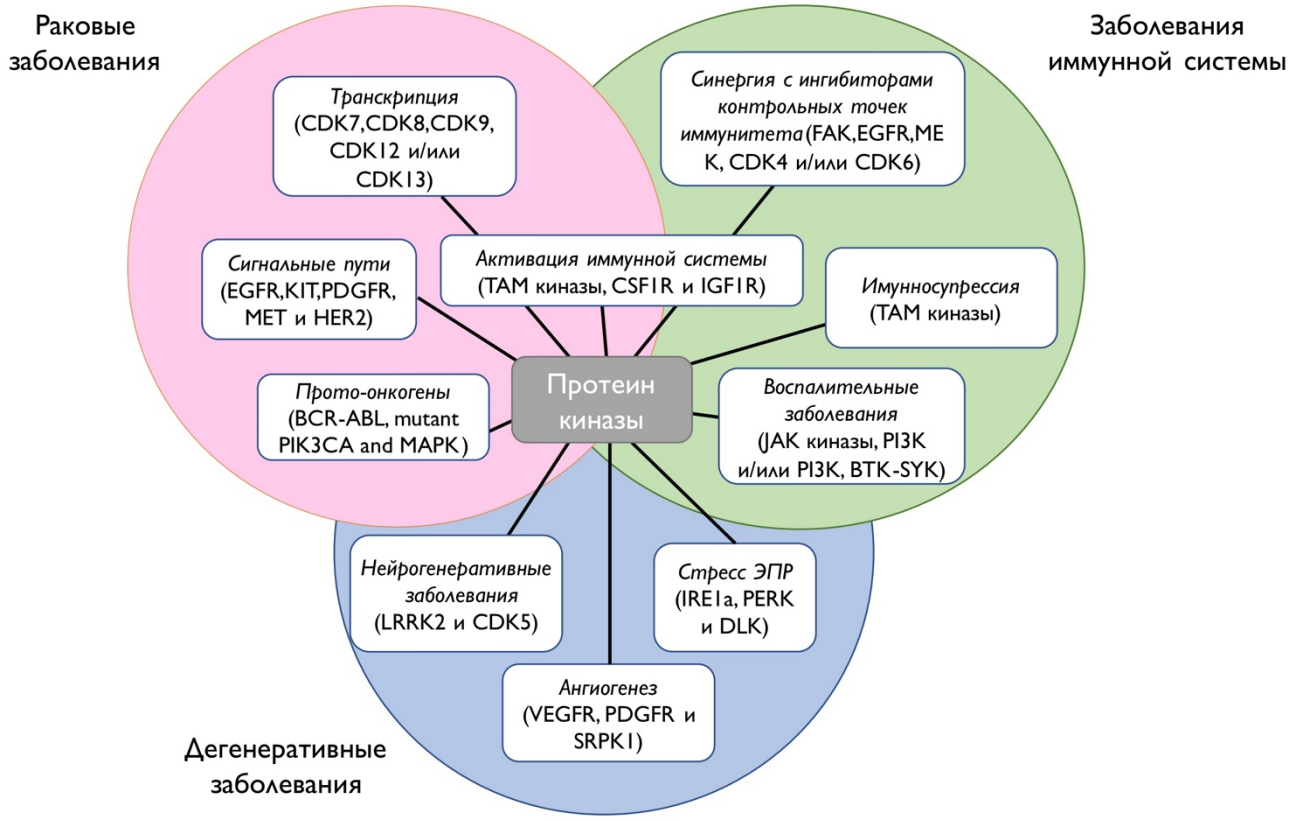


Рисунок 1 - Роль ПК в различных биологических процессах. В контексте раковых заболеваний (розовый) ПК участвуют в регуляции транскрипции, активации сигнальных путей и могут действовать как прото-онкогены. В иммунных заболеваниях (зеленый) они играют центральную роль в иммунной активации, регулируют воспалительные процессы и взаимодействуют с ингибиторами контрольных точек иммунного ответа. В дегенеративных заболеваниях (синий) протеинкиназы участвуют в механизмах нейродегенерации, стрессовых реакциях эндоплазматического ретикулума и ангиогенезе [41]

Многие киназы до сих пор мало изучены, и их совокупность часто называется "темным киномом", что указывает на относительную незрелость данной области [45,46]. Изначально терапевтические стратегии были сосредоточены на ингибировании онкогенных киназ из семейства рецепторных тирозинкиназ. Однако с появлением дополнительных механизмов резистентности к разработанным лекарствам [47], возникла необходимость найти альтернативные мишени, являющиеся важными для выживания злокачественных клеток [48]. Несмотря на расширяющееся применение ингибиторов киназ в лечении различных заболеваний, существует ряд проблем:

1) Развитие специфических ингибиторов для мутантных ПК увеличивает шансы на положительный терапевтический исход, но может привести к резистентности при длительном применении, что уменьшает эффективность лечения. Так, недавно одобренный ингибитор третьего поколения, осимертиниб, нацеленный на мутацию $EGFR^{T790M}$, оказался неэффективным в случае появления дополнительной мутации $EGFR^{C797S}$ [49].
2) Из-за высокой степени консервативности АТФ-связывающего сайта, селективность остается одной из главных проблем при разработке новых лекарственных средств [50].

1.3 Структурная характеристика каталитического киназного домена

Каталитический домен ПК является крайне консервативным как в виде одномерной линейной последовательности аминокислотных остатков, так и в трехмерном представлении [14,51] и содержит два больших функциональных элемента - верхний N-концевой фрагмент, ответственный за правильную координацию положения АТФ в связывающем сайте, и нижний С-концевой фрагмент, определяющий связывание с белковым субстратом. Обе части соединены подвижной шарнирной петлей, а также между ними образуется карман, который взаимодействует с АТФ (рисунок 2А). Внутри связывающего сайта аденин АТФ зажат между гидрофобными группами сверху и снизу, а его полярный край образует несколько водородных связей с аминокислотными остатками шарнирной области (рисунок 2Б). Как правило, АТФ-конкурентные ингибиторы имеют фрагмент, взаимодействующий с шарнирным регионом подобно взаимодействию аденина АТФ (рисунок 2В).

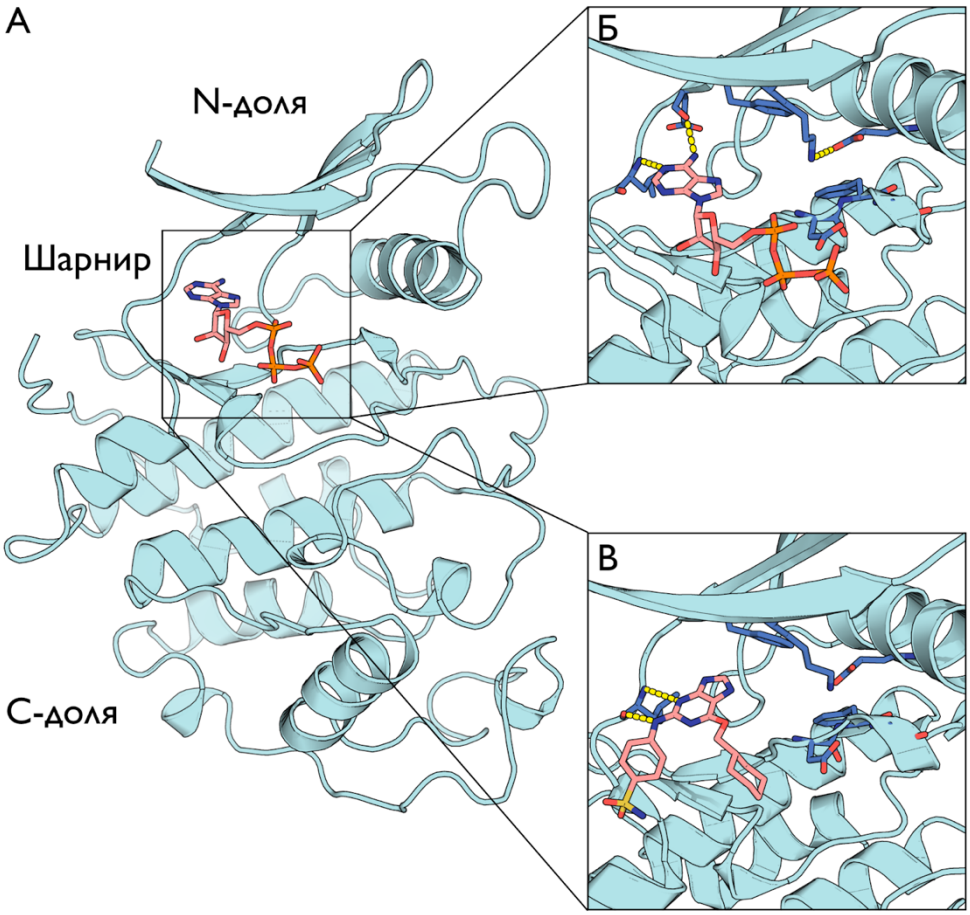


Рисунок 2 - Структурная характеристика каталитического домена ПК в комплексе с АТФ и АТФ-конкурентным ингибитором. А - Каталитический домен CDK9, состоящий из двух долей, связанных подвижной шарнирным петлей (PDB ID: 3BLQ), Б - Взаимодействие АТФ с АТФ-связывающим сайтом ПК (PDB ID: 3BLQ), В - Взаимодействие АТФ-конкурентного ингибитора NU6102 с сайтом ПК (PDB ID: 5LQF)

Фосфорилирование субстрата осуществляется только тогда, когда каталитический домен киназ находится в активной конформации. Это состояние имеет общие структурные черты для всех ПК, тогда как неактивные состояния могут значительно различаться в зависимости от семейства киназ [14]. Выделяется ряд структурных элементов домена, взаимное расположение которых обусловливает активные и неактивные конформации домена.

На N-конце шарнирной петли расположена аминокислота-привратник, размер и конформация которой значительно влияет на доступ к заднему гидрофобному карману сайта связывания. В этом кармане находятся два структурных элемента, определяющие активность домена. В начале активационной

петли домена находится DFG-мотив (Asp-Phe-Gly). Боковая цепь аспарагиновой кислоты определяет положение иона Mg^{2+}, который в свою очередь необходим для координации β- и γ-фосфатных группы молекулы АТФ. Ориентация фенилаланина контролирует доступ к внутренней части гидрофобного кармана. В активной конформации данный аминокислотный остаток повернут к внутренней части кармана и такое состояние называют DFG-in, а для неактивной - по направлению к АТФ и такое состояние охарактеризовано как DFG-out. Глутамин αC-спирали и лизин β3-структуры формируют солевой мостик в активной конформации (αC-in), стабилизируя α- и β-фосфатные группы при связывании АТФ (рисунок 3).

Конформационная характеристика каталитического домена ПК важна для разработки конформационно-специфических ингибиторов, которые могут избирательно связываться с активной или неактивной конформацией домена.

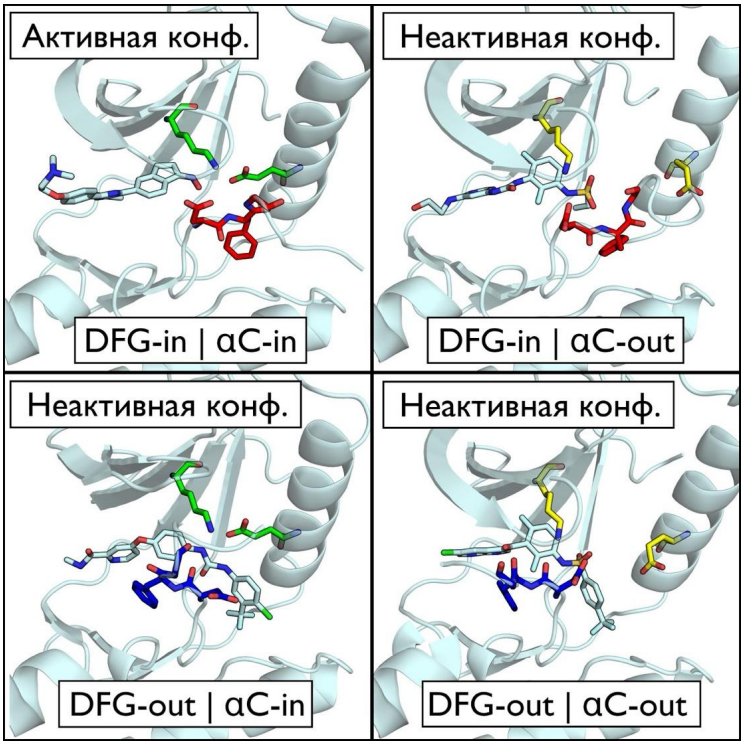


Рисунок 3 - Структурная характеристика элементов, определяющих конформацию каталитического домена ПК. Активная конформация домена определяется внутренними конформациями мотива DFG (DFG-in, красный) и спирали αC (αC-in, зеленый). В иных вариантах, конформация домена считается неактивной. Использованные структуры были получены из PDB - 2FB8, 4EHG, 1UWH и 4XV9

В 2014 году van Linden et al. формализовали определение АТФ-связывающего кармана для того, чтобы выявить закономерности, определяющие

активность разработанных ингибиторов. Согласно их определению, киназный домен имеет 85 ключевых аминокислотных остатков, которые формируют связывающий сайт и определяют положение регулирующих конформацию структурных элементов [52–54]. Также были выявлены три основных региона, участвующих во взаимодействии с низкомолекулярными соединениями: 1) передний карман, имеющий три под-кармана в которых располагаются аденин, рибозный и фосфатные фрагменты АТФ, 2) карман привратника, состоящий из двух под-карманов, которые подразумевают контакт с аминокислотой привратника и DFG-мотивом, 3) задний карман, который имеет две конфигурации на основе конформации DFG-мотива – в DFG-in имеется три под-кармана, а в DFG-out - пять (рисунок 4).

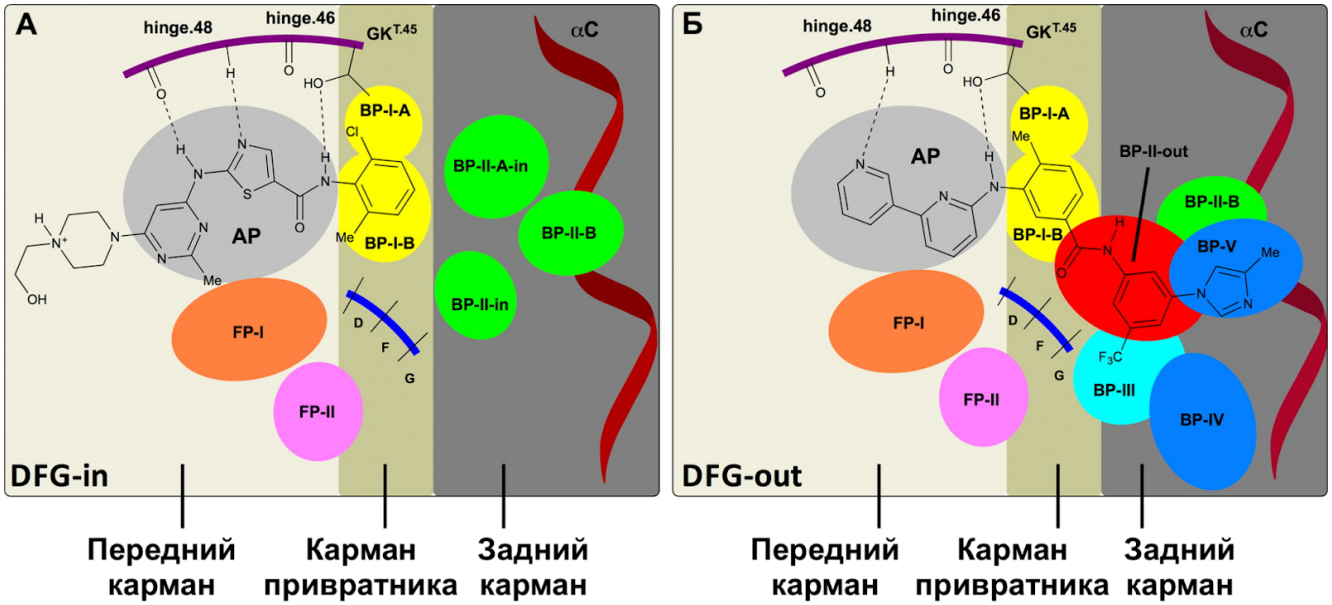


Рисунок 4 - Схематическое представление доступных для взаимодействия с низкомолекулярными соединениями карманов АТФ-связывающего сайта. Конфигурация карманов при А) DFG-in конформации и Б) DFG-out конформации каталитического домена. FP-* обозначают доступные для взаимодействия регионы в переднем кармане, BP-* - в заднем кармане [54]

В работе Modi V. и соавторов был предложен подход для автоматического определения конформации домена ПК [14], основанный на анализе совокупности двугранных углов DFG-мотива с использованием карты Рамачандрана. Активная конформация домена была отнесена к одному классу (BLAMinus), тогда как неактивная конформация включала пять различных классов (BLBplus, ABAminus, BLBminus, BLBtrans, BLAplus).

1.4 Типы киназных ингибиторов

По первоначальной классификации было выделено три типа киназных ингибиторов [55].

Ингибиторы I типа. Представляют собой АТФ-конкурентные соединения, которые взаимодействуют только с активной конформацией домена ПК (рисунок 5А). Часто эти ингибиторы подражают взаимодействию АТФ с АТФ-связывающим сайтом, занимая передний карман и образуя водородные связи с шарнирной петлей. Такие соединения обладают высокой эффективностью, однако их селективность, как правило, ограничена высокой консервативностью АТФ-связывающего сайта. Примерами ингибиторов I типа являются дасатиниб, который ингибирует BCR-ABL и Src.

Ингибиторы II типа. Представляют собой соединения, которые взаимодействуют с АТФ-связывающим карманом ПК, подобно ингибиторам I типа. Однако их взаимодействие осуществляется исключительно с неактивной конформацией мишени (рисунок 5Б). В этом состоянии DFG-мотив находится в «наружном» положении, что открывает доступ к заднему гидрофобному карману. Благодаря этому ингибиторы II типа могут иметь более высокую селективность, поскольку задний карман обладает меньшей степенью консервативности. Тем не менее, разработка таких ингибиторов требует глубокого изучения конформационных особенностей, характерных для неактивного состояния изучаемых киназ. Примерами ингибиторов такого типа являются сорафениб, которые используется для ингибирования RAF.

Ингибиторы III типа. Это аллостерические ингибиторы, которые взаимодействуют с регионами, расположенными рядом с АТФ-связывающим карманом, не перекрывая его (рисунок 5В). Это позволяет ингибиторам взаимодействовать с киназой, стабилизируя определенную, как правило, неактивную конформацию домена ПК, что обеспечивает эффективную регуляцию активности фермента без прямой конкуренции с АТФ. Одним из ключевых преимуществ ингибиторов III типа является их высокая селективность, поскольку они взаимодействуют с менее консервативными структурными элементами

домена, характерными для конкретной мишени. Примером такого ингибитора является траметиниб, который избирательно ингибирует MEK1 и MEK2 [56].

Позже были выделены дополнительные типы киназных ингибиторов, основанных на аллостерическом и ковалентном взаимодействии с мишенью.

Ингибиторы IV типа. Представляют собой аллостерические соединения, которые взаимодействуют с участками домена, расположенными вне АТФ-связывающего сайта (рисунок 5Г). Одним из ключевых преимуществ ингибиторов данного типа также является их высокая селективность. Однако разработка таких соединений требует более детального изучения доступных аллостерических карманов. Примером ингибитора IV типа является сиролимус (рапамицин), который избирательно ингибирует активность mTOR, делая его эффективным средством для терапии онкологических заболеваний [57].

Ингибиторы V типа. Данные ингибиторы представляют собой бивалентные соединения, которые состоят из двух функциональных компонентов: один предназначен для взаимодействия с АТФ-связывающим карманом, а другой - с любым иным регионом домена (например, с аллостерическим или субстрат-связывающим сайтом). Эти компоненты связаны длинным подвижным линкером, который обеспечивает их одновременное взаимодействие с мишенью, что значительно повышает селективность и эффективность ингибиторов [58–60]. Тем не менее, их разработка сопряжена с проблемой проницаемости из-за большого размера полученных соединений. Одним из примеров ингибиторов V типа является RapaLink-1, эффективно подавляющий активность mTOR [61].

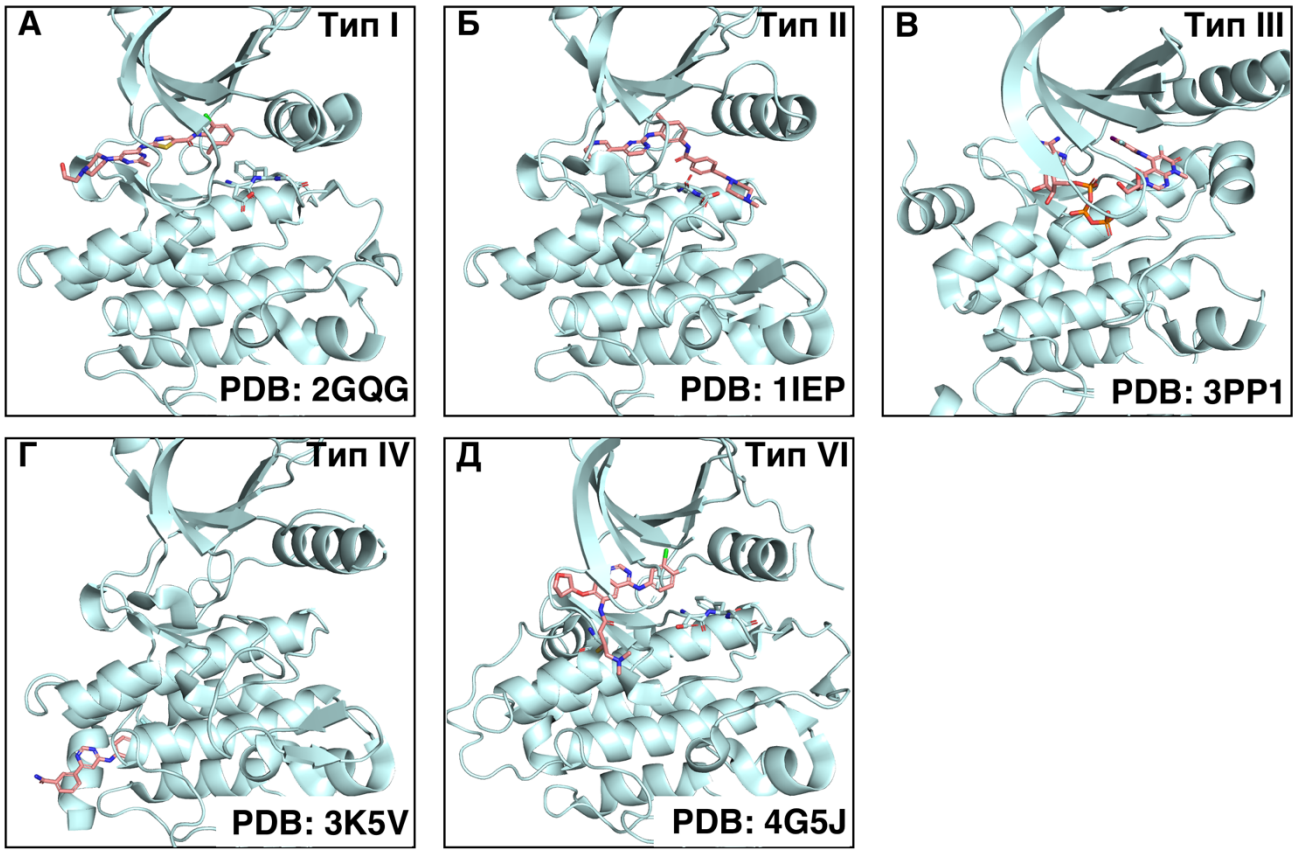


Рисунок 5 - Способ взаимодействия различных типов ингибиторов (розовый цвет) с каталитическим доменом ПК (голубой цвет). А) ингибитор I типа - дасатиниб, Б) ингибитор II типа - иматиниб, В) ингибитор III типа - TAK733, Г) ингибитор IV типа - GNF2, Д) ингибитор VI типа – афатиниб

Ингибиторы VI типа. Представляют собой ковалентные соединения, которые необратимо связываются с аминокислотными остатками, расположенными вблизи активного сайта каталитического домена ПК (рисунок 5E). Такое взаимодействие обеспечивает длительное ингибирование активности фермента, что делает эти соединения эффективными в лечении заболеваний, связанных с гиперактивностью киназ. Ключевое преимущество ингибиторов такого типа заключается в их высокой специфичности, достигаемой благодаря избирательному взаимодействию с неконсервативными реакционноспособными аминокислотными остатками. Однако их необратимый механизм действия может увеличить риск токсичности, особенно при воздействии на здоровые клетки. К числу примеров ингибиторов VI типа относятся дакомитиниб, используемый для ингибирования EGFR, и ибрутиниб, направленный на взаимодействие с BTK [62].

Протеолиз-таргетированные химеры (PROTAC). В последние годы PROTAC стали важной стратегией регуляции активности белковых мишеней через их деградацию [63–65]. Эти молекулы связывают мишень с убиквитинлигазой E3, что инициирует процесс деградации целевого белка.

PROTAC состоят из двух функциональных частей: одна предназначена для связывания с мишенью, другая — для связывания с лигазой E3. Эти две части соединены длинным гибким линкером, который существенно влияет на эффективность и селективность образования комплекса между мишенью и лигазой E3 [26,66,67]. В результате взаимодействия молекулы PROTAC формируется тройной комплекс мишень-PROTAC-лигаза E3. Когда мишень приближается к лигазе E3, происходит убиквитинирование белка-мишени, что в конечном итоге приводит к его деградации, как показано на рисунке 6.

Ключевым преимуществом этого метода является высокая селективность, обусловленная множеством факторов, влияющих на образование тройного комплекса. Такие аспекты, как положение белок-белковых взаимодействий между мишенью и лигазой, а также жесткость линкера PROTAC, вносят дополнительный вклад в избирательность действия ингибитора. Например, в одной из работ, в качестве функциональной части PROTAC для взаимодействия с мишенью был использован форетиниб, известный ингибитор ПК. Сам по себе форетиниб является неселективным и, по разным оценкам, ингибирует около четверти всего кинома человека [26,68]. Однако после разработки двух структур PROTAC, взаимодействующих с различными лигазами E3 — VHL и CRBN — было обнаружено, что эффективность деградации ПК для этих молекул различается. Часть из исследованных киназ деградировалась более эффективно под воздействием первой молекулы PROTAC, связывающей VHL, в то время как другая часть активнее деградировала под действием PROTAC, рекрутирующей CRBN [26]. Этот результат подтверждает гипотезу, что молекулы PROTAC обеспечивают дополнительную селективность благодаря особенностям образования тройного комплекса.

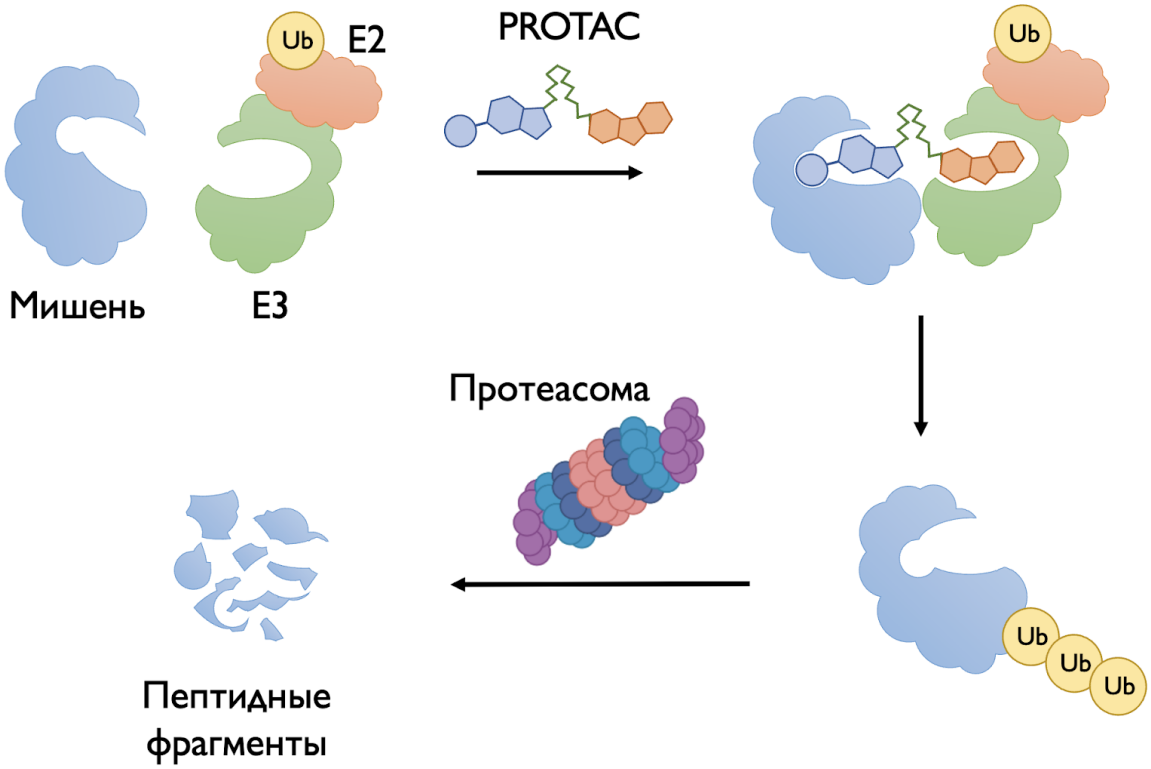


Рисунок 6 - Схематичное представление принципа работы низкомолекулярных ингибиторов PROTAC. При взаимодействии PROTAC с мишенью (голубой цвет) и лигазой E3 (зеленый цвет) образуется тройной комплекс. Затем, благодаря пространственному сближению, лигаза E3 способствует убиквитинированию мишени, что ведет к последующей деградации при помощи протеасомы

Покрытие кинома клинически-одобренными ингибиторами

Большинство клинически одобренных низкомолекулярных соединений разработаны для ингибирования тирозиновых киназ (рисунок 7). Это свидетельствует о том, что многие другие группы киназ по-прежнему остаются недооцененными при разработке новых ингибиторов, и на данный момент терапевтически покрыто лишь 10-15% кинома человека. Около 300 киназ всё ещё не имеют специфических ингибиторов, а для 200 из них отсутствует какая-либо структурная информация, что значительно затрудняет разработку и поиск новых соединений.

Как уже упоминалось, одна из главных сложностей при разработке новых ингибиторов киназ заключается в высокой консервативности АТФ-связывающего сайта. Это создает проблемы с селективностью новых соединений и может

приводить к множеству нежелательных побочных эффектов. Тем не менее, многие исследователи считают, что широкое ингибирование целых семейств ПК может быть полезным для преодоления лекарственной резистентности при онкологических заболеваниях, так как подавляется передача сигнала в обширных регуляторных сетях.

Для успешной разработки новых, более эффективных и селективных ингибиторов киназ необходимы методы, которые могут выявить даже мельчайшие различия в киназных доменах и использовать их для создания новых препаратов. Одним из таких подходов являются методы виртуального скрининга, которые способны удовлетворить эти требования.

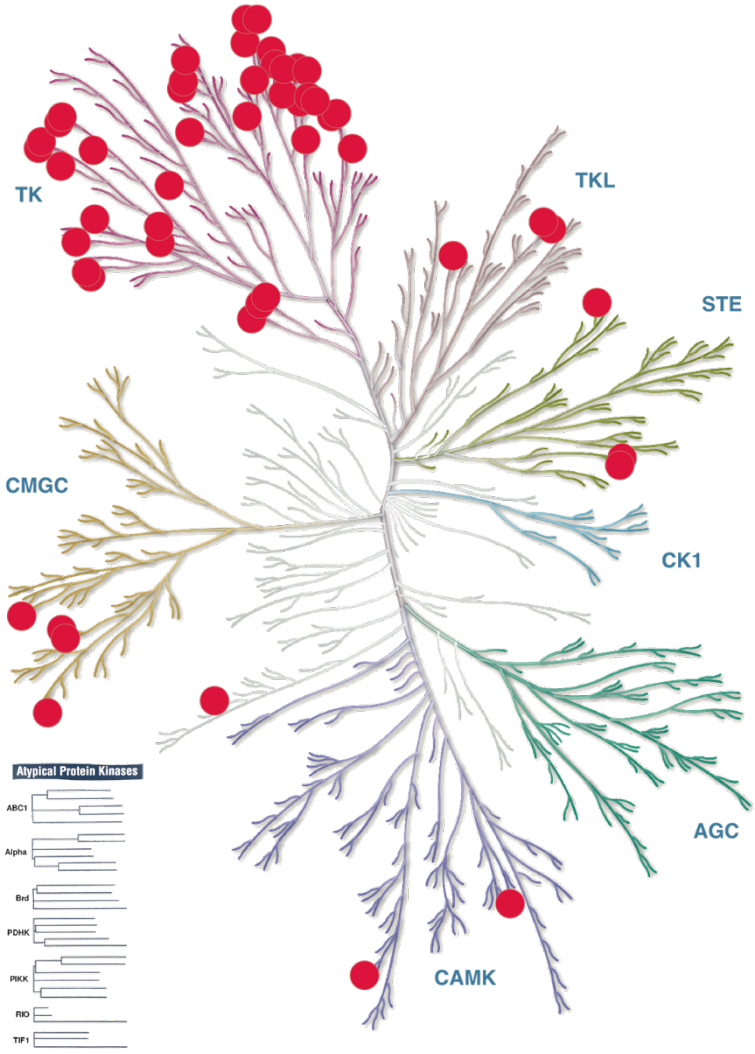


Рисунок 7 - Филогенетическое дерево ПК человека, покрытое клинически одобренными ингибиторами (красный). Значительное количество соединений было разработано для ингибирования ПК семейства ТК

1.5 Виртуальный скрининг в разработке лекарственных средств

Разработка лекарственных средств представляет собой один из наиболее перспективных подходов к лечению различных заболеваний, однако этот процесс характеризуется значительной временно́й и финансовой затратностью, а также требует технологичных ресурсов. Традиционный многоэтапный процесс включает в себя идентификацию соединений-кандидатов, их оптимизацию, а также проведение доклинических и клинических испытаний, которые предшествуют выходу препарата на рынок [69]. Этапы поиска перспективных соединений и их последующей оптимизации могут продолжаться более двух лет. Высок риск, что обнаруженное соединение не пройдет клинические испытания, что приведет к дополнительным затратам. Улучшение начальных этапов отбора перспективных соединений может существенно повысить качество последующих шагов разработки лекарственных средств.

По современным оценкам химическое пространство содержит порядка 10^{30}-10^{60} синтезируемых соединений [20,70]. Экспериментальная проверка даже 0,001% от этого объема для определенной мишени является крайне трудо- и ресурсозатратной. Альтернативным решением является виртуальный скрининг, который использует современные вычислительные методы. Так, в одной из работ посвященных поиску новых хемотипов для β-лактамазы AmpC и дофаминового рецептора D4, был проведен докинг библиотеки из 170 миллионов соединений, что привело к обнаружению новых структур с пико- и наномолярной активностью [71].

Несмотря на огромный потенциал современных вычислительных машин, прямой скрининг без какой-либо предобработки химических библиотек является неэффективным и все еще требует больших трудозатрат. Поскольку химическое пространство огромно, ученые пытаются применить различные эвристические фильтры, которые позволили бы сузить круг поисков. Часто применяются параметры, способные оценить структурное сходство новых химических соединений с уже известными лекарственными средствами. Например, если сравнить структуры трех соединений с обезболивающим эффектом - морфин, кодеин и героин, то можно заметить, что все три соединения имеют общий каркас

(“скаффолд”), подтверждающий, что структурно похожие соединения должны иметь похожий механизм действия [72]. Однако в настоящее время этот подход имеет некоторые сложности из-за проблем с интеллектуальной собственностью, поскольку новые соединения могут иметь минорные изменения в структуре, что ведет к правовому конфликту с уже запатентованными ингибиторами.

Другим способом, как ограничить исследование химического пространства, является разработка сфокусированных комбинаторных библиотек соединений, которые могут быть созданы из доступных реагентов (“строительных блоков”). Очевидно, что разнообразие соединений в данном подходе ограничено начальным набором реагентов. Следовательно, результат виртуального скрининга будет сильно зависеть от размера библиотеки, а также от количества и разнообразия строительных блоков [19,71]. Тем не менее, данный подход показывает достаточно неплохие результаты при локальном поиске новых соединений в химическом пространстве.

Виртуальный скрининг можно представить как последовательный процесс поиска и разработки новых лекарств, включающий следующие основные этапы:

1. Генерация новых низкомолекулярных соединений:
    a) Построение молекул из заранее определенных групп атомов или фрагментов путем простого комбинирования между собой. Хотя данный метод способен значительно обогатить библиотеки, но он имеет проблемы с плохой синтезируемостью сгенерированных молекул.
    b) Проведение виртуальных химических реакций с использованием шаблонов, которые трансформируют строительные блоки в соединения, обладающие физико-химическими и структурными свойствами присущими лекарствам.
    c) Обучение генеративных нейросетевых моделей. В последние несколько лет данный метод начал активно применяться при разработке лекарств, поскольку сложные цепи вычислений, лежащие в основе моделей глубокого обучения, способны “выучить” синтаксис

химического языка и затем использовать его при построении новых соединений.

2. Оценка и отбор соединений по определенным параметрам:
    a) Отсев нежелательных химических соединений может быть выполнен как по физико-химическим свойствам (количество атомов, молекулярный вес, растворимость, подобность лекарственным низкомолекулярным соединениям) [73–75], так и при помощи различных структурных фильтров, которые способны распознать наличие нежелательных реактивных групп [76] или охарактеризовать общую синтезируемость соединений [77,78].
    b) При помощи классических методов виртуального скрининга, таких как молекулярный докинг можно достаточно быстро и точно оценить сродство сгенерированных соединений к мишени [79].
    c) Параллельно применяются модели машинного и глубокого обучения для предсказания активности полученных соединений [80–82]. Данные методы можно разделить на два вида относительно постановки проблемы: классификационные методы, т.е. модель отделяет активные соединения от неактивных и на регрессионные методы - для предсказания количественной меры эффективности найденных соединений.
3. Поиск лучших молекул, опираясь на знания, полученных на предыдущих двух шагах. На данном этапе можно рассматривать эту задачу как поиск количественных соотношений структура-свойство (англ. quantitative structure–activity relationship, QSAR), используя классические [83] и современные алгоритмы оптимизации свойств лекарственных соединений [84,85].

1.6 Фрагментно-ориентированная разработка лекарственных средств

За последние два десятилетия высокопроизводительный скрининг химических соединений (англ. high-throughput screening, HTS) значительно

изменил процесс поиска новых лекарств, что сказалось на скорости, цене и эффективности отобранных кандидатов. Однако покрытие доступного химического пространства органических соединений остается ничтожным, поскольку совокупность всех соединений с возможными лекарственными средствами содержит порядка 10^{30-60} [20,86].

В последние полтора десятилетия вместе с HTS развивается другая парадигма - фрагментно-ориентированная разработка лекарств, которая подразумевает исследование отдельных небольших фрагментов с постепенным увеличением и усложнением их структуры до размеров обычных химических соединений, подчиняющихся правилам лекарственных средств (например, “правило пяти” Липинского [75]). По сравнению с HTS, фрагментная разработка лекарств имеет три явных преимущества. Первое — это более плотное покрытие доступного химического пространства. Например, по некоторым оценкам, набор из 1000 фрагментов примерно равен покрытию 1000000 обычных химических соединений с лекарственными свойствами [87]. Второе - малый размер фрагментов позволяет найти больше хитов, а следовательно, больше возможностей для дальнейшей оптимизации. Кроме того, они имеют лучшую растворимость и чаще удовлетворяют фармакокинетическим, метаболическим и токсикологическим свойствам. Третье - оптимизированные соединения имеют гораздо лучшую эффективность [88]. Сам процесс фрагментно-ориентированной разработки лекарств включает два основных этапа: 1) скрининг низкомолекулярных библиотек для выбора перспективных фрагментов, 2) многошаговая оптимизация выбранных фрагментов для улучшения сродства с мишенью [88,89].

Создание фрагментной библиотеки часто начинается с разделения уже существующих ингибиторов на более простые части с помощью методов, таких как RECAP (Retrosynthetic Combinatorial Analysis Procedure), DAIM (Decomposition And Identification of Molecules) или BRICS (Breaking of Retrosynthetically Interesting Chemical Substructures) [88]. Полученные фрагменты имеют менее 20 атомов и подчиняются “правилу трех”: 1) молекулярный вес $\leq$ 300 Да, 2) число атомов-доноров водородной связи ≤ 3, а также акцепторов ≤ 3, 3) коэффициент разделения

LogP ≤ 3 [74]. Помимо этого, часто учитывают количество вращающихся связей, площадь полярной поверхности [88,90].

Из-за небольшого размера фрагментов их сродство к мишеням обычно находится в микромолярном диапазоне, и для их идентификации используются методы, такие как ядерно-магнитный резонанс и рентгеновская кристаллография. С развитием технологий массивно применяются методики масс-спектрометрии, изотермической титрационной калориметрии, поверхностного плазмонного резонанса, капиллярного электрофореза и других [88,90,91]. Параллельно с экспериментальным методами, часто применяются различные методы виртуального скрининга фрагментов, что позволяет существенно уменьшить количество кандидатов для экспериментальной проверки и дать некоторое понимание того, как будет взаимодействовать тот или иной фрагмент с мишенью. В зависимости от того какая информация доступна, ученые применяют как обычный докинг, который стремится найти наиболее вероятное положение фрагмента относительно мишени, используя некоторую оценочную функцию [79], так и поиск по молекулярному подобию [92]. Однако, из-за малого размера фрагментов, взаимодействие может происходить с множеством карманов, что создает трудности для точного предсказания. Также оценочные функции обычно разработаны для скрининга полноразмерных лекарственных соединений, что создает некоторую предвзятость при скрининге фрагментов [88].

Как уже упоминалось, фрагменты обычно обладают более низким сродством к мишеням, что может создавать проблемы для выявления действующих соединений, поскольку возникает сложность в различении перспективных кандидатов от ложноположительных [88]. Для решения этой проблемы был разработан коэффициент эффективности лиганда (англ. ligand efficiency, LE), который позволяет оценить вклад каждого атома фрагмента для взаимодействия с мишенью [93]. Существует несколько модификаций этого показателя, которые применяются в зависимости от целей проекта: энтальпийная эффективность для оценки термодинамического профиля связывания, эффективность размера

фрагмента за счет молекулярного веса или площади полярной поверхности, липофильная эффективность и т.д.

После отбора перспективных кандидатов из библиотеки необходимо определить стратегию "расширения" фрагмента (рисунок 8), чтобы преобразовать его в более крупное соединение, улучшив его эффективность или селективность [87,88,90,94]. Существуют три основных подхода:

1) Рост (англ. growing). Наиболее простой и популярный в использовании подход, который подразумевает постепенное наращивание дополнительных функциональных групп на основной фрагмент. Для этого пытаются обнаружить векторы роста - точки прикрепления новых фрагментов, которые дадут максимальный прирост к эффективности.
2) Связывание (англ. linking). Объединение неконкурирующих фрагментов, которые связываются с мишенью в близких, но разных участках кармана, при помощи линкера, т.е. специально подобранной линейной комбинации атомов. Поиск совместимого линкера, который не изменит положения фрагментов, является достаточно трудной задачей, поскольку необходимо учесть геометрию линкера для того, чтобы избежать различных пространственных конфликтов с мишенью.
3) Слияние (англ. merging). Имеет много общего со стратегией "связывания", но в данном случае объединение фрагментов происходит, если оба фрагмента имеют идентичные перекрывающиеся части, которые играют роль линкера.

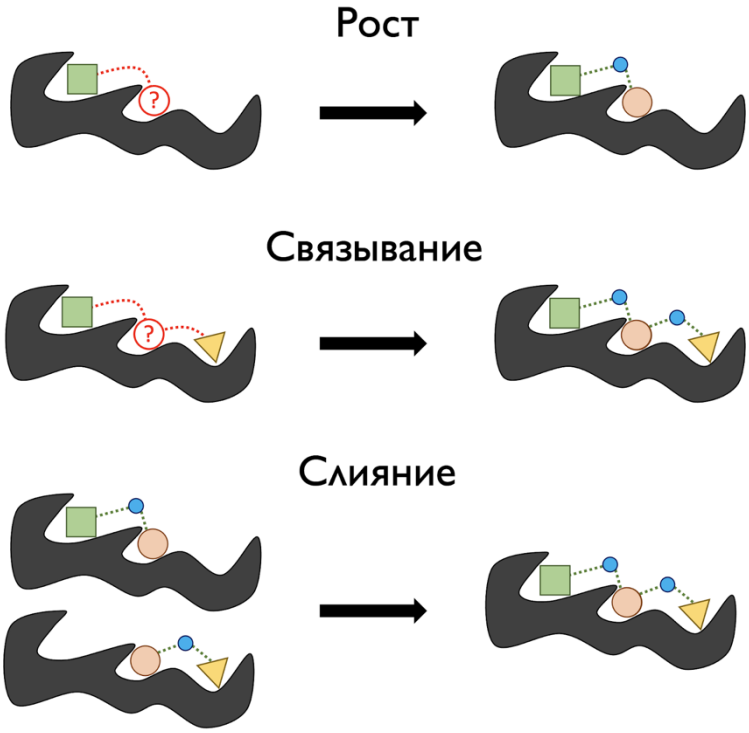


Рисунок 8 - Стратегии для «расширения» соединений во фрагментно-ориентированной разработке лекарств. Для стратегии “Рост” усложнение соединение происходит пошагово, т.е. при усложнении структуры проверяются различные векторы роста и затем отбираются перспективные для последующей оптимизации. Для стратегии “Связывание” выполняется параллельный поиск фрагментов, взаимодействующих с определенным связывающим сайтом, и потом подбирается линкер, который связывает их в одно соединение. Стратегия “Слияние” является аналогичной стратегии “Связывание”, но в данном случае объединение выполняется по одинаковой части, присутствующей в объединяемых фрагментах

Фрагментно-ориентированная стратегия все больше используется в разработке новых ингибиторов киназ. Так, с помощью этой методологии были созданы вемурафениб и эрдафитиниб, для ингибирования BRAF и FGFR, соответственно. Эти лекарства получили одобрение для лечения рака от Управления по санитарному надзору за качеством пищевых продуктов и медикаментов США в 2011 и 2019 годах [94].

1.7 Модели машинного обучения в разработке лекарственных средств

Машинное обучение играет ключевую роль в современном процессе разработки лекарственных средств, значительно ускоряя поиск и дизайн новых соединений [95,96]. Традиционные методы виртуального скрининга широко применяются для идентификации молекул, взаимодействующих с мишенями,

однако они часто страдают от низкой точности предсказаний и высокой доли ложноположительных результатов [97]. В то же время современные модели машинного обучения способны эффективно различать истинные лиганды от ложных, что существенно повышает качество отбора кандидатов для дальнейших исследований [98,99].

Прежде чем модели машинного обучения могут быть использованы для предсказаний, они должны пройти обучение на заранее подготовленном тренировочном наборе данных, состоящем из репрезентативных примеров, которые модель использует для выявления закономерностей. Важным этапом перед обучением является кодирование дескрипторов, то есть преобразование входных данных в числовые представления, описывающих различные параметры в предсказанных моделях. В данном случае могут использоваться различные параметры, описывающие структуру соединений, мишени и их взаимодействие друг с другом. Этот шаг является критически важным, поскольку качество и информативность извлеченных дескрипторов напрямую влияют на эффективность и точность предсказаний, выполняемых моделью.

Одним из примеров успешного применения машинного обучения в разработке лекарств является вычислительный подход vScreenML [31], разработанный в 2020 году для решения задачи классификации потенциально активных соединений от неактивных. Эта модель была использована для поиска новых ингибиторов ацетилхолинэстеразы человека (AChE), где из каталога потенциально доступных химических соединений компании Enamine было извлечено 15 миллионов соединений. Затем их структуры были состыкованы с кристаллической структурой AChE, и 20 000 лучших моделей были отобраны для ранжирования с использованием vScreenML. Для проверки *in vitro* были выбраны 23 соединения, из которых 10 показали IC50 ниже 50 мкМ, а лучшее соединение имело значение K_i 173 нМ. Анализ 2D-структурного сходства между известными ингибиторами AChE и выбранными соединениями показал минимальное сходство, что подтверждает способность модели решать проблемы низкого качества предсказаний.

2 МАТЕРИАЛЫ И МЕТОДЫ

2.1 Разработка новых киназных ингибиторов с использованием фрагментно-ориентированного дизайна

Создан новый вычислительный метод, позволяющий в кратчайшее время генерировать и анализировать многомиллионные коллекции потенциальных ингибиторов исследуемой киназной мишени. Основной задачей данного метода является оптимизация уже известных ингибиторов с низкой эффективностью путем замены прежних периферийных фрагментов в его структуре на новые.

Прямой перебор функциональных групп является неэффективным, поскольку в итоге библиотека может содержать множество структурно несовместимых с мишенью соединений. Например, в исследовании был выбран определенный структурный каркас, который можно оптимизировать, заменяя два боковых фрагмента. В процессе виртуального комбинаторного синтеза использовались два набора реагентов (строительных блоков): один для оптимизации первой функциональной группы (R1), другой — для второй (R2). Каждый набор содержал по 100 строительных блоков, что потенциально дает 100 R1 x 100 R2 = 10 000 новых соединений. Во время виртуального скрининга оказалось, что некоторые полученные соединения имеют слишком большой размер, что привело к их несовместимости с мишенью. После детального анализа выяснилось, что 15 блоков R1 и 20 блоков R2 имели чрезмерно большие функциональные группы, что составило 35% напрасно проверенных соединений из-за неэффективности использованной комбинаторной стратегии.

Более эффективным подходом является предварительная проверка совместимости отдельных строительных блоков с мишенью с последующим отбором лучших из них для комбинаторного синтеза. Поскольку положение шарнир-связывающего скаффолда, как правило, крайне консервативно и уже известно, то позиционирование исследуемых строительных блоков может быть решено путем их слияния с этим скаффолдом. Таким образом, на первом этапе

метод подходит под стратегию "Рост", когда к шарнир-связывающему каркасу присоединяются различные группы и проверяется их эффективность против мишени. На втором этапе метод переходит к стратегии "Слияние", когда лучшие фрагменты объединяются друг с другом через общий скаффолд для получения полноразмерных соединений.

Далее каждый шаг разработанного метода описан в деталях и изображен на рисунке 9А:

1. Анализ взаимодействия ингибитора с киназным доменом. На первом шаге необходимо выбрать соединение, которое будет использовано для оптимизации, выявить центральный скаффолд, который взаимодействует с шарнирной петлей каталитического домена, и определить, какие боковые группы можно использовать для дальнейшей оптимизации (рисунок 9Б).

2. Ретросинтетический анализ. Затем необходимо разработать схему синтеза, которая будет использоваться для создания новых соединений (рисунок 10В). В данной работе в качестве отправных точек для оптимизации были использованы известные ингибиторы киназ, что позволило применить уже опубликованные схемы синтеза данных соединений. Однако для этого также можно использовать модели глубокого обучения для реконструкции путей синтеза [100,101].

3. Создание библиотеки фрагментов. На следующем этапе формируются наборы строительных блоков, которые будут использоваться для создания библиотеки фрагментов (строительных блоков, объединенных с шарнир-связывающим скаффолдом). Для этого схемы синтеза, полученные на шаге 2, были обобщены и преобразованы в текстовое представление SMARTS для поиска подходящих строительных блоков [102] (Приложение 1). Затем, с использованием этих данных, был проведен поиск в базе данных химических соединений PubChem.

4. Скрининг фрагментов. Для каждого фрагмента в библиотеке был создан набор трехмерных конформаций (конформеров) с использованием программного обеспечения OMEGA [103]. Чтобы упростить размещение фрагментов в АТФ-связывающем кармане каталитического домена, все сгенерированные конформеры

фрагментов были структурно выравнены относительно шарнир-связывающего скаффолда исходного соединения. В случае АТФ-конкурентных ингибиторов взаимодействие между соединением и шарнирной петлей домена является очень консервативным, и часто дизайн новых киназных ингибиторов начинается с подбора, подходящего шарнир-связывающего скаффолда, который будет определять положение боковых фрагментов. На следующем этапе каждый конформер фрагментов был объединен со структурой мишени, и полученные лиганд-белковые комплексы подвергались минимизации с помощью программного пакета для биомолекулярного моделирования Rosetta [104,105]. В данном случае это подразумевает оптимизацию структуры модели взаимодействия низкомолекулярного лиганда и белка с целью уменьшения общей энергии, рассчитанной на основе оценочной функции в Rosetta. Этот процесс основан на градиентной оптимизации, при которой положение атомов лиганда и мишени последовательно изменяется для достижения энергетического минимума комплекса. Для каждого фрагмента отбирался конформер с наилучшим показателем взаимодействия и сохранением "родительского" взаимодействия с шарнирным регионом.

5. Слияние лучших фрагментов. На этом этапе осуществляется попарное слияние отобранных фрагментов в более крупные соединения через общий шарнир-связывающий скаффолд. Однако из-за раздельной минимизации фрагментов существует вероятность, что положение этого скаффолда может быть смещено и значительно отличаться между фрагментами. Для решения этой проблемы добавлен фильтр, определяющий среднеквадратичное отклонение (англ. root mean square deviation, RMSD) положения шарнир-связывающего скаффолда в паре фрагментов. Если это значение ниже определенного порога, то эти фрагменты объединяются в новое соединение. В противном случае, если положение общего скаффолда сильно отличается, то такие соединения не объединяются, т.к. структура объединенных фрагментов плохо предсказуема. Важно отметить, что стратегия "Слияние" предполагает аддитивность энергии взаимодействия с мишенью для объединенного соединения, что позволяет представить ее как сумму

энергий взаимодействия отдельных фрагментов. Это позволяет идентифицировать лучшие соединения уже на стадии скрининга фрагментов, без явной проверки объединенных фрагментов. Различия в положении общего скаффолда, а также конформационные изменения в структуре мишени, возникшие в процессе минимизации, могут влиять на точность вычисления аддитивной энергии взаимодействия “объединенного” соединения. Поэтому на последнем этапе каждое соединение, образованное из комбинации фрагментов, снова минимизируется для того, чтобы убедиться, что фрагменты действительно совместимы друг с другом.

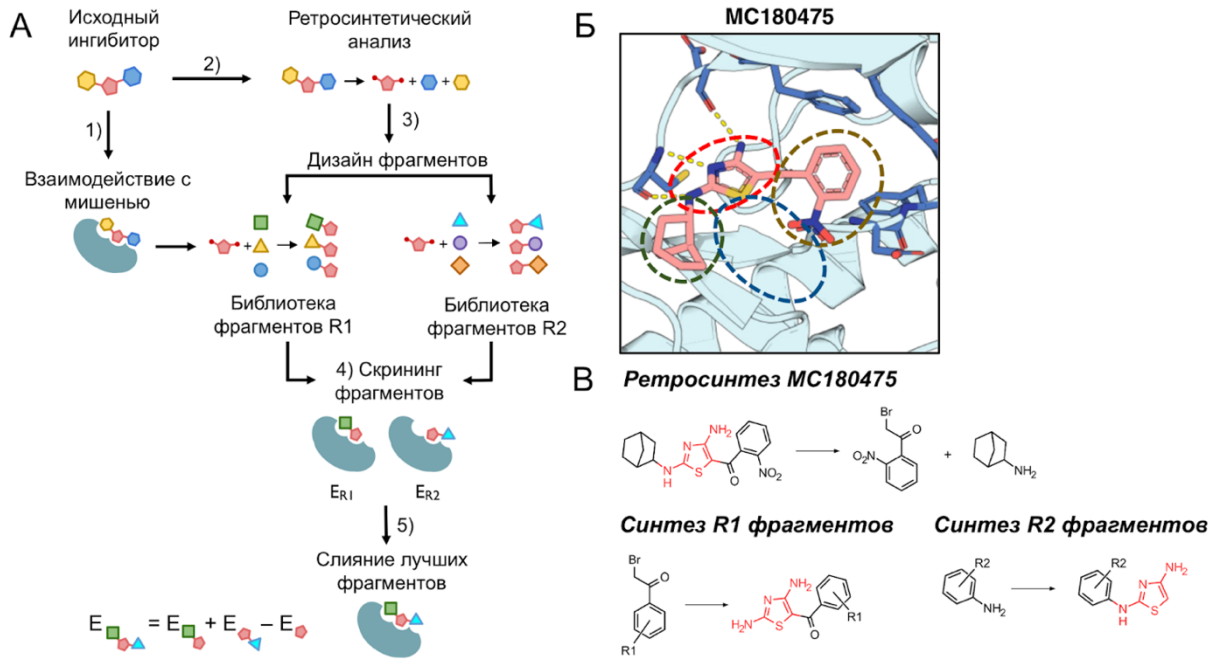


Рисунок 9 - Разработка новых киназных ингибиторов при помощи фрагментно-ориентированного дизайна. А - схематичное представление разработанного метода. Б - расположение фрагментов типичного ингибитора (на примере MC180295) в АТФ-связывающей кармане киназы CDK9 (красный – взаимодействует с шарнирным регионом, коричневый - с задним гидрофобным карманом, зеленый - с открытым карманом, синий - потенциальный фрагмент рибозного кармана). В - декомпозиция структуры MC180295 на отдельные фрагменты (верхняя панель) и шаблоны синтеза новых фрагментов (нижние панели), состоящих из центрального скаффолда (красный) и оптимизируемых периферийных групп (черный)

2.2 Моделирование взаимодействия мишень-PROTAC-лигаза E3

Структура PROTAC состоит из трех компонентов: 1) функциональная часть, связывающая мишень, 2) функциональная часть, взаимодействующая с лигазой E3, и 3) линкер, объединяющий две эти части. При создании PROTAC часто

используются известные низкомолекулярные лиганды, взаимодействующие с исследуемыми белками. Это предполагает, что связывание каждой функциональной части не изменяется при формировании тройного комплекса. Следовательно, основная задача виртуального скрининга заключается в предсказании взаимного расположения мишени, лигазы Е3 и линкера между ними. Эту задачу можно решить двумя способами. Первый способ предполагает, что если конформация линкера известна заранее, то положение тройного комплекса можно предсказать, выравнивая каждый белок относительно фрагмента, ответственного за его связывание [106]. Альтернативный подход заключается в построении тройного комплекса на основе известного взаимодействия мишени и лигазы Е3 с последующим предсказанием конформации линкера.

Данная работа была сосредоточена на втором подходе, так как он более перспективен для оптимизации структуры линкера при разработке новых молекул PROTAC. Суть метода заключается в создании всех возможных моделей взаимодействия между мишенью и лигазой Е3, а затем применении различных структурных и энергетических фильтров для отбора структур, совместимых с данным линкером.

Далее каждый шаг метода описан в деталях и изображен на рисунке 10:

1. Выбор объектов для моделирования. На начальном этапе необходимо выбрать двумерную структуру PROTAC, функциональные части которой подходят для взаимодействия с мишенью и лигазой Е3.

2. Предсказание взаимодействия между мишенью и лигазой Е3. Полученные трехмерные структуры используются для предсказания всех возможных способов взаимодействия между мишенью и лигазой Е3. При моделировании тройного комплекса белки должны быть расположены так, чтобы их функциональные лиганды были близки друг к другу, т.к. в ином случае линкер не сможет объединить их в одну структуру. Предсказание выполняется при помощи белок-белкового докинга, реализованного в программном пакете для биомолекулярного моделирования PyRosetta путем изменения положения одного белка по отношению к другому с сохранением внутреннего положения лигандов в связывающем сайте.

В ходе работы генерируется около 50 000 вариантов взаимодействия комплекса мишень-лигаза E3, что дает большую коллекцию моделей с разнообразными способами связывания. Затем полученные модели ранжируются по энергии взаимодействия между белками и 10% лучших из них используются для построения будущего тройного комплекса. Также вычисляется медианное значение энергий взаимодействия относительно всего набора сгенерированных моделей, которое будет использовано для определения совместимых моделей тройного комплекса мишень-PROTAC-лигаза E3.

3. Создание возможных конформаций линкера. Параллельно с докингом белков необходимо сгенерировать все возможные трехмерные низкоэнергетические конформации линкера, который будет использован для объединения лигандов мишени и лигазы E3 в одну структуру PROTAC. Для этого конформеры линкера выравниваются относительно функциональных лигандов при помощи небольших фрагментов этих же лигандов, присоединенных к концам линкера (так называемые “концевые заглушки”). Для генерации трехмерных конформаций используется программное обеспечение OMEGA [107,108]. Затем полученные конформеры отбираются на основе структурного разнообразия для обеспечения равномерного распределения возможных конфигураций линкера. В ходе работы генерируется около 1000 конформаций для каждого линкера. Таким образом, на этом шаге происходит моделирование гибкости молекулы PROTAC, которая необходима для образования тройного комплекса.

4. Построение всех возможных моделей тройного комплекса. На следующем этапе выполняется комбинирование всех предсказанных комплексов мишень-лигаза E3 и конформеров линкера для построения возможных моделей тройного комплекса, т.е. для каждого комплекса белков из шага 2 и линкера из шага 3 проверяется совместимость и возможность формирования полноразмерной структуры PROTAC. Этот шаг является ключевым для разработки новых PROTAC, поскольку он включает детальное исследование того, как линкер в сочетании с конкретным расположением двух белков приводит к образованию тройного комплекса. Полученные модели комплекса отфильтровываются по RMSD

положения “концевых заглушек” линкера относительно атомов лигандов в белках. Как правило, значения RMSD менее 0.4 Å гарантируют правильное формирование модели PROTAC. Затем для каждой возможной пары модели взаимодействия белков и конформации линкера строится полная модель тройного комплекса, при котором структура выровненного линкера объединяется с функциональными лигандами. Полученные модели подвергаются минимизации энергии с использованием протокола в PyRosetta для исправления возможных пространственных конфликтов между линкером и белками тройного комплекса.

5. Анализ образованных тройных моделей. Для каждой модели, полученной на последнем шаге, вычисляется энергия взаимодействия между всеми компонентами тройного комплекса мишень-PROTAC-лигаза Е3. Минимизированные структуры, у которых энергия взаимодействия лучше порогового значения (вычисленного на шаге 2), используются для дальнейшего анализа. Этот фильтр предназначен для выявления моделей тройного комплекса, у которых энергия взаимодействия лучше медианного значения для всех возможных положений белков. Затем вычисляется доля полностью совместимых комплексов (ДСК), то есть отношение количества совместимых тройных комплексов (# СК) к общему числу предсказанных белковых комплексов (# МЕ3) и конформаций линкера (# Л). Эта оценка помогает понять, насколько хорошо структурные особенности PROTAC сочетаются с формированием белковых комплексов мишень-лигаза Е3.

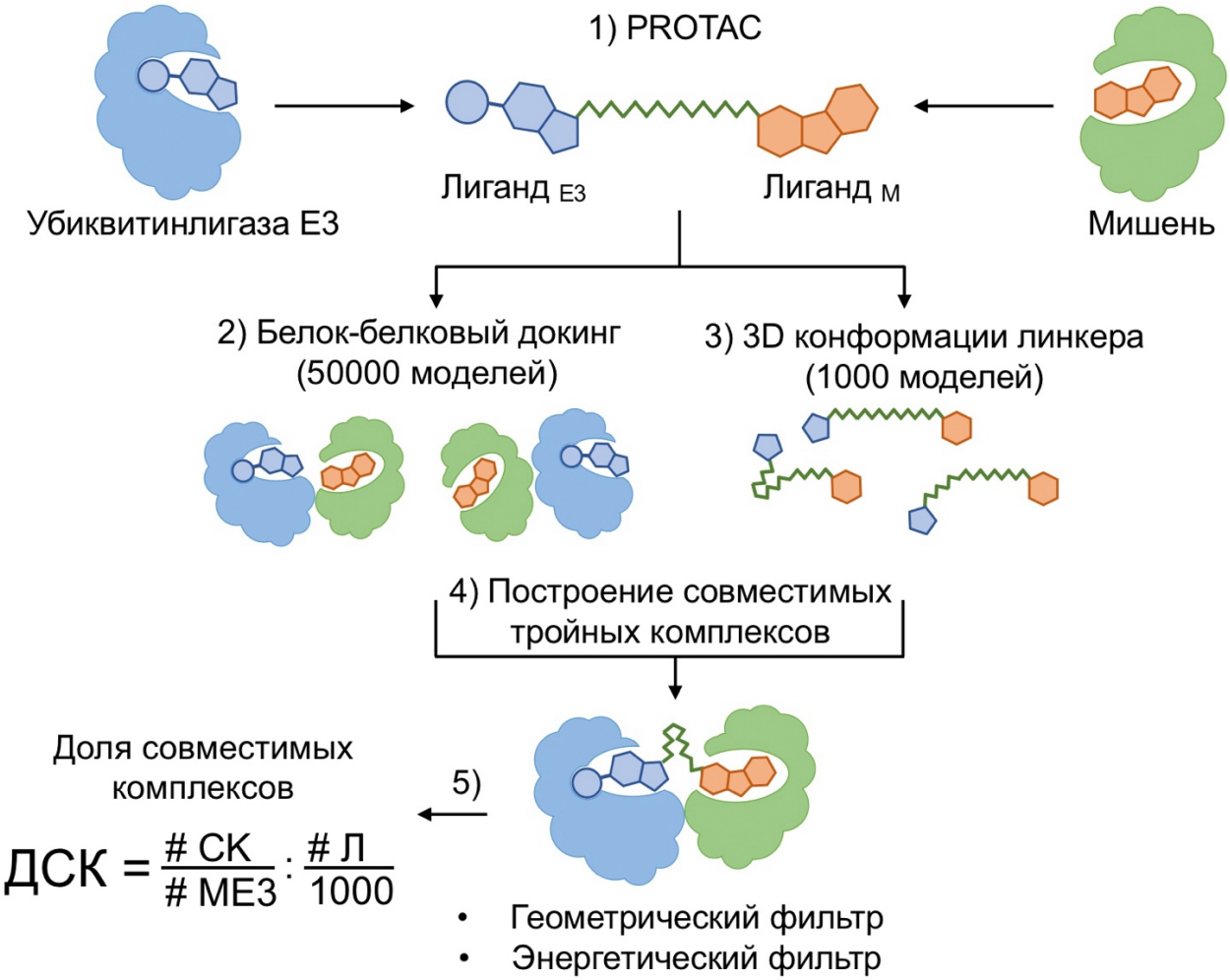


Рисунок 10 - Схематичное представление метода для предсказания моделей тройного комплекса мишень-PROTAC-лигаза E3: 1) выбор подходящей структуры мишени и лигазы E3, связанных с лигандами, которые будут использованы для построения PROTAC, 2) белок-белковый докинг мишени и лигазы E3, 3) генерация всех возможных трехмерных конформаций линкера, 4) выравнивание конформеров линкера относительно лигандов, связанных с белками, и слияние их в одну структуру, 5) вычисление теоретической эффективности PROTAC через долю полностью совместимых комплексов (ДСК), которая определяется как отношение количества образованных совместимых комплексов (# СК) к числу сгенерированных белок-белковых моделей (# МЕ3), и нормализованное на реально использованное количество конформеров линкера (# Л поделенное на тысячу)

2.3 Поиск киназных ингибиторов традиционными методами виртуального скрининга

Несмотря на развитие специализированных методов дизайна ингибиторов, классические методы виртуального скрининга по-прежнему широко используются для поиска новых соединений. В рамках этих методов выделяются три основных

этапа: 1) компиляция набора низкомолекулярных соединений для скрининга, 2) предсказание наиболее вероятного положения каждого соединения в комплексе со структурой мишени, 3) ранжирование полученных моделей с использованием оценочной функции и отбор лучших из них. В данном случае был выбран поиск по молекулярному подобию в качестве метода виртуального скрининга, так как он имеет более высокую точность по сравнению с молекулярным докингом. Это объясняется тем, что метод поиска по молекулярному подобию использует информацию об известных работающих ингибиторах для точного позиционирования исследуемых соединений в связывающем кармане, что повышает вероятность успешного предсказания активных соединений.

Каждый этап метода описан ниже и изображен на рисунке 11:

1. Компиляция библиотеки соединений. Для этого из каталога Enamine была извлечена специализированная библиотека соединений для ингибирования протеинкиназ, содержащая соединения с аминотиазолом, потенциально способным взаимодействовать с шарнирной петлей каталитического домена.

2. Генерация конформеров. Соединения в каталогах обычно представлены в двухмерном формате SMILES [109]. Для следующего шага необходимо создать набор трехмерных конформеров с помощью программного обеспечения OMEGA [103,107].

3. Выравнивание конформеров. Механизм взаимодействия АТФ-конкурентных ингибиторов с каталитическим доменом киназы является консервативным, поскольку он воспроизводит взаимодействие с аденином АТФ в области шарнирной петли. Таким образом, изучаемые соединения могут быть расположены в связывающем кармане путем структурного выравнивания относительно известных ингибиторов киназ с использованием программного обеспечения ROCS [108].

4. Минимизация и оценка энергии взаимодействия. В этом этапе выполняется оптимизация расположения выровненных соединений в комплексе с целевой структурой с помощью PyRosetta, после чего вычисляется энергия взаимодействия между лигандом и рецептором для выбора наиболее перспективных кандидатов.

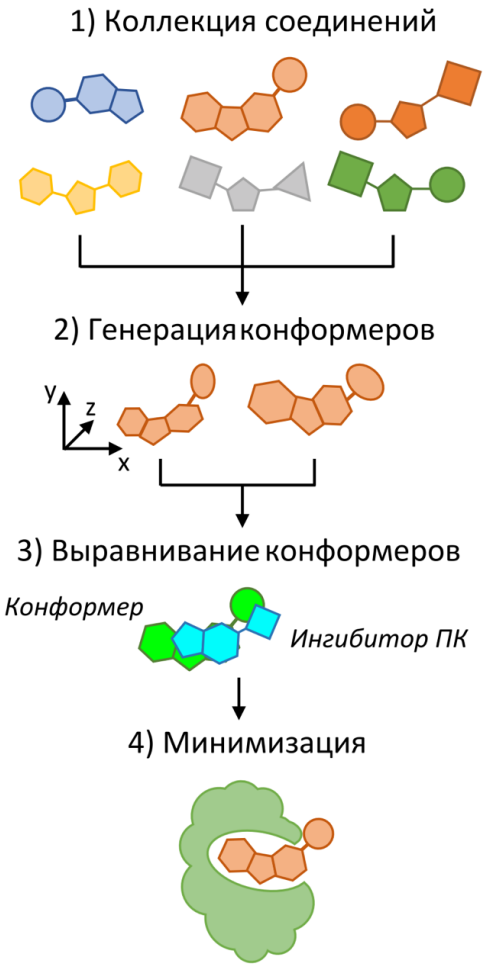
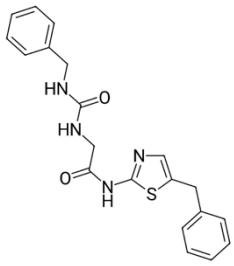
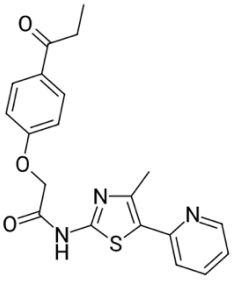
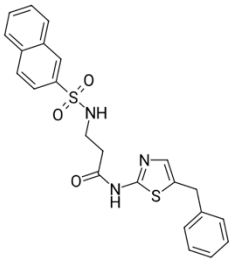
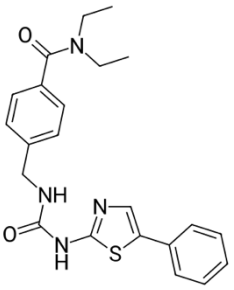
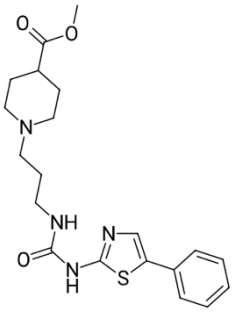
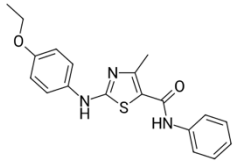
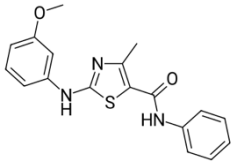
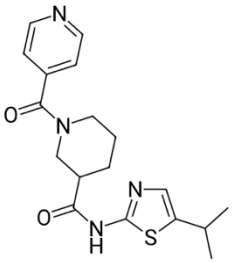
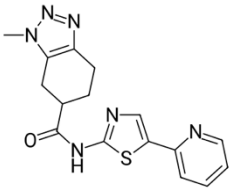
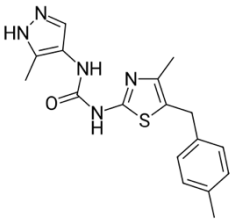
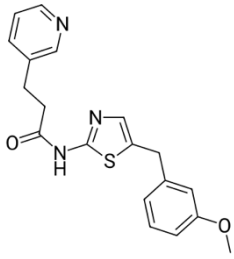
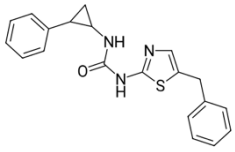
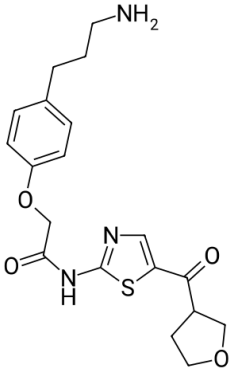
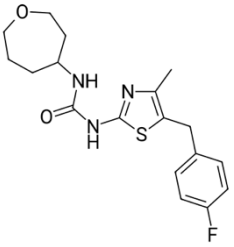
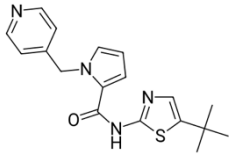


Рисунок 11 - Схематичное представление работы метода молекулярного подобия: 1) компиляция библиотеки соединений, 2) генерация трехмерных конформеров для отобранных соединений, 3) выравнивание конформеров относительно структуры известных ингибиторов и 4) минимизация в комплексе с мишенью и с последующие ранжирование с использованием энергии взаимодействия

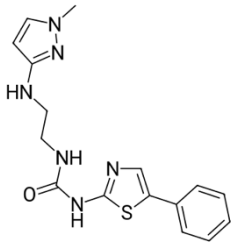
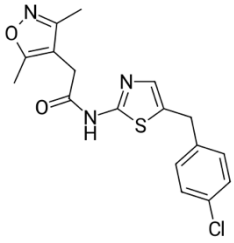
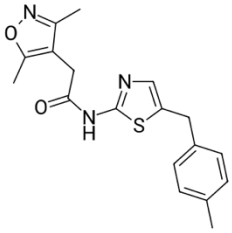
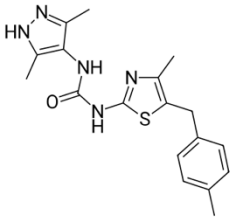
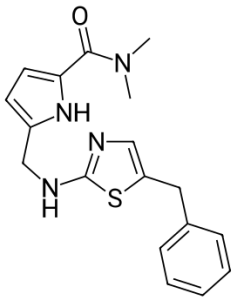
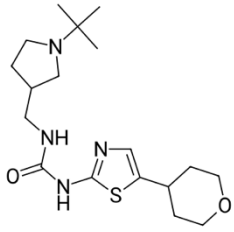
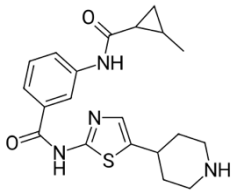
2.4 Синтез киназных ингибиторов и определение их активности

Отобранные 22 соединения (таблица 1) в ходе виртуального скрининга с использованием традиционных методов было были синтезированы компанией Enamine (США).

Таблица 1 - Соединения отобранные для экспериментальной проверки

				
Z169536760	Z238190978	Z212912350	Z411664372	Z859023978
				
Z48861625	Z48853635	Z1265432563	Z2177403824	Z1995613226
				
Z238156308	Z826248620	Z4452246830	Z2294424157	Z1275191743

Продолжение таблицы 1

 Z1911365952	 Z67764668	 Z70864235	 Z1995612474	 Z1673531755
 Z2624229394	 Z2378678900			

Эффективность отобранных ингибиторов была проверена при помощи радиометрического анализа HotSpot™ Kinase Assay компании ReactionBiology (США). Этот метод позволяет точно измерить ингибирующую потенциал соединений за счет прямого мониторинга фосфорилирования субстрата киназой. В ходе эксперимента тестируемые соединения были инкубированы с исследуемой киназой, субстратом, и меченным ^{33}P-АТФ. После завершения инкубации измеряли количество меченого радиоизотопа, который был перенесен на субстрат мишени, что позволяло оценить эффективность ингибирования (рисунок 12).

Эксперимент был проведен в одном повторе при концентрациях ингибитора от 100 мкМ до 0,005 мкМ с шагом разведения в 3 раза. Реакционная смесь содержала 10 мкМ АТФ и буфер: 20 мМ HEPES (pH=7,5), 10 мМ $MgCl_2$, 1 мМ EGTA, 0,01% Brij35, 0,02 мг/мл BSA, 0,1 мМ Na_3VO_4, 2 мМ DTT, 1% DMSO. Необходимые кофакторы добавлялись индивидуально для каждой киназы.

Для определения концентрации полумаксимального ингибирования (IC50) был проведен анализ зависимости остаточной активности киназы от концентрации ингибитора с применением нелинейной регрессии, основанной на четырехпараметрической логистической модели, реализованной в программном обеспечении GraphPad.

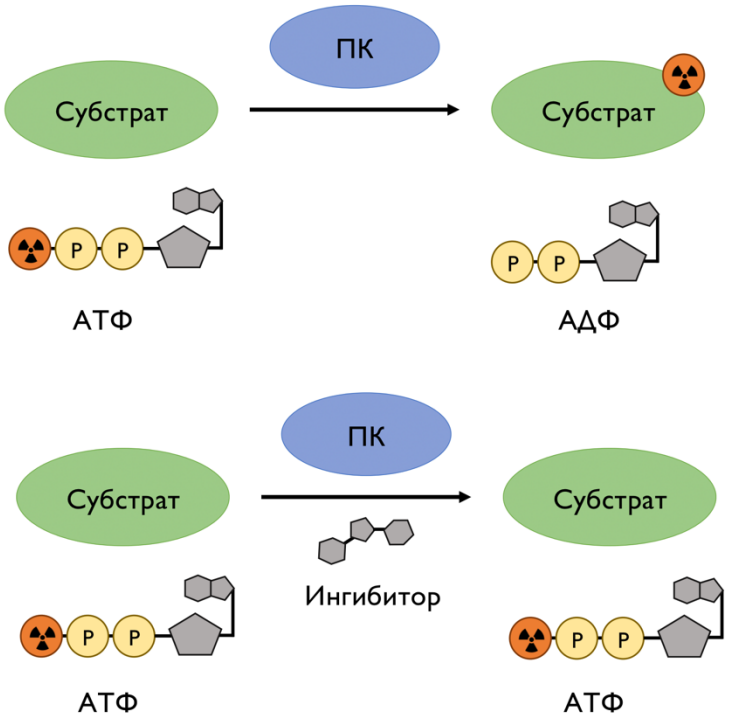


Рисунок 12 - Измерение активности киназных ингибиторов. В ходе эксперимента определяется количество меченого субстрата, образовавшегося в результате переноса меченого фосфата с ^{33}P-АТФ на субстрат, катализируемого киназой. При добавлении ингибитора активность киназы уменьшается, что приводит к снижению количества меченого субстрата

2.5 Подготовка тренировочного набора и обучение vScreenML

vScreenML — это модель машинного обучения, предназначенная для минимизации количества ложноположительных результатов, часто возникающих при использовании традиционных методов виртуального скрининга. В данном случае произвольный метод виртуального скрининга выполняет роль "генератора" гипотез, формируя множество моделей взаимодействия лигандов с мишенью, в то время как vScreenML служит "дискриминатором", стремясь сократить количество ложноположительных моделей. Таким образом, это сочетание технологий

машинного обучения с традиционными методами виртуального скрининга значительно улучшает точность и эффективность поиска новых лекарственных соединений.

Основой vScreenML выступает обученная модель машинного обучения, использующая подготовленный тренировочный набор данных, включающий как истинно активные соединения, так и ложные приманки. Во время обучения модель анализирует числовые дескрипторы, характеризующие взаимодействия между лигандом и мишенью, что позволяет эффективно отличать активные соединения от неактивных. Подготовка тренировочного набора и обучение модели описаны ниже и проиллюстрированы на рисунке 13A:

1. Подготовка обучающего набора. Набор активных лигандов (положительный класс) был получен из PDB с использованием следующих критериев: 1) метод определения структуры — рентгеновская дифракция, 2) разрешение менее 2.5 Å, 3) тип полимерного объекта — белок. Полученные структуры были отфильтрованы по физико-химическим и структурным характеристикам, представленным в Приложении 4. Затем каждая отфильтрованная структура проходила обработку с использованием Protoss для предсказания положения атомов водорода и протонирования аминокислотных остатков. Если кристаллическая структура имела несколько копий, то она была разделена на отдельные экземпляры для обогащения обучающего набора. Затем обработанные структуры были минимизированы в PyRosetta. Набор ложных приманок (отрицательный класс) был получен с помощью веб-сервера Useful Decoys Enhanced (DUD-E), который для каждого активного лиганда сгенерировал 50 приманок со сходными физико-химическими, но разными топологическими свойствами. Затем полученные приманки были отфильтрованы аналогично активным соединениями, и для каждого из них был сгенерирован набор трехмерных конформаций с помощью OMEGA. Затем эти конформации были выровнены относительно структуры соответствующего активного лиганда при помощи ROCS, и в результате были отобраны три наиболее похожих приманки по показателю TanimotoCombo. Отобранные приманки были объединены со

структурой белка соответствующего активного лиганда, состояние протонирования было уточнено с помощью Protoss, и структуры были минимизированы в PyRosetta.

2. Кодирование дескрипторов обучающего набора. Полученный набор был охарактеризован с различных аспектов для комплексной оценки взаимодействий между целевым белком и лигандом. Оценка включала следующие аспекты: 1) энергетические, для оценки взаимодействий белка и лиганда при помощи программы Rosetta, 2) структурные, определение доступности связывающих карманов осуществлялось с помощью PocketDruggability, 3) физико-химические свойства анализировались с помощью RDKit, 4) различные типы белково-лигандных взаимодействий с помощью программ Rosetta, Binana, RF-score и LUNA. Эти дескрипторы позволили всесторонне охватить все аспекты взаимодействия между целевым белком и лигандом, обеспечивая глубокое понимание их взаимодействий.

Для выбора наиболее информативных дескрипторов применялся Sequential Backward Floating Selection (SBFS) из программного пакета mlxtend [110]. Этот метод последовательно анализирует все дескрипторы, удаляя на каждом шаге наименее значимые, пока не будет достигнут оптимальный набор. Качество предсказаний на каждом этапе оценивалось с помощью k-fold кросс-валидации (рисунок 13Б), при которой обучающая выборка разделяется на k непересекающихся частей. Процесс обучения повторяется k раз, каждый раз используя k−1 частей для обучения и оставшуюся часть для тестирования. Разбиение данных основывалось на кластеризации аминокислотных последовательностей тренировочного набора с помощью MMSeq2 [111]. В качестве метрики для оценки качества классификации применялся коэффициент корреляции Мэтьюса, реализованный в библиотеке scikit-learn.

3. Обучение модели. Полученный тренировочный набор был использован для обучения модели машинного обучения Extreme Gradient Boosting для бинарной классификации. Подбор гиперпараметров модели, таких как n_estimators, max_depth и learning_rate, был выполнен с помощью Optuna [112].

4. Предсказание класса. Для предсказания класса предсказанной модели необходимо выполнить следующие шаги (рисунок 13В): 1) минимизация для того, чтобы значения дескрипторов принадлежали распределению значений в тренировочном наборе, 2) вычисление дескрипторов, 3) предсказание, где vScreenML выдает вероятность от 0 (неактивные лиганд) до 1 (активный лиганд).

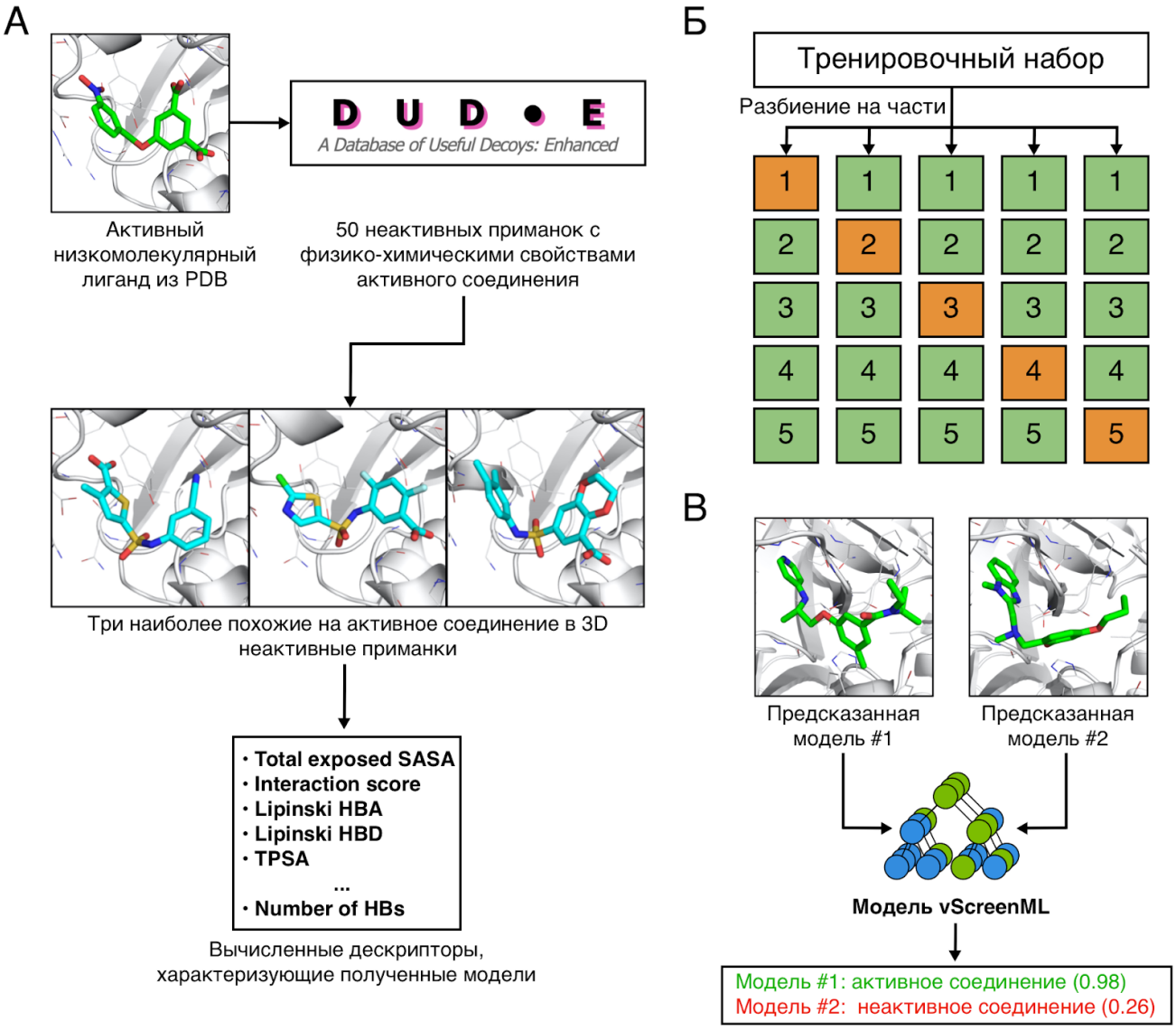


Рисунок 13 - Схема подготовки обучающего набора, тренировки модели и предсказания класса. А - тренировочный набор состоит из активных лигандов из PDB и неактивных приманок из DUD-E, Б - пример k-fold валидации при k = 5, где тренировочный набор делится на пять частей: четыре используются для обучения модели (зеленый), одна для тестирования (оранжевый), В - предсказание класса для двух структур: 1) первый пример имеет вероятность 0,98 (т.е. высокий шанс активности соединения), второй - 0,26 (т.е. низкий шанс активности)

3 РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

3.1 Фрагментно-ориентированная разработка ингибиторов ПК

Для оценки эффективности разработанного метода были выбраны шесть известных ингибиторов киназ в качестве исходных соединений для дальнейшей оптимизации. Каждое из них несет определенный мотив взаимодействия с шарнирной областью каталитического домена ПК, как показано на рисунке 14.

1. MC180295 (PDB-код: 6W9E) - селективный ингибитор CDK9, основанный на диаминотиазольном скаффолде [113,114]. Имеет две боковые функциональные группы для оптимизации.
2. BDBM50091276 (PDB-код: 4RVK) - ингибитор CHK1, построенный на скаффолде 1,7-диазакарбазола [115]. Имеет две боковые функциональные группы для оптимизации.
3. BDBM7773 (PDB-код: 1KE7) - ингибитор CDK2 с оксиндольным центральным скаффолдом [116]. Имеет две боковые функциональные группы для оптимизации.
4. Дакомитиниб (PDB-код: 4I24) - ковалентный ингибитор $EGFR^{T790M}$, построенный на скаффолде хиназолина [117]. Имеет две боковые функциональные группы для оптимизации.
5. CHIR-124 (PDB-код: 2GDO) - ингибитор CHK1, построенный на хинолиноновом скаффолде [118]. Имеет три боковые функциональные группы для оптимизации.
6. BDBM50246162 (PDB-код: 3EQR), ингибитор ACK1, построенный на скаффолде пиразолопиримидин-3,6-диамина [119]. Имеет три боковые функциональные группы для оптимизации.

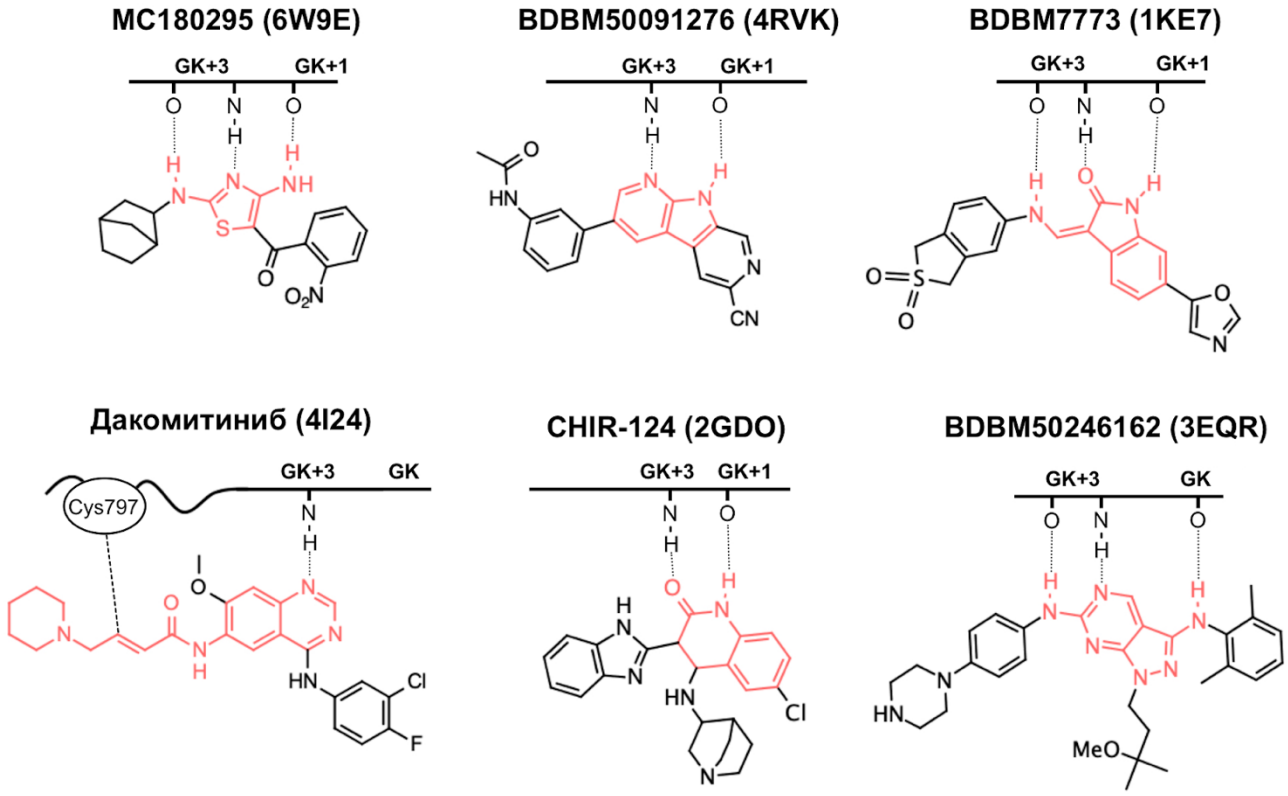


Рисунок 14 - Киназные ингибиторы, использованные для оптимизации с помощью разработанного фрагментно-ориентированного метода. Каждое соединение имеет центральную часть (красный), которая взаимодействует с шарнирной петлей (нумерация начинается от аминокислоты-привратника (англ. gatekeeper, GK)). Ковалентное взаимодействие дакомитиниба отмечено пунктирной линией

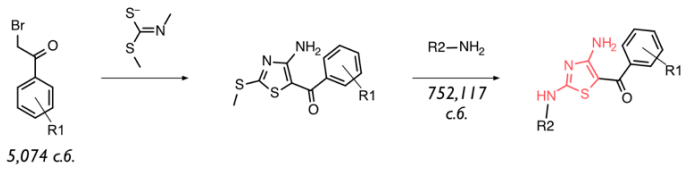
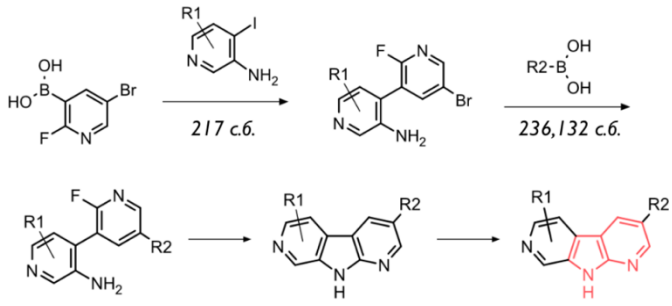
3.1.1 Создание библиотек фрагментов

Для создания виртуальных библиотек фрагментов были использованы опубликованные в литературе схемы синтеза каждого из шести исследуемых ингибиторов. Структуры подходящих для реакции строительных блоков были извлечены из базы данных PubChem и применены для виртуального синтеза. Из-за комбинаторного характера процесса получения новых соединений итоговые библиотеки оказались чрезвычайно большими. Например, для библиотеки, созданной на основе MC180295, было найдено около 5 000 2-бромацетофенонов и 750 000 первичных аминов, что привело к образованию около 4 миллиардов возможных соединений после их попарного сочетания.

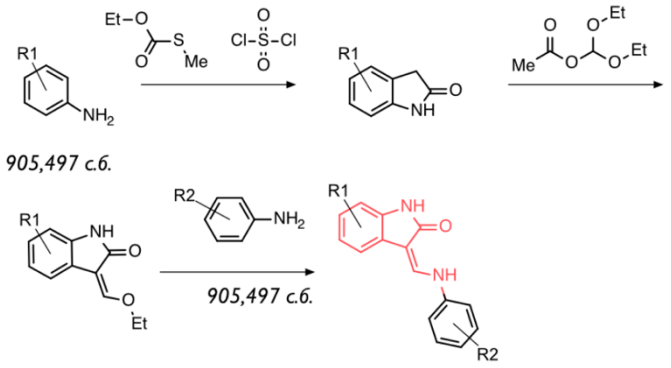
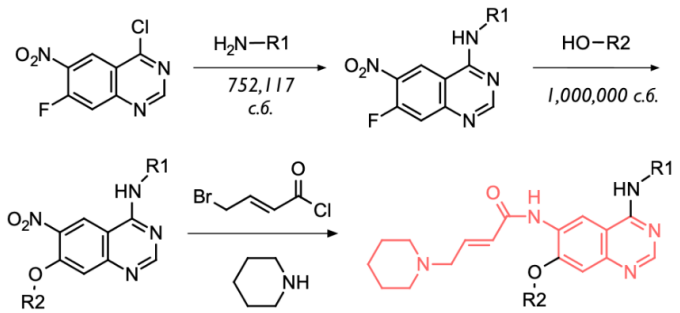
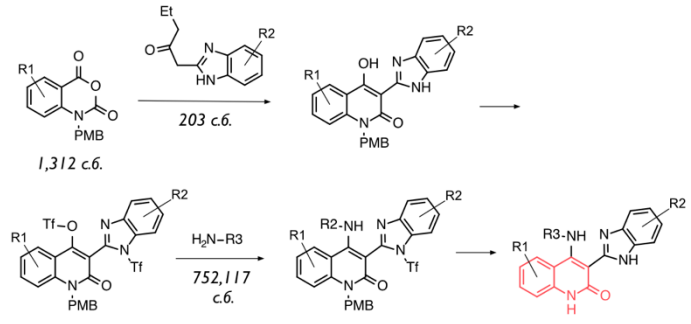
Важно отметить, что размер библиотек сильно зависит от сложности синтеза и простоты используемых строительных блоков. В случаях, когда схема синтеза включает первичные амины, бороновые кислоты или ариламины (как в синтезах MC180295, BDBM50091276 и BDBM7773), количество возможных строительных блоков становится особенно значительным.

Таблица 2 содержит информацию о примененных шаблонах синтеза, количестве строительных блоков, использованных для виртуального синтеза, и размере библиотек, которые могут быть получены путем комбинаторного слияния. Также была оценена эффективность работы фрагментного метода по сравнению с традиционным скринингом комбинаторных библиотек, которая вычислялась как отношение произведения всех строительных блоков к сумме всех блоков, использованных для создания библиотеки.

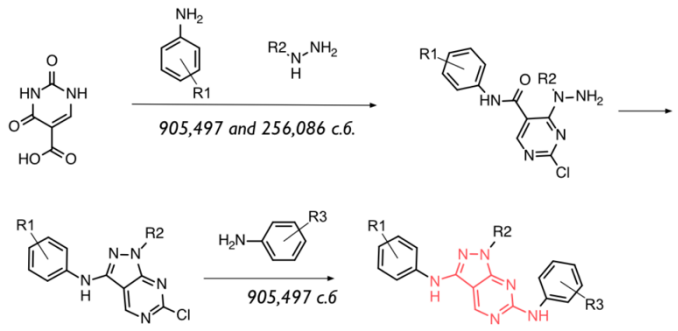
Таблица 2 - Библиотеки соединений, созданные при помощи методов виртуальной комбинаторной химии

Обобщенная схема синтеза	**Размер комбинаторной библиотеки**	**Эффективность нового метода**
MC180295 (CDK9)		
	5 074 x 752 117 = 3,8 млрд доступных соединений	5 040x
BDBM50091276 (CHK1)		
	217 x 236 132 = 51 млн доступных соединений	217x

Продолжение таблицы 2

Обобщенная схема синтеза	Размер комбинаторной библиотеки	Эффективность нового метода
BDBM7773 (CDK2)		
	905 497 x 905 497 = 819 млрд доступных соединений	452 749x
Дакомитиниб (EGFR^{T790M})		
	752 117 x 1 000 000 = 752 млрд доступных соединений	429 262x
CHIR-124 (CHK1)		
	1 312 x 203 x 752 117 = 200 млрд доступных соединений	265 801x

Продолжение таблицы 2

Обобщенная схема синтеза	Размер комбинаторной библиотеки	Эффективность нового метода
BDBM50246162 (ACK1)		
	905 497 x 256 086 x 905 497 = 2 x 10^{17} доступных соединений	10^{11} x

3.1.2 Скрининг фрагментов и последующее слияние

Перед масштабным скринингом сгенерированных фрагментов была проведена предварительная проверка метода на способность восстанавливать структуру исходного соединения. Для каждого фрагмента, представляющего структурную часть этого соединения, был сгенерирован набор трехмерных конформеров, которые затем выравнивались относительно шарнир-связывающего скаффолда "родительского" соединения. После этого каждый конформер объединялся со структурой мишени, и образованный комплекс подвергался минимизации для уточнения положения лиганда в связывающем сайте. Затем модели фрагментов с наилучшей энергией взаимодействия использовались для последующего попарного объединения.

В большинстве случаев фрагменты принимали конформации, близкие к тем, что наблюдаются в кристаллических структурах исходных ингибиторов. После слияния отдельных фрагментов и последующей минимизации объединенные модели сохраняли значительное сходство с исходным положением и конформацией оригинальных ингибиторов, что подтверждается значениями RMSD: MC180295 – 2,74 Å, BDBM50091276 – 2,60 Å, BDBM7773 – 2,51 Å, дакомитиниб – 0,79 Å, CHIR-124 – 2,28 Å, BDBM50246162 – 1,13 Å. Тем не менее, в ходе анализа предсказанных моделей было выявлено несколько отклонений,

связанных с изменением положения отдельных групп исследуемых соединений. У BDBM50091276, BDBM7773 и BDBM50246162 боковая группа, расположенная в переднем кармане, является перевернутой, в то время как у MC180295 некорректную конформацию приобрела группа, находящаяся в заднем кармане. В случае CHIR-124, несмотря на высокую степень сходства конформаций (RMSD 0,29 Å), лиганд оказался существенно смещенным относительно исходной структуры. Эти отклонения, вероятно, возникают из-за ограничений, присущих моделированию в PyRosetta, которое игнорирует конформационную энтропию лиганда и не учитывает присутствие растворитель явно. При этом, несмотря на небольшое снижение точности предсказаний, подобное упрощение позволяет существенно ускорить отбор потенциальных комбинаций фрагментов для экспериментальной проверки.

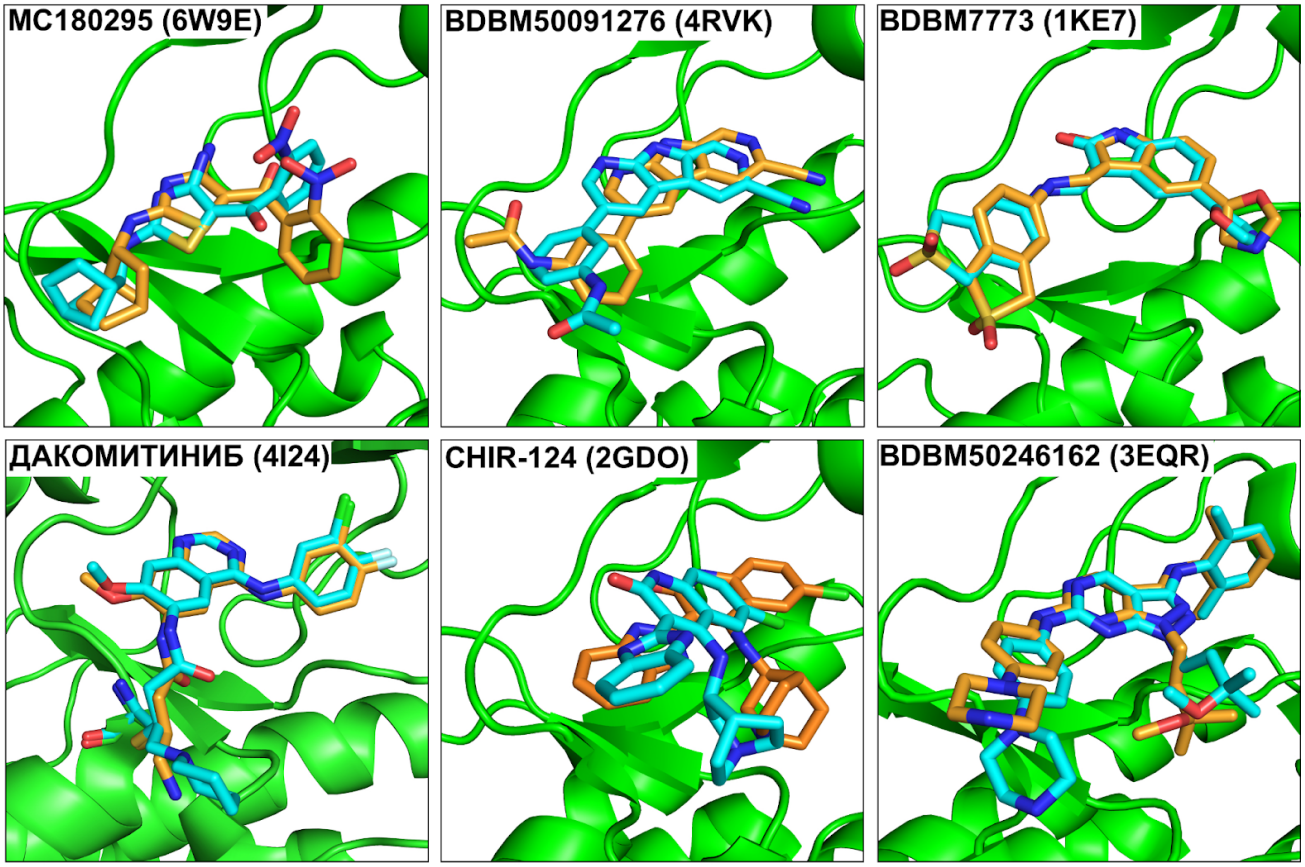


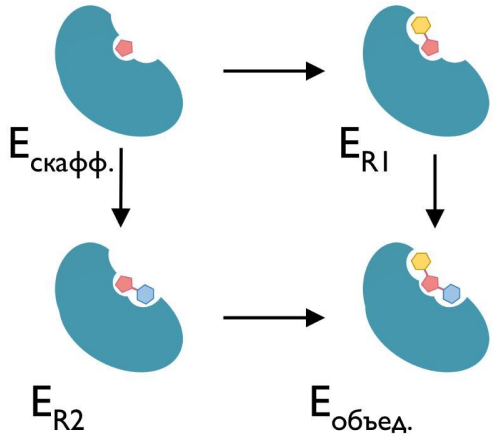
Рисунок 15 - Моделирование положения известных ингибиторов в связывающем сайте каталитического домена разработанным методом. Отдельные фрагменты были минимизированы и затем объединены в одну структуру. Затем положение объединенной пары после минимизации (голубой) было сопоставлено с кристаллической структурой исходного ингибитора (оранжевый)

3.1.3 Попарная аддитивность энергий взаимодействия

Затем разработанный фрагментно-ориентированный метод был применен к созданным библиотекам фрагментов для проведения оптимизации функциональных групп, присоединенных к центральному шарнир-связывающему скаффолду исходных соединений. Одним из преимуществ применения стратегии слияния в фрагментно-ориентированной разработке является аддитивность энергий взаимодействия фрагментов, составляющих полноразмерную структуру соединения. Это означает, что можно предсказать энергию взаимодействия всего соединения, используя только значения составляющих его фрагментов. Однако точность такого предсказания сильно зависит от множества факторов. Например, известно, что положение общего шарнир-связывающего скаффолда в фрагментах значительно влияет на точность аддитивной энергии [120–123], как и пространственное расположение функциональных групп, которые могут быть структурно несовместимы друг с другом. Тем не менее, в случае используемых исходных соединений функциональные группы присоединяются с противоположных сторон шарнир-связывающего каркаса, поэтому маловероятно, что они будут оказывать значительное влияние друг на друга. Основываясь на этом наблюдении, было предположено, что точность предсказания энергии взаимодействия объединенных фрагментов зависит только от положения общего скаффолда.

Для проверки был проведен сравнительный эксперимент, в котором точность предсказанной аддитивной энергии взаимодействия фрагментов оценивалась относительно явно вычисленной энергии полноразмерной структуры ингибитора после слияния фрагментов и минимизации. Поскольку проведение такого эксперимента для всех пар фрагментов в библиотеках требует огромных вычислительных ресурсов, были подготовлены репрезентативные выборки фрагментов для каждой библиотеки (Приложение 2). Затем эти фрагменты были попарно объединены с друг другом при помощи центрального скаффолда, и аддитивная энергия фрагментов была сравнена с энергией полноразмерного соединения.

В данном случае аддитивная энергия была вычислена с использованием стандартного двойного мутантного цикла, как показано на рисунке 16 [124]. Его суть заключается в том, что предсказанная энергия полноразмерного соединения ($E_{объед.}$) может быть представлена как сумма энергий составляющих его фрагментов (E_{R1}, E_{R2}) за вычетом энергии шарнир-связывающего каркаса ($E_{скафф.}$).



$$E_{объед.} = E_{R1} + E_{R2} - E_{скафф.}$$

Рисунок 16 - Двойной мутантный химический цикл для оценки энергии взаимодействия, полученной из фрагментов. В соответствии с идеей аддитивности фрагментов, энергия взаимодействия объединенного соединения ($E_{объед.}$) определяется суммой энергий взаимодействия фрагментов (E_{R1}, E_{R2}) за вычетом энергии взаимодействия шарнир-связывающего скаффолда ($E_{скафф.}$), который присутствует в обоих фрагментах

Результаты данного эксперимента показали, что для MC180295, BDBM50091276, BDBM7773 и дакомитиниба предсказанная аддитивная энергия взаимодействия из отдельных фрагментов достаточно близка к явно вычисленной энергии взаимодействия объединенных фрагментов (рисунок 17). Это особенно верно в тех случаях, когда положение общего шарнир-связывающего скаффолда двух объединяемых молекул практически совпадает. Коэффициент корреляции Спирмена составил 0,76 ($p < 2.2 \times 10^{-16}$) для MC180295, 0,64 ($p < 2.2 \times 10^{-16}$) для BDBM50091276, 0,51 ($p < 2.2 \times 10^{-16}$) для BDBM7773 и 0,7 ($p < 2.2 \times 10^{-16}$) для дакомитиниба. Однако при оптимизации трех групп одновременно точность предсказаний аддитивной энергии снижается, что было продемонстрировано на примере двух библиотек — CHIR-124 и BDBM50246162, где коэффициенты корреляции составили 0,26 ($p < 2.2 \times 10^{-16}$) и 0,1 ($p = 9.9 \times 10^{-8}$) соответственно

(рисунок 17). Значения RMSD для фрагментов первой библиотеки оказались низкими, что позволяет предположить, что некоторые функциональные группы расположены слишком близко друг к другу, приводя к структурной несовместимости. В случае второй библиотеки лишь незначительное количество комбинаций фрагментов имеют схожее положение общего скаффолда, что затрудняет нахождение сочетаний трёх фрагментов с похожим расположением скаффолда. Это, в свою очередь, подразумевает, что подвижность шарнир-связывающего скаффолда необходимо учитывать при выборе исходного соединения и последующей оптимизации.

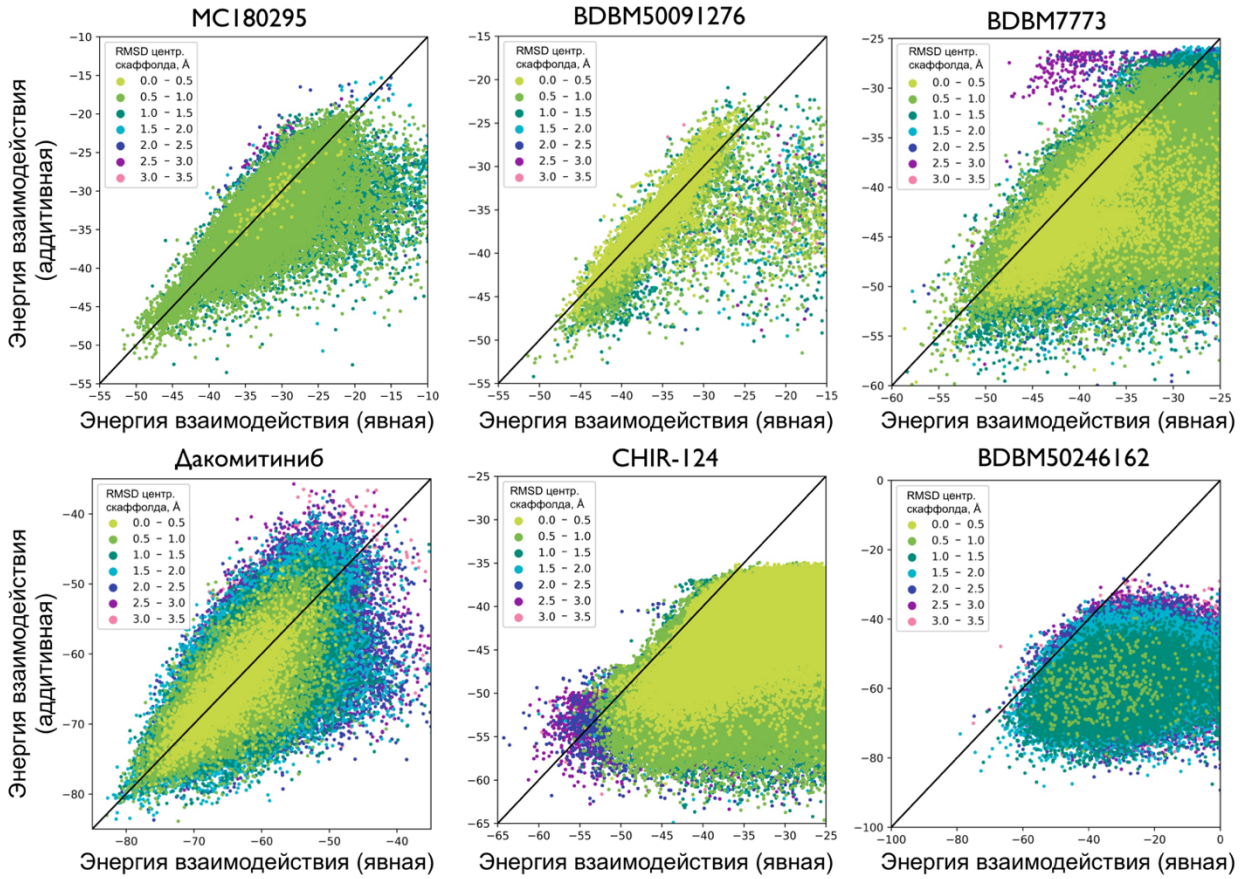


Рисунок 17 - Сравнение энергий взаимодействия полноразмерного соединения предсказанной из суммы энергий фрагментов (ось ординат) и после явного объединения пары фрагментов (ось абсцисс) в зависимости от RMSD центрального скаффолда фрагментов. Значения энергии взаимодействия указаны в энергетических единицах PyRosetta (ЭЕР)

3.1.4 Селективность из аддитивной энергии фрагментов

На следующем этапе была проверена способность разработанного метода предсказывать селективность взаимодействия с различными мишенями на основе

аддитивной энергии. Для этого использовались соединения из библиотеки MC180295 (Приложение 2) и десять ПК семейства CDK. Для каждой комбинации фрагментов сначала была предсказана аддитивная энергия взаимодействия, а затем вычислена явная энергия взаимодействия для объединенных пар. На следующем шаге были определены как предсказанная, так и явно вычисленная селективность каждого соединения относительно всех исследуемых CDK.

Для пяти случайно выбранных соединений из библиотеки было отмечено, что аддитивная энергия, рассчитанная на основе фрагментов, демонстрирует тенденцию к корреляции с точно вычисленными значениями энергии как представлено о на рисунке 18А. Среднее значение коэффициента корреляции Спирмена для этих соединений составило 0,74. Небольшие отклонения, выявленные в отдельных случаях, могут быть обусловлены упрощениями в ходе моделирования, описанными в предыдущем разделе. Кроме того, была проведена оценка распределения селективности соединений из библиотеки относительно набора исследуемых ПК (рисунок 18Б). Анализ показал, что избирательность соединений варьировала от киназы к киназе. Наибольшее число соединений (более 25 000) демонстрировали высокую эффективность против CDK8, тогда как лишь немногие соединения были селективны в отношении других киназ: менее 3 000 соединений – против CDK2, менее 3 500 – против CDK4 и CDK7, а также менее 4 000 – против CDK11.

Кроме того, предсказание энергии взаимодействия позволило правильно определить селективность в среднем по всем киназам в 78% случаев. Например, из 26 053 соединений, предсказанных как селективные ингибиторы CDK8, 22 793 из них были подтверждены при помощи явно вычисленной энергии полноразмерного соединения, т.е. точность предсказаний составила 87,5% для данной киназы.

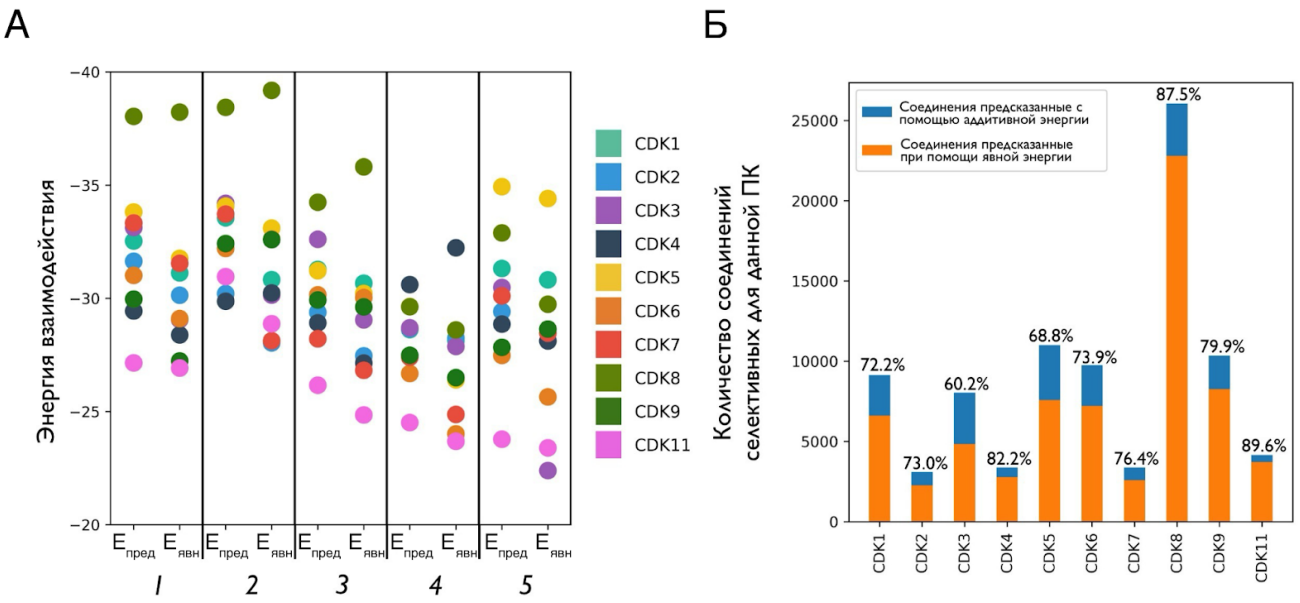


Рисунок 18 - Определение селективности соединений из библиотеки MC180295 для киназ семейства CDK при помощи разработанного метода. А - энергия взаимодействия 5 случайных соединений, рассчитанная при помощи аддитивной энергии фрагментов ($E_{пред}$) и явно вычисленной энергией ($E_{явн}$) полноразмерного соединения. Значения энергии взаимодействия указаны в энергетических единицах PyRosetta (ЭЕР), Б - распределение предсказанной селективности соединений для ПК семейства CDK, вычисленной по аддитивной энергии фрагментов (голубой), и ее проверка с помощью явно вычисленной энергии (оранжевый). Проценты отражают долю совпадений между предсказанной и явно вычисленной селективностью

Таким образом, наш метод позволяет быстро оценить селективность соединений в огромной библиотеке, используя в качестве приближения лишь попарное приближение предсказанных энергий взаимодействия.

3.1.5 Скрининг масштабных химических библиотек

Высокая точность аддитивной энергии полноразмерного соединения, вычисленной на основе энергий его фрагментов, может существенно ускорить процесс виртуального скрининга больших комбинаторных библиотек. Это позволяет значительно сократить количество вычислений: вместо оценки каждого возможного соединения в библиотеке достаточно проверить совместимость отдельных фрагментов с мишенью и отобрать лучшие из них для дальнейшего слияния. Такой подход позволяет предсказать максимально эффективные объединенные соединения в кратчайшее время.

Например, для каждого фрагмента было сгенерировано до 10 конформеров в библиотеке MC180295. Всего было рассмотрено 2914 фрагментов (и сгенерировано 29 012 конформеров) для замещаемой группы R1 и 9399 фрагментов (и 90 085 конформеров) для заместителя R2. Каждый конформер был минимизирован в комплексе со структурой мишени, что в конечном итоге дало 119 097 отдельных вычислений. Каждая минимизация в среднем занимает 2.5 минуты на современном процессоре среднего ценового сегмента, что суммарно дает 4962 часа на выполнение данной задачи. На вычислительном кластере в 256 потоках время вычисления составит всего лишь около 19 часов. Таким образом, используя данную стратегию скрининга, возможно эффективно предсказать энергию взаимодействия для всех 27 миллионов соединений (2 914 R1 x 9 399 R2), из которых состоит сгенерированная библиотека. Следует признать, что новые алгоритмы докинга, работающие на графических видеокартах, показывают значительные улучшения в скорости вычислений [71]. Тем не менее, скрининг 27 миллионов соединений даже для такого сочетания самых современных алгоритмов и процессоров требует гораздо бóльших вычислительных ресурсов. Кроме того, многие программы докинга игнорируют подвижность структуры белка, в то время как использованная нами минимизация позволяет изменять конформацию структуры мишени в ответ на различные структуры лигандов.

Чтобы убедиться, что разработанный метод действительно способен найти и приоритизировать активные соединения среди огромного множества сгенерированных соединений, в библиотеку были добавлены экспериментально изученные структурные аналоги для всех шести исходных ингибиторов, использованных в данной работе. Критериями включения служили как относительно высокая эффективность (с IC50 < 1 мкМ), молекулярную масса не более 500 Да, а также отсутствие в составе веществ химических элементов, которые обычно не используются для разработки лекарственных соединений (т.е. C, H, O, N, S, F, Cl и Br). В ходе эксперимента мы обнаружили, что оценочная функция, используемая для вычисления энергии взаимодействия в программном пакете для биомолекулярного моделирования PyRosetta, имеет тенденцию отдавать

предпочтение более крупным соединениям, содержащих более 30 неводородных атомов. В целом это отражает подобную же тенденцию, проявляющуюся для экспериментально измеренных аффинностей связывания, т.е. более крупные соединения часто имеют лучшую эффективность [125]. Тем не менее, использование такой метрики для отбора соединений без всяких поправок может быть проблематичным, поскольку более крупные соединения могут оказаться менее перспективными из-за своих фармакокинетических свойств [126]. Поэтому во многих работах по оптимизации структур ингибиторов используется метрика эффективности лиганда, которое представляет нормализацию активности соединения на его размер и отражает вклад каждого атома во взаимодействие с мишенью [125]. В связи с этим именно такая метрика была использована нами для ранжирования соединений в сгенерированных библиотеках.

Поскольку каждый опубликованный структурный аналог является результатом тщательной экспериментальной оптимизации для достижения высокой эффективности против мишени, то ожидается, что они должны иметь наилучшие значения эффективности лиганда по сравнению с другими (гипотетически возможными) соединениями в библиотеке. Таким образом, если наш метод предложит слишком много теоретически возможных соединений с более высокими показателями эффективности, чем у проверенных аналогов, то это, возможно, будет указывать на его неспособность отличать активные соединения от неактивных. С другой стороны, если список лучших соединений будет состоять в основном из аналогов, это будет свидетельствовать о том, что структура большинства сгенерированных соединений не подходит для взаимодействия с мишенью, и метод хорошо различает эти два класса соединений. Следует отметить, однако, что в данном эксперименте все сгенерированные соединения имеют один и тот же шарнир-связывающий скаффолд, что и реально существующие аналоги. Кроме того, многие функциональные группы фрагментов обладают такими же размерами и химическим составом, что и входящие в состав реальных ингибиторов. По этой причине ожидается, что часть сгенерированных соединений

могут быть в действительности сопоставимы с лучшими из экспериментально изученных ингибиторов по активности.

Для оценки того, насколько быстро разработанный метод может находить активные соединения (в данном случае — аналоги исходных соединений) среди остальной части сгенерированных соединений в библиотеках, мы использовали анализ обогащения. Этот метод часто применяется для оценки эффективности новых алгоритмов виртуального скрининга с целью понять, насколько хорошо метод способен отличать «активные» соединения от «неактивных» приманок [127–131]. Стандартным допущением метода является, отсутствие активности у большей части сгенерированных соединений, и, таким образом, основная задача метода — приоритетно находить экспериментально изученные активные соединения.

Результаты эксперимента показали, что активные соединения были обнаружены значительно быстрее, чем при случайном переборе (рисунок 19). При случайном поиске ожидалось, что для нахождения хотя бы одного экспериментально изученного аналога потребуется проверить 5–20% исходной библиотеки. Однако наш метод позволил решить эту задачу при анализе всего 1% лучших соединений для большинства библиотек, ранжированных по значению эффективности лиганда.

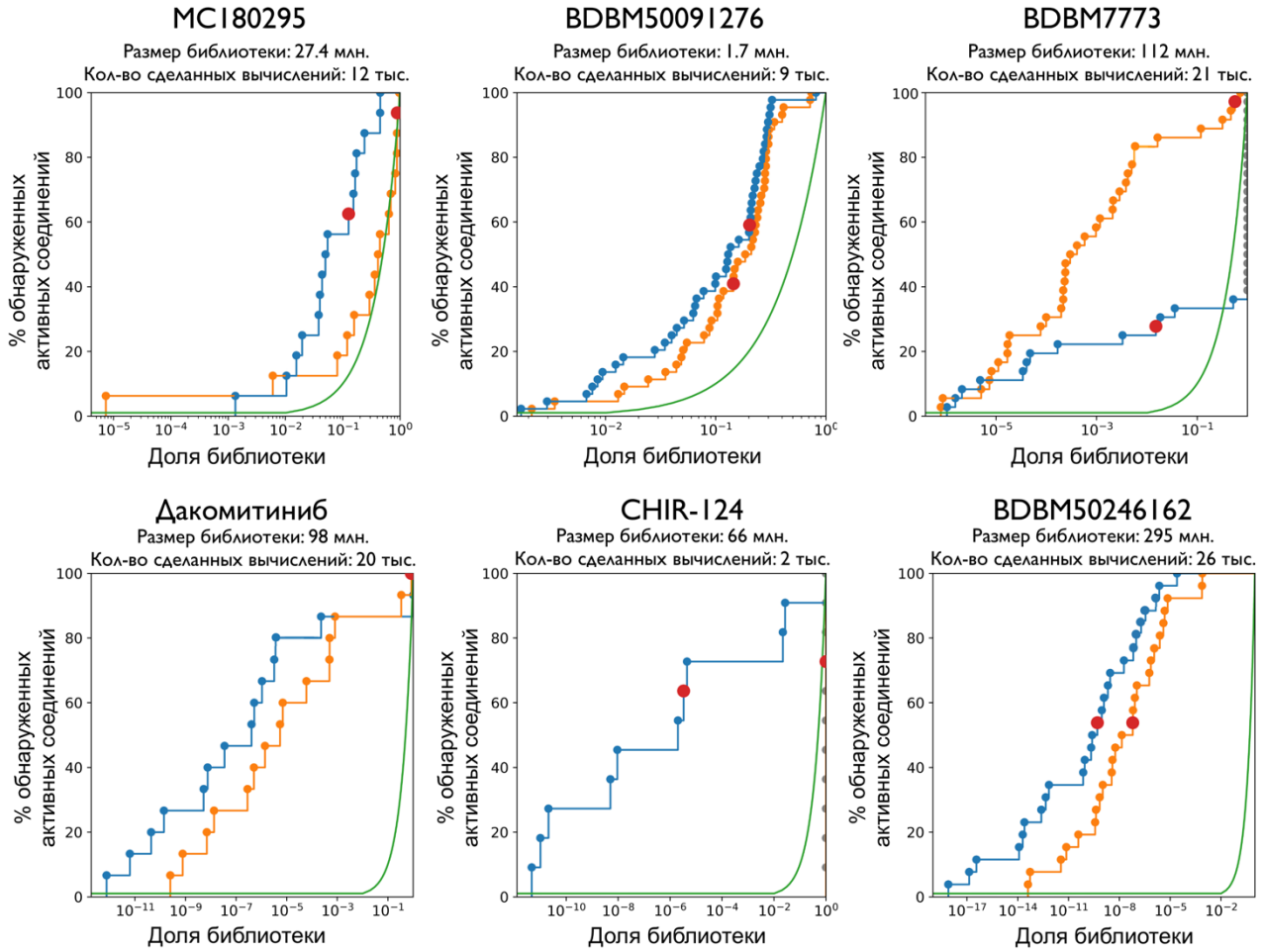


Рисунок 19 - Ранжирование активных соединений (красные точки) и их аналогов (голубые и оранжевые точки) относительно сгенерированных химических библиотек. В ходе скрининга были использованы две структуры мишени – оптимизированная под исходный ингибитор (голубая линия) и неоптимизированная, связанная с АТФ (оранжевая линия), которые затем были сравнены со случайным перебором (зеленая линия). Соединениям, утратившим взаимодействие с шарнирной петлей ходе минимизации, был присвоен самый низкий ранг (серые точки)

Мы также проверили эффективность метода в условиях, когда оптимизированная под ингибитор структура мишени недоступна из литературных данных. Для этого использовалась новая структура белка в комплексе с АТФ или АДФ, которые гораздо более доступны. Результаты показали, что производительность метода снизилась на 1-2 порядка (рисунок 19), за исключением одного случая: единственный аналог фрагмента R2 из библиотеки CHIR-124 не взаимодействовал с шарнирной областью ПК, в результате чего все аналоги были отсеяны в ходе скрининга.

3.2 Моделирование структуры тройного комплекса

К сожалению, количество кристаллических тройных комплексов мишень-PROTAC-лигаза E3, опубликованных в PDB, все еще недостаточно для проведения полномасштабного сравнительного анализа методов моделирования взаимодействия PROTAC с белками. Также остается открытым вопрос о том, насколько репрезентативной является кристаллизованная структура тройного комплекса, который часто характеризуется высокой подвижностью и может иметь более одного способа взаимодействия. В связи с этим разработанный метод был использован для сравнения предсказанной совместимости PROTAC, мишени и лигазы E3 с экспериментальными данными по деградации мишени. В первой части исследования было изучено влияние структуры линкера PROTAC на эффективность деградации $Brd4^{BD1}$, а во второй — анализ селективности разработанных соединений на деградацию киназ c-Met, EphA2 и STK10 [132,133].

3.2.1 Влияние длины линкера PROTAC на эффективность деградации

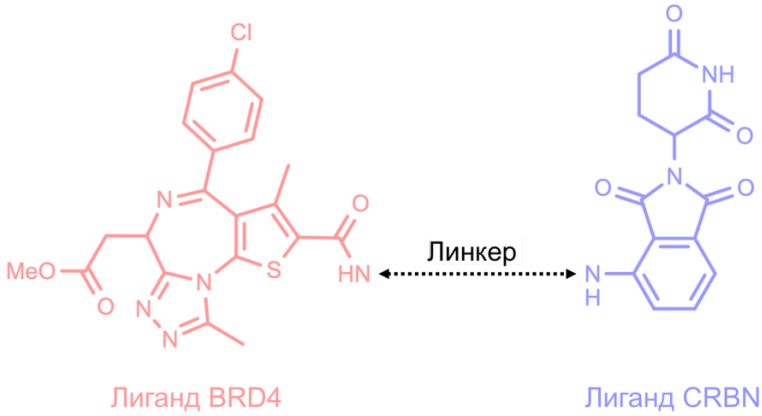
Одной из главных проблем при разработке новых молекул PROTAC является оптимизация структуры линкера, который соединяет функциональные фрагменты, предназначенные для взаимодействия с мишенью и лигазой E3. Как правило, этот этап требует кропотливой проверки множества аналогов одного и того же соединения с незначительно измененной структурой линкера. В связи с этим была поставлена задача понять, способен ли разработанный метод облегчить этот трудоемкий процесс, то есть возможно ли с его помощью определить совместимые линкеры, которые демонстрируют наилучшую эффективность в экспериментах. Для этого была использована коллекция PROTAC (табл. 3), разработанных для деградации $Brd4^{BD1}$ с помощью лигазы E3 CRBN [132,133].

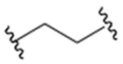
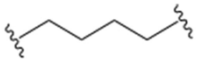
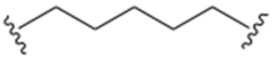
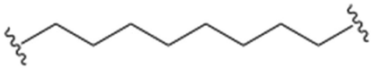
В ходе работы метода было отобрано около 5 000 предсказанных моделей взаимодействия $Brd4^{BD1}$ и CRBN. Затем для каждого линкера был сгенерирован набор трехмерных конформеров, которые выравнивались относительно функциональных фрагментов PROTAC, находящихся в комплексе с белками. На следующем этапе была проведена проверка для каждой пары модели

взаимодействия белков и линкера на предмет пространственной и энергетической совместимости. Количество моделей, прошедших все фильтры, использовалось для определения доли полностью совместимых комплексов (ДСК).

В итоге было выявлено, что значение ДСК связано со значениями экспериментально определенной деградации мишени, где PROTAC с более высокими значениями ДСК соответствуют сильной деградации $Brd4^{BD1}$ (таблица 3).

Таблица 3 - Соответствие между эффективностью деградации и предсказанным значением доли совместимых комплексов (ДСК)



PROTAC	**Линкер**	**Кол-во совместимых комплексов (кол-во конформеров линкера)**	**ДСК**	**Деградация $Brd4^{BD1}$ (EC50, нМ)**
ZXH-3-27		0 (237)	0,00	> 10000
ZXH-2-147		21 (606)	0,01	10 - 100
ZXH-2-184		520 (1 000)	0,10	10 - 100
ZXH-3-26		1 214 (1 000)	0,24	1 - 10
dBet70		1 257 (1 000)	0,25	1 - 10

3.2.2 Селективность PROTAC в деградации киназных мишеней

Разработанный метод также был применен для изучения PROTAC, разработанных для деградации киназных мишеней. Для этого были использованы данные из исследования, в котором несколько PROTAC были построены на основе неселективного ингибитора киназ — форетиниба, использованного в качестве низкомолекулярного лиганда мишени [68]. При слиянии структуры соединения с линкером и лигандами, связывающими либо VHL, либо CRBN, результаты показали, что некоторые киназы гораздо более эффективно деградируются PROTAC, рекрутирующей VHL (PROTAC1), в то время как другие мишени более значимо деградируются другой молекулой PROTAC, связывающей CRBN (PROTAC2) [26]. На основании того, что для связывания мишени использовался одинаковый функциональный фрагмент, было выдвинуто предположение, что эта селективность обусловлена различиями в формировании тройных комплексов. Таким образом, если это предположение верно, то значение ДСК разработанного метода может это подтвердить. Для данного исследования из PDB были извлечены кристаллические структуры трёх киназ — c-Met, EphA2 и STK10 в комплексе с форетинибом (рисунок 20) [134,135], а также кристаллические структуры VHL и CRBN, связанные со своими функциональными фрагментами [136]. Затем разработанный метод был применен для предсказания моделей взаимодействия тройного комплекса и оценки эффективности деградации мишени через ДСК для каждого PROTAC в паре с каждой киназой.

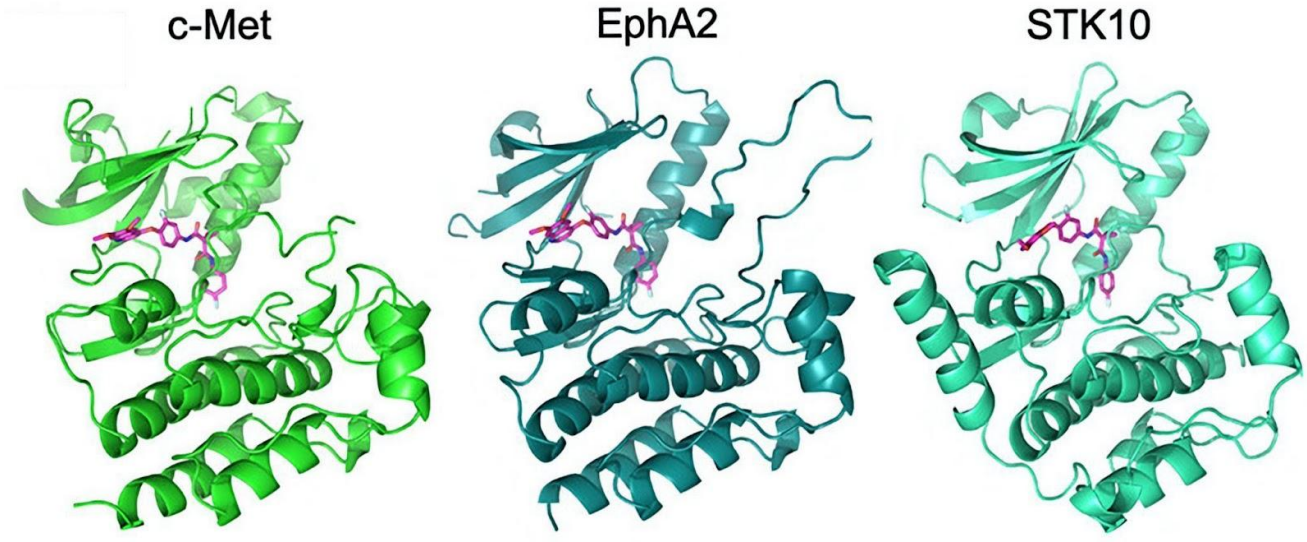
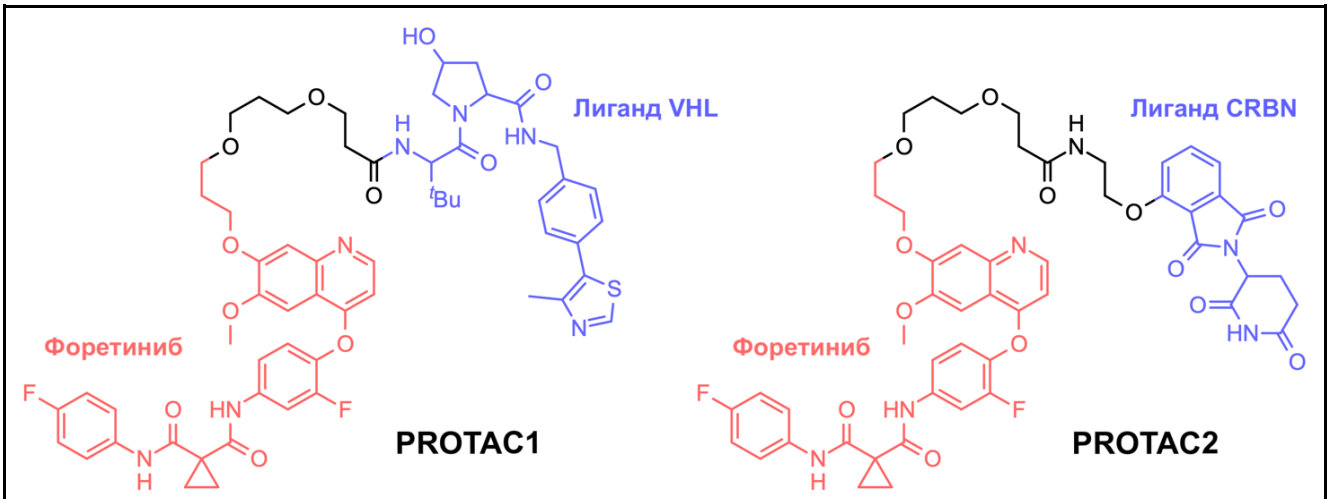


Рисунок 20 - Положение низкомолекулярного каталитического ингибитора форетиниба (розовый) в комплексе с c-Met (PDB ID: 6SD9), EphA2 (PDB ID: 5IA4) и STK10 (PDB ID: 6I2Y). Все три ПК имеют высокую консервативность АТФ-связывающего кармана и положение ингибитора является практически идентичным между всеми тремя киназами

Для PROTAC1, связывающегося с VHL, было зафиксировано более высокое значение ДСК, равное 0,66 для c-Met, что согласуется с экспериментальными данными по деградации киназ [26]. В то же время PROTAC2, взаимодействующий с CRBN, продемонстрировал повышенные значения ДСК для c-Met и EphA2, составляющие 0,23 и 0,27 соответственно, что также подтверждается экспериментальными результатами (таблица 4).

Таблица 4 – Доля совместимых комплексов для PROTAC, основанных на структуре форетиниба



Тройной комплекс	Кол-во совместимых комплексов (кол-во конформеров линкера)	ДСК	% деградации (1 мкМ PROTAC)
PROTAC1			
c-Met–PROTAC1–VHL	3 308 (1 000)	0,66	38
EphA2–PROTAC1–VHL	809 (1 000)	0,16	4
STK10–PROTAC1–VHL	879 (1 000)	0,18	1
PROTAC2			
c-Met–PROTAC2–CRBN	1 141 (1 000)	0,23	35
EphA2–PROTAC2–CRBN	1 339 (1 000)	0,27	41
STK10–PROTAC2–CRBN	8 (1 000)	0,00	0

Также в моделях PROTAC1 с c-Met и VHL наблюдалось множество способов взаимодействия мишени с лигазой E3 (рисунок 21А), хотя положение форетиниба оставалось одинаковым во всех моделях. В свою очередь это позволяет предположить, что структурные особенности внешней поверхности домена киназных мишеней могут определять эффективность распознавания лигазами VHL и CRBN. Взаимодействие VHL с c-Met включает неконсервативные области внешней поверхности семейства киназ Met (рисунок 23Б), что согласуется с предположением о том, что белок-белковые взаимодействия с участием этих

областей могут объяснить возникшую селективность PROTAC. В данном случае, даже среди трёх рассматриваемых здесь киназ отсутствуют консервативные участки на поверхности вне АТФ-связывающего сайта (рисунок 21В).

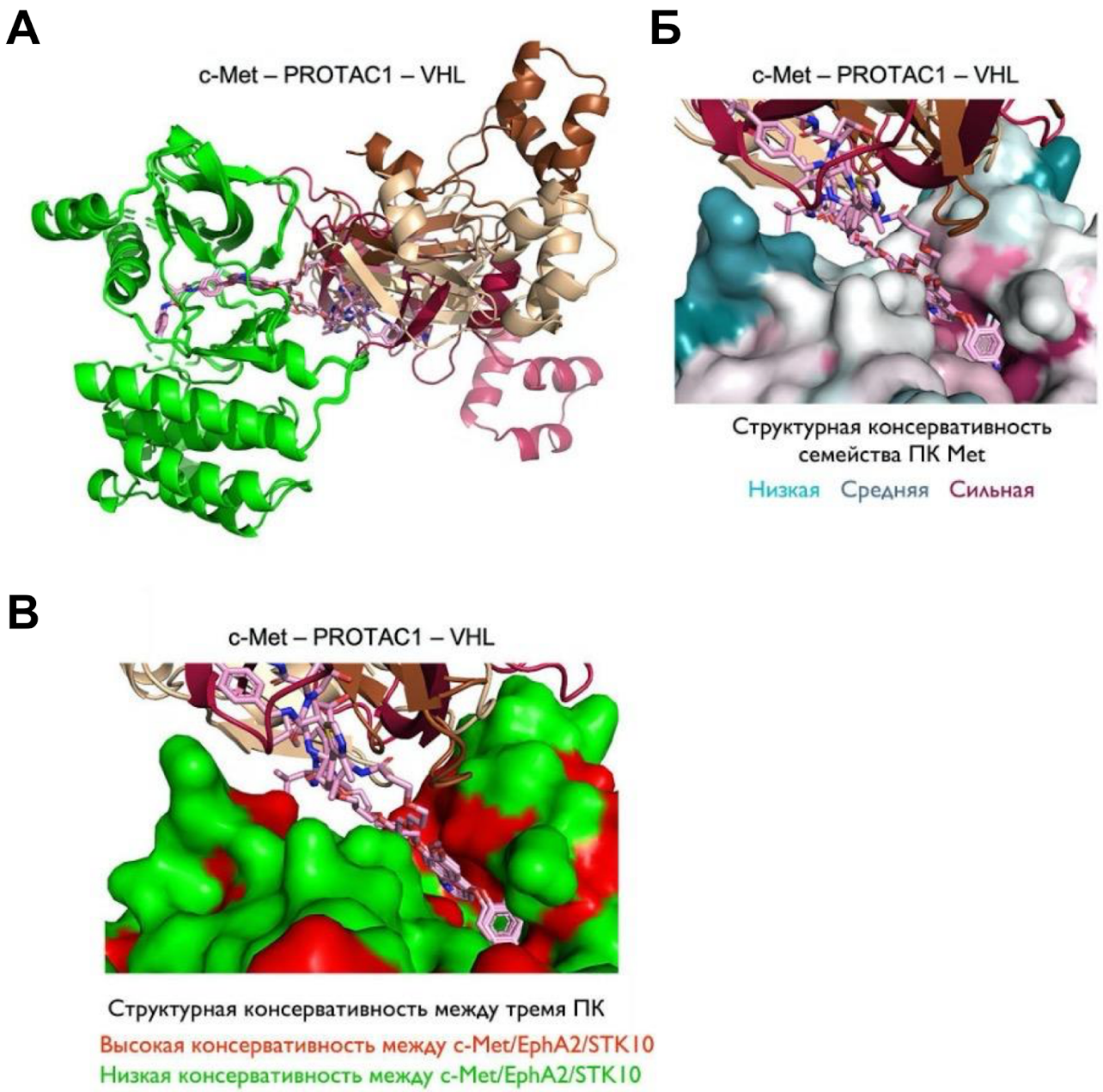


Рисунок 21 - Моделирование тройного комплекса c-Met-PROTAC1-VHL. А - предсказанные геометрически и энергетически совместимые положения c-Met (зеленый), PROTAC1 (светло-розовый) и VHL (светло- и темно-коричневый, и бордовый), Б - консервативность аминокислотных остатков на внешней и внутренней поверхности каталитического домена внутри семейства Met, В - консервативность аминокислотных остатков в каталитическом домене киназ c-Met, EphA2 и STK10

Таким образом, было выявлено, что взаимодействия тройного комплекса могут быть крайне разнообразными, при этом сохраняя положение лиганда в

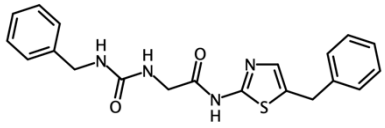
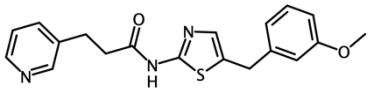
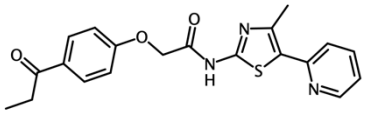
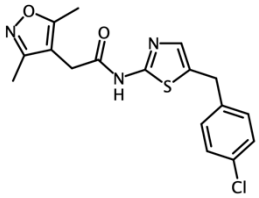
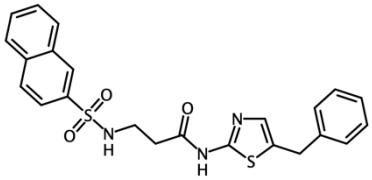
связывающем сайте мишени. Белок-белковые взаимодействия в тройном комплексе возникают на менее консервативных участках внешней поверхности киназ. Это, в свою очередь, приводит к селективной активности PROTAC, которая может значительно отличаться от селективности ингибиторов киназ, использованных для построения PROTAC. Также следует отметить, что данная селективность также зависит от рекрутируемой лигазы E3, поскольку разные лигазы будут по-разному взаимодействовать с мишенями.

3.3 Поиск новых ингибиторов с помощью традиционных методов виртуального скрининга

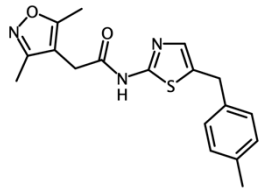
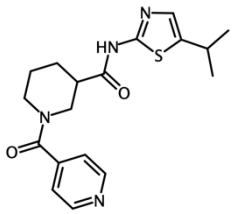
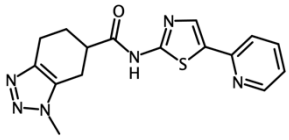
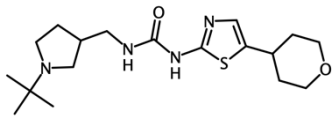
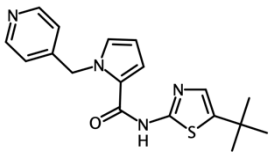
Для оценки предсказательной способности разработанного фрагментно-ориентированного метода была создана база для сравнения с уже зарекомендовавшими себя подходами. Для ее создания мы провели поиск ингибиторов ПК с использованием традиционных методов виртуального скрининга. Из каталога компании Enamine было извлечено 7 809 структур соединений с аминотиазольным центральным фрагментом. Для каждого соединения был сгенерирован набор трехмерных конформеров, которые затем выравнивались относительно структур известных ингибиторов ПК в PDB. Поскольку АТФ-конкурентные ингибиторы занимают консервативное положение в связывающем сайте, особенно в области шарнирной петли, предполагалось, что новые соединения будут иметь аналогичное расположение. Выяснилось, что большинство соединений демонстрируют сходство с шаблонами в диапазоне 1,0–1,2 по коэффициенту TanimotoCombo, что указывает на наличие структурных мотивов, повышающих вероятность взаимодействия. Затем мы провели минимизацию выровненных конформаций в комплексах со структурами CDK2 (PDB: 3QHW) и CDK9 (PDB: 3BLQ). Распределение энергий взаимодействия оказалось схожим: от -5 до -42 ЭЕР (медиана -23,78 ЭЕР) для CDK2 и от -5 до -37 ЭЕР (медиана -24,18 ЭЕР) для CDK9. Около 43% соединений показали лучшую энергию взаимодействия с CDK2, в то время как 57% — с CDK9.

Из этих соединений были отобраны 22 перспективных кандидата, обладающих благоприятными значениями энергии взаимодействия с мишенями и предполагаемой специфичностью к CDK2 и CDK9. В ходе экспериментального биохимического определения ингибирующей способности соединений-кандидатов было установлено, что только 10 соединений продемонстрировали значения IC50 ниже 100 мкМ, и лишь одно соединение — ниже 1 мкМ. Тенденция селективности к CDK2 и CDK9 была корректно предсказана только для двух из них.

Таблица 5 - Результаты виртуального скрининга соединений из специализированной библиотеки для ПК

Название	**Структура**	**E_{CDK2}, ЭЕР**	**E_{CDK9}, ЭЕР**	**$IC50_{CDK2}$, мкМ**	**$IC50_{CDK9}$, мкМ**
Z169536760		-31,08	-33,38	61,1	> 100,0
Z238156308		-32,62	-33,66	14,30	9,40
Z238190978		-28,24	-31,72	51,10	> 100,00
Z67764668		-35,44	-35,12	9,19	74,30
Z212912350		-36,26	-36,77	16,00	> 100,00

Продолжение таблицы 5

Название	**Структура**	**E_{CDK2}, ЭЕР**	**E_{CDK9}, ЭЕР**	**$IC50_{CDK2}$, мкМ**	**$IC50_{CDK9}$, мкМ**
Z70864235		-36,74	-36,89	5,81	32,30
Z1265432563		-30,31	-27,45	1,96	0,58
Z2177403824		-26,96	-32,24	55,40	9,21
Z2624229394		-30,78	-30,04	30,10	12,60
Z1275191743		-32,73	-32,91	> 100,00	40,80

3.4 Классификация предсказанных моделей при помощи vScreenML

Традиционные методы виртуального скрининга, широко применяемые в дизайне лекарственных препаратов, часто сталкиваются с проблемой низкой точности предсказаний. Методы машинного обучения играют ключевую роль в повышении эффективности виртуального скрининга, так как они способны значительно улучшить точность предсказаний, обучаясь различать истинные взаимодействия между лигандом и мишенью от ложных.

Перед использованием модель машинного обучения необходимо обучить на тренировочном наборе данных, чтобы она могла корректно интерпретировать влияние различных численных дескрипторов, описывающих предсказанные

комплексы, на их принадлежность к определенному классу. Этот набор может быть произвольным, и, как правило, в него включают все доступные параметры, вычисляемые различными программами виртуального скрининга. Например, могут быть включены энергетические характеристики, описывающие взаимодействие между низкомолекулярным лигандом и мишенью, а также структурные особенности связывающего кармана, с которым взаимодействует соединение. Большое количество параметров, полученных с помощью таких программ, позволяет более детально описывать предсказанные модели взаимодействия, что, в свою очередь, повышает точность предсказаний с использованием методов машинного обучения.

Сам процесс предсказания можно разделить на три этапа: 1) виртуальный скрининг, в ходе которого формируется набор предсказанных моделей взаимодействия между исследуемыми соединениями и мишенью, 2) вычисление дескрипторов, характеризующих предсказанные комплексы, и 3) предсказание с использованием предобученной модели машинного обучения.

В 2020 году был разработан действующий прототип ПО vScreenML [31], эффективность которого была продемонстрирована посредством корректной идентификации 10 из 23 соединений с концентрацией полумаксимального ингибирования (IC50) ацетилхолинэстеразы менее 50 мкМ. Однако установка и использование нулевой версии программы были неудобны для конечного пользователя, а тренировочный набор модели машинного обучения, подготовленный четыре года назад, нуждался в обновлении. В целях предоставления возможности для широкого научного сообщества пользоваться наиболее современным и удобным инструментом улучшения точности предсказаний была выпущена усовершенствованная версия программы vScreenML, устраняющая недостатки прототипа.

Подготовка нового тренировочного набора для обучения модели состояла из двух этапов. На первом этапе были отобраны истинные активные взаимодействия низкомолекулярных лигандов и мишеней. Для этого из базы данных PDB было извлечена информация о 132 146 кристаллических комплексов, которая затем была

отфильтрована по качеству определения трехмерных структур и физико-химическим свойствам основного низкомолекулярного лиганда, чтобы они соответствовали возможным лекарственным свойствам, как описано в приложении 3. Затем отобранные структуры были обработаны с помощью веб-сервера Protoss [137] для предсказания расположения атомов водорода в комплексах мишень-лиганд и определения наиболее вероятного состояния протонирования. В результате был получен набор из 1 321 структуры. Затем каждая копия комплекса, представленного в отобранных структурах, была извлечена для увеличения размера обучающего набора, что привело к созданию 1 806 объектов. Эти структуры затем были минимизированы с использованием PyRosetta. На втором этапе с помощью веб-сервера DUD-E [138] для активных соединений были сгенерированы ложные низкомолекулярные приманки со сходными физико-химическими, но разными топологическими свойствами. Этот набор был отфильтрован аналогично активным лигандам, чтобы избежать систематических различий между двумя классами соединений. Затем для приманок были сгенерированы трехмерные конформеры, которые были выровнены относительно структуры минимизированного активного лиганда. Три наиболее похожие приманки для каждого активного соединения были отобраны и минимизированы вместе с соответствующей структурой белка.

Кроме того, мы улучшили подбор дескрипторов для нашей модели, принимая во внимание и необходимость избавить конечного пользователя от зависимости от платных программ. В частности, cxcalc от ChemAxon был заменен на RDKit [Landrum, 2010], а SZYBKI от OpenEye был удален из-за низкой информативности для предсказания. Были добавлены дескрипторы свободных полярных атомов, удаленных от растворителя (англ. buried unsatisfied polar atoms) для лиганда, вычисляемые с помощью PyRosetta и RDKit. Также был добавлен набор, характеризующих различные типы взаимодействия между соединением и мишенью, вычисляемые с помощью LUNA [139], а также набор геометрических параметров связывающего кармана мишени, вычисляемых с помощью PocketDruggability. В то же время были сохранены энергетические параметры,

вычисляемые Rosetta, взаимодействия между лигандом и белком, вычисляемые с помощью BINANA [140] и RF-score [141]. Общее количество дескрипторов составило 165 (Приложение 4).

Для сокращения времени, необходимого для вычисления и уменьшения вероятности переобучить модель, был проведен отбор наиболее информативных для предсказания дескрипторов. Для этого использовался метод пошагового отбора совместно с k-fold кросс-валидацией. В данном случае k было равно 269, по количеству кластеров последовательностей мишеней в обучающем наборе. В качестве метрики эффективности классификации использовался коэффициент корреляции Мэтьюса (ККМ), который является устойчивым к несбалансированным данным. Благодаря этому был выбран набор из 49 параметров (рисунок 22А), которые дали максимальное значение точности классификации среди всех возможных комбинаций параметров.

После оптимизации гиперпараметров модели машинного обучения, дискриминирующая способность новой версии на новом обучающем наборе продемонстрировала улучшенные результаты по сравнению с прототипом: точность составила – 0,95, прецизионность – 0,95, полнота – 0,89 и ККМ – 0,89 (рисунок 22Б).

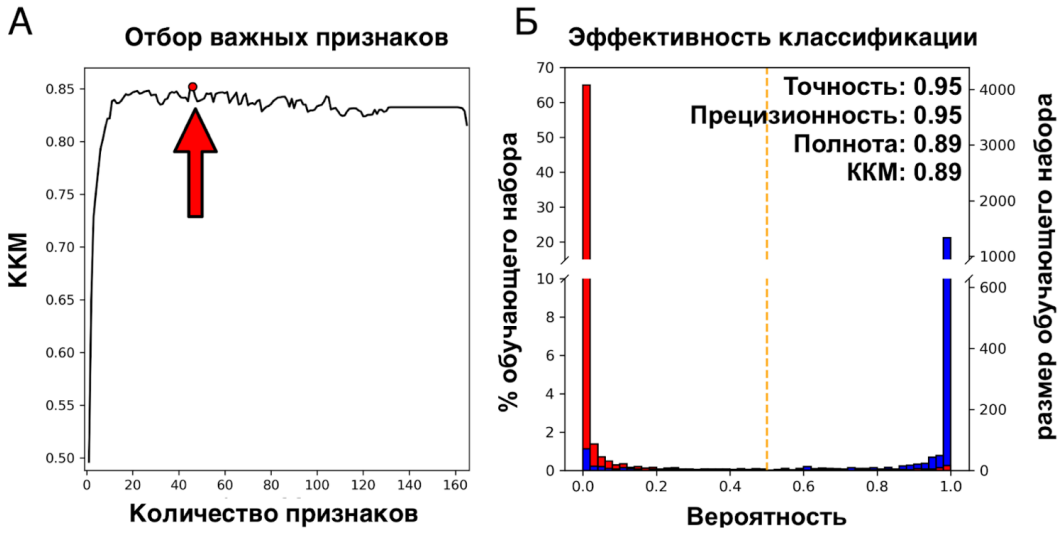


Рисунок 22 - Эффективность классификации на новом тренировочном наборе. А - отбор наиболее информативных дескрипторов с помощью метода Sequential Backward Floating Selection, Б - распределение вероятностей того, что лиганд в предсказанном комплексе с мишенью является активным, в наборе приманок (красный) и истинных лигандов (синий)

Дополнительная проверка предсказательной способности новой версии vScreenML была проведена на наборе DEKOIS2 [142], который часто используется для оценки эффективности различных функций оценки и инструментов докинга. Этот набор включает от 30 до 40 истинных и от 800 до 1 200 ложных лигандов в формате SMILES для каждой из 81 мишеней. Для того, чтобы избежать дополнительных вычислений, подготовленный набор DEKOIS2 со структурами мишеней и состыкованными к ним истинными и ложными лигандами был получен из работы, посвященной разработке нового метода докинга — KarmaDock [143,144]. Затем этот набор был минимизирован аналогично процессу подготовки нового обучающего набора vScreenML. Для оценки эффективности классификации был использован фактор обогащения EF1%. Сначала для каждого соединения предсказывалась вероятность активности, после чего все соединения сортировались по убыванию этого показателя. Затем определялось количество истинно активных соединений среди 1% наиболее вероятных кандидатов, и этот показатель сравнивался с общей долей активных соединений в наборе.

Для сравнения были изначально выбраны оценочные функции таких моделей глубокого обучения, как KarmaDock, RTMScores [145] и gnina (CNNaffinity) [146], а также эмпирические оценочные функции AA-Score [147] и AutoDock Vina [146]. Однако в случае моделей глубокого обучения существует риск утечки данных, когда тестируемая выборка перекрывается с обучающим набором, что может приводить к некорректно высоким оценкам классификации. Например, в предсказаниях KarmaDock 6 структур PYGL-in были представлены в обучающем наборе, что могло необоснованно завысить качество результатов предсказания; в то же самое время структура PYGL-out не входила в обучающий набор и модель продемонстрировала гораздо худшие результаты (приложение 5). После тщательного анализа всех оценочных функций для сравнения были выбраны только AA-Score и AutoDock Vina, имплементированная в gnina, а также CNNaffinity из gnina, поскольку у них было меньшее перекрытие между обучающим и тестовым наборами.

Новая версия vScreenML превзошла эмпирические функции оценки AA-Score и AutoDock Vina (рисунок 23), показав медианный EF1% 7,64% по сравнению с 5,02% у AA-Score и 2.61% у AutoDock Vina. Кроме того, 36 из 81 цели достигли EF1% выше 10 с vScreenML, тогда как AA-Score достиг этого только для 13 целей, а AutoDock Vina — для 8. Точность CNNaffinity оказалась схожей с точностью vScreenML, и значения EF1% были практически идентичны для 35 мишеней (разница между значениями менее 3%). В то же время новая версия vScreenML показала значительно лучшую точность для 28 мишеней, тогда как CNNaffinity — только для 18 целей.

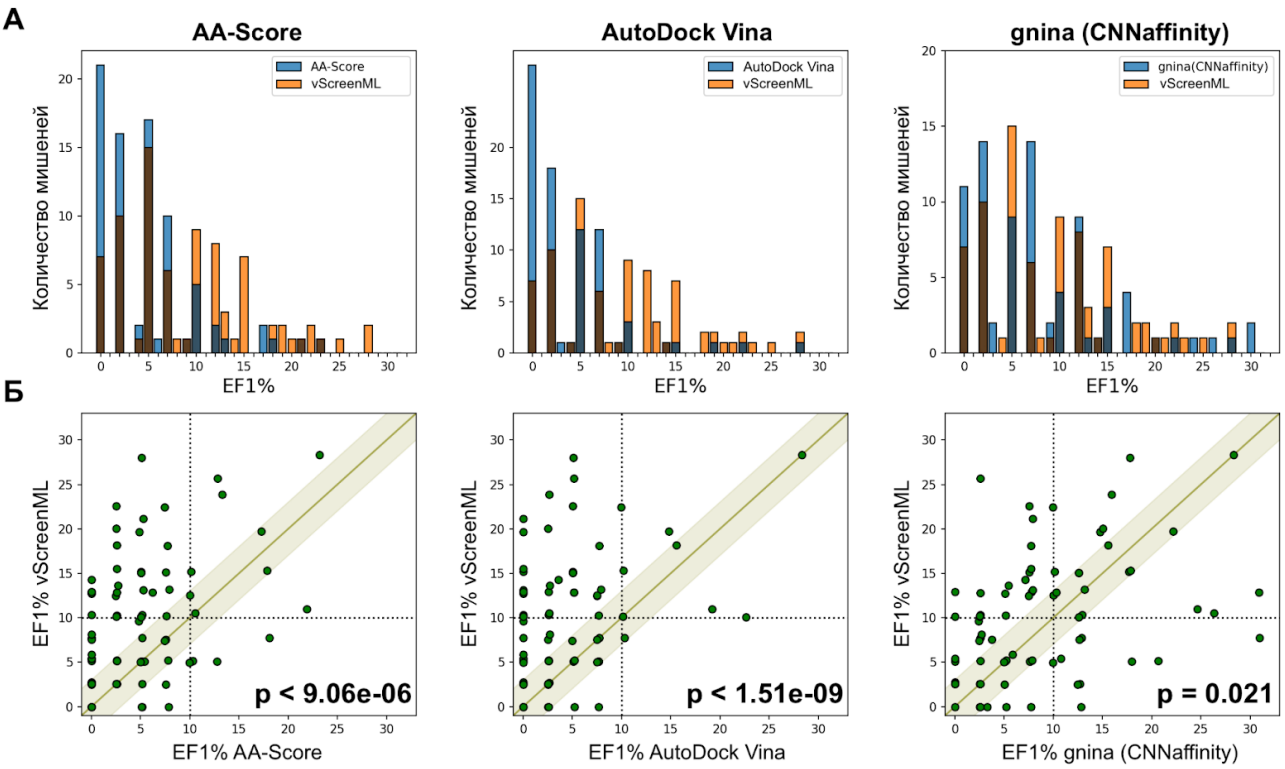


Рисунок 23 - Сравнительная характеристика производительности vScreenML, AA-Score, AutoDock Vina и gnina (CNNaffinity) на наборе DEKOIS2. А - распределение значений EF1% для AA-Score, AutoDock Vina и CNNaffinity (синий) и vScreenML (оранжевый). Значения столбцов начинаются от нуля и в случае наложения друг на друга имеют темно-синий и коричневые цвета, Б - соответствие значений EF1% между разными программами. Зеленая диагональная линия отражает равенство значений EF1% между сравниваемыми методами (с допуском 3% различия между значениями). Значение p было вычислено с помощью биномиального теста

На завершающем этапе исследования была проведена оценка эффективности классификации 22 ингибиторов CDK2 и CDK9, отобранных для экспериментальной проверки, рассмотренных в предыдущем разделе. Для этого 44 полученные модели были минимизированы, параметризованы, и для каждой из них рассчитана оценка активности vScreenML. Для сопоставления предсказаний с экспериментальными данными соединения были разделены на два класса: неактивные (IC50 > 100 мкМ) и активные (IC50 < 100 мкМ).

Результаты показали, что обновленная версия vScreenML классифицировала с точностью 79,54% на всех моделях, при этом 45,45% из них правильно определены как неактивные, а 34,09% — как активные. Анализ предсказаний по отдельности для CDK2 и CDK9 продемонстрировал точность 86,36% и 72,23% соответственно. Ошибочно активными были предсказаны два соединения для CDK2 (9,09%) и пять — для CDK9 (22,73%), тогда как по одному соединению (4,55%) для каждой киназы ошибочно классифицировано как неактивное.

ЗАКЛЮЧЕНИЕ

В последние годы наблюдается интерес к киназам как терапевтическим мишеням, и количество разработанных ингибиторов постоянно растет. В связи с этим активно развиваются новые методы структурной оптимизации уже известных ингибиторов, в том числе и с помощью вычислительных инструментов. Одним из примеров является недавняя работа, в которой коммерчески доступные соединения были деконструированы на отдельные фрагменты для создания новой коллекции, сфокусированной на взаимодействии с шарнирным регионом каталитического домена [148]. Затем эти фрагменты были перекомбинированы между собой и была получена библиотека из 39 миллионов соединений. Однако трехмерная совместимость этих соединений с сайтом связывания в данной работе не была проверена. В другой работе был проведен анализ для всех кристаллических структур ПК, связанных с низкомолекулярными лигандами для того, чтобы выявить фрагменты, которые характерны для карманов каталитического домена ПК. На основе этих данных авторы получили библиотеку из почти 7 миллионов новых соединений [149]. Однако в обоих случаях синтетическая доступность созданных библиотек не была оценена.

Наш метод позволяет генерировать более масштабные библиотеки, так как использует все коммерчески доступные строительные блоки, а не только фрагменты из известных лекарственных соединений. Хотя некоторые соединения в этих библиотеках могут оказаться синтетически недоступными, вероятность успешного синтеза остается выше благодаря тому, что каждое соединение создаётся с учётом экспериментально проверенных схем синтеза. Идея нашего метода больше всего близка к другой работе, посвященной разработке подхода к созданию разнообразных и фокусированных соединений DOTS [150]. Этот метод предполагает многошаговую оптимизацию соединений, где используется исходный фрагмент и набор правил для его структурной трансформации. Затем оптимизированные варианты исходного фрагмента отбираются с помощью различных методов виртуального скрининга, после чего синтезируются и

тестируются экспериментально. Эффективные соединения переходят на следующий этап оптимизации, и процесс продолжается до достижения желаемого уровня эффективности.

Расширение методов комбинаторной химии позволило поставщикам химических соединений увеличить каталоги доступных соединений. Например, еще несколько лет назад компания Enamine могла предложить до 170 миллионов соединений для использования в виртуальном скрининге [71], а недавно этот показатель достиг 19 миллиардов, что превышает возможности традиционного виртуального скрининга. В ответ был разработан скрининг на основе синтонов [151], где каждое соединение было разобрано на отдельные фрагменты в соответствии с протоколами синтеза соединений в компании Enamine. Затем получившиеся фрагменты были состыкованы со структурой мишени и лучшие из них были отобраны для дальнейшей оптимизации. На следующем шаге к выбранным фрагментам присоединялись новые функциональные группы, и полученные соединения снова были состыкованы со структурой мишени и лучшие из них отбирались для экспериментальной проверки. Хотя этот метод похож на разработанный нами, он ограничен схемами синтеза из каталога Enamine и не может использовать широкий спектр поставщиков химических соединений, что в конечном итоге может повлиять на покрытие химического пространства.

Наш метод позволяет быстро и надежно оценивать энергию взаимодействия большого числа соединений благодаря аддитивности энергий фрагментов. Это является ключевым преимуществом, позволяющим нам достаточно быстро оценить энергии взаимодействия соединений с мишенью в многомиллионных и даже многомиллиардных библиотеках без явной необходимости проверять каждое из них. Однако, следует напомнить о границах применимости данного подхода - аддитивность будет проявляться только в том случае, если общий скаффолд не меняет своего положения. Если соединение изменяет способ взаимодействия при замене функциональной группы, итоговый эффект становится непредсказуемым [122], что полностью согласуется с нашими наблюдениями о аддитивности. Необходимо отметить, что разработанный метод игнорирует некоторые

физические аспекты при взаимодействии с мишенью, включая конформационную энтропию лиганда или отсутствие явного растворителя. Соответственно, от метода нельзя ожидать, что энергия взаимодействия, использованная в исследовании, будет сильно коррелировать со экспериментальными значениями активности соединений. Наилучшим решением было бы решением включить этот метод в многоуровневый процесс, при котором размер библиотеки постепенно сокращается за счет использования все более точных, но ресурсозатратных методов, таких как молекулярная динамика и вычисление возмущения свободной энергии [152].

Однако разработанный фрагментно-ориентированный метод не решает проблему низкой селективности соединений, поскольку АТФ-связывающий сайт ПК является крайне консервативным. Для решения этой проблемы мы разработали другой подход, который предназначен для облегчения разработки новых молекул PROTAC. Мы разработали метрику доли полностью совместимых комплексов (ДСК), которая имеет явную корреляцию с экспериментально определенной эффективностью деградации мишени. В результате наш метод создает множество моделей возможного расположения членов тройного комплекса, которые соответствуют друг другу с точки зрения геометрического и энергетического взаимодействия.

Следует отметить, что значение ДСК нельзя использовать в качестве определения активности PROTAC. Данная метрика скорее будет полезна при сравнительном анализе, когда необходимо отобрать только несколько перспективных соединений PROTAC для дальнейшей проверки. Как было показано в результатах, метод способен помочь выбрать подходящую структуру для линкера. Кроме того, метод может помочь в решении как именно линкера должен прикрепляться к функциональным фрагментам PROTAC, что также существенно влияет на вероятность формирования тройного комплекса. Предсказание белок-белковых взаимодействий между мишенью и различными лигазами Е3 с использованием нашего метода позволяет точно определить, какие лигазы лучше подходят для деградации конкретной мишени. Например, наши

предсказания подтвердили экспериментальные данные, показывающие, что EphA2 более эффективно деградируется с участием CRBN, а не VHL. Таким образом, наш подход позволяет оптимизировать выбор компонентов PROTAC, повышая их эффективность в таргетной деградации белков.

При сравнительной анализе было показано, что предсказательная эффективность разработанных вычислительных методов значительно выше по сравнению с уже зарекомендовавшими себя подходами виртуального скрининга. Фрагментно-ориентированный метод способен проанализировать многомиллионные библиотеки соединений значительно быстрее и тщательнее, что в конечном итоге значительно повышает шансы найти работающие соединения. Второй же метод значительно улучшает наши возможности разработки селективных ингибиторов киназ, поскольку использует принципиально иной механизм ингибирования. Тем не менее, традиционные методы виртуального скрининга в сочетании с методами машинного обучения остаются важными инструментами, особенно в решении таких критических проблем, как высокий уровень ложноположительных результатов. Обучаясь на тренировочном наборе, состоящем из истинно активных соединений и ложных приманок, модели машинного обучения могут эффективно различать активные и неактивные соединения. Это подтверждено высокой точностью классификации, достигнутой моделью vScreenML при отборе потенциальных ингибиторов CDK2 и CDK9, что демонстрирует её применимость в контексте виртуального скрининга.

Однако, несмотря на достижения в области глубокого обучения, такие приложения, как KarmaDock, RTMScore и gnina с функцией оценки CNNaffinity, которые в некоторых аспектах превосходят vScreenML, вызывают определенные опасения относительно своей способности к обобщению. В частности, существует риск, что модели, такие как KarmaDock и RTMScore, могут не выявлять общие закономерности, присущие активным соединениям, а лишь "заучивать" особенности активных комплексов в обучающем наборе. Эта проблема была особенно заметна при анализе производительности на мишени PYGL-out в наборе

DEKOIS2, что указывает на потенциальные ограничения моделей глубокого обучения в способности обобщать за пределами своих обучающих данных.

ВЫВОДЫ

1. Разработанный фрагментно-ориентированный метод обеспечил значительное ускорение анализа многомиллионных библиотек соединений (продемонстрированное на примере циклин-зависимых киназ 2 и 9, киназы контрольной точки 1, рецептора эпидермального фактора роста с мутацией T790M и активированной CDC42-ассоциированной киназы 1) за счет предварительного скрининга отдельных фрагментов и использования аддитивного подхода к предсказанию энергии взаимодействия полноразмерных соединений с мишенью.
2. Созданный метод моделирования взаимного положения компонентов тройного комплекса предсказал совместимость серий протеолиз-таргетированных химер с различными конфигурациями линкера, предназначенных для одновременного взаимодействия с первым бромодоменом бромодоменсодержащего белка 4 и цереблоном, а также селективность действия протеолиз-таргетированных химер, взаимодействующих с тирозиновой протеинкиназой Met, рецептором эфрина типа A2 и серин/треониновой протеинкиназой 10. Разработанная нами метрика доли совместимых комплексов показала высокую согласованность с экспериментальными данными об эффективности деградации протеинкиназ.
3. Традиционные методы виртуального скрининга продемонстрировали недостаточно высокую точность в ходе предсказания эффективных ингибиторов циклин-зависимых киназ 2 и 9: лишь 45% соединений показали ингибирующий эффект *in vitro*. Внедрение модели машинного обучения vScreenML повысило точность предсказаний до 79,5%, значительно сократив количество ложноположительных результатов.

СПИСОК СОКРАЩЕНИЙ И ОБОЗНАЧЕНИЙ

АТФ	-	аденозинтрифосфорная кислота
ГТФ	-	гуанозинтрифосфорная кислота
ДСК	-	доля совместимых комплексов, отражающая насколько PROTAC совместима с мишенью и убиквитинлигазой E3
ККМ	-	коэффициент корреляции Мэтьюса (англ. Matthews correlation coefficient)
ПК	-	протеинкиназа
ЭЕР	-	энергетические единицы Rosetta
ACK1	-	активированная CDC42-ассоциированная киназа 1 (англ. activated CDC42 kinase 1)
Brd4^{BD1}	-	бромодоменсодержащий белок 4 с бромодоменом 1 (англ. bromodomain containing protein 4 bromodomain 1)
Brij35	-	полиоксиэтилен(23)-лауриловый эфир (англ. polyoxyethylene (23) lauryl ether)
BSA	-	бычий сывороточный альбумин (англ. bovine serum albumin)
c-Met	-	тирозиновая протеинкиназа Met (англ. tyrosine-protein kinase Met)
CDK2	-	циклинзависимая киназа 2 (англ. cyclin-dependent kinase 2)
CDK9	-	циклинзависимая киназа 9 (англ. cyclin-dependent kinase 9)
CHK1	-	киназа контрольной точки 1 (англ. checkpoint kinase 1)
CRBN	-	убиквитин лигаза E3 цереблон (англ. Cereblon)
DFG	-	мотив Asp-Phe-Gly, который находится в начале активационной петли
DMSO	-	диметилсульфоксид (англ. dimethyl sulfoxide)
DTT	-	дитиотреитол (англ. dithiothreitol)
EC50	-	полумаксимальная эффективная концентрация (англ. Half Maximal Effective Concentration)
EF1%	-	фактор обогащения на 1% (англ. enrichment factor in 1%)
EGFR	-	рецептор фактора роста эндотелия (англ. epidermal growth factor receptor)

EGTA	-	этиленгликоль-бис(β-аминоэтиловый эфир)-N,N,N',N'-тетрауксусная кислота (англ. ethylene glycol-bis(β-aminoethyl ether)-N,N,N',N'-tetraacetic acid)
EphA2	-	рецептор эфрина типа A2 (англ. ephrin type-A receptor 2)
GK	-	аминокислота-привратник шарнирного региона (англ. gatekeeper)
HEPES	-	4-(2-гидроксиэтил)-1-пиперазинэтансульфоновая кислота (англ. 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)
HTS	-	высокопроизводительный скрининг химических соединений (англ. high-throughput screening)
IC50	-	полумаксимальная концентрация ингибирования (англ. half Maximal Inhibition Concentration)
PDB	-	банк данных трёхмерных структур белков и нуклеиновых кислот (англ. Protein Data Bank)
PROTAC	-	протеолиз-таргетированная химера (англ. PROteolysis TArgeting Chimera)
RMSD	-	среднеквадратическое отклонение (англ. Root Mean Square Deviation)
SMARTS	-	произвольная спецификация шаблона SMILES (англ. SMILES arbitrary target specification)
SMILES	-	упрощенная система молекулярного ввода и линейного ввода (англ. Simplified Molecular-Input Line-Entry System)
STK10	-	серин/треониновая протеинкиназа 10 (англ. Serine/threonine proteinkinase 10)
VHL	-	опухолевый супрессор Гиппеля—Линдау (англ. Von Hippel–Lindau tumor suppressor)

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Hanks S.K. Genomic analysis of the eukaryotic protein kinase superfamily: a perspective // Genome Biol. 2003. Vol. 4, № 5. P. 111.
2. Manning G., Whyte D.B., Martinez R., Hunter T., Sudarsanam S. The protein kinase complement of the human genome // Science. American Association for the Advancement of Science (AAAS), 2002. Vol. 298, № 5600. P. 1912–1934.
3. Cohen P. The origins of protein phosphorylation // Nat. Cell Biol. Springer Science and Business Media LLC, 2002. Vol. 4, № 5. P. E127-30.
4. Davis R.J. Transcriptional regulation by MAP kinases // Mol. Reprod. Dev. Wiley, 1995. Vol. 42, № 4. P. 459–467.
5. Ding L., Cao J., Lin W., Chen H., Xiong X., Ao H., Yu M., Lin J., Cui Q. The roles of cyclin-dependent kinases in cell-cycle progression and therapeutic strategies in human breast cancer // Int. J. Mol. Sci. MDPI AG, 2020. Vol. 21, № 6. P. 1960.
6. Gonçalves A.P., Chow K.M., Cea-Sánchez S., Glass N.L. WHI-2 regulates intercellular communication via a MAP kinase signaling complex // Front. Microbiol. Frontiers Media SA, 2019. Vol. 10. P. 3162.
7. Götz C., Montenarh M. Protein kinase CK2 in development and differentiation // Biomed. Rep. Spandidos Publications, 2017. Vol. 6, № 2. P. 127–133.
8. Lu Z., Hunter T. Metabolic kinases moonlighting as protein kinases // Trends Biochem. Sci. 2018. Vol. 43, № 4. P. 301–310.
9. Rauch J., Volinsky N., Romano D., Kolch W. The secret life of kinases: functions beyond catalysis // Cell Commun. Signal. Springer Nature, 2011. Vol. 9, № 1. P. 23.
10. Bhullar K.S., Lagarón N.O., McGowan E.M., Parmar I., Jha A., Hubbard B.P., Rupasinghe H.P.V. Kinase-targeted cancer therapies: progress, challenges and future directions // Mol. Cancer. 2018. Vol. 17, № 1. P. 48.
11. PKI [Electronic resource] // Blue Ridge Institute for Medical Research | an independent non-profit organization founded in 2006. The Blue Ridge Institute for Medical Research, 2022. URL: https://brimr.org/protein-kinase-inhibitors/ (accessed: 26.10.2024).

12. Roskoski R. Jr. Properties of FDA-approved small molecule protein kinase inhibitors: A 2024 update // Pharmacol. Res. Elsevier BV, 2024. Vol. 200, № 107059. P. 107059.
13. Gani O.A., Thakkar B., Narayanan D., Alam K.A., Kyomuhendo P., Rothweiler U., Tello-Franco V., Engh R.A. Assessing protein kinase target similarity: Comparing sequence, structure, and cheminformatics approaches // Biochim. Biophys. Acta. Elsevier BV, 2015. Vol. 1854, № 10 Pt B. P. 1605–1616.
14. Modi V., Dunbrack R.L. Jr. Defining a new nomenclature for the structures of active and inactive kinases // Proc. Natl. Acad. Sci. U. S. A. Proceedings of the National Academy of Sciences, 2019. Vol. 116, № 14. P. 6818–6827.
15. Manning B.D., Toker A. AKT/PKB signaling: Navigating the network // Cell. Elsevier BV, 2017. Vol. 169, № 3. P. 381–405.
16. Gagic Z., Ruzic D., Djokovic N., Djikic T., Nikolic K. In silico methods for design of kinase inhibitors as anticancer drugs // Front. Chem. Frontiers Media SA, 2019. Vol. 7. P. 873.
17. Maia E.H.B., Assis L.C., de Oliveira T.A., da Silva A.M., Taranto A.G. Structure-based virtual screening: From classical to artificial intelligence // Front. Chem. Frontiers Media SA, 2020. Vol. 8. P. 343.
18. Mouchlis V.D., Afantitis A., Serra A., Fratello M., Papadiamantis A.G., Aidinis V., Lynch I., Greco D., Melagraki G. Advances in de Novo drug design: From conventional to machine learning methods // Int. J. Mol. Sci. MDPI AG, 2021. Vol. 22, № 4. P. 1676.
19. Hoffmann T., Gastreich M. The next level in chemical space navigation: going far beyond enumerable compound libraries // Drug Discov. Today. Elsevier BV, 2019. Vol. 24, № 5. P. 1148–1156.
20. Polishchuk P.G., Madzhidov T.I., Varnek A. Estimation of the size of drug-like chemical space based on GDB-17 data // J. Comput. Aided Mol. Des. Springer Science and Business Media LLC, 2013. Vol. 27, № 8. P. 675–679.

21. Southall N.T., Ajay. Kinase patent space visualization using chemical replacements // J. Med. Chem. American Chemical Society (ACS), 2006. Vol. 49, № 6. P. 2103–2109.
22. Murray C.W., Rees D.C. The rise of fragment-based drug discovery // Nat. Chem. Springer Science and Business Media LLC, 2009. Vol. 1, № 3. P. 187–192.
23. de Souza Neto L.R., Moreira-Filho J.T., Neves B.J., Maidana R.L.B.R., Guimarães A.C.R., Furnham N., Andrade C.H., Silva F.P. Jr. In silico strategies to support fragment-to-lead optimization in drug discovery // Front. Chem. Frontiers Media SA, 2020. Vol. 8. P. 93.
24. den Besten W., Lipford J.R. Prospecting for molecular glues // Nature chemical biology. Springer Science and Business Media LLC, 2020. Vol. 16, № 11. P. 1157–1158.
25. Tian C., Burgess K. PROTAC compatibilities, degrading cell-surface receptors, and the sticky problem of finding a molecular glue // ChemMedChem. Wiley, 2021. Vol. 16, № 2. P. 316–318.
26. Bondeson D.P., Smith B.E., Burslem G.M., Buhimschi A.D., Hines J., Jaime-Figueroa S., Wang J., Hamman B.D., Ishchenko A., Crews C.M. Lessons in PROTAC design from selective degradation with a promiscuous warhead // Cell Chem. Biol. Elsevier BV, 2018. Vol. 25, № 1. P. 78-87.e5.
27. Su M., Yang Q., Du Y., Feng G., Liu Z., Li Y., Wang R. Comparative Assessment of Scoring Functions: The CASF-2016 Update // J. Chem. Inf. Model. 2019. Vol. 59, № 2. P. 895–913.
28. Carpenter K.A., Huang X. Machine Learning-based virtual Screening and its applications to Alzheimer's drug discovery: A review // Curr. Pharm. Des. Bentham Science Publishers Ltd., 2018. Vol. 24, № 28. P. 3347–3358.
29. Moshawih S., Bu Z.H., Goh H.P., Kifli N., Lee L.H., Goh K.W., Ming L.C. Consensus holistic virtual screening for drug discovery: a novel machine learning model approach // J. Cheminform. Springer Science and Business Media LLC, 2024. Vol. 16, № 1. P. 62.

30. Ricci-Lopez J., Aguila S.A., Gilson M.K., Brizuela C.A. Improving structure-based virtual screening with ensemble docking and machine learning // J. Chem. Inf. Model. American Chemical Society (ACS), 2021. Vol. 61, № 11. P. 5362–5376.
31. Adeshina Y.O., Deeds E.J., Karanicolas J. Machine learning classification can reduce false positives in structure-based virtual screening // Proc. Natl. Acad. Sci. U. S. A. 2020. Vol. 117, № 31. P. 18477–18488.
32. Niefind K., Pütter M., Guerra B., Issinger O.-G., Schomburg D. GTP plus water mimic ATP in the active site of protein kinase CK2 // Nat. Struct. Biol. Springer Science and Business Media LLC, 1999. Vol. 6, № 12. P. 1100–1103.
33. Duong-Ly K.C., Peterson J.R. A high-throughput radiometric kinase assay // Methods Mol. Biol. 2016. Vol. 1360. P. 87–95.
34. Devaiah B.N., Lewis B.A., Cherman N., Hewitt M.C., Albrecht B.K., Robey P.G., Ozato K., Sims R.J. III, Singer D.S. BRD4 is an atypical kinase that phosphorylates Serine2 of the RNA Polymerase II carboxy-terminal domain // Proc. Natl. Acad. Sci. U. S. A. Proceedings of the National Academy of Sciences, 2012. Vol. 109, № 18. P. 6927–6932.
35. Cheng H.-C., Qi R.Z., Paudel H., Zhu H.-J. Regulation and function of protein kinases and phosphatases // Enzyme Res. Hindawi Limited, 2011. Vol. 2011. P. 794089.
36. Lahiry P., Torkamani A., Schork N.J., Hegele R.A. Kinase mutations in human disease: interpreting genotype-phenotype relationships // Nat. Rev. Genet. Springer Science and Business Media LLC, 2010. Vol. 11, № 1. P. 60–74.
37. Xu F., Du P., Shen H., Hu H., Wu Q., Xie J., Yu L. Correlated mutation analysis on the catalytic domains of serine/threonine protein kinases // PLoS One. Public Library of Science (PLoS), 2009. Vol. 4, № 6. P. e5913.
38. Chavda J., Bhatt H. Systemic review on B-RafV600E mutation as potential therapeutic target for the treatment of cancer // Eur. J. Med. Chem. Elsevier BV, 2020. Vol. 206, № 112675. P. 112675.
39. Zaman A., Wu W., Bivona T.G. Targeting oncogenic BRAF: Past, present, and future // Cancers (Basel). MDPI AG, 2019. Vol. 11, № 8. P. 1197.

40. Zhao K., Zhou X., Ding M. Molecular insight into mutation-induced conformational change in metastasic bowel cancer BRAF kinase domain and its implications for selective inhibitor design // J. Mol. Graph. Model. 2018. Vol. 79. P. 59–64.
41. Ferguson F.M., Gray N.S. Kinase inhibitors: the road ahead // Nat. Rev. Drug Discov. 2018. Vol. 17, № 5. P. 353–377.
42. Feng F.Y., Kothari V. Driven to metastasize: Kinases as potential therapeutic targets in prostate cancer // Proceedings of the National Academy of Sciences of the United States of America. Proceedings of the National Academy of Sciences, 2016. Vol. 113, № 3. P. 473–475.
43. Levitzki A. Protein kinase inhibitors as a therapeutic modality // Acc. Chem. Res. American Chemical Society (ACS), 2003. Vol. 36, № 6. P. 462–469.
44. Müller S., Chaikuad A., Gray N.S., Knapp S. The ins and outs of selective kinase inhibitor development // Nat. Chem. Biol. Springer Science and Business Media LLC, 2015. Vol. 11, № 11. P. 818–821.
45. Botta M. New frontiers in kinases: special issue // ACS Med. Chem. Lett. American Chemical Society (ACS), 2014. Vol. 5, № 4. P. 270.
46. Fedorov O., Müller S., Knapp S. The (un)targeted cancer kinome // Nat. Chem. Biol. Springer Science and Business Media LLC, 2010. Vol. 6, № 3. P. 166–169.
47. Alexander P.B., Wang X.-F. Resistance to receptor tyrosine kinase inhibition in cancer: molecular mechanisms and therapeutic strategies // Front. Med. Springer Science and Business Media LLC, 2015. Vol. 9, № 2. P. 134–138.
48. Villicaña C., Cruz G., Zurita M. The basal transcription machinery as a target for cancer therapy // Cancer Cell Int. Springer Science and Business Media LLC, 2014. Vol. 14, № 1. P. 18.
49. Rangachari D., To C., Shpilsky J.E., VanderLaan P.A., Kobayashi S.S., Mushajiang M., Lau C.J., Paweletz C.P., Oxnard G.R., Jänne P.A., Costa D.B. EGFR-mutated lung cancers resistant to osimertinib through EGFR C797S respond to first-generation reversible EGFR inhibitors but eventually acquire EGFR T790M/C797S in preclinical models and clinical samples // J. Thorac. Oncol. Elsevier BV, 2019. Vol. 14, № 11. P. 1995–2002.

50. Davis M.I., Hunt J.P., Herrgard S., Ciceri P., Wodicka L.M., Pallares G., Hocker M., Treiber D.K., Zarrinkar P.P. Comprehensive analysis of kinase inhibitor selectivity // Nat. Biotechnol. Springer Science and Business Media LLC, 2011. Vol. 29, № 11. P. 1046–1051.
51. Modi V., Dunbrack R.L. Jr. A structurally-validated multiple sequence alignment of 497 human protein kinase domains // Sci. Rep. Springer Science and Business Media LLC, 2019. Vol. 9, № 1. P. 19790.
52. Kanev G.K., de Graaf C., de Esch I.J.P., Leurs R., Würdinger T., Westerman B.A., Kooistra A.J. The landscape of atypical and eukaryotic protein kinases // Trends Pharmacol. Sci. Elsevier BV, 2019. Vol. 40, № 11. P. 818–832.
53. Kanev G.K., de Graaf C., Westerman B.A., de Esch I.J.P., Kooistra A.J. KLIFS: an overhaul after the first 5 years of supporting kinase research // Nucleic Acids Res. Oxford University Press (OUP), 2021. Vol. 49, № D1. P. D562–D569.
54. van Linden O.P.J., Kooistra A.J., Leurs R., de Esch I.J.P., de Graaf C. KLIFS: a knowledge-based structural database to navigate kinase-ligand interaction space // J. Med. Chem. American Chemical Society (ACS), 2014. Vol. 57, № 2. P. 249–277.
55. Dar A.C., Shokat K.M. The evolution of protein kinase inhibitors from antagonists to agonists of cellular signaling // Annu. Rev. Biochem. Annual Reviews, 2011. Vol. 80, № 1. P. 769–795.
56. Zhao Z., Xie L., Bourne P.E. Insights into the binding mode of MEK type-III inhibitors. A step towards discovering and designing allosteric kinase inhibitors across the human kinome // PLoS One. Public Library of Science (PLoS), 2017. Vol. 12, № 6. P. e0179936.
57. Gavrin L.K., Saiah E. Approaches to discover non-ATP site kinase inhibitors // Medchemcomm. Royal Society of Chemistry (RSC), 2013. Vol. 4, № 1. P. 41–51.
58. Lamba V., Ghosh I. New directions in targeting protein kinases: focusing upon true allosteric and bivalent inhibitors // Curr. Pharm. Des. 2012. Vol. 18, № 20. P. 2936–2945.
59. Okamoto K., Ikemori-Kawada M., Jestel A., von König K., Funahashi Y., Matsushima T., Tsuruoka A., Inoue A., Matsui J. Distinct binding mode of

multikinase inhibitor lenvatinib revealed by biochemical characterization // ACS Med. Chem. Lett. American Chemical Society (ACS), 2015. Vol. 6, № 1. P. 89–94.

60. Gower C.M., Chang M.E.K., Maly D.J. Bivalent inhibitors of protein kinases // Crit. Rev. Biochem. Mol. Biol. Informa UK Limited, 2014. Vol. 49, № 2. P. 102–115.
61. Lee S., Kim J., Jo J., Chang J.W., Sim J., Yun H. Recent advances in development of hetero-bivalent kinase inhibitors // Eur. J. Med. Chem. Elsevier BV, 2021. Vol. 216, № 113318. P. 113318.
62. Abdeldayem A., Raouf Y.S., Constantinescu S.N., Moriggl R., Gunning P.T. Advances in covalent kinase inhibitors // Chem. Soc. Rev. Royal Society of Chemistry (RSC), 2020. Vol. 49, № 9. P. 2617–2687.
63. Sun X., Gao H., Yang Y., He M., Wu Y., Song Y., Tong Y., Rao Y. PROTACs: great opportunities for academia and industry // Signal Transduct. Target. Ther. Springer Science and Business Media LLC, 2019. Vol. 4, № 1. P. 64.
64. Toure M., Crews C.M. Small-molecule PROTACS: New approaches to protein degradation // Angew. Chem. Int. Ed Engl. Wiley, 2016. Vol. 55, № 6. P. 1966–1973.
65. Walczak M.J., Petzold G., Thomä N.H. Targeted protein degradation: You can glue it too! // Nature chemical biology. 2017. Vol. 13, № 5. P. 452–453.
66. Bai N., Miller S.A., Andrianov G.V., Yates M., Kirubakaran P., Karanicolas J. Rationalizing PROTAC-mediated ternary complex formation using Rosetta // J. Chem. Inf. Model. American Chemical Society (ACS), 2021. Vol. 61, № 3. P. 1368–1382.
67. Smith B.E., Wang S.L., Jaime-Figueroa S., Harbin A., Wang J., Hamman B.D., Crews C.M. Differential PROTAC substrate specificity dictated by orientation of recruited E3 ligase // Nat. Commun. Springer Science and Business Media LLC, 2019. Vol. 10, № 1. P. 131.
68. Qian F., Engst S., Yamaguchi K., Yu P., Won K.-A., Mock L., Lou T., Tan J., Li C., Tam D., Lougheed J., Yakes F.M., Bentzien F., Xu W., Zaks T., Wooster R., Greshock J., Joly A.H. Inhibition of tumor cell growth, invasion, and metastasis by EXEL-2880 (XL880, GSK1363089), a novel inhibitor of HGF and VEGF receptor

tyrosine kinases // Cancer Res. American Association for Cancer Research (AACR), 2009. Vol. 69, № 20. P. 8009–8016.

69. Chen J., Luo X., Qiu H., Mackey V., Sun L., Ouyang X. Drug discovery and drug marketing with the critical roles of modern administration // Am. J. Transl. Res. 2018. Vol. 10, № 12. P. 4302–4312.
70. Reymond J.-L. The chemical space project // Acc. Chem. Res. American Chemical Society (ACS), 2015. Vol. 48, № 3. P. 722–730.
71. Lyu J., Wang S., Balius T.E., Singh I., Levit A., Moroz Y.S., O'Meara M.J., Che T., Algaa E., Tolmachova K., Tolmachev A.A., Shoichet B.K., Roth B.L., Irwin J.J. Ultra-large library docking for discovering new chemotypes // Nature. 2019. Vol. 566, № 7743. P. 224–229.
72. Martin Y.C., Kofron J.L., Traphagen L.M. Do structurally similar molecules have similar biological activity? // J. Med. Chem. American Chemical Society (ACS), 2002. Vol. 45, № 19. P. 4350–4358.
73. Bickerton G.R., Paolini G.V., Besnard J., Muresan S., Hopkins A.L. Quantifying the chemical beauty of drugs // Nat. Chem. 2012. Vol. 4, № 2. P. 90–98.
74. Congreve M., Carr R., Murray C., Jhoti H. A 'Rule of Three' for fragment-based lead discovery? // Drug Discov. Today. Elsevier BV, 2003. Vol. 8, № 19. P. 876–877.
75. Lipinski C.A., Lombardo F., Dominy B.W., Feeney P.J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings // Adv. Drug Deliv. Rev. Elsevier BV, 2012. Vol. 64. P. 4–17.
76. Baell J.B., Holloway G.A. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays // J. Med. Chem. American Chemical Society (ACS), 2010. Vol. 53, № 7. P. 2719–2740.
77. Coley C.W., Rogers L., Green W.H., Jensen K.F. SCScore: Synthetic complexity learned from a reaction corpus // J. Chem. Inf. Model. 2018. Vol. 58, № 2. P. 252–261.

78. Ertl P., Schuffenhauer A. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions // J. Cheminform. Springer Science and Business Media LLC, 2009. Vol. 1, № 1. P. 8.
79. Trott O., Olson A.J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading // J. Comput. Chem. Wiley, 2010. Vol. 31, № 2. P. 455–461.
80. Bitencourt-Ferreira G., de Azevedo W.F. Jr. Machine learning to predict binding affinity // Methods Mol. Biol. 2019. Vol. 2053. P. 251–273.
81. Jones D., Kim H., Zhang X., Zemla A., Stevenson G., Bennett W.F.D., Kirshner D., Wong S.E., Lightstone F.C., Allen J.E. Improved protein-ligand binding affinity prediction with structure-based deep fusion inference // J. Chem. Inf. Model. American Chemical Society (ACS), 2021. Vol. 61, № 4. P. 1583–1592.
82. Kundu I., Paul G., Banerjee R. A machine learning approach towards the prediction of protein-ligand binding affinity based on fundamental molecular properties // RSC Adv. Royal Society of Chemistry (RSC), 2018. Vol. 8, № 22. P. 12127–12137.
83. Korb O., Stützle T., Exner T.E. PLANTS: Application of ant colony optimization to structure-based drug design // Ant Colony Optimization and Swarm Intelligence. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006. P. 247–258.
84. David L., Arús-Pous J., Karlsson J., Engkvist O., Bjerrum E.J., Kogej T., Kriegl J.M., Beck B., Chen H. Applications of deep-learning in exploiting large-scale and heterogeneous compound data in industrial pharmaceutical research // Front. Pharmacol. Frontiers Media SA, 2019. Vol. 10. P. 1303.
85. Segler M.H.S., Kogej T., Tyrchan C., Waller M.P. Generating focused molecule libraries for drug discovery with recurrent neural networks // ACS Cent. Sci. American Chemical Society (ACS), 2018. Vol. 4, № 1. P. 120–131.
86. Bohacek R.S., McMartin C., Guida W.C. The art and practice of structure-based drug design: A molecular modeling perspective // Med. Res. Rev. Wiley, 1996. Vol. 16, № 1. P. 3–50.
87. Chevillard F., Rimmer H., Betti C., Pardon E., Ballet S., van Hilten N., Steyaert J., Diederich W.E., Kolb P. Binding-site compatible fragment growing applied to the

design of β2-adrenergic receptor ligands // J. Med. Chem. 2018. Vol. 61, № 3. P. 1118–1129.

88. Kumar A., Voet A., Zhang K.Y.J. Fragment based drug design: from experimental to computational approaches // Curr. Med. Chem. 2012. Vol. 19, № 30. P. 5128–5147.
89. Bissaro M., Sturlese M., Moro S. The rise of molecular simulations in fragment-based drug design (FBDD): an overview // Drug Discov. Today. Elsevier BV, 2020. Vol. 25, № 9. P. 1693–1701.
90. Kirsch P., Hartman A.M., Hirsch A.K.H., Empting M. Concepts and core principles of fragment-based drug design // Molecules. MDPI AG, 2019. Vol. 24, № 23. P. 4309.
91. Troelsen N.S., Clausen M.H. Frontispiece: Library design strategies to accelerate fragment-based drug discovery // Chemistry. Wiley, 2020. Vol. 26, № 50.
92. Liu X., Ouyang S., Yu B., Liu Y., Huang K., Gong J., Zheng S., Li Z., Li H., Jiang H. PharmMapper server: a web server for potential drug target identification using pharmacophore mapping approach // Nucleic Acids Res. Oxford University Press (OUP), 2010. Vol. 38, № Web Server issue. P. W609-14.
93. Hopkins A.L., Keserü G.M., Leeson P.D., Rees D.C., Reynolds C.H. The role of ligand efficiency metrics in drug discovery // Nat. Rev. Drug Discov. Springer Science and Business Media LLC, 2014. Vol. 13, № 2. P. 105–121.
94. Bancet A., Raingeval C., Lomberget T., Le Borgne M., Guichou J.-F., Krimm I. Fragment linking strategies for structure-based drug design // J. Med. Chem. American Chemical Society (ACS), 2020. Vol. 63, № 20. P. 11420–11435.
95. Réda C., Kaufmann E., Delahaye-Duriez A. Machine learning applications in drug development // Comput. Struct. Biotechnol. J. 2020. Vol. 18. P. 241–252.
96. Vamathevan J., Clark D., Czodrowski P., Dunham I., Ferran E., Lee G., Li B., Madabhushi A., Shah P., Spitzer M., Zhao S. Applications of machine learning in drug discovery and development // Nat. Rev. Drug Discov. 2019. Vol. 18, № 6. P. 463–477.
97. Wang Z., Sun H., Yao X., Li D., Xu L., Li Y., Tian S., Hou T. Comprehensive evaluation of ten docking programs on a diverse set of protein-ligand complexes: the

prediction accuracy of sampling power and scoring power // Phys. Chem. Chem. Phys. 2016. Vol. 18, № 18. P. 12964–12975.

98. Boldini D., Friedrich L., Kuhn D., Sieber S.A. Machine Learning Assisted Hit Prioritization for High Throughput Screening in Drug Discovery // ACS Cent Sci. 2024. Vol. 10, № 4. P. 823–832.
99. Shi S., Fu L., Yi J., Yang Z., Zhang X., Deng Y., Wang W., Wu C., Zhao W., Hou T., Zeng X., Lyu A., Cao D. ChemFH: an integrated tool for screening frequent false positives in chemical biology and drug discovery // Nucleic Acids Res. 2024. Vol. 52, № W1. P. W439–W449.
100. Coley C.W., Rogers L., Green W.H., Jensen K.F. Computer-assisted retrosynthesis based on molecular similarity // ACS Cent. Sci. American Chemical Society (ACS), 2017. Vol. 3, № 12. P. 1237–1245.
101. Lee A.A., Yang Q., Sresht V., Bolgar P., Hou X., Klug-McLeod J.L., Butler C.R. Molecular Transformer unifies reaction prediction and retrosynthesis across pharma chemical space // Chem. Commun. (Camb.). Royal Society of Chemistry (RSC), 2019. Vol. 55, № 81. P. 12152–12155.
102. Daylight theory: SMARTS - A language for describing molecular patterns [Electronic resource]. URL: https://www.daylight.com/dayhtml/doc/theory/theory.smarts.html (accessed: 26.10.2024).
103. Hawkins P.C.D., Skillman A.G., Warren G.L., Ellingson B.A., Stahl M.T. Conformer generation with OMEGA: algorithm and validation using high quality structures from the Protein Databank and Cambridge Structural Database // J. Chem. Inf. Model. 2010. Vol. 50, № 4. P. 572–584.
104. Alford R.F., Leaver-Fay A., Jeliazkov J.R., O'Meara M.J., DiMaio F.P., Park H., Shapovalov M.V., Renfrew P.D., Mulligan V.K., Kappel K., Labonte J.W., Pacella M.S., Bonneau R., Bradley P., Dunbrack R.L. Jr, Das R., Baker D., Kuhlman B., Kortemme T., et al. The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design // J. Chem. Theory Comput. 2017. Vol. 13, № 6. P. 3031–3048.

105. Park H., Zhou G., Baek M., Baker D., DiMaio F. Force Field Optimization Guided by Small Molecule Crystal Lattice Data Enables Consistent Sub-Angstrom Protein-Ligand Docking // J. Chem. Theory Comput. 2021. Vol. 17, № 3. P. 2000–2010.

106. Zorba A., Nguyen C., Xu Y., Starr J., Borzilleri K., Smith J., Zhu H., Farley K.A., Ding W., Schiemer J., Feng X., Chang J.S., Uccello D.P., Young J.A., Garcia-Irrizary C.N., Czabaniuk L., Schuff B., Oliver R., Montgomery J., et al. Delineating the role of cooperativity in the design of potent PROTACs for BTK // Proc. Natl. Acad. Sci. U. S. A. 2018. Vol. 115, № 31. P. E7285–E7292.

107. Hawkins P.C.D., Nicholls A. Conformer generation with OMEGA: learning from the data set and the analysis of failures // J. Chem. Inf. Model. American Chemical Society (ACS), 2012. Vol. 52, № 11. P. 2919–2936.

108. Hawkins P.C.D., Skillman A.G., Nicholls A. Comparison of shape-matching and docking as virtual screening tools // J. Med. Chem. 2007. Vol. 50, № 1. P. 74–82.

109. Weininger D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules // J. Chem. Inf. Comput. Sci. American Chemical Society (ACS), 1988. Vol. 28, № 1. P. 31–36.

110. Raschka S. MLxtend: Providing machine learning and data science utilities and extensions to Python’s scientific computing stack // J. Open Source Softw. The Open Journal, 2018. Vol. 3, № 24. P. 638.

111. Steinegger M., Söding J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets // Nat. Biotechnol. 2017. Vol. 35, № 11. P. 1026–1028.

112. Akiba T., Sano S., Yanase T., Ohta T., Koyama M. Optuna: A Next-generation Hyperparameter Optimization Framework // arXiv [cs.LG]. 2019.

113. Kirubakaran P., Morton G., Zhang P., Zhang H., Gordon J., Abou-Gharbia M., Issa J.-P.J., Wu J., Childers W., Karanicolas J. Comparative modeling of CDK9 inhibitors to explore selectivity and structure-activity relationships // bioRxiv. bioRxiv, 2020.

114. Zhang H., Pandey S., Travers M., Sun H., Morton G., Madzo J., Chung W., Khowsathit J., Perez-Leal O., Barrero C.A., Merali C., Okamoto Y., Sato T., Pan J., Garriga J., Bhanu N.V., Simithy J., Patel B., Huang J., et al. Targeting CDK9

reactivates epigenetically silenced genes in cancer // Cell. Elsevier BV, 2018. Vol. 175, № 5. P. 1244-1258.e26.

115. Gazzard L., Williams K., Chen H., Axford L., Blackwood E., Burton B., Chapman K., Crackett P., Drobnick J., Ellwood C., Epler J., Flagella M., Gancia E., Gill M., Goodacre S., Halladay J., Hewitt J., Hunt H., Kintz S., et al. Mitigation of acetylcholine esterase activity in the 1,7-diazacarbazole series of inhibitors of checkpoint kinase 1 // J. Med. Chem. American Chemical Society (ACS), 2015. Vol. 58, № 12. P. 5053–5074.

116. Bramson H.N., Corona J., Davis S.T., Dickerson S.H., Edelstein M., Frye S.V., Gampe R.T. Jr, Harris P.A., Hassell A., Holmes W.D., Hunter R.N., Lackey K.E., Lovejoy B., Luzzio M.J., Montana V., Rocque W.J., Rusnak D., Shewchuk L., Veal J.M., et al. Oxindole-based inhibitors of cyclin-dependent kinase 2 (CDK2): design, synthesis, enzymatic activities, and X-ray crystallographic analysis // J. Med. Chem. American Chemical Society (ACS), 2001. Vol. 44, № 25. P. 4339–4358.

117. Gajiwala K.S., Feng J., Ferre R., Ryan K., Brodsky O., Weinrich S., Kath J.C., Stewart A. Insights into the aberrant activity of mutant EGFR kinase domain and drug recognition // Structure. Elsevier BV, 2013. Vol. 21, № 2. P. 209–219.

118. Ni Z.-J., Barsanti P., Brammeier N., Diebes A., Poon D.J., Ng S., Pecchi S., Pfister K., Renhowe P.A., Ramurthy S., Wagman A.S., Bussiere D.E., Le V., Zhou Y., Jansen J.M., Ma S., Gesner T.G. 4-(Aminoalkylamino)-3-benzimidazole-quinolinones as potent CHK-1 inhibitors // Bioorg. Med. Chem. Lett. Elsevier BV, 2006. Vol. 16, № 12. P. 3121–3124.

119. Kopecky D.J., Hao X., Chen Y., Fu J., Jiao X., Jaen J.C., Cardozo M.G., Liu J., Wang Z., Walker N.P.C., Wesche H., Li S., Farrelly E., Xiao S.-H., Kayser F. Identification and optimization of N3,N6-diaryl-1H-pyrazolo[3,4-d]pyrimidine-3,6-diamines as a novel class of ACK1 inhibitors // Bioorg. Med. Chem. Lett. Elsevier BV, 2008. Vol. 18, № 24. P. 6352–6356.

120. Barelier S., Cummings J.A., Rauwerdink A.M., Hitchcock D.S., Farelli J.D., Almo S.C., Raushel F.M., Allen K.N., Shoichet B.K. Substrate deconstruction and the

nonadditivity of enzyme recognition // J. Am. Chem. Soc. American Chemical Society (ACS), 2014. Vol. 136, № 20. P. 7374–7382.

121. Jencks W.P. On the attribution and additivity of binding energies // Proc. Natl. Acad. Sci. U. S. A. Proceedings of the National Academy of Sciences, 1981. Vol. 78, № 7. P. 4046–4050.

122. Kramer C., Fuchs J.E., Liedl K.R. Strong nonadditivity as a key structure-activity relationship feature: distinguishing structural changes from assay artifacts // J. Chem. Inf. Model. American Chemical Society (ACS), 2015. Vol. 55, № 3. P. 483–494.

123. Nasief N.N., Hangauer D. Additivity or cooperativity: which model can predict the influence of simultaneous incorporation of two or more functionalities in a ligand molecule? // Eur. J. Med. Chem. Elsevier BV, 2015. Vol. 90. P. 897–915.

124. Cockroft S.L., Hunter C.A. Chemical double-mutant cycles: Dissecting non-covalent interactions // ChemInform. Wiley, 2007. Vol. 38, № 17.

125. Bembenek S.D., Tounge B.A., Reynolds C.H. Ligand efficiency and fragment-based drug discovery // Drug Discov. Today. Elsevier BV, 2009. Vol. 14, № 5–6. P. 278–283.

126. van De Waterbeemd H., Smith D.A., Beaumont K., Walker D.K. Property-based design: optimization of drug absorption and pharmacokinetics // J. Med. Chem. American Chemical Society (ACS), 2001. Vol. 44, № 9. P. 1313–1333.

127. Ballester P.J. Selecting machine-learning scoring functions for structure-based virtual screening // Drug Discov. Today Technol. Elsevier BV, 2019. Vol. 32–33. P. 81–87.

128. Chaput L., Martinez-Sanz J., Saettel N., Mouawad L. Benchmark of four popular virtual screening programs: construction of the active/decoy dataset remains a major determinant of measured performance // J. Cheminform. Springer Science and Business Media LLC, 2016. Vol. 8, № 1. P. 56.

129. Lagarde N., Zagury J.-F., Montes M. Benchmarking data sets for the evaluation of virtual ligand screening methods: Review and perspectives // J. Chem. Inf. Model. American Chemical Society (ACS), 2015. Vol. 55, № 7. P. 1297–1307.

130. Réau M., Langenfeld F., Zagury J.-F., Lagarde N., Montes M. Decoys Selection in Benchmarking Datasets: Overview and Perspectives // Front. Pharmacol. 2018. Vol. 9. P. 11.

131. Stein R.M., Yang Y., Balius T.E., O'Meara M.J., Lyu J., Young J., Tang K., Shoichet B.K., Irwin J.J. Property-unmatched decoys in docking benchmarks // J. Chem. Inf. Model. American Chemical Society (ACS), 2021. Vol. 61, № 2. P. 699–714.

132. Nowak R.P., DeAngelo S.L., Buckley D., He Z., Donovan K.A., An J., Safaee N., Jedrychowski M.P., Ponthier C.M., Ishoey M., Zhang T., Mancias J.D., Gray N.S., Bradner J.E., Fischer E.S. Plasticity in binding confers selectivity in ligand-induced protein degradation // Nat. Chem. Biol. Springer Science and Business Media LLC, 2018. Vol. 14, № 7. P. 706–714.

133. Qin C., Hu Y., Zhou B., Fernandez-Salas E., Yang C.-Y., Liu L., McEachern D., Przybranowski S., Wang M., Stuckey J., Meagher J., Bai L., Chen Z., Lin M., Yang J., Ziazadeh D.N., Xu F., Hu J., Xiang W., et al. Discovery of QCA570 as an exceptionally potent and efficacious proteolysis targeting chimera (PROTAC) degrader of the bromodomain and extra-terminal (BET) proteins capable of inducing complete and durable tumor regression // J. Med. Chem. 2018. Vol. 61, № 15. P. 6685–6704.

134. Collie G.W., Koh C.M., O'Neill D.J., Stubbs C.J., Khurana P., Eddershaw A., Snijder A., Mauritzson F., Barlind L., Dale I.L., Shaw J., Phillips C., Hennessy E.J., Cheung T., Narvaez A.J. Structural and molecular insight into resistance mechanisms of first generation cMET inhibitors // ACS Med. Chem. Lett. American Chemical Society (ACS), 2019. Vol. 10, № 9. P. 1322–1327.

135. Heinzlmeir S., Kudlinzki D., Sreeramulu S., Klaeger S., Gande S.L., Linhard V., Wilhelm M., Qiao H., Helm D., Ruprecht B., Saxena K., Médard G., Schwalbe H., Kuster B. Chemical proteomics and structural biology define EPHA2 inhibition by clinical kinase drugs // ACS Chem. Biol. American Chemical Society (ACS), 2016. Vol. 11, № 12. P. 3400–3411.

136. Galdeano C., Gadd M.S., Soares P., Scaffidi S., Van Molle I., Birced I., Hewitt S., Dias D.M., Ciulli A. Structure-guided design and optimization of small molecules targeting the protein-protein interaction between the von Hippel-Lindau (VHL) E3 ubiquitin ligase and the hypoxia inducible factor (HIF) alpha subunit with in vitro nanomolar affinities // J. Med. Chem. American Chemical Society (ACS), 2014. Vol. 57, № 20. P. 8657–8663.

137. Bietz S., Urbaczek S., Schulz B., Rarey M. Protoss: a holistic approach to predict tautomers and protonation states in protein-ligand complexes // J. Cheminform. 2014. Vol. 6. P. 12.

138. Mysinger M.M., Carchia M., Irwin J.J., Shoichet B.K. Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking // J. Med. Chem. 2012. Vol. 55, № 14. P. 6582–6594.

139. Fassio A.V., Shub L., Ponzoni L., McKinley J., O'Meara M.J., Ferreira R.S., Keiser M.J., de Melo Minardi R.C. Prioritizing Virtual Screening with Interpretable Interaction Fingerprints // J. Chem. Inf. Model. 2022. Vol. 62, № 18. P. 4300–4318.

140. Durrant J.D., McCammon J.A. BINANA: a novel algorithm for ligand-binding characterization // J. Mol. Graph. Model. 2011. Vol. 29, № 6. P. 888–893.

141. Wójcikowski M., Zielenkiewicz P., Siedlecki P. Open Drug Discovery Toolkit (ODDT): a new open-source player in the drug discovery field // J. Cheminform. 2015. Vol. 7. P. 26.

142. Bauer M.R., Ibrahim T.M., Vogel S.M., Boeckler F.M. Evaluation and optimization of virtual screening workflows with DEKOIS 2.0 – A public library of challenging docking benchmark sets // J. Chem. Inf. Model. American Chemical Society (ACS), 2013. Vol. 53, № 6. P. 1447–1462.

143. Zhang X., Zhang O., Shen C., Qu W., Chen S., Cao H., Kang Y., Wang Z., Wang E., Zhang J., Deng Y., Liu F., Wang T., Du H., Wang L., Pan P., Chen G., Hsieh C.-Y., Hou T. Efficient and accurate large library ligand docking with KarmaDock // Nat Comput Sci. 2023. Vol. 3, № 9. P. 789–804.

144. Zhang X. DEKOIS2.0 for KarmaDock. Zenodo, 2023.

145. Shen C., Zhang X., Deng Y., Gao J., Wang D., Xu L., Pan P., Hou T., Kang Y. Boosting Protein-Ligand Binding Pose Prediction and Virtual Screening Based on Residue-Atom Distance Likelihood Potential and Graph Transformer // J. Med. Chem. 2022. Vol. 65, № 15. P. 10691–10706.

146. McNutt A.T., Francoeur P., Aggarwal R., Masuda T., Meli R., Ragoza M., Sunseri J., Koes D.R. GNINA 1.0: molecular docking with deep learning // J. Cheminform. 2021. Vol. 13, № 1. P. 43.

147. Pan X., Wang H., Zhang Y., Wang X., Li C., Ji C., Zhang J.Z.H. AA-Score: a New Scoring Function Based on Amino Acid-Specific Interaction for Molecular Docking // J. Chem. Inf. Model. 2022. Vol. 62, № 10. P. 2499–2509.

148. Yang Y., Zhang Y., Hua Y., Chen X., Fan Y., Wang Y., Liang L., Deng C., Lu T., Chen Y., Liu H. In silico design and analysis of a kinase-focused combinatorial library considering diversity and quality // J. Chem. Inf. Model. American Chemical Society (ACS), 2020. Vol. 60, № 1. P. 92–107.

149. Sydow D., Schmiel P., Mortier J., Volkamer A. KinFragLib: Exploring the kinase inhibitor space using subpocket-focused fragmentation and recombination // J. Chem. Inf. Model. American Chemical Society (ACS), 2020. Vol. 60, № 12. P. 6081–6094.

150. Hoffer L., Voitovich Y.V., Raux B., Carrasco K., Muller C., Fedorov A.Y., Derviaux C., Amouric A., Betzi S., Horvath D., Varnek A., Collette Y., Combes S., Roche P., Morelli X. Integrated strategy for lead optimization based on fragment growing: The diversity-oriented-target-focused-synthesis approach // J. Med. Chem. American Chemical Society (ACS), 2018. Vol. 61, № 13. P. 5719–5732.

151. Sadybekov A.A., Sadybekov A.V., Liu Y., Iliopoulos-Tsoutsouvas C., Huang X.-P., Pickett J., Houser B., Patel N., Tran N.K., Tong F., Zvonok N., Jain M.K., Savych O., Radchenko D.S., Nikas S.P., Petasis N.A., Moroz Y.S., Roth B.L., Makriyannis A., et al. Synthon-based ligand discovery in virtual libraries of over 11 billion compounds // Nature. Springer Science and Business Media LLC, 2022. Vol. 601, № 7893. P. 452–459.

152. Wang L., Chambers J., Abel R. Protein–ligand binding free energy calculations with FEP+ // Methods in Molecular Biology. New York, NY: Springer New York, 2019. Vol. 2022. P. 201–232.

ПРИЛОЖЕНИЕ А

Схемы синтеза, использованные для создания библиотек фрагментов

Библиотека	SMARTS
MC180295	
R1	[#6:4]-[#6:2](=[O:1])-[#6:3]Br>>[#6:4]-[#6:2](=[O:1])-[#6:3]-1=[#6](-[#7])-[#7]=[#6](-[#7])-[#16]-1
R2	[#6:1]-[#7H2:2]>>[#6:1]-[#7:2]-[#6]-1=[#7]-[#6](-[#7])=[#6]-[#16]-1
BDBM50091276	
R1	[#8]-[#5](-[#8])-[*:1]>>[*:1]-[#6]-1=[#6]-[#6]-2=[#6](-[#7]-[#6]-3=[#6]-[#7]=[#6]-[#6]=[#6]-2-3)-[#7]=[#6]-1
R2	[#7:6]-[#6:5]-1=[#6:7]-[#7:1]=[#6:2]-[#6:3]=[#6:4]-1I>>[#7:6]-1-[#6:5]-2=[#6:4](-[#6:3]=[#6:2]-[#7:1]=[#6:7]-2)-[#6]-2=[#6]-1-[#7]=[#6]-[#6]=[#6]-2
BDBM7773	
R1	[#7:1]-[#6:2]-1=[#6:3]-[#6:4]=[#6:5]-[#6:6]=[#6:7]-1>>O=[#6]-1-[#6]-[#6:7]-2=[#6:6]-[#6:5]=[#6:4]-[#6:3]=[#6:2]-2-[#7:1]-1
R2	[#7:1]-[#6:2]-1=[#6:3]-[#6:4]=[#6:5]-[#6:6]=[#6:7]-1.[#6H3:10]-[#7:9]-[*:8]>>[*:8]-[#7:9]\[#6:10]=[#6]-1/[#6](=O)-[#7:1]-[#6:2]-2=[#6:3]-[#6:4]=[#6:5]-[#6:6]=[#6:7]-1-2
Дакомитиниб	
R1	[#6:2]-[#7:1]>>[#6:2]-[#7:1]-[#6]-1=[#7]-[#6]=[#7]-[#6]-2=[#6]-1-[#6]=[#6](-[#7]-[#6](=O)-[#6]-[#6]-[#6]-[#7]-1-[#6]-[#6]-[#6]-[#6]-[#6]-1)-[#6]=[#6]-2
R2	[#6:2]-[#8:1]>>[#6:2]-[#8:1]-[#6]-1=[#6]-[#6]-2=[#6](-[#6]=[#7]-[#6]=[#7]-2)-[#6]=[#6]-1-[#7]-[#6](=O)-[#6]-[#6]-[#6]-[#7]-1-[#6]-[#6]-[#6]-[#6]-[#6]-1
CHIR-124	
R1	[O:6]=[#6:5]-1-[#6:7]~[#6:8]-[#7:1]-[#6:2](=[O:3])-[#8:4]-1>>[#8:6]-[#6:5]-1=[#6:4]-[#6:2](=[O:3])-[#7:1]-[#6:8]~[#6:7]-1

R2	[#6]-[#8]-[#6:3](=[O:4])-[#6:2]-[*:1]>>[#8]-[#6]-1=[#6:2](-[*:1])-[#6:3](=[O:4])-[#7]-[#6]-2=[#6]-[#6]=[#6]-[#6]=[#6]-1-2
BDBM50246212	
R1	[#7;A;H2,H3+,H4+:2][*:1]>>[#7;A][#6]-1=[#7]-[#6]=[#6]-2-[#6](-[#7H1:2]-[*:1])=[#7]-[#7]-[#6]-2=[#7]-1
R2	[#7;A;H2,H3+,H4+:2][*:1]>>[#7;A][#6]-1=[#7]-[#7]-[#6]-2=[#7]-[#6](-[#7:2]-[*:1])=[#7]-[#6]=[#6]-1-2
R3	[#7;A][#7;A;H1,H2+,H3+:2][*:1]>>[#7;A][#6]-1=[#7]-[#7:2](-[*:1])-[#6]-2=[#7]-[#6]([#7;A])=[#7]-[#6]=[#6]-1-2

ПРИЛОЖЕНИЕ Б

Размер библиотек фрагментов, использованных для эксперимента по аддитивности

Библиотека	Количество R1	Количество R2	Количество R3	Размер библиотеки
MC180295	77	2 221	-	171 017
BDBM50091276	9	2 500	-	22 500
BDBM7773	650	650	-	422 500
Дакомитиниб	1 162	1 045	-	1 214 290
CHIR-124	286	33	1 563	14 751 594
BDBM50246212	150	100	45	675 000

ПРИЛОЖЕНИЕ В

Фильтры, использованные для нового тренировочного набора vScreenML

Фильтр	Описание	Количество отфильтрованных
Физико-химические свойства		
Молекулярная масса (МА)	300 < МА < 500	31 036
Количество атомов (КА)	КА > 19	29 219
LogP	0 < LogP < 5	14 496
2D структурные свойства		
Наименьший размер цикла (НРЦ)	0 < НРЦ <= 6	12 153
Количество объединенных колец (КОК)	КОК < 4	11 556
Размер неразветвленной цепи	Отсутствие подструктуры `[D2&!R]~[D2&!R]~[D2&!R]~[D1,D2&!R]`	9 837
Наличие углевода	Отсутствие подструктуры `[#6&R1]-[#8&R]-[#6&R1]-[$([#6&R]~[#8&H&!R])]-[$([#6&R]~[#8&H&!R])]`	9 725
Наличие АТФ-подобной структуры	Отсутствие нуклеотида, сахара и фосфатных групп	9 465
Наличие атомов азота и кислорода	По крайней мере наличие одного атома азота и одного атома кислорода	8 407
Наличие нежелательных атомов	Только C, N, O, S, F, Cl or Br могут быть представлены в структуре	7 879

Количество галогенных атомов (КГФ)	КГА < 4	7 666
3D структурные свойства		
Полнота структурная (ПС)	ПС = 1.0	7 427
Ранжирование соответствия модели (РСМ)	РСМ > 0.35	6 077
Ранжирование геометрии модели (РГМ)	РГМ > 0.2	3 484
Ближайшая дистанция лиганда к любому цистеину (d)	d > 1.9 A	3 387
Ближайшая дистанция лиганда к другому лиганду (d)	d > 12.0 A	1 411
Наличие нуклеиновых кислот (ННК)	ННК = False	1 407

ПРИЛОЖЕНИЕ Г

Набор дескрипторов для vScreenML

Зеленым цветом отмечены наиболее информативные параметры, которые были использованы для обучения финальной модели.

PyRosetta		
TotalExposedSasa	FaAtrInteraction	HBondScInteraction
TotalBSA	FaRepInteraction	GenBonded
InterfaceHydrophobicSasa	FaSolInteraction	HBInterface
InterfacePolarSasa	FaElecInteraction	InterfaceUnsat
InteractionScore	HBondBbScInteraction	
PyRosetta and RDKit		
NH1	NH3	carbonyl O
NH2	OH	carboxylate O
BINANA		
SideFlexAlpha	PiPi	TotalHphobics
SideFlexBeta	TStacking	
SideFlexOther	CationPi	
BackFlexAlpha	SaltBridge	
BackFlexBeta	TotalElec	
BackFlexOther	TotalHBond	
RF-score		
6.6	9.6	17.6
6.7	9.7	17.7
6.8	9.8	17.8
6.16	9.16	17.16
7.6	15.6	35.6
7.7	15.7	35.7
7.8	15.8	35.8

7.16	15.16	35.16
8.6	16.6	
8.7	16.7	
8.8	16.8	
8.16	16.16	
RDKit		
exactmw	NumHeterocycles	chi1n
amw	NumAromaticHeterocycles	chi2n
lipinskiHBA	NumSaturatedHeterocycles	chi3n
lipinskiHBD	NumAliphaticHeterocycles	chi4n
NumRotatableBonds	NumSpiroAtoms	hallKierAlpha
NumHBD	NumBridgeheadAtoms	kappa1
NumHBA	labuteASA	kappa2
NumHeavyAtoms	tpsa	kappa3
NumAtoms	CrippenClogP	Phi
NumHeteroatoms	CrippenMR	
NumAmideBonds	chi0v	
FractionCSP3	chi1v	
NumRings	chi2v	
NumAromaticRings	chi3v	
NumAliphaticRings	chi4v	
NumSaturatedRings	chi0n	
LUNA		
Proximal	Chalcogen bond	Face-to-face pi-stacking
Hydrogen bond	Chalcogen-pi	Face-to-edge pi-stacking
Ionic	Halogen-pi	Face-to-slope pi-stacking

Salt bridge	Orthogonal multipolar	Edge-to-edge pi-stacking
Cation-pi	Parallel multipolar	Edge-to-face pi-stacking
Hydrophobic	Antiparallel multipolar	Edge-to-slope pi-stacking
Halogen bond	Tilted multipolar	Displaced face-to-face pi-stacking
Repulsive	Multipolar	Displaced face-to-edge pi-stacking
Water-bridged hydrogen bond	Cation-nucleophile	Displaced face-to-slope pi-stacking
Amide-aromatic stacking	Anion-electrophile	
Weak hydrogen bond	Unfavorable anion-nucleophile	
Covalent bond	Unfavorable cation-electrophile	
Atom overlap	Unfavorable nucleophile-nucleophile	
Van der Waals clash	Unfavorable electrophile-electrophile	
Van der Waals	Pi-stacking	
PocketDruggability		
C_RESIDUE	hydrophobic_kyte	p_aliphatic_residue
INERTIA_3	hydrophobicity_pocket	p_aromatic_residue
SMALLEST_SIZE	p_Ccoo	p_negative_residue
SURFACE_HULL	p_N_atom	
VOLUME_HULL	p_Ooh	

ПРИЛОЖЕНИЕ Д

Значения EF1% для KarmaDock для DEKOIS2

Желтом цветом отмечены мишени, которые указывают на возможную утечку тестируемого набора в тренировочный.

Мишень	EF1%	Последовательность представлена в тренировочном наборе	Структура представлена в тренировочном наборе
11betahsd1	2,51	-	-
17betahsd1	0	+	-
a2a	10,23	+	+
ace	15,44	+	+
ace2	0	-	-
ache	10,17	+	+
adam17	12,81	+	+
adrb2	9,59	+	-
akt1	15,15	+	+
alr2	0	-	-
ar	15,05	+	+
aurka	25,85	+	+
aurkb	27,84	+	+
bcl2	10,19	+	+
braf	18,08	+	+
catl	5,14	+	+
cdk2	14,95	+	-
cox1	2,61	-	-
cox2	5,14	+	-
ctsk	10,29	+	+
cyp2a6	0	-	-

dhfr	20,4	+	-
egfr	10,3	+	-
ephb4	25,29	+	+
er-beta	14,29	+	-
erbb2	29,6	+	+
fgfr1	27,02	+	+
fkbp1a	27	+	-
fxa	30,59	+	+
gba	5,15	+	+
gr	5,1	+	+
gsk3b	17,59	+	+
hdac2	10,12	+	+
hdac8	2,57	+	+
hiv1pr	30,13	+	+
hiv1rt	0	+	-
hmgr	23,19	+	-
hsp90	30,35	+	+
igf1r	2,48	+	+
inha	5,1	-	-
itk	7,61	+	+
jak3	29,93	+	+
jnk1	13,13	+	+
jnk2	17,51	+	+
jnk3	15,5	+	+
kif11	22,67	+	-
lck	20,07	+	+
mdm2	7,49	+	+
mk2	13,11	+	+
mmp2	2,55	-	-

na	30,98	+	+
p38-alpha	30,55	+	+
parp-1	27,61	+	+
pde4b	12,83	+	+
pde5	8,07	+	+
pdk1	31,74	+	+
pi3kg	20,68	+	+
pim-1	20,15	+	+
pim-2	18,07	+	-
pnp	30,83	+	+
ppara	8,17	+	-
pparg	12,95	+	+
pr	0	+	+
prkcq	26,32	+	+
pygl-in	25,67	+	-
pygl-out	10,3	+	-
qpct	18,05	+	+
rock-1	15,09	+	+
rxr	23,33	+	-
sars-hcov	2,57	+	+
sirt2	2,54	+	-
src	15,53	+	-
thrombin	29,21	+	-
tie2	14,99	+	+
tk	26,37	-	-
tp	0	+	+
tpa	30,55	+	+
ts	0	+	+
upa	29,9	+	+

vegfr1	16,33	+	+
vegfr2	23,46	+	+