

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение высшего образования
«Тольяттинский государственный университет»

Институт химии и энергетики

(наименование института полностью)

Центр медицинской химии

(наименование)

04.03.01 Химия

(код и наименование направления подготовки)

Медицинская и фармацевтическая химия

(направленность (профиль))

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА (БАКАЛАВРСКАЯ РАБОТА)

на тему Моделирование тройного комплекса ROI-PROTAC-CRBN

Обучающийся

А.С. Дронова

(Инициалы Фамилия)

(личная подпись)

Руководитель

д. х. н., доцент, З. П. Белоусова

(ученая степень (при наличии), ученое звание (при наличии), Инициалы Фамилия)

Консультант

А.Ю. Кудинов

(ученая степень (при наличии), ученое звание (при наличии), Инициалы Фамилия)

Аннотация

Целью данной работы стала валидация ряда подходов, направленных на генерацию тройных комплексов POI-PROTAC-CRBN. Разработка химер, нацеленных на протеолиз (PROTAC, *англ.* PROteolysis TArgeting Chimera) является актуальной задачей в контексте целенаправленной белковой деградации, а также *in silico* подходы являются наиболее востребованным в наши дни методом генерации тройных комплексов.

В литературном обзоре описаны концепция направленной деградации белков, структура и механизм действия PROTAC и рациональный дизайн химер, нацеленных на протеолиз, а именно дизайн лигандов POI и E3, дизайн линкеров и методы создания тройных комплексов.

Во второй части описана подготовка белковых структур и лигандов, результаты работы алгоритма IFD-MD, отчет о симуляции молекулярной динамики полученного комплекса, результаты работы рабочего процесса Protein Degradation Sampler и энергетическая оценка сгенерированных комплексов с помощью модели Prime.

Экспериментальная часть содержит подробное описание процессов подготовки структур мишеней и лигандов, протоколов индуцированного докинга, запуска рабочего процесса Protein Degradation Sampler, энергетической оценки Prime и метода возмущения свободной энергии.

На основании полученных данных были сделаны выводы, что алгоритмы IFD-MD и Protein Degradation Sampler могут быть использованы в контексте моделирования тройных комплексов POI-PROTAC-E3, и, что, основываясь на энергетической модели Prime можно ранжировать полученные комплексы.

Работа изложена на 91 странице, содержит 40 рисунков и 12 таблиц.

Abstract

The title of the graduation work is Modeling of ternary complex POI-PROTAC-CRBN.

The graduation work consists of an introduction, three chapters, 40 figures, 12 tables, a conclusion and the list of 84 references including foreign sources.

The aim of this graduation project is the validation of the workflows implemented in Schrodinger software aimed at generating POI-PROTAC-E3 ternary complexes.

The object of this graduation work is the concept of targeted protein degradation in the context of PROTAC.

The subject of the graduation work is modelling of the POI-PROTAC-E3 ternary complex based on the benzotriazole analogue dBET6.

The graduation work may be divided into several logically connected parts which are literature review, results and their discussion and experimental part.

The first part describes the concept of targeted protein degradation, the structure and mechanism of PROTAC action and rational design of chimeras targeting proteolysis, namely POI and E3 ligand design, linker design and methods for generating ternary complexes.

The second part describes the preparation of protein structures and ligands, the results of running the IFD-MD algorithm, the report of molecular dynamics simulation of the produced complex and the results of running the Protein Degradation Sampler workflow.

The third part comprises the methodology of computational studies.

Based on the data obtained, it can be concluded that the IFD-MD and Protein Degradation Sampler algorithms can be used in the context of modelling POI-PROTAC-E3 ternary complexes, and it is shown that the resulting structures can be ranked based on the Prime energy model.

Содержание

Введение.....	6
1 Литературный обзор	8
1.1 Убиквитин-протеасомная система	8
1.2 Принципиальное строение и механизм действия PROTAC	9
1.3.1 Дизайн лигандов POI и рекрутеров E3-лигазы	13
1.3.2 Дизайн линкеров	14
1.3.3 Генерация тройных комплексов POI-PROTAC-E3	18
1.3.4 Подходы к ранжированию тройных комплексов	33
2 Результаты и обсуждение.....	37
2.1 Дизайн исследования.....	39
2.2 Подготовка структуры белков и лигандов	40
2.3 Получение тройного комплекса BRD4-PROTAC-CRBN для бензотриазольного аналога dBET6 с помощью алгоритма IFD-MD	41
2.4 Молекулярная динамика комплекса, полученного с помощью алгоритма IFD-MD.....	63
2.5 Метод возмущения свободной энергии.....	70
2.6 Генерация тройных комплексов с помощью Protein Degradation Sampler. 72	
3 Экспериментальная часть.....	78
3.1 Компьютерное моделирование	78
3.1.1 Подготовка белков	78
3.1.2 Подготовка лигандов	78
3.1.3 Индуцированный докинг.....	78
3.1.4 Симуляция молекулярной динамики	79
3.1.5 Binding Pose Metadynamics.....	79
3.1.6 FEP+.....	80
3.1.7 Protein Degradation Sampler.....	80
3.1.8 Энергетическая оценка Prime	81
Заключение	82
Список используемой литературы	83

Определения, обозначения и сокращения

В настоящей работе применяют следующие термины с соответствующими определениями:

ALK	киназа анапластической лимфомы
BTK	тирозинкиназа Брутона
BRD4	бромодоменсодержащий белок 4
CBD	карбоксиконцевой домен
CRBN	цереблон
CRL4	куллин-RING-E3-убиквитинлигаза 4
Cul4	куллин 4
DDB1	белок, связывающий повреждения ДНК
ER- α	рецептор эстрогена альфа
FAK	киназа фокальной адгезии
GNN	графовая нейронная сеть
HBD	домен α -спирального пучка
IC ₅₀	концентрация полумаксимального ингибирования
ITC	изотермическая титрационная калориметрия
IMiDs	иммуномодулирующие препараты
MMGBSA	молекулярная механика с обобщенной Борновской и поверхностной сольватацией
NTD	N-концевой домен
PK / PD	фармакокинетические/фармакодинамические параметры
POI	белок интереса
RMSD	среднеквадратичное отклонение атомных координат
RNN	рекуррентная нейронная сеть
SAR	взаимосвязь структуры и активности
SMARCA2	связанный с матрицей, актин-зависимый регулятор хроматина
SPR	поверхностный плазмонный резонанс
STAT3	преобразователь сигнала и активатор транскрипции
VHL	E3-лигаза фон Гиппеля – Линдау
PCA	рентгеноструктурный анализ

Введение

Концепция направленной белковой деградации (TPD, *англ.* Targeted protein degradation,) является многообещающей терапевтической стратегией для регуляции количества макромолекул, связанных с развитием различных заболеваний. Широкую распространённость данный подход получил за счет способности воздействовать на мишени, которые ранее считались «нелекарственными» (*англ.* undruggable) ввиду отсутствия возможности регуляции их активности классическими малыми молекулами, например, путем ингибирования. Также стоит отметить, что деградация, или полное разрушение терапевтически интересных белков, позволяет модулировать как ферментативные, так и структурные функции.

Одним из подходов TPD является использование технологии химер, направленных на протеолиз (PROTAC, *англ.* PROteolysis TArgeting Chimera). Молекулы PROTAC представляют собой гетеробифункциональные соединения, состоящие из двух лигандов (т.н боеголовок, *англ.* warheads), один из которых направлен на связывание с интересующей мишенью, а другой с E3 убиквитинлигазой – ключевым компонентом убиквитинпротеосомной системы. Боеголовки, в свою очередь, соединены линкером, что позволяет им одновременно рекрутировать и E3-лигазу и белок-интереса (*англ.* POI, Protein of Interest). Такое взаимодействие опосредует пространственное сближение макромолекул, что в конечном итоге приводит к полиубиквитинированию и деградации POI в протеосоме 26S.

Однако, к сожалению, рациональный дизайн и оптимизация структур PROTAC всё еще является сложной задачей. А слепой эмпирический подход к SAR (*англ.* Structure-activity relationships) путем многостадийного синтеза больших библиотек деградеров и их последующее тестирование отнимает очень много временных и материальных ресурсов.

В настоящее время решить проблему дизайна химер пытаются с помощью *in silico* подходов. Их разработка, модификация и валидация является актуальной задачей.

Таким образом, целью данной работы стала валидация ряда подходов, направленных на генерацию тройных комплексов POI-PROTAC-CRBN.

Перед нами были поставлены следующие задачи:

1. Генерация тройного комплекса BRD4-PROTAC-CRBN на основе аналога dBET6 помощью IFD-MD.

2. Выполнение молекулярно-динамической симуляции IFD-MD-поз и сравнение структурных данных с таковыми для кристаллического комплекса dBET6.

3. Подбор оптимальных параметров запуска альтернативного алгоритма Protein Degradation Sampler для получения тройных комплексов.

4. Оценка применимости энергетических характеристик, рассчитываемых с помощью модели Prime, для идентификации кристаллоподобных решений Protein Degradation Sampler.

1 Литературный обзор

1.1 Убиквитин-протеасомная система

Убиквитин-протеасомная система (UPS, *англ.*, Ubiquitin–Proteasome System) представляет собой основной внутриклеточный способ деградации белков. UPS также участвует в дифференцировке клеток и репарации ДНК, деградируя при этом неправильно свернутые или поврежденные белки высокоспецифическим образом [1].

В процессе убиквитинирования принимает участие ряд ферментов – E1, E2 и E3 [2]. Сначала E1 активирует убиквитин, который впоследствии переносится на цистеиновый остаток E2. После этого E2 способствует взаимодействию Ub с субстратами благодаря убиквитинлигазе E3, которая распознает и связывает различные мишени (рисунок 1) [3, 4].

Протеасома 26S состоит из коровой 20S, содержащей активные каталитические центры и регуляторных 19S субъединиц, ответственных за перенос отмеченных белков в 20S и последующее деубиквитинирование [5]. Субъединица 20S состоит из гептамерных колец, сложенных друг на друга, обеспечивающих взаимодействие с 19S и протеолитическую активность [6, 7].

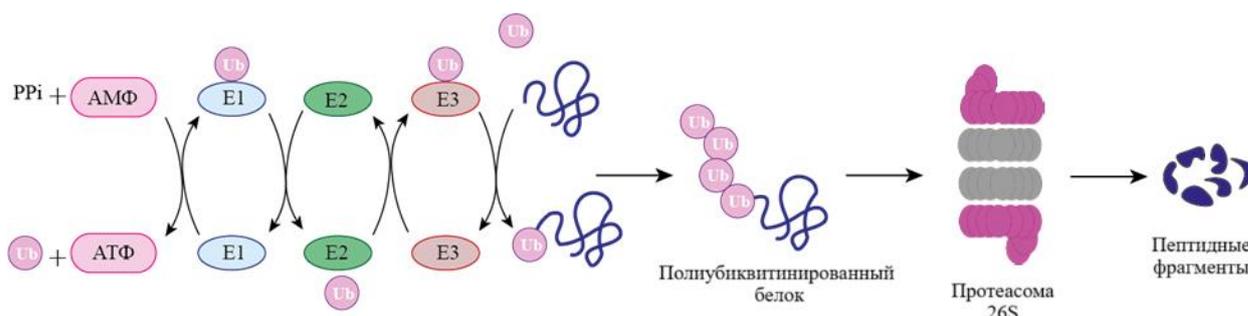


Рисунок 1 – Процесс деградации белков посредством UPS

1.2 Принципиальное строение и механизм действия PROTAC

В течение последних 10 лет в области разработки и открытия лекарств особое внимание уделялось технологии направленного протеолиза, и, в частности, химер, нацеленных на протеолиз (PROTAC, *англ.* PROTeolysis TArgeting Chimera).

PROTACs состоят из лиганда, который связывается с белком интереса (POI, *англ.* Protein of Interest), лиганда, рекрутирующего E3-лигазу, и соединяющего их линкера [8]. Бифункциональные молекулы связываются с POI одним концом, в то время как другим – с E3, индуцируя образование тройного комплекса. Рекрутированная лигаза затем опосредует перенос убиквитина. Тройной комплекс диссоциирует, убиквитинированный POI деградируется протеасомой 26S и PROTAC может связываться с другой мишенью (при условии, что взаимодействие с белком интереса является обратимым) (рисунок 2) [9].

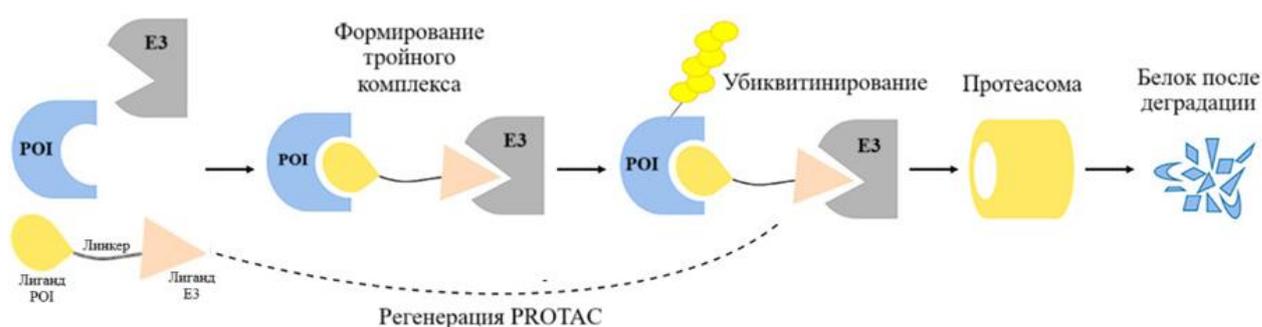


Рисунок 2 – Образование тройного комплекса и его диссоциация

Данная технология стала инновационным открытием в области разработки лекарственных средств и предоставила новые возможности как для промышленности, так и для научных кругов, которые проявлялись во множестве аспектов.

Во-первых, концепция направленной деградации белков позволила преодолеть лекарственную устойчивость рака. За счет полного разрушения

целевых белков, PROTACs во многом превосходят традиционную терапию низкомолекулярными ингибиторами [10]. Например, деградер ARCC-4 **1**, направленный на деградацию AR, ингибирует пролиферацию клеток опухоли предстательной железы, путем разрушения белков и клинически значимых точечных мутаций. И, в отличие от энзалутамида, ингибитора AR, сохраняет антипролиферативный эффект в среде с высоким содержанием андрогенов [11], а L18I **2**, нацеленный на ВТК, может преодолеть мутацию С481S (рисунок 3) [12].

Во-вторых, химеры способствуют подавлению неферментативных функций белков. Так, группа Крейга М. Крюса [13] разработала высокоэффективный и селективный деградер FAK **3**, проявляющий высокую активность в отношении миграции клеток и инвазии.

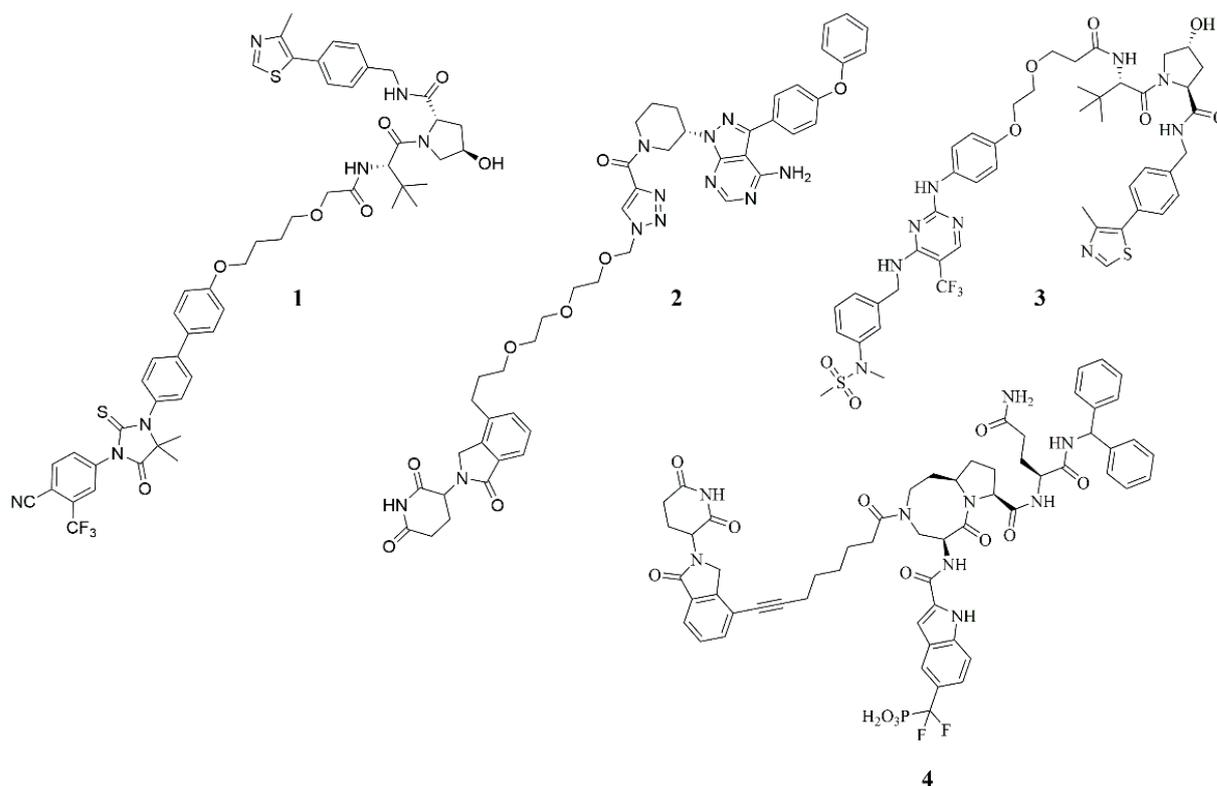
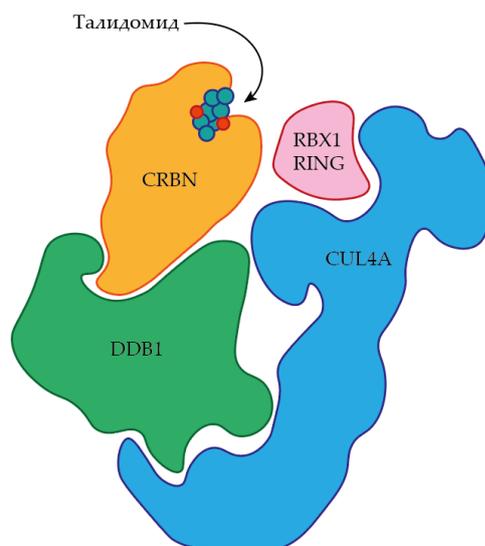


Рисунок 3 – Структуры PROTAC

В-третьих, PROTACs позволяют деградировать так называемый «неуловимый» протеом. В настоящее время лишь 25% от всех известных белков служат мишенями для «классических» лекарственных соединений. Остальные белки, которые не обладают каталитической активностью или у которых отсутствует активный сайт связывания, до недавних пор считались "нелекарственными" мишенями [14]. Примером такой мишени является STAT3, для которой отсутствуют активные лиганды. Группа Шаоменг Ванга сообщила о PROTAC, нацеленном на деградацию STAT3 4, который обладает мощной биологической активностью *in vitro* и *in vivo* (рисунок 3) [15].

Но несмотря на то, что технология PROTAC имеет большой потенциал, у нее также есть и ряд недостатков: низкая проницаемость и биодоступность при пероральном приеме ввиду большой молекулярной массы и других неблагоприятных физико-химических свойств [16]; сложность синтеза и отсутствие единой стратегии дизайна тройных комплексов; проблема в подборе методов достоверной и качественной оценки фармакокинетических и фармакодинамических параметров [17]; «эффект крючка» (*англ.* Hook effect), который в снижении эффективности образования тройного комплекса при высоких концентрациях за счет преимущественного образования двойных комплексов PROTAC-E3-лигаза или PROTAC-POI [18]

В геноме человека кодируется более 600 лигаз E3, но лишь несколько из них могут быть использованы при создании PROTAC (VHL, CRBN, MDM2, cIAP) [19]. Среди них CRBN является одной из наиболее широко используемых в области направленной деградации белков E3-лигаз. В данной работе в качестве лигазы был использован именно цереблон, следовательно, нужно подробнее рассмотреть строение этого комплекса. Он представляет собой белок размером примерно 200 кДа и имеет три домена; NTD, HBD и CBD, среди которых последний содержит сайт связывания цинка [20]. CRBN взаимодействует с белком DDB1 и Cul4A или Cul4B, образуя функциональный комплекс E3-убиквитинлигаз CRL4 (рисунок 4) [21,22].



Каркасный белок (голубой), RING-домен (розовый), цереблон (жёлтый) с талидомидом, адаптерный белок DDB1 (зеленый)

Рисунок 4 – Схематичное строение белкового комплекса CRL4

Наиболее распространенными лигандами к CRBN являются IMiDs, такие как талидомид **5**, помалидомид **6** и леналидомид **7** (рисунок 5) [23-25].

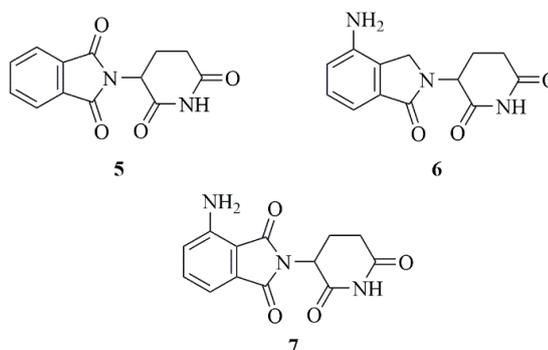


Рисунок 5 – Структурные формулы IMiDs

Было запатентовано большое количество PROTAC, нацеленных на различные POI, которые можно разделить на 4 категории, а именно PROTACs, нацеленные на киназы, ядерные рецепторы, эпигенетические механизмы и неправильно свернутые белки [26-28].

1.3 Рациональный дизайн PROTAC

Эффективность деградации белков-мишеней, посредством PROTACs зависит от выбора и оптимизации лигандов POI/E3-лигаз и линкеров. В настоящее время SAR-оптимизация в основном основывается на эмпирическом анализе и включает традиционный синтез объемных библиотек и биологические тестирования.

Дизайн линкера ключевой стадией в разработке деградеров, поскольку он обеспечивает эффективность деградации белков и стабильность ТК. Разработка ингибиторов/лигандов POI и рекрутеров E3-лигаз в свою очередь не представляет сложности, т.к. лежит в области привычного нам классического дизайна малых молекул [29, 30].

Другими препятствиями в области рационального дизайна являются создание, скрининг и анализ конформационного пространства структур тройных комплексов POI-PROTAC-E3. Точное знание конфигураций ТК необходимо для структурного проектирования, сравнения различных PROTAC и отбора хитов. На сегодняшний день существует несколько алгоритмов, которые предлагают подходы к генерации и оценке ТК. Некоторые из них будут обсуждаться в данном разделе.

Таким образом, разработка эффективных PROTAC представляет собой сложный и длительный процесс, требующий обширных химических испытаний и оптимизации. Для решения этой проблемы могут быть применены вычислительные подходы, которые включают в себя белок-белковый докинг, моделирование с помощью молекулярной динамики и методы, основанные на машинном обучении и искусственном интеллекте.

1.3.1 Дизайн лигандов POI и рекрутеров E3-лигазы

В основном, рекрутеры POI в составе PROTAC представляют собой хорошо известные и описанные ранее ингибиторы. При их выборе в основном учитывают такие факторы как аффинность связывания с белком и физико-

химические свойства. Также конструированию деградеров, с точки зрения дизайна лигандов ROI, способствует структурная информация, полученная из анализа кристаллических структур белково-лигандных комплексов [31]. Такой анализ позволяет идентифицировать точки крепления и так называемые *вектора выхода* линкеров. Правильный выбор точки присоединения влияет на аффинность связывания.

Так, авторы работы на начальном этапе дизайна нового деградера ALK использовали кристаллографический комплекс соответствующего белка и алектиниба (PDB ID: 3АОХ) [32]. После анализа способа и места связывания было установлено, что морфолиновый фрагмент, направленный в область растворителя, является наиболее благоприятным местом конъюгации линкера (рисунок 6).

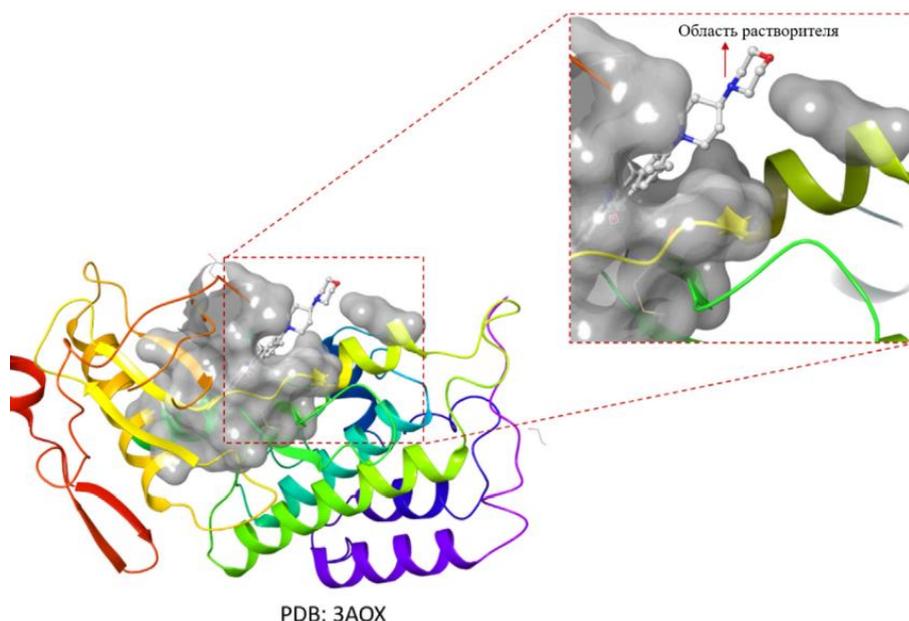


Рисунок 6 – Кристаллическая структура ALK в комплексе с алектинибом

1.3.2 Дизайн линкеров

Дизайн линкера является ключевым шагом в разработке PROTAC, поскольку его структура и длина определяет эффективность, специфичность и фармакокинетику полученных химер. Следовательно, правильно

подобранный линкерный мотив необходим для обеспечения максимальной аффинности связывания, эффективной деградации целевого белка и минимизации нецелевых эффектов [33]. Основными направлениями для оптимизации являются длина, конформационная подвижность, химический состав, гидрофильность и гидрофобность.

На первом этапе дизайна линкера обычно оптимизируют длину для обеспечения оптимального и достаточного расстояния между двумя другими составными частями деградера. Так, например, Сайрус и др в своей работе пришли к выводу, что линкер, состоящий из 16 атомов, является оптимальным для деградации ER- α [34].

Оптимизация конформационной подвижности линкерного мотива является не менее важным шагом. Это свойство позволяет двум лигандам принять правильную ориентацию для дальнейшей эффективной деградации. Жесткость в свою очередь достигается путем добавления структурных звеньев, которые могут ограничивать конформационную подвижность деградера. Так, например, в недавних исследованиях [35,36] авторам удалось разработать стабильный и эффективный тройной комплекс PROTAC путем добавления в состав линкера пиперазинового и пипиридинового звеньев. Данная оптимизация привела к изменению физико-химических свойств PROTAC, а именно к улучшению растворимости и клеточной проницаемости.

Мэйпл и др. собрали базу данных, состоящую более чем из 400 опубликованных деградеров, в которой на сегодняшний день наиболее распространенными линкерными мотивами PROTAC, являются ПЭГ и алкильные цепи различной длины, они являются единственными мотивами примерно в 55% и 30% линкеров соответственно. Около 65% структур в базе сочетают в себе как алкильный, так и ПЭГ фрагмент (рисунок 7) [37].

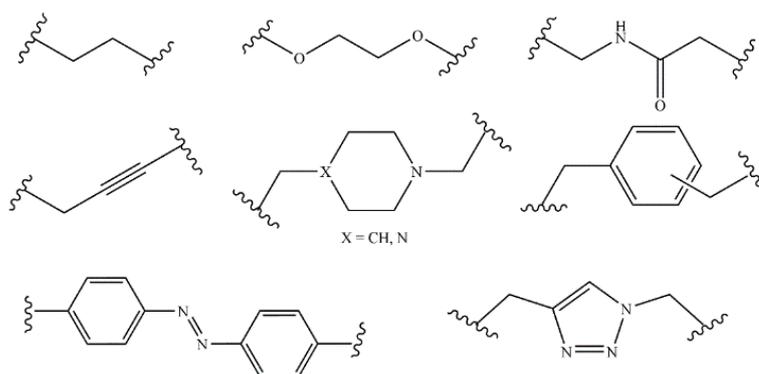


Рисунок 7 – Примеры линкерных мотивов

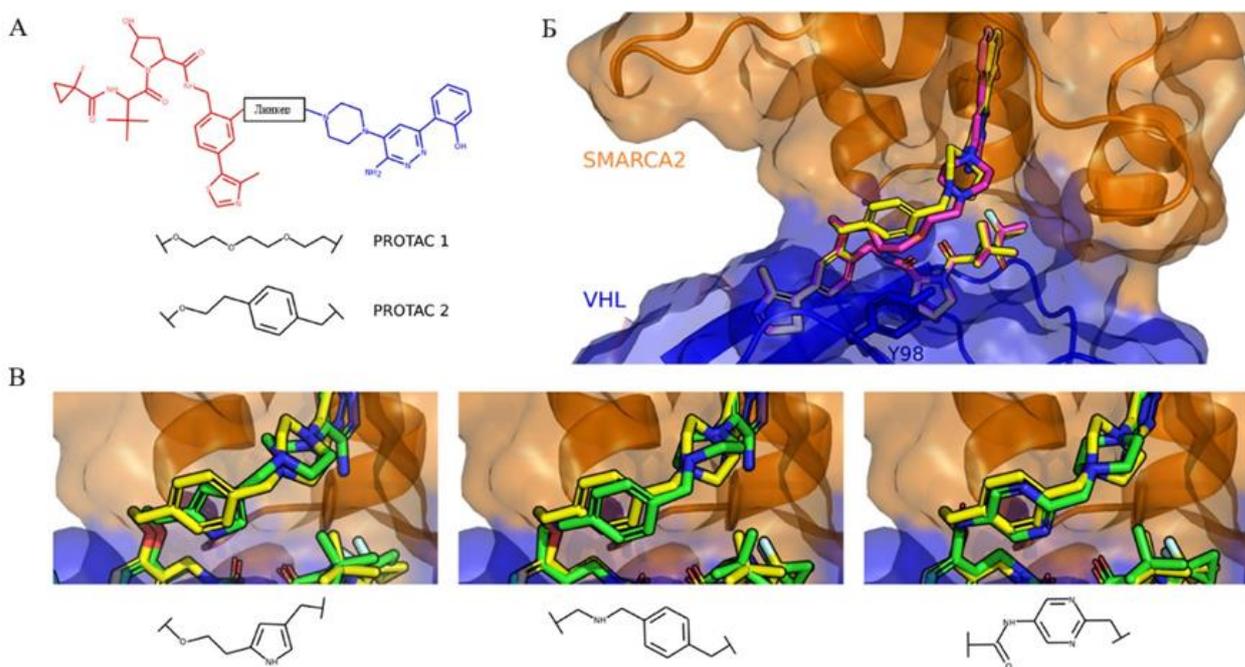
Подходы к конструированию линкеров, описанные в литературе, в основном заключаются в генерации библиотек деградеров, обладающих заданными физико-химическими свойствами и степенью конформационной подвижности. Так, например, Link-INVENT, разработанный Го и др., эффективно исследует оптимальную длину линкера в определенном диапазоне физико-химических свойств, а также контролирует линейность и гибкость путем выбора различных фрагментов, контроля соотношения вращающихся связей [38]. Другой алгоритм – AIMLinker на основе нейронных сетей – использует структурную информацию, полученную от родственных фрагментов, для создания линкеров, способных вместить и включить эти фрагменты, а затем отфильтровывает нелекарственные структуры, ориентируясь на белково-белковые комплексы, чтобы конечные молекулы были похожи на лекарственные препараты [39].

Еще один метод – DeLinker – представляет собой генеративную модель на основе графов, которая, используя трехмерную структурную информацию о структуре тройного комплекса, полученной из белок-белковой стыковки или PDB, позволяет создавать *de novo* или заменять имеющийся линкер между двумя фрагментами лигандов. Разработчики использовали алгоритм для поиска новых уникальных линкеров для тройного комплекса SMARCA2-PROTAC-VHL (рисунок 8). DeLinker произвел 2150 уникальных соединений,

3 из которых в точности воспроизводили геометрию линкера в нативном PROTAC [40].

В заключение отметим, что все три метода требуют информацию о способах связывания боеголовок PROTAC в сайтах POI и E3 и могут быть использованы для получения структур тройных комплексов.

Некоторые другие вычислительные алгоритмы, используемые для генерации библиотек линкеров и молекул PROTAC представлены в таблице 1.



А – Структуры PROTAC 1 и 2; Б – наложение тройных комплексов PROTAC 1 (PDB ID 6HAY) и PROTAC 2 (PDB ID 6HAX) с SMARCA2 (оранжевый) и VHL (синий); В – Наложения трех линкеров с разными каркасами

Рисунок 8 – Структуры, полученные с использованием DeLinker [40]

Таблица 1 – Методы генерации библиотеки линкеров

Модель	Метод	Описание	Источник
Link- INVENT	RNN	Создает оптимальные линкеры, соединяющие две молекулы; Гибкая скринговая функция	[38]
AIMLinker	GNN	Включает 3D-информацию о структуре; Возможность фильтрации нелекарственных структур.	[39]
DeLinker	Генеративная модель на основе графов	Использует 3D-структурные данные для создания линкера, соединяющего два лиганда; Процесс генерации можно регулировать, указав длину линкера и соединяемые фрагменты.	[40]
DRlinker	Глубокое обучение с подкреплением	Генерация соединений с определенными свойствами; Оптимизация прогнозируемой биоактивности	[41]
ShapeLinker	Обучение с подкреплением	Воспроизводит линкеры с удовлетворяющими 2D- и 3D-критериями	[42]
PROTAC- INVENT	Обучение с подкреплением	Может генерировать трехмерную предполагаемую конформацию связывания с белком-мишенью и E3-лигазой.	[43]

1.3.3 Генерация тройных комплексов POI-PROTAC-E3

В настоящее время все больше экспериментально определенных тройных комплексов POI-PROTAC-E3 депонируются в PDB, но, к сожалению, подходы к предсказанию таких структур все еще скудны.

Одним из наиболее популярных методов для генерации тройных комплексов является белок-белковый докинг (*англ.* Protein-protein Docking, PPD). Предсказание структур в этом методе зависит от двух ключевых этапов: выборки конформационного ландшафта белок-белкового комплекса и оценки полученных поз [44]. Чаще всего протоколы PPD представляют белки как жесткие тела, игнорируя при этом их конформационную подвижность, что является существенным ограничением.

Пионерскими в области *in silico* дизайна деградеров стали работы Драммонда и др [45]. В своей работе коллеги описали четыре вычислительных

подхода, которые основывались на белок-белковом докинге и на моделировании молекулярной динамики (рисунок 9).

Все методы требуют начальной конформации PROTAC, а также точно определенных способов связывания рекрутеров, которые могут быть получены исходя из данных PCA или «классического» докинга.

В первом методе осуществляется генерация конформаций молекулы деградера, при этом мотивы, отвечающие за рекрутирование POI и E3, жестко фиксируются в связывающих конформациях. Используя координаты рекрутирующих лигандов в качестве шаблонов, на втором этапе структуры белков просто накладываются на каждый из конформеров PROTAC. Возникающие при этом стерические столкновения разрешаются путем модификации вращающихся связей деградера до значений, которые минимизируют перекрытие. Все полученные конформации впоследствии сохраняются и оцениваются.

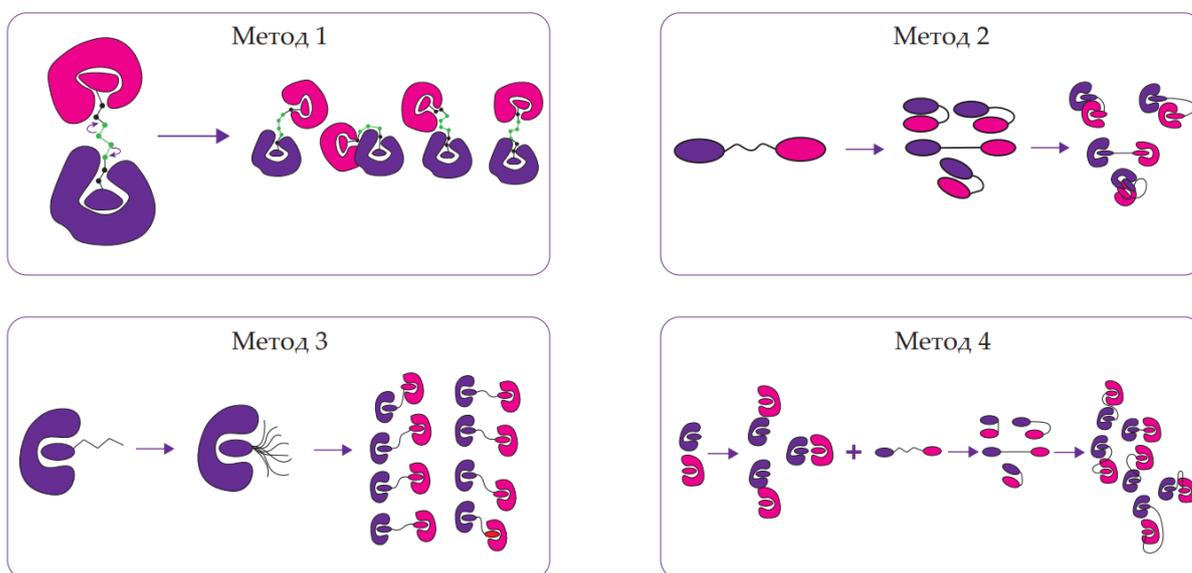


Рисунок 9 – Графическое представление четырех вычислительных методов генерации тройных комплексов

Второй метод похож на предыдущий тем, что также происходит генерация конформеров PROTAC и наложение структур белков, однако релаксации полученных ТК не происходит и отбираются лишь те позы, которые не продемонстрировали ни одного столкновения.

Третий метод требует структуры одного из белков (POI или E3), связанного с лигандом и линкером. Выполняется симуляция молекулярной динамики, позволяющая получить ансамбль конформеров. После завершения поиска второй белок накладывается.

Входные данные для четвертого алгоритма не отличаются от предыдущих. На первом этапе происходит белок-белковый докинг двух белково-лигандных комплексов. В ходе этой процедуры боковые цепи белков автоматически перестраиваются, а предполагаемые позы минимизируются. Затем стабильные конформации отбираются и совмещаются с конформациями PROTAC.

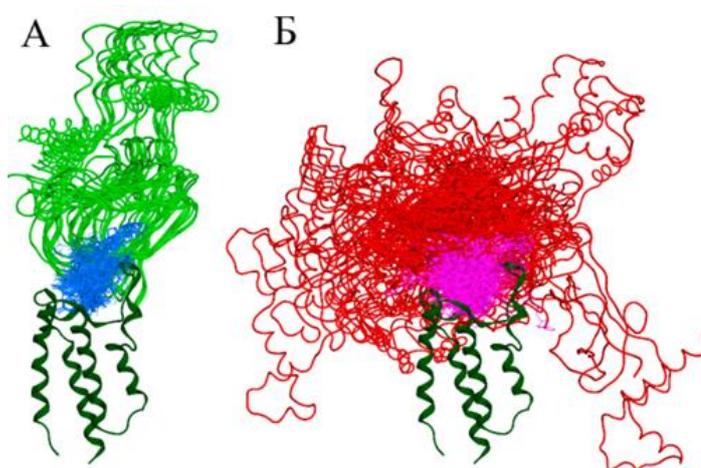
После была проведена валидация описанных выше алгоритмов. В качестве референсной структуры был выбран тройной комплекс, содержащий E3-лигазу VHL, белок BRD4 и PROTAC MZ1 (PDB ID: 5T35). Результаты работы авторов представлены в таблице ниже (таблица 2). Значение $C\alpha$ -RMSD и RMSD PROTAC использовалось в качестве метрики для оценки кристаллоподобности структуры. Все позы, обладающие $C\alpha$ -RMSD < 10 Å, считались кристаллоподобными.

Также были сгенерировали метрики, а именно 50 дескрипторов для оценки белок-белковых взаимодействий, которые в общем виде принимали вид – RecM_type1_covered_by_RecN_type2, где M и N – это или POI/E3, а type 1(2) – могли принимать одно из пяти значений: hyd, neg, pos, any, all.

Таблица 2 – Обзор результатов воспроизведения структуры 5T35 с помощью четырех методов

Метод	Фильтр 1	Фильтр 2	Кристалло-подобные позы	Общее количество сгенерированных комплексов	Hit rate (%)
1	–	–	0	1055	0
2	Hyd_hyd $\geq 100 \text{ \AA}$	S_all $\leq 525 \text{ \AA}^3$	22	496	4.44
3	Hyd_hyd $\geq 100 \text{ \AA}$	S_all $\leq 525 \text{ \AA}^3$	14	177	7.91
4	Hyd_any $\geq 100 \text{ \AA}$	Core RMSD $\leq 0.8 \text{ \AA}$	20	51	39.22

Фильтр S_all обозначает общий объем перекрытия между атомами белков, а фильтр Hyd_hyd в свою очередь есть ничто иное как общий объем перекрытия гидрофобных областей ROI и E3. Стоит отметить, что атомы PROTAC не участвовали в данном расчёте, т.к. эти дескрипторы использовались исключительно для более детального изучения белок-белковых взаимодействий.



А – все позы наложены на мишень Brd4 (темно-зеленые ленты), атомы PROTAC (голубой); Б – остальные позы, которые не находятся в пределах 10 \AA RMSD от VHL (красные ленты), атомы PROTAC (пурпурный)

Рисунок 10 – Результаты наложение структур, генерируемых методом 4

Таким образом, четвертый метод продемонстрировал самую высокую степень обогащения позами, близкими к нативным (рисунок 10). Авторы

отмечают, что некристаллоподобные тройные комплексы, генерируемые методами 1-4, не обязательно являются ошибочными, т.к. индуцируемые PROTAC ТК могут быть представлены целым ансамблем низкоэнергетических конформаций.

Анализируя полученные результаты, можно прийти к выводу, что даже несмотря на применение методов машинного обучения и компьютерного моделирования определение кристаллоподобных поз все еще является не тривиальной задачей. Но описанный выше метод 4 и его результаты могут способствовать дальнейшему развитию рационального дизайна PROTAC.

Другими алгоритмами, которые используют белок-белковый докинг для генерации тройных комплексов являются MOE [46], PRosettaC [47] и ICM [48] (рисунок 11).

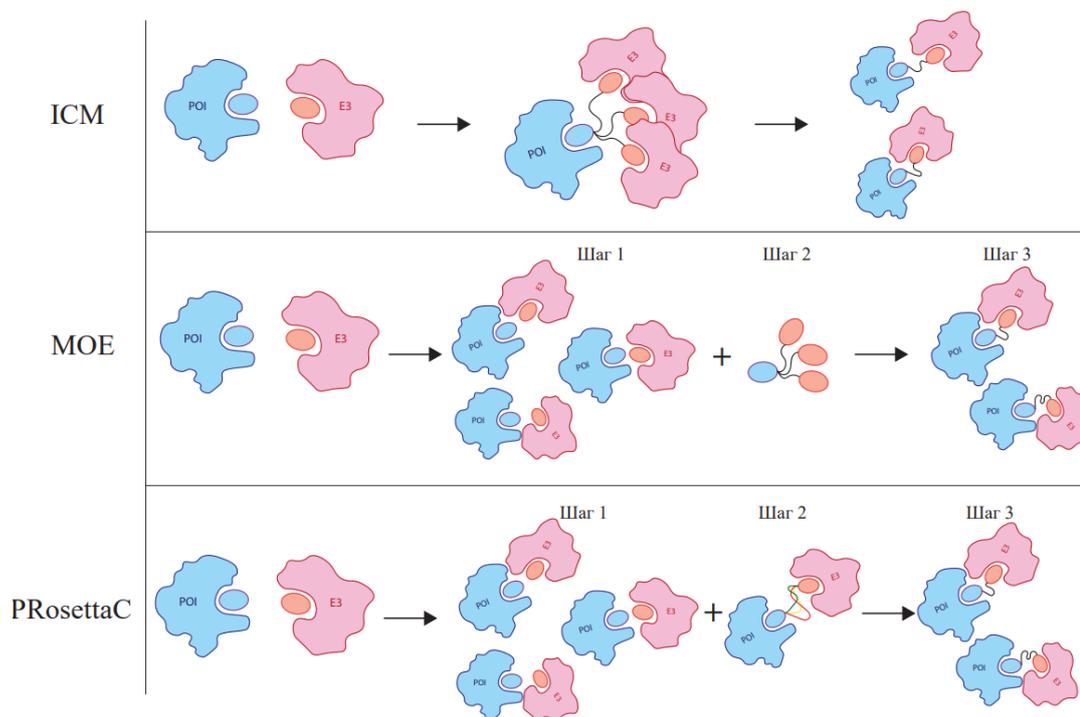


Рисунок 11 – Графическое представление алгоритмов предсказания структур тройных комплексов

Алгоритм MOE и PROsetta выполняют независимый белок-белковый докинг, затем проводят конформационный поиск PROTAC. После чего определяются конформации PROTAC, совместимые со структурами белковых комплексов, которые впоследствии объединяются в кластеры (PROsetta генерирует конформации PROTAC, основываясь на ограничениях, задаваемых той или иной позой). Размер кластера используется в качестве метрики для ранжирования ТК. Метод ICM Molsoft подразумевает выполнение моделирования молекулярной динамики молекул PROTAC, связанных с POI или E3 по отдельности, затем – наложение структур белков, а после – дополнительную минимизацию энергии тройного комплекса с гибким PROTAC. В данном алгоритме функция оценки PPD, наряду с предсказанной энергией комплекса, используется для ранжирования [48].

Более детально хотелось бы рассмотреть протокол, предложенный Зайдмана и др [47] (рисунок 12). PROsettaC является новым методом моделирования тройных комплексов PROTAC. В качестве исходных данных пользователю необходимо предоставить информацию о структуре POI и E3-лигазы в комплексе с их лигандами, строку SMILES всей молекулы деградера и информацию о двух якорных атомах¹.

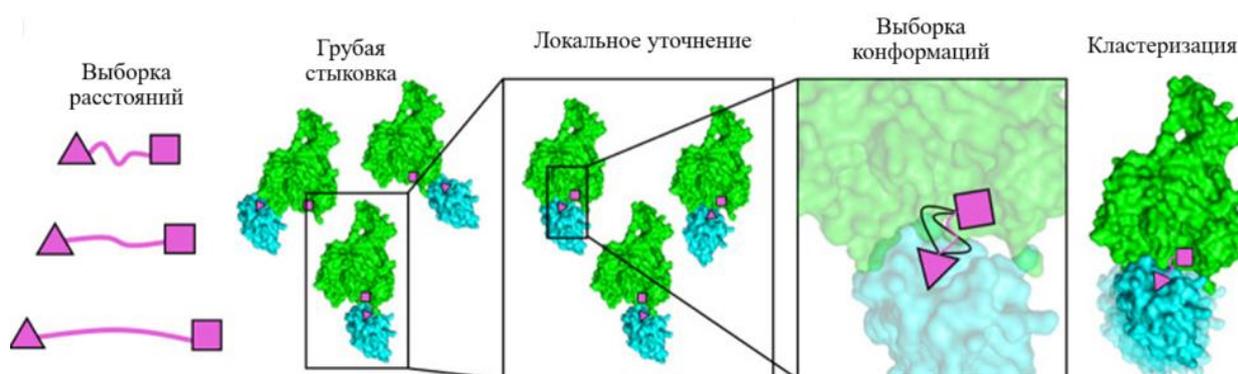


Рисунок 12 – Схематичный алгоритм действия PROsettaC

¹ Якорной атом (*англ.* Anchor atom, anchor point) называют атомы, входящие в состав рекрутеров POI и E3, к которым происходит прикрепление линкера.

На первом этапе, начиная с 1Å и с шагом в 1Å устанавливается до 21 значения попарных расстояний между якорными атомами. Для каждого расстояния происходит генерация 200 случайных взаимных ориентаций рекрутеров, для которых, происходит получение случайной конформации всего PROTAC. После, успешно сгенерированные конформации (без стерических столкновений) складываются для каждого интервала расстояний. Ограничения расстояний сохраняются в соответствии с распределением успешных конформаций.

На втором шаге с помощью алгоритма жесткой стыковки PatchDock происходит непосредственно белок-белковый докинг. Расстояния между якорными атомами вынуждено оставаться в пределах значений, сохранённых на предыдущем шаге, что существенно ограничивает область поиска.

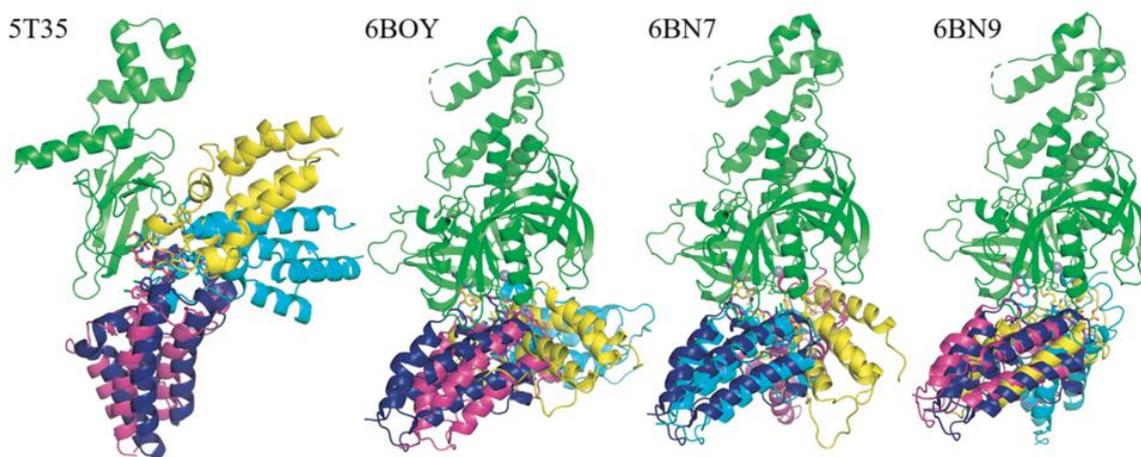
На третьем этапе авторы предлагают использовать алгоритм RosettaDock для создания 50 моделей тройного комплекса для каждого стыковочного решения, полученного на предыдущем этапе, с помощью подходов гибкого PPD. В конце производится окончательное моделирование линкера (до 100 вариантов) и кластеризация полученных результатов. Количество сгенерированных поз на промежуточных этапах и количество окончательных ТК представлены в таблице 3. Конформации, в которых нарушается заданный порог для якорных расстояний, или конформации, в которых невозможно соединений рекрутеры данным линкером, очевидно, будут отброшены.

Таблица 3 – Статистика количества сгенерированных моделей

Структура	Обучающая выборка					Тестовая выборка				
	5T35	6BOY	6BN7	6BN8	6BN9	6BN8	6SIS	6HAX	6HAY	6HR2
Общее количество поз PatchDock	555	1367	366	10161	65	74	1272	648	788	359
Количество поз Rosetta	17517	36103	13530	38242	2704	2102	32106	21143	26320	13012
Количество сгенерированных	8065	27805	10686	37696	1872	527	4061	7213	11789	4778

конформационных ансамблей										
---------------------------	--	--	--	--	--	--	--	--	--	--

Как можно увидеть на рисунке ниже, алгоритм PRosettaC способен успешно воспроизводить тройные комплексы (авторы сообщают о получении кристаллоподобных ТК в 6 случаях из 10) (рисунок 13).



Нативные структуры VHL (зеленый), нативная структура BRD4-PROTAC (темно-синий), тройной комплекс первого кластера (голубой), тройной комплекс второго кластера (розовый), тройной комплекс третьего кластера (жёлтый)

Рисунок 13 – Три лучших кластера для 5T35, 6BOY, 6BN7, 6BN9

В 2023 году был предложен метод создания высокоточных структурных моделей тройных комплексов E3-PROTAC-POI, основанный также на докинге, но с помощью алгоритма быстрого преобразования Фурье (FTT) [49].

Помимо этого, была проведена валидация алгоритма на ретроспективных моделях путем построения структурных моделей и предсказания активности PROTAC, которую они сравнивали с экспериментально определенными значениями (ITC и SPR). Схематично данный алгоритм представлен на рисунке ниже (рисунок 14) [52].

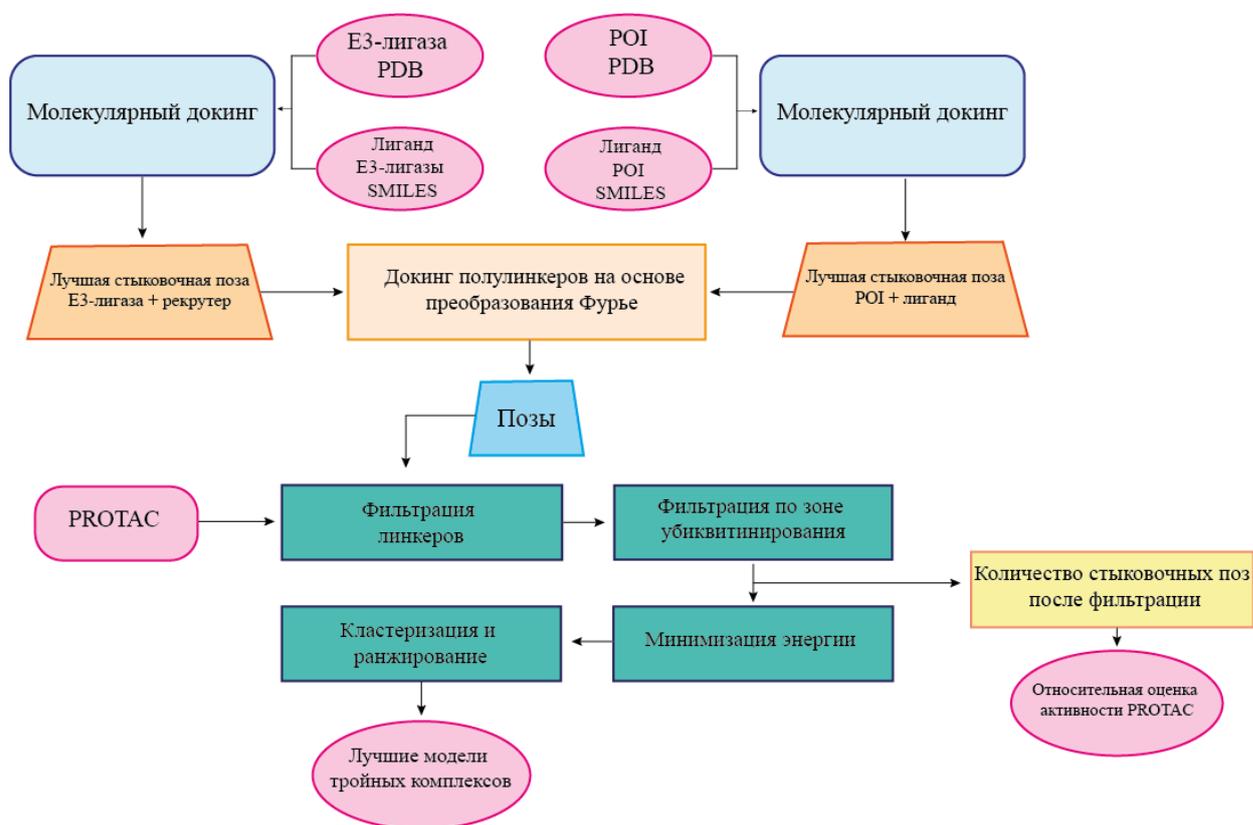


Рисунок 14 – Блок-схема предлагаемого метода

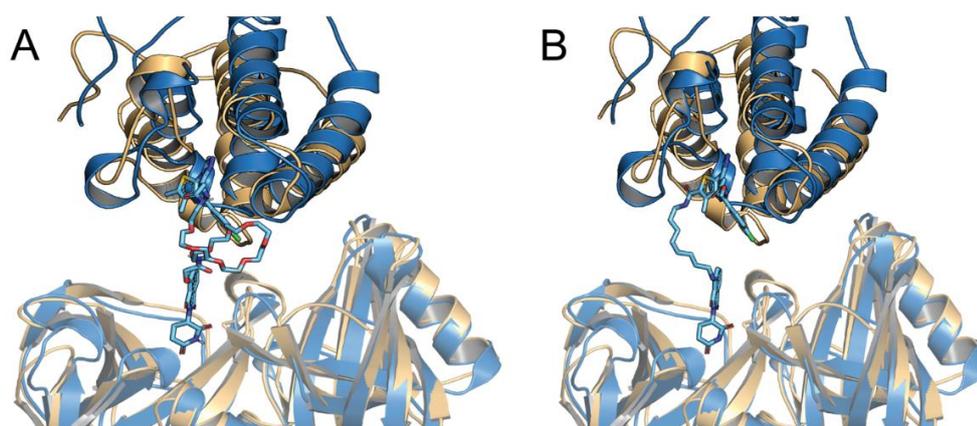
Моделирование ТК основывалось на отдельно закристаллизованных структурах POI и E3, предоставленных в виде файла PDB, а также химической структуры лигандов в формате SMILES. Первым шагом алгоритма являлся обычный докинг малых молекул. На следующем шаге генерировалось 10 000 конформаций линкера отдельно от лигандов, а затем каждая полученная конформация разделялась на две половины, одна из которых конъюгировалась с рекрутером E3, а другая с лигандом POI. Ансамбль таких структур назывался облаком полулинкеров (*англ.* half-linker clouds) и фильтровался от конформаций, приводящих к стерическим столкновениям.

Основным и наиболее инновационным шагом этого метода являлось расширение алгоритма преобразования Фурье, который использует дополнительный член свертки для генерации только тех конформаций белок-белкового комплекса, у которых концы полулинкеров расположены близко друг к другу.

Идентификация комплекса между E3-лигазой и белком-мишенью, содержащего связанные лиганды с полулинкерами, которые встречаются на соответствующих концах, указывала на то, что соответствующий PROTAC может эффективно связываться как с белком-мишенью, так и с E3-лигазой.

Таким образом, результатом жесткого докинга белков с их связанными лигандами и полулинкерами являлся тройной комплекс, представленный в виде низкоэнергетических поз двух белков с потенциально выгодной геометрией линкера.

Сгенерированные позы затем оцениваются, чтобы определить, может ли молекула Ub связываться POI, т.к. направленная деградация белка с участием лигазы E3 VHL или CRBN возможна лишь в том случае, если остаток Cys фермента E2 лигазного комплекса находится на расстоянии от 50 до 60 Å от поверхностного остатка Lys на целевом белке. Данная оценка исключала некоторые позы, которые в противном случае считались бы благоприятными. Затем концевые атомы полулинкеров соединялись, оставшиеся структуры уточнялись путем минимизации энергии и объединялись в кластеры для дальнейшего формирования моделей полных тройных комплексов (рисунок 15).



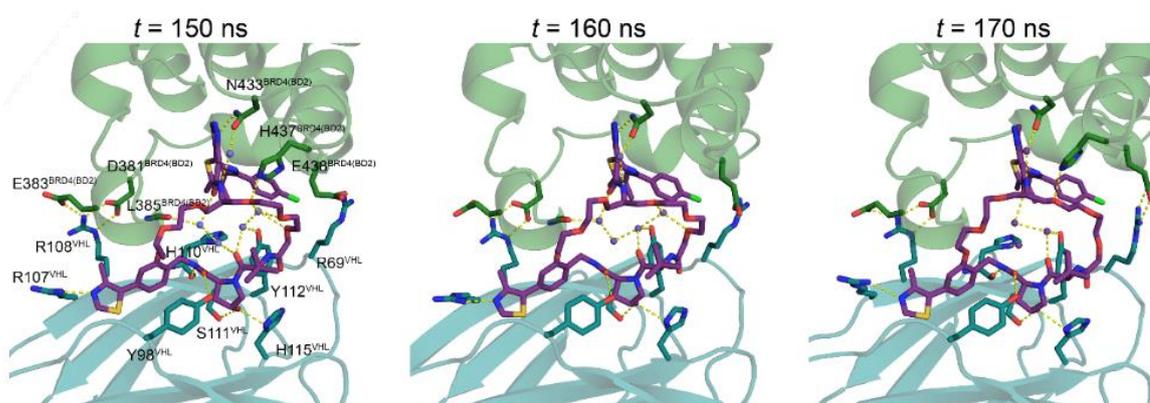
А – Наложение предсказанной модели CRBN-dBET6-BRD4BD1 (синий) и его рентгеновской структуры (желтый, PDB ID: 6BOY); В – Наложение предсказанной модели CRBN-dBET23-BRD4BD1 (синий) и его рентгеновской структуры (желтый) (PDB ID: 6BN7)

Рисунок 15 – Сравнение полученных результатов анализа

В 2020 году группой Теста [50] было предложено использовать симуляцию МД для предсказания структур ТК и для определения их стабильности.

Стоит отметить, что авторы разрабатывали макроциклический PROTAC, основываясь на том стремлении воссоздать геометрию связывания активной конформации. Подход к циклизации MZ1 (PDB ID: 5T35) основывался на определении подходящего вектора выхода и линкера, которые сохраняли бы ключевые взаимодействия нативного комплекса и в то же время поддерживали бы индуцированные PPI, способствующие формированию стабильного ТК, который лежит в основе деградации BRD4.

Впоследствии была проведена симуляция МД комплекса BRD4-MZ1-VHL. Тройной комплекс оставался стабильным на протяжении всего моделирования, с сохранением ключевых взаимодействий (рисунок 16).



Желтые пунктирные линии – белок-белковые и белок-лигандные взаимодействия

Рисунок 16 – Взаимодействия в ТК в ходе МД-симуляции

На рисунке 17 изображен график зависимости времени симуляции от RMSD остатков относительно начальных координат. Снижение амплитуды колебаний значений (выход на плато) указывает на стабильность полученной структуры.

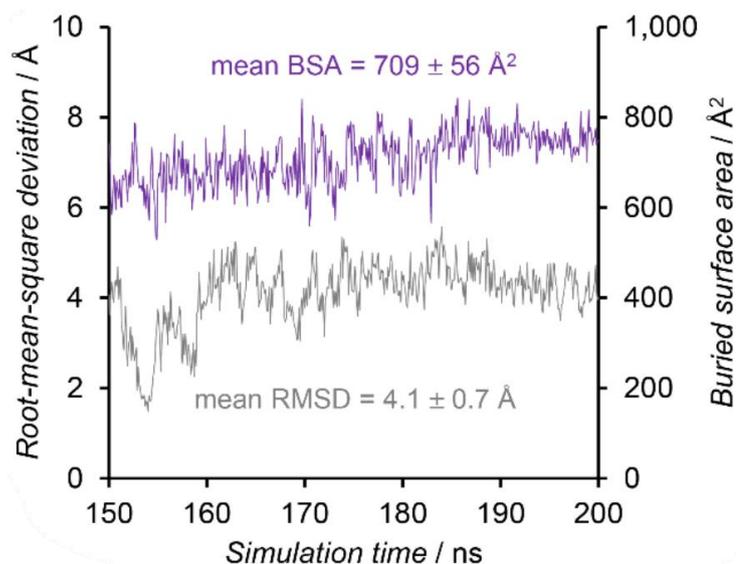


Рисунок 17 – Изменение BSA и $C\alpha$ -RMSD VCB-MZ1-BRD4 во время симуляции относительно координат кристаллического комплекса после наложения на кристаллическую структуру

Анализируя результаты симуляции тройного комплекса за последние 50 нс можно заметить, что минимальные расстояния между ключевыми аминокислотами, участвующих в белок-белковых взаимодействиях в VHL и BRD4 не изменялись, что говорит опять уже указывает на стабильность и кристаллоподобность полученного ТК (рисунок 18).

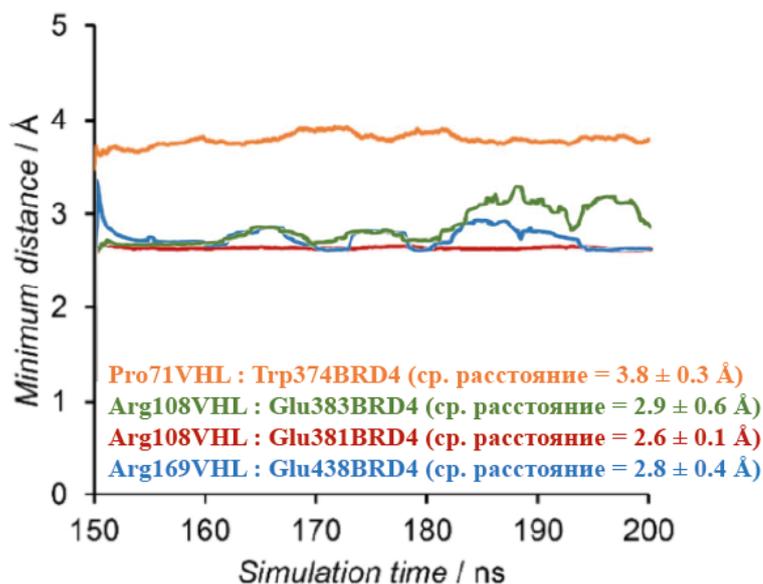


Рисунок 18 – Результаты симуляции молекулярной динамики

Ключевым недостатком описанного метода являлось то, что он не давал количественной оценки для ранжирования полученных кандидатов по их способности образовывать стабильные тройные комплексы.

Так, например, в работе Ляо и др. [51] описали подход, основанный на симуляции МД и кластеризации полученных ТК на основе MMGBSA. Схематичное описание алгоритма представлено на рисунке ниже (рисунок 19).

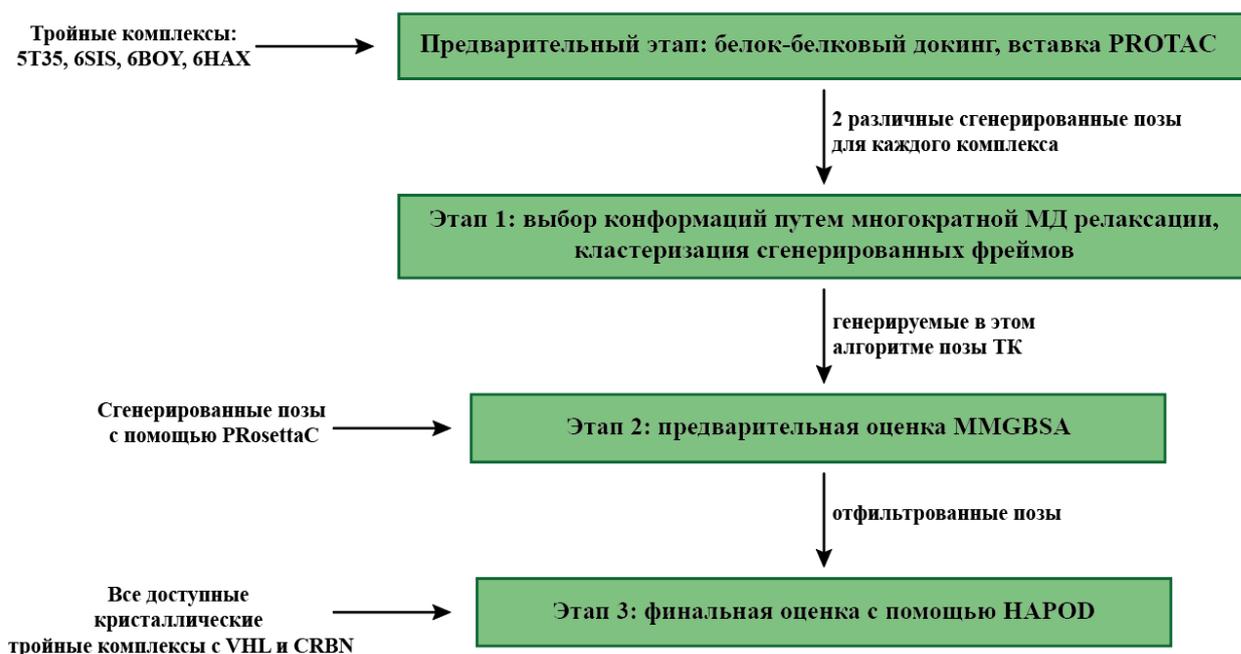


Рисунок 19 – Схематичное изображение рабочего алгоритма

На первом этапе выполнялся белок-белковый докинг POI и E3 без учета лигандов. Поскольку, в дальнейшем ожидалось, что релаксация, посредством симуляции МД, будет выполнять только локальный конформационный поиск, после докинга выбиралось не менее двух поз. Также на этом этапе отклонялись те варианты, в которых белки находились слишком далеко для связывания с PROTAC. После дополнительного визуального осмотра расстояний было выбрано по два докинг-решения для каждой кристаллической структуры.

Затем из молекулы PROTAC вырезался линкер, полученные конформации белков накладывались на лиганды и линкер вновь устанавливался вручную. Чтобы избежать искажение длины связей и углов

присоединенного линкера выполнялась минимизация с помощью метода скорейшего спуска и сопряженных градиентов.

После было проведено предварительное тестирование для проверки возможностей МД вместе с общим выбором настроек (выбрано подходящее силовое поле, представление воды, назначены частичные атомные заряды), прежде чем был запущен полный рабочий процесс и исследованы остальные комплексы PROTAC. Подтверждая тем самым, что некристаллическая поза может воспроизвести нативную и что кристаллический ТК стабилен на протяжении всей симуляции. В предварительном тесте была предпринята попытка поиска и оценки ТК BRD4-MZ1-VHL (PDB: 5T35) с использованием классического МД и MMGBSA. В контрольной симуляции поза кристалла оставалась стабильной в течение 110 нс (рисунок 20). α -BRD4 были выровнены по позе кристалла для каждого фрейма, RMSD PROTAC построен на основе ЕЗ-лигазы VHL с лигандом и RMSD VHL был построен на основе его α .

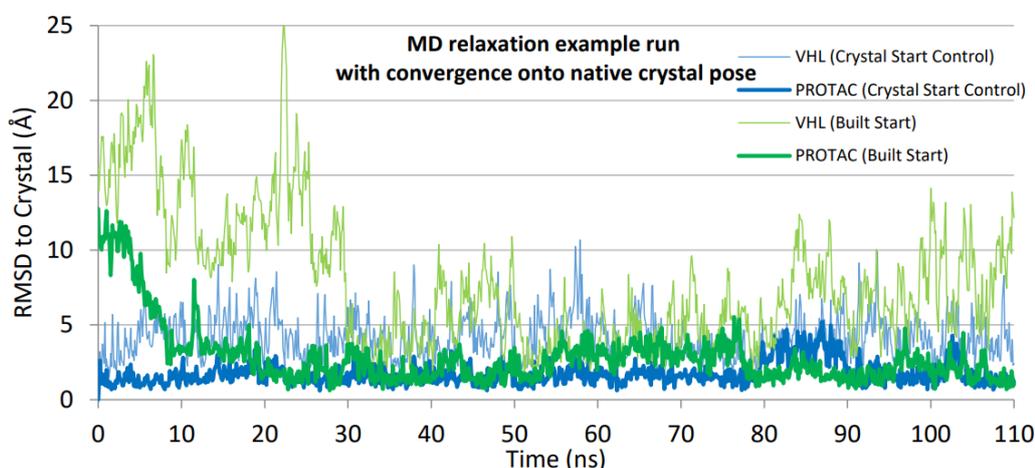


Рисунок 20 – Симуляция молекулярной динамики нативного тройного комплекса и позы 2-MD3

Затем авторы рассчитывали методом MMGBSA ΔH образования ТК и $\Delta\Delta H$ (кооперативность связывания). Среди 4 экспериментов, которые приводили к воспроизведению кристаллической позы, а также одного

эксперимента, который начинался с нативного ТК, в пяти из пяти случаях значения энтальпии связывания (ΔH) оказались благоприятными (таблица 4). Этот обнадеживающий результат позволяет предположить, что быстрое локальное МД-уточнение потенциально способно находить стабилизированные ТК, а классический показатель связывания MMGBSA может быть полезен для ранжирования нативной позы в верхние положения кластеров.

Таблица 4 – Моделирование при 350 К в неявном растворителе с последующим анализом энергии MMGBSA

Начальная поза	Сходимость с нативной позой	$\Delta H_{\text{связывания, ТК}}$, ккал/моль	ΔH ранг	$\Delta\Delta H$, ккал/моль	$\Delta\Delta H$ ранг
кристаллическая структура	контроль	-117.5 ± 0.6	5.5	-25.3 ± 0.8	4.5
1-MD1	да	-119.1 ± 0.1	5	-26.4 ± 1.5	4
1-MD4	да	-119.1 ± 0.9	4	-32.1 ± 1.2	2
2-MD3	да	-122.1 ± 0.9	3	-30.2 ± 1.5	3
2-MD4	да	-122.3 ± 0.8	2	-32.7 ± 1.1	1
1-MD2	нет	-112.6 ± 0.8	8	-14.5 ± 1.1	9
1-MD3	нет	-105.3 ± 0.7	10	-11.7 ± 0.7	10

Основываясь на этом предварительном успехе, было принято решение расширить область предсказания, включив в нее другие ТК, в том числе те, которые не удавалось предсказать ранее описанными методами. Докинг в этом случае начинался с независимых структур белков, связанных с лигандами, чтобы имитировать ситуацию, когда информация о кристаллической структуре ТК отсутствует.

1.3.4 Подходы к ранжированию тройных комплексов

Несмотря на большое количество различных подходов к получению ТК и способности этих алгоритмов генерировать в том числе структуры, близкие к кристаллическим, до сих пор открыт вопрос поиска метрики, способной эффективно выявлять кристаллоподобные решения.

Одним из таких подходов является кластеризация полученных поз тройных комплексов. Так, Драмонд и др. в своей работе предположили, что существует корреляция между размером кластера и обогащением его кристаллоподобными позами, т.е. наиболее многочисленный кластер будет в целом обогащен кристаллоподобными комплексами [52].

К конечному набору смоделированных ТК были применены два отдельных алгоритма кластеризации: один - к самим белкам, другой - к конформациям PROTAC, используемым в этих ТК. Основывалось разделение на $C\alpha$ RMSD и RMSD для белков и конформаций PROTAC соответственно. После чего происходила процедура так называемой двойной кластеризации (совмещение полученных результатов). Эту процедуру повторяли для всех возможных комбинаций двух отдельных кластерных порогов, т.е. 5, 10 и $15 \text{ \AA} \times 1, 1,5, 2, 2,5$ и 3 \AA для кластеров белка и PROTAC соответственно. Конкретные пороговые значения для оптимальной процедуры двойной кластеризации были выбраны на основе того, какая комбинация пороговых значений давала наилучшую частоту попадания кристаллоподобных поз в наиболее густонаселенном двойном кластере (таблица 5).

В целом, процедура двойной кластеризации позволяет получить целевой набор смоделированных ТК, существенно обогащенных кристаллоподобными позами. Для тройных комплексов, использующих VHL, более 40 % поз в наиболее густонаселенном двойном кластере являются по крайней мере "приемлемыми" при сравнении с нативной кристаллической структурой. Результат для 5T35 являлся наилучшим, т.к. почти каждый смоделированный комплекс в самом большом двойном кластере является приемлемым.

Таблица 5 – Полученные данные кластеризации

PDB	E3- лигаза	Высокий/средний/приемлемый процент попадания
5T35	VHL	50/950/198 (70.8%)
6BN7	CRBN	0/5/15 (5.1%)
6BOY	CRBN	0/4/0 (0.4%)
6HAX	VHL	88/36/71 (37.1%)
6HAY	VHL	0/36/144 (42.9%)
6HR2	VHL	0/67/32 (22.1%)

Ли и др в своей работе изучали применение более физически обоснованного моделирования посредством молекулярной динамики и механики в сочетании с MMGBSA для моделирования тройных комплексов PROTAC [53].

В качестве референсных структур авторы взяли два закристаллизованных комплекса включающих BRD4 в качестве POI, E3-лигазу VHL и структуру PROTAC MZ1 и PROTAC1 (PDB ID: 5T35 и 6SIS). Кристаллическая структура 5T35 в дальнейшем использовалась в качестве референса для построения тройных комплексов.

На начальном этапе структуры были минимизированы, а затем нагреты до 300 К 200 пс. с временным шагом в 1 фс. После при той же температуре было произведено NVT-моделирование с шагом 2 фс до 240 нс.

Энергия связывания MMGBSA (ΔG_b) определялась по формуле следующим образом (формула 1):

$$\Delta G_b = G_{comp} - (G_{rec} + G_{lig}) \approx \Delta E_{elec} + \Delta E_{vdW} + \Delta E_{GB} + \Delta E_{surf} \quad (1)$$

где G_{comp} – свободная энергия комплекса;

G_{rec} – свободная энергия рецептора;

G_{lig} – свободная энергия лиганда;

E_{elec} – энергия связи при электростатическом взаимодействии;

E_{vdw} – энергия Ван-дер-Ваальсовых взаимодействий;

ΔE_{GB} и ΔE_{surf} – энергия полярной и неполярной сольватации.

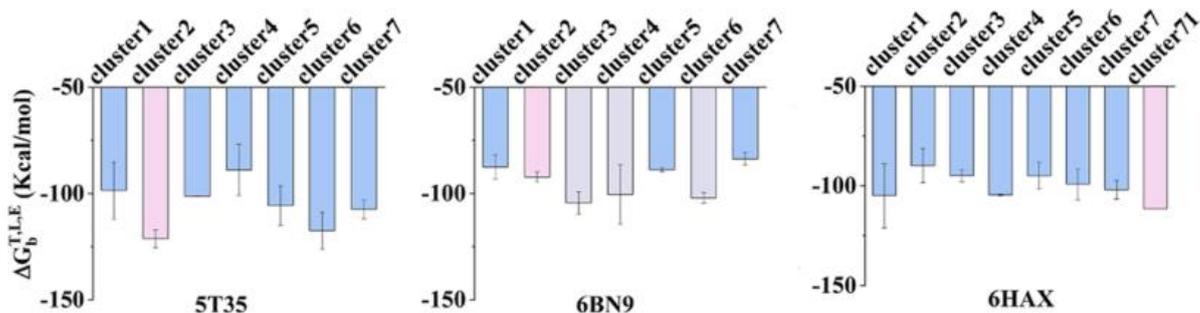
Затем авторы дополнительно оценили каждый кластер, взятый из исследования Зайдмана [47], а также заново ранжировали их используя свой метод. Чтобы снизить вычислительные затраты, вместо использования всех конформаций в каждом кластере были выбраны и использованы только репрезентативные, т.е. те которые охватывают все конформационное разнообразие в каждом кластере, для моделирования МД и расчета MMGBSA.

В качестве «двойной проверки» использовались два критерия отбора. Первый – отбор на основе RMSD атомов $C\alpha$ в подвижной цепи каждого комплекса, определенного в PRosettaC. Эталонная структура выбиралась в каждом кластере. Затем конформации отбирались вручную по наиболее распространенному значению RMSD в каждом кластере.

Затем каждая выбранная конформация подвергалась 100 нс МД-симуляции, на основании которой рассчитывалась MMGBSA энергия связывания ΔG_b тройного комплекса, а кластерная энергия рассчитывалась как среднее значение всех значений энергии в каждом кластере (рисунок 21).

Чтобы избежать необъективности при отборе, был также протестирован другой критерий отбора, основанный на одноточечной энергии MMGBSA. Сначала каждая конформация в кластере подвергалась минимизации энергии, а затем рассчитывалась одноточечная энергия MMGBSA. Кроме того, все конформации из докинга были отсортированы по этой энергии. После чего в каждом кластере конформации сортировали от самого низкого значения энергии до самого высокого и производили отбор.

Таким образом, было доказано, что метод MMGBSA позволяет прогнозировать стабильность связывания тройных комплексов PROTAC, посредством ранжирования соединений.



Розовая полоса представляет собой кристаллоподобный кластер. Серая полоса указывает на кластер с самой низкой энергией, если он не является кристаллоподобным

Рисунок 21 – Результаты ранжирования поз с использованием критериев выбора на основе ΔG

Благодаря все более детальному пониманию того, какие параметры играют главную роль в образовании стабильных тройных комплексов, в настоящее время авторы предлагают использовать альтернативные методы оценки и ранжирования полученных ТК.

2 Результаты и обсуждение

Как уже обсуждалось выше, рациональный дизайн PROTAC является не тривиальной задачей ввиду того, что он представляет собой трудоемкую и длительную работу, включающую множество итераций синтеза, оптимизаций и биологических испытаний. По этим причинам разработка *in silico* подходов, которые позволяют идентифицировать эффективные деградеры, является актуальной задачей.

Существующие в настоящий момент вычислительные методы опираются на предположение о том, что эффективный PROTAC должен приводить к образованию стабильного тройного комплекса. Таким образом, *in silico* алгоритмы должны решать две задачи: с одной стороны, генерировать структуры возможных ТК для различных деградеров, а с другой стороны, оценивать и идентифицировать среди них наиболее энергетически выгодные. Несмотря на то, что с первой задачей большая часть доступных вычислительных инструментов справляется отлично, вопрос идентификации и оценки является не решенным. В основном из-за отсутствия единой метрики, которую можно было бы использовать при скоринге. Кроме того, анализ большей части соответствующих профильных публикаций позволяет сделать вывод о том, что еще на этапе валидации таких алгоритмов попытки предсказать структуру закристаллизованного тройного комплекса приводят к неутешительным результатам – кристаллоподобные позы присутствуют в выдаче результатов, но по-прежнему имеют низкие ранги.

Таким образом, целью данной работы стала валидация ряда подходов, направленных на генерацию тройных комплексов POI-PROTAC-CRBN.

Перед нами были поставлены следующие задачи:

1. Генерация тройного комплекса BRD4-PROTAC-CRBN на основе аналога dBET6 помощью IFD-MD;

2. Выполнение молекулярно-динамической симуляции IFD-MD-поз и сравнение структурных данных с таковыми для кристаллического комплекса dBET6;

3. Подбор оптимальных параметров запуска альтернативного алгоритма Protein Degradation Sampler для получения тройных комплексов;

4. Оценка применимости энергетических характеристик, рассчитываемых с помощью модели Prime, для идентификации кристаллоподобных решений Protein Degradation Sampler.

В данной работе в качестве референсного тройного комплекса был выбран BRD4-dBET6-CRBN (PDB ID: 6BOY) по нескольким причинам:

1. Наиболее распространенными E3-лигазами, которые используются при моделировании структур ТК являются CRBN и VHL. Ключевым фактором, повлиявшим на наш выбор, стал размер комплексов, нацеленных на рекрутирование таких E3-лигаз. ТК с CRBN существенно меньше по размеру, что значительно сокращает вычислительные затраты при дальнейшей работе с ними.

2. Закристаллизованный комплекс 6BOY имеет приемлемое разрешение (3.33 Å), не содержит в себе молекул воды, а также каких-либо серьезных структурных ошибок.

3. PROTAC dBET6 представляет собой структуру с оптимальной длиной линкера.

4. Центр медицинской химии, на базе которого проводился эксперимент, специализируется на разработке химерных молекул, ориентированных на рекрутирование CRBN. А также в нем велась разработка бензотриазольного аналога dBET6.

Схематично dBET6 и его бензотриазольный аналог (BT-PROTAC) представлены на рисунке ниже (рисунок 22).

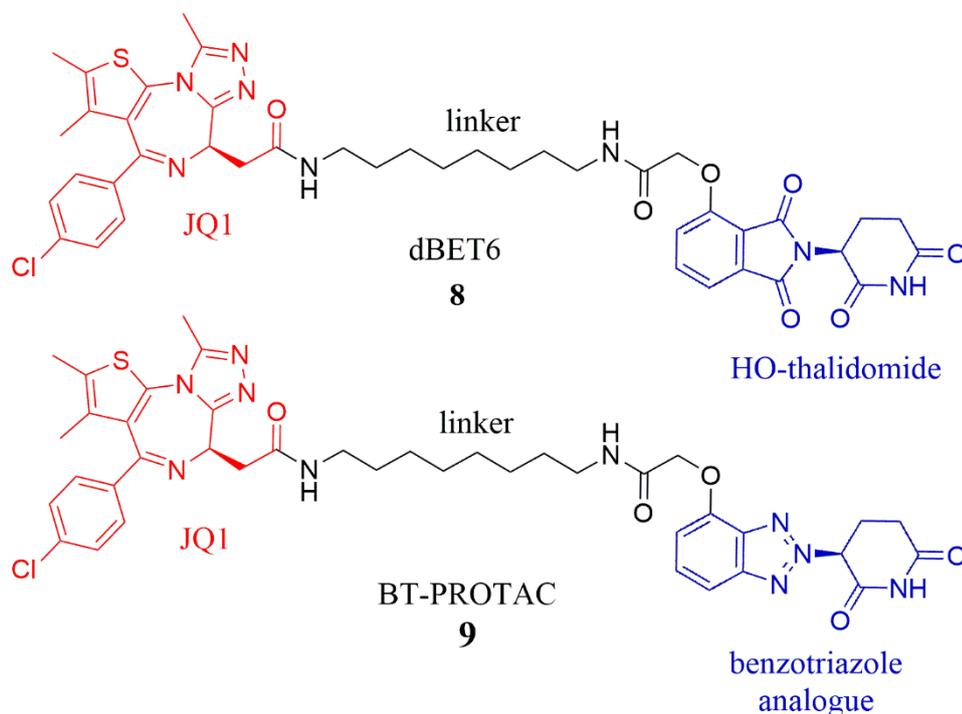


Рисунок 22 – Структуры PROTACs dBET6 и бензотриазольного аналога

2.1 Дизайн исследования

В данной работе мы предположили, что для прогнозирования структуры тройного комплекса и предсказания конформации деградера можно использовать алгоритмы индуцированной стыковки (*англ.* IFD, induced fit docking), позволяющие моделировать конформационные изменения белка при связывании лиганда [54]. Полученные таким образом ТК в дальнейшем могут быть использованы для задач рационального дизайна химер, направленных на протеолиз, а также в качестве входных данных для более точных методов молекулярного моделирования – например, расчета аффинности методом возмущения свободных энергий (*англ.* FEP, free energy perturbation) [55]. Однако стоит понимать, что данный алгоритм работает только в случае близкородственных PROTAC, в частности, потому что в качестве входных

данных необходима информация о кристаллической структуре тройного комплекса, которых в настоящее время не так много.

Мы использовали алгоритм IFD-MD [56], потому что он позволяет моделировать структуры высокого качества, а также эффективно предсказывать способы связывания белков с лигандами.

На последнем этапе исследования нами также был применен и альтернативный подход прогнозирования структур тройных комплексов, а именно алгоритм Protein Degrader Sampler.

2.2 Подготовка структуры белков и лигандов

Кристаллическая структура тройного комплекса BRD4-CRBN-DDB1-dBET6 (PDB ID: 6BOY), обладающая разрешением 3.33 Å, была загружена из Protein Data Bank [57]. Тройной комплекс был представлен как гетеротример, включающий в себя цепь А (DDB1), цепь В (CRBN), цепь С (BRD4), а также лиганд, представленный молекулой PROTAC (dBET6), и ион цинка Zn^{2+} . Перед началом подготовки белка из структуры была удалена цепь, включающая DDB1 для сокращения времени дальнейших расчетов.

Затем белки были подготовлены с помощью модуля Protein Preparation Wizard [58], интегрированного в пакет ПО Schrodinger Suite. Структура не содержала серьезных ошибок, а именно отсутствующих боковых цепей, петель или остатков, имеющих альтернативные ориентации.

Структура WT-PROTAC была загружена, а затем подготовлена с помощью модуля LigPrep [59].

2.3 Получение тройного комплекса BRD4-PROTAC-CRBN для бензотриазольного аналога dBET6 с помощью алгоритма IFD-MD

Во всех современных подходах структурно-ориентированного дизайна лекарств (*англ.* SBDD, Structure-Based Drug Discovery) основополагающим фактором является наличие или предсказание способа связывания лигандов с белками-мишенями. В настоящий момент наиболее точными являются экспериментальные методы, такие как криоэлектронная микроскопия [60], PCA [61] и ЯМР [62]. Однако, такие эксперименты являются дорогими и долгими. В случае, когда конформация рецептора значительно не меняется при связывании с новым лигандом, проблема предсказания структуры становится решаемой с помощью алгоритмов молекулярного докинга.

Ситуация сильно осложняется, когда имеют место быть эффекты индуцированного соответствия, влекущие за собой сильные изменения конформации белка, которые «классический» молекулярный докинг не учитывает.

Одним из алгоритмов, позволяющих решить проблемы конформационного изменения структуры белков является IFD, разработанный Schrodinger. Основная суть заключается в том, чтобы объединить протокол стыковки Glide [63] и моделирование белков посредством Prime [64]. Оптимизация белков происходит в несколько этапов. Исходя из значений В-фактора или площади, доступной для растворителя, выбирается ряд аминокислотных остатков, боковые цепи которых затем удаляются.

В дальнейшем выполняется «классический» докинг, а затем достраивание отсутствующих боковых цепей и минимизация комплексов.

Но, к сожалению, на начальном этапе докинга иногда не идентифицируется ни одной позы близкой к нативной, а также оценочная функция не ранжирует лучшую позу в число 5-10 наиболее подходящих.

Для решения этой проблемы в 2020 году командой Schrodinger был создан новый алгоритм, направленный на повышение точности и надежности результатов, получаемых с помощью IFD. Коллеги решили объединить ряд различных технологий моделирования, а именно фармакофорный докинг [65], WScore [66], метадинамическое моделирование [67], уточнение посредством симуляции молекулярной динамики, с известными алгоритмами Glide и Prime.

Помимо того, что данный алгоритм объединяет в себя большое количество инструментов молекулярного моделирования, используемых в вычислительной химии, он отличается этапом оптимизации белковых структур. На этапе уточнения боковых цепей с помощью Prime, для конформации лиганда, получившейся в ходе фармакофорного докинга, выбираются боковые цепи, с атомами которых образуются стерические столкновения.

Несмотря на вычислительную эффективность, этап оценки результатов является более сложным, т.к приходится учитывать не только RMSD лиганда относительно кристаллической структуры, но и точность индуцированных изменений самого белка.

В качестве входных структур требуется информация о двух белково-лигандных комплекса R_A и R_B одного и того же белка интереса. Цель алгоритма состоит в том, чтобы состыковать лиганд комплекса R_A к белку комплекса R_B , с получением точной позы. Рабочий процесс IFD-MD представлен на рисунке (рисунок 23).

Как уже отмечалось выше, изначально реализуется грубый фармакофорный докинг с использованием Phase [68]. Генерируется библиотека конформеров лиганда интереса и осуществляется итеративная процедура наложения полученных фармакофорных моделей на фармакофор референсного лиганда в сайте. В процессе такого совмещения неизбежно будут возникать стерические столкновения атомов лиганда с боковыми цепями белка, однако они затем будут устранены. Формируется набор из 1000 поз лиганда интереса на основе степени наложения фармакофоров.

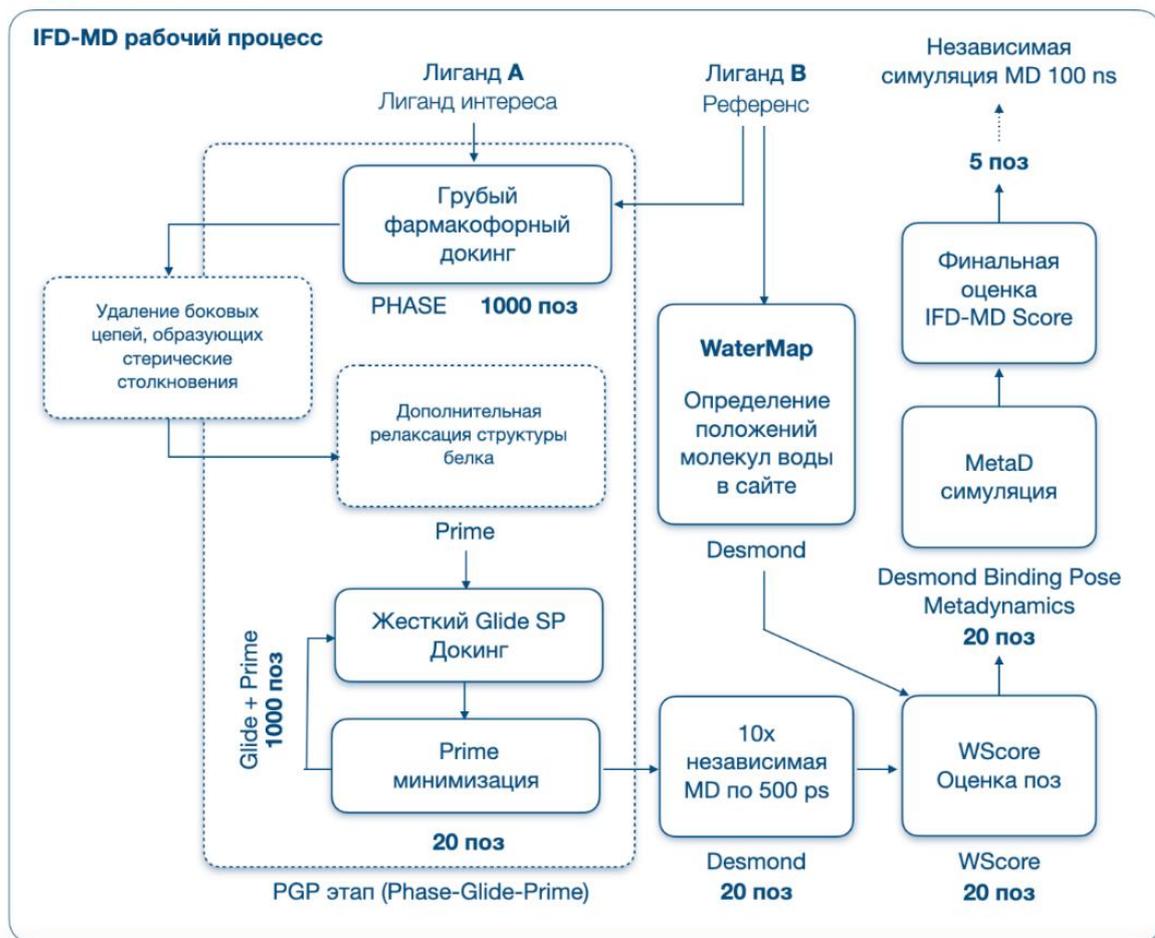


Рисунок 23 – Рабочий процесс IFD-MD [56]

Таким образом, для VT-PROTAC было получено 109267 конформеров, на основе которых были построены фармакофорные модели, распределенные на 58413 кластеров, центр каждого кластера использовался для наложения на фармакофор dBET6. Наложение фармакофоров лучшей оцененной алгоритмом позы и референсного PROTAC в связывающей конформации представлено на рисунке ниже (рисунок 24).

Как видно из данных, представленных на рисунке, самое лучшее совпадение демонстрирует акцепторно-донорная-акцепторная триада глутаримидного кольца сравниваемых молекул. Несмотря на различие в конформации и векторе линкера, атомы кислорода карбонильных фрагментов

также накладываются. В свою очередь, BRD-связывающая боеголовка претерпела зеркальную инверсию ориентации, тем не менее обеспечивая соответствие кольцевых фармакофорных признаков.

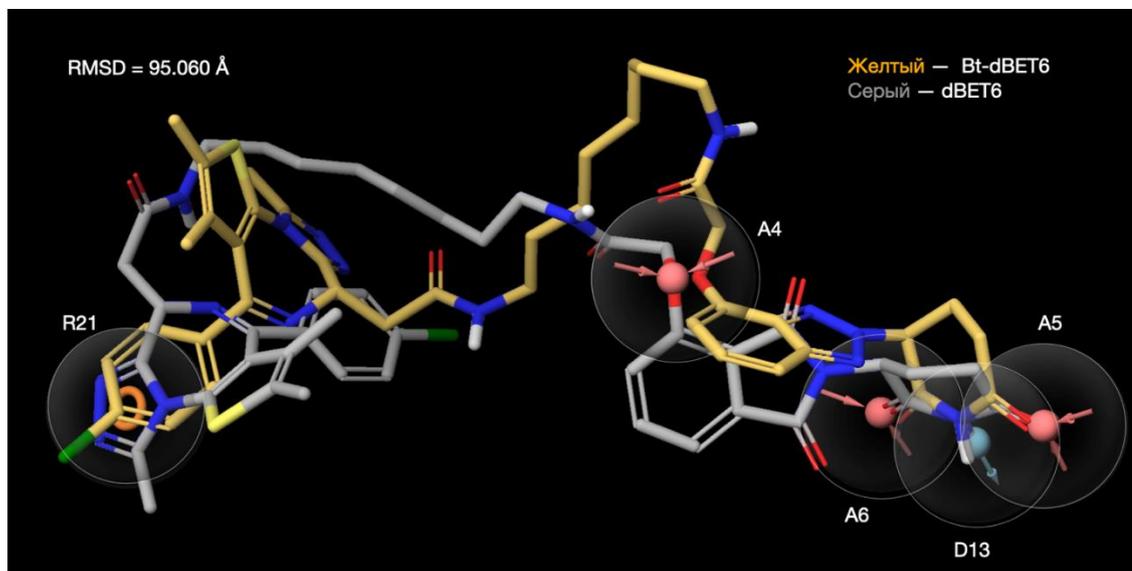


Рисунок 24 – Наложение структур Vt-PROTAC (желтый) и dBET6 (серый) с демонстрацией совпавших фармакофорных сайтов, определенных Phase

Алгоритм также произвел расчет RMSD атомных координат между генерируемыми позами и референсным лигандом. Среди всех полученных результатов фармакофорного докинга минимальное отклонение $RMSD_{min}$ составило 89.037 Å, максимальное $RMSD_{max}$ – 99.498 Å.

Все отобранные конформации демонстрировали стерические конфликты с атомами белков.

В качестве примера приведем уже описанную ранее позу – атомы CRBN-связывающего бензотриазольного лиганда показывают перекрывание Ван-дер-Ваальсовых радиусов с атомами таких остатков как Trp380 (по большей части), Trp386, Trp400, Phe402.

Наличие таких клэшей (*англ.* clash) также, как и больших значений RMSD, не критично – в ходе последующих процедур неизбежно осуществится минимизация структуры комплекса (рисунок 25).

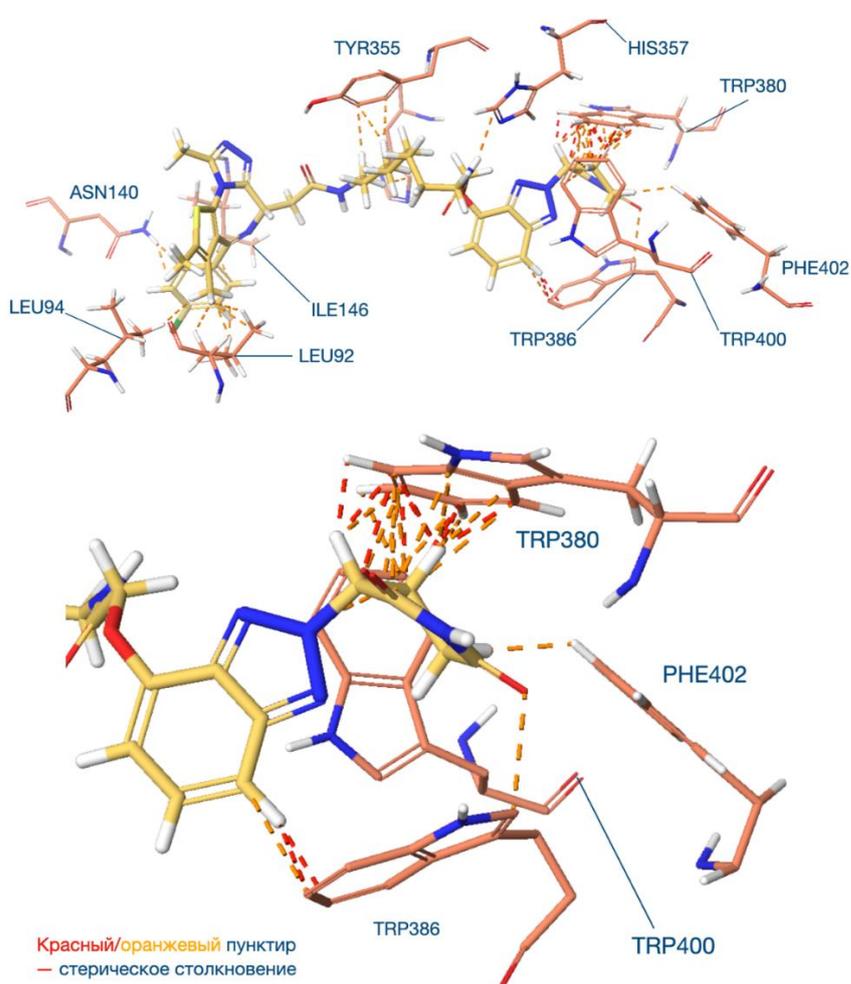


Рисунок 25 – Стерические столкновения между атомами WT-PROTAC и бВΟΥ (поза № 1 по результатам Phase-докинга)

Затем каждая из полученных поз подвергалась процедуре разрешения столкновений и релаксации с помощью Prime.

Вкратце, Prime исследует конформационное пространство остатков, используя библиотеку ротамеров. Аминокислотные остатки выбранных цепей оптимизируются или заменяются на аланин. Данная процедура происходит в присутствии лиганда, в отличие от стандартного IFD, где «обрезка» боковых цепей (*англ.*, *trimming side chains*) происходит без учета лиганда.

Комплекс был минимизирован, в том числе оптимизации подвергалась и конформация лиганда.

После осуществлялся следующий процесс – в каждой из 1000 поз боковые цепи аминокислот, участвующих в стерических конфликтах, удалялись, выполнялась жесткая стыковка Glide SP лиганда интереса, а конформации аминокислотных остатков предсказывалась *de novo* с использованием Prime.

В измененный сайт вновь осуществлялась стыковка, и результирующие комплексы минимизировались.

Значения GlideScore – суммарной эмпирической оценочной функции Glide SP, и GlideECoul – термина GlideScore, кодирующего вклад в энергию связывания кулоновских взаимодействий, использовались для отбора 1000 лучших кандидатов на каждом из этапов стыковки.

Полученные позы повторно оценивались с помощью параметра PGP_{Score} (*англ.* Phase-Glide-Prime) – линейной комбинации Prime Energy, GlideScore и GlideECoul с коэффициентами, показанными ниже (формула 2).

$$PGP_{score} = 1.0PrimeEnergy + 9.057GlideScore + 1.428GlideECoul, \quad (2)$$

где PrimeEnergy – полная потенциальная энергия минимизированного лиганд-белкового комплекса;

GlideScore – значение оценочной функции Glide SP;

GlideECoul – вклад кулоновских взаимодействий в Glide SP

На этапе валидации алгоритма авторами было показано, что поза, соответствующая/близкая кристаллической, как правило, находилась среди топ-20 результатов при ранжировании по PGP_{Score} .

В нашем случае алгоритм также отобрал 20 поз для следующего этапа. Значения PGP_{Score} для топ-20 отобранных поз представлены в таблице ниже (таблица 6).

Таблица 6 – значения PGP_{Score} для топ-20 поз PROTAC на стадии PGP

Поза	GlideECoul, ккал/моль	PrimeEnergy, ккал/моль	GlideGScore, ккал/моль	PGP_{Score} , ккал/моль	$RMSD_{Lig}^*$, Å
1	-20.365	347690004	-18.122	347496.7918	6.999
2	-12.985	347683.234	-16.116	347518.7288	5.328
3	-18.229	347692.346	-15.452	347526.3662	7.105
4	-13.847	347691.517	-15.951	347527.2753	7.513
5	-10.134	347681.346	-15.077	347530.3223	12.550
6	-15.830	347696.916	-15.847	347530.7845	9.189
7	-12.380	347697.466	-16.271	347532.4209	7.843
8	-15.110	347695.415	-14.940	347538.5263	7.043
9	-10.596	347703.770	-16.424	347539.8867	13.463
10	-8.879	347690.635	-15.238	347539.9452	13.269
11	-14.193	347705.789	-15.724	347543.1091	6.792
12	-13.773	347700.459	-14.910	347545.7513	6.130
13	-16.318	347707.377	-15.226	347546.1730	8.799
14	-17.454	347709.636	-15.240	347546.6830	9.423
15	-16.306	347708.571	-15.259	347547.0853	7.251
16	-16.879	347701.515	-14.387	347547.1087	9.132
17	-17.524	347711.156	-15.317	347547.4057	6.552
18	-9.485	347687.377	-13.891	347548.0216	13.210
19	-13.871	347700.068	-14.469	347549.2145	5.776
20	-9.680	347696.633	-14.641	347550.2064	13.070

* – RMSD координат атомов лигандов измерялось между позой, полученной на финальной оптимизации Prime и кристаллической конформацией с помощью модуля Superpostion.

Стоит отметить, что уже на этом этапе конформация ВТ- PROTAC в сайте, образованном BRD4 и CRBN, довольно сильно напоминала связывающую конформацию dBET6 по данным PCA. Исходя из рассчитанных значений RMSD, несмотря на то, что данная поза имеет лучшее значение PGP_{Score} , при наложении ее на нативный лиганд, можно заметить, что расположение линкера, как и боеголовок, значительно отличается. Вторая поза, имеющая значение PGP_{Score} незначительно больше, при наложении показывала RMSD меньше, чем у предыдущей. На рисунке ниже можно заметить, что линкер и боеголовки имеют конформации, близкие к нативной.

Также при анализе результатов, нами было установлено, что среди неотобранных результатов присутствует поза 37, которая имеет значение RMSD немного меньше, чем поза 2, хотя ее значение $PGP\ Score = 347564.372$, $GlideScore = -14.050$, $E_{coul} = -12.029$, $Prime = 347708.797$ сильно отличаются от поз в топе 20. Наложение данных поз представлено на рисунке ниже (рисунок 26).

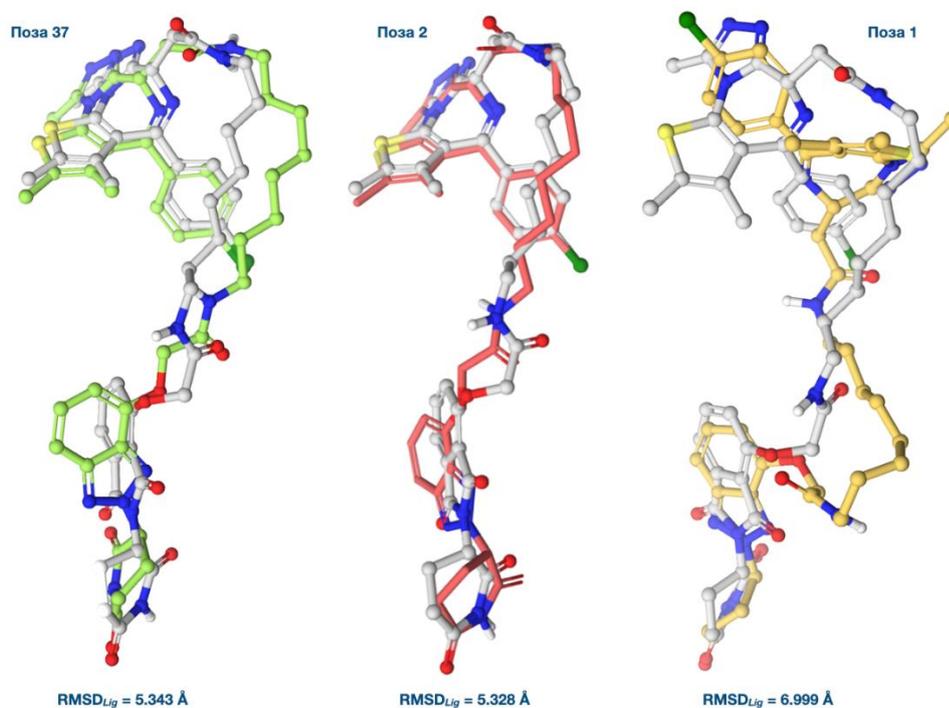


Рисунок 26 – Наложение поз 37 (зеленый), 2 (красный) и 1 (желтый), полученных после стадии PGP, и связывающей конформации dBET6 (серый)

Исходя из полученных результатов можно сделать вывод, что уже на ранних этапах генерации конформаций PROTAC можно получать достаточно оптимистичные результаты, т.е. полученные конформеры являются уже достаточно близкими с нативному dBET6, но все же имеют значительные отклонения, которые будут устранены на последующих этапах, включающих МД.

Двадцать отобранных поз затем были подвергнуты серии из десяти независимых симуляций молекулярной динамики по 500 пс каждая. Суть этого этапа заключается в том, что моделирование молекулярной динамики позволяет исключить любые близкие контакты молекул или другие существенные проблемы, которые присутствуют в структуре комплекса, в случае если они не были устранены на предыдущих этапах. Также использование такой симуляции перед оценкой WScore может улучшить исходную структуру IFD и устранить неправильную десольватацию полярных и заряженных групп. Проведение нескольких независимых МД-исследований также необходимо для того, чтобы избежать ложного назначения штрафов для нативных структур.

После проведения MD для всех поз был проведен расчет WScore (в режиме *score-in-place*) – скоринговой функции, позволяющей оценить способ связывания малой молекулы в сайте с учетом эффектов явно представленных молекул воды. Согласно оригинальной публикации, WScore имеет следующий вид (формула 3):

$$WScore = cE_{vdw} + E_{hydrophob} + E_{HB} + E_{MR} + E_{deloc} + E_{strain} + E_{desolv} + E_{reorg} + S_{complex}, \quad (3)$$

где cE_{vdw} – энергетический вклад Ван-дер-Ваальсовых взаимодействий, умноженный на весовой коэффициент c (который идентичен оптимизированному значению, используемому в скоринговой функции Glide XP [69]);

$E_{hydrophob}$ – энергетический вклад гидрофобных взаимодействий, вычисляемый попарно между всеми неполярными атомами;

E_{HB} – энергетический вклад образования водородных связей и солевых мостиков;

E_{MR} – энергетический вклад терминов «молекулярного распознавания» (*англ.*, molecular recognition (MR)). Включает π -катионные и ароматические взаимодействия;

E_{strain} – штраф за локализованную энергию деформации, который включает в себя электростатические отталкивания между белком и акцепторными группами лиганда;

E_{desolv} – штраф за локализованную десольватацию полярных или заряженных групп лиганда или белка;

E_{deloc} – штраф за эффекты делокализованной деформации и десольватации. Делокализованная деформация может включать стерические столкновения между белком и лигандом, умеренно неблагоприятные энергии конформаций и электростатические эффекты, не учитывающиеся выше;

E_{reorg} – штраф за свободную энергию реорганизации белка. Эта величина формально определяется как свободная энергия, необходимая для изменения конформации белка из апо-состояния (несвязанное) в холо-состояние, индуцированное соответствующим лигандом;

$S_{complex}$ – приближенная обработка энтропии комплекса белок-лиганд. Когда лиганд получает благоприятные оценки в нескольких рецепторах, это указывает на то, что большая часть фазового пространства комплекса энергетически доступна для лиганда, и, следовательно, общая энтропия комплекса выше.

Для расчета параметров $E_{hydrophob}$ и E_{desolv} алгоритму требуется информация о термодинамических параметрах молекул воды, размещенных в

сайте связывания, которая получается в ходе расчета WaterMap [70]. WaterMap выполняет явное моделирование молекулярной динамики растворителя и молекулы воды из каждого сохраненного кадра симуляции группируются в сферы радиусом 1 Å (т. н. сайты гидратации). Каждый сайт анализируется с применением теории неоднородной сольватации (англ. IST, Inhomogeneous solvation theory) [71] для вычисления локальных энтальпийных (ΔH) и энтропийных ($-T\Delta S$) свойств воды, занимающей тот или иной участок. Затем свободная энергия (ΔG) воды в каждом участке гидратации берется как сумма вкладов ΔH и $-T\Delta S$ [72,73]. Полученное значение ΔG можно трактовать как свободную энергию перемещения воды из данного участка сайта связывания в объемный растворитель, и те молекулы, которые обладают большими значениями, являются нестабильными – вытеснение их из сайта при связывании маломолекулярного лиганда приведет к улучшению аффинности.

Алгоритм IFD-MD выполняет моделирование WaterMap для референсного комплекса и рассчитывает все обсуждаемые выше термодинамические характеристики. На рисунке ниже представлен результат расчета (рисунок 27).

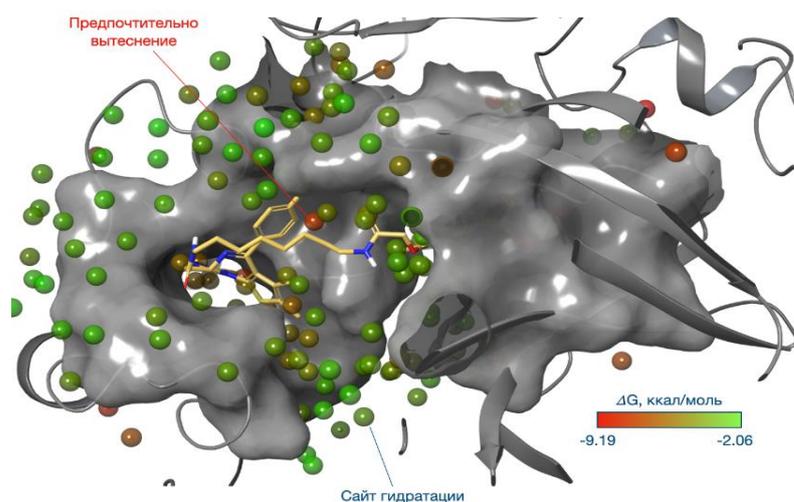


Рисунок 27 – Результаты расчета WaterMap

В таблице 7 приведены представлена полная характеристика всех сайтов гидратации, идентифицированных алгоритмом.

Таблица 7 Стерические и термодинамические параметры сайтов гидратации WaterScore

№	Occup	Overlap	ΔH	$-T\Delta S$	ΔG	№	Occup	Overlap	ΔH	$-T\Delta S$	ΔG	№	Occup	Overlap	ΔH	$-T\Delta S$	ΔG	№	Occup	Overlap	ΔH	$-T\Delta S$	ΔG
9	0,91	1,00	0,52	3,52	4,04	121	0,36	0,00	1,65	1,03	2,68	52	0,61	0,00	-2,76	2,03	-0,73	108	0,37	0,00	-1,03	1,06	0,03
13	0,89	1,00	6,08	3,11	9,19	122	0,36	0,00	1,60	1,02	2,62	54	0,60	0,00	1,26	1,97	3,23	109	0,37	0,00	-0,88	1,08	0,2
23	0,74	1,00	0,29	2,33	2,62	123	0,35	0,00	-0,49	1,01	0,52	55	0,60	0,00	-2,16	1,89	-0,27	110	0,37	0,00	-0,14	0,99	0,85
40	0,66	1,00	1,62	1,98	3,6	124	0,35	0,00	-0,29	1,01	0,72	56	0,60	0,00	0,24	1,70	1,94	111	0,37	0,00	0,19	1,07	1,26
51	0,61	1,00	4,44	1,93	6,37	126	0,35	0,00	-0,42	1,02	0,6	57	0,59	0,00	-1,46	1,77	0,31	112	0,37	0,00	-1,53	1,10	-0,43
53	0,61	1,00	-0,43	1,77	1,34	127	0,34	0,00	0,03	0,93	0,96	58	0,59	0,00	0,70	1,78	2,48	113	0,36	0,00	0,71	1,00	1,71
71	0,48	1,00	4,17	1,30	5,47	128	0,34	0,00	0,13	0,95	1,08	59	0,58	0,00	2,58	1,77	4,35	115	0,36	0,00	0,83	1,04	1,87
77	0,46	1,00	2,37	1,26	3,63	129	0,33	0,00	0,20	1,05	1,25	60	0,58	0,00	0,69	1,74	2,43	116	0,36	0,00	2,26	1,02	3,28
81	0,45	1,00	0,07	1,28	1,35	130	0,33	0,00	0,12	1,14	1,26	61	0,58	0,00	-0,05	1,77	1,72	117	0,36	0,00	-1,06	0,97	-0,09
92	0,42	1,00	0,63	1,15	1,78	131	0,33	0,00	-0,49	0,92	0,43	62	0,57	0,00	0,14	1,69	1,83	118	0,36	0,00	0,46	1,00	1,46
99	0,39	1,00	2,00	1,15	3,15	132	0,33	0,00	-0,16	0,93	0,77	63	0,57	0,00	0,25	1,73	1,98	24	0,74	0,00	1,59	2,41	4
106	0,38	1,00	1,56	1,14	2,7	134	0,33	0,00	-0,25	1,10	0,85	64	0,54	0,00	-2,18	1,68	-0,5	25	0,74	0,00	-2,27	2,32	0,05
120	0,36	1,00	1,95	0,90	2,85	135	0,33	0,00	-0,85	0,91	0,06	65	0,53	0,00	1,11	1,85	2,96	26	0,72	0,00	-1,25	2,35	1,1
125	0,35	1,00	1,64	1,01	2,65	136	0,32	0,00	-0,84	0,89	0,05	66	0,53	0,00	2,55	1,63	4,18	27	0,71	0,00	3,02	2,79	5,81
133	0,33	1,00	2,68	0,97	3,65	137	0,32	0,00	1,48	0,85	2,33	67	0,52	0,00	-0,56	1,45	0,89	28	0,71	0,00	2,10	2,22	4,32
148	0,30	1,00	1,92	0,89	2,81	138	0,31	0,00	1,35	0,87	2,22	68	0,51	0,00	0,38	1,84	2,22	29	0,70	0,00	2,55	2,83	5,38
149	0,30	1,00	3,23	0,83	4,06	139	0,31	0,00	-2,09	1,21	-0,88	69	0,50	0,00	0,98	1,78	2,76	30	0,70	0,00	-2,03	2,37	0,34
157	0,29	1,00	1,70	0,76	2,46	140	0,31	0,00	-0,24	0,97	0,73	70	0,49	0,00	-0,16	1,49	1,33	31	0,70	0,00	-2,46	2,39	-0,07
166	0,28	0,81	4,35	0,74	5,09	141	0,31	0,00	0,72	0,83	1,55	72	0,48	0,00	-1,61	1,49	-0,12	32	0,69	0,00	-2,62	2,20	-0,42
114	0,36	0,80	0,11	1,05	1,16	142	0,31	0,00	0,23	0,83	1,06	73	0,47	0,00	-1,93	1,52	-0,41	33	0,69	0,00	-0,81	2,23	1,42
119	0,36	0,64	0,15	1,04	1,19	144	0,31	0,00	0,26	0,85	1,11	74	0,47	0,00	-0,32	1,34	1,02	34	0,69	0,00	-1,06	2,09	1,03
78	0,46	0,15	1,07	1,25	2,32	145	0,31	0,00	0,35	0,81	1,16	75	0,47	0,00	0,13	1,45	1,58	35	0,68	0,00	0,54	2,53	3,07

Продолжение таблицы 7

№	Occup	Overlap	ΔH	$-T\Delta S$	ΔG	№	Occup	Overlap	ΔH	$-T\Delta S$	ΔG	№	Occup	Overlap	ΔH	$-T\Delta S$	ΔG	№	Occup	Overlap	ΔH	$-T\Delta S$	ΔG
82	0,45	0,10	1,59	1,30	2,89	146	0,30	0,00	0,98	0,84	1,82	76	0,46	0,00	-0,56	1,38	0,82	36	0,68	0,00	1,20	1,99	3,19
155	0,29	0,08	0,48	0,84	1,32	147	0,30	0,00	-2,26	0,94	-1,32	79	0,46	0,00	-3,42	1,36	-2,06	37	0,68	0,00	-1,09	2,00	0,91
80	0,45	0,07	-0,82	1,35	0,53	150	0,30	0,00	0,49	0,84	1,33	83	0,45	0,00	-0,62	1,27	0,65	38	0,67	0,00	-0,87	2,12	1,25
143	0,31	0,03	0,04	0,88	0,92	151	0,30	0,00	0,34	0,90	1,24	84	0,44	0,00	-0,59	1,29	0,7	39	0,66	0,00	-0,80	1,97	1,17
1	1,00	0,00	-0,76	4,18	3,42	152	0,30	0,00	-1,93	0,88	-1,05	85	0,43	0,00	-1,11	1,36	0,25	41	0,66	0,00	-0,69	2,16	1,47
2	1,00	0,00	0,51	4,61	5,12	153	0,30	0,00	0,99	0,84	1,83	86	0,43	0,00	-0,18	1,25	1,07	42	0,65	0,00	-0,79	2,30	1,51
3	1,00	0,00	-3,66	4,94	1,28	154	0,29	0,00	0,76	0,80	1,56	87	0,43	0,00	0,06	1,20	1,26	43	0,65	0,00	-2,76	2,20	-0,56
4	1,00	0,00	3,35	4,00	7,35	156	0,29	0,00	-0,40	0,83	0,43	88	0,43	0,00	-0,83	1,30	0,47	44	0,65	0,00	-0,94	1,94	1
5	1,00	0,00	0,11	4,52	4,63	158	0,29	0,00	0,81	0,91	1,72	89	0,42	0,00	-1,26	1,27	0,01	45	0,64	0,00	-1,46	2,00	0,54
6	0,99	0,00	1,34	4,65	5,99	159	0,29	0,00	0,48	0,79	1,27	90	0,42	0,00	0,07	1,18	1,25	46	0,64	0,00	-3,21	2,37	-0,84
7	0,92	0,00	0,86	3,79	4,65	160	0,29	0,00	1,23	0,85	2,08	91	0,42	0,00	-0,79	1,21	0,42	47	0,64	0,00	-2,22	2,11	-0,11
8	0,92	0,00	1,78	3,41	5,19	161	0,29	0,00	-0,71	0,80	0,09	93	0,41	0,00	0,17	1,24	1,41	48	0,64	0,00	-1,49	1,90	0,41
10	0,90	0,00	0,19	3,76	3,95	162	0,29	0,00	0,39	0,80	1,19	94	0,41	0,00	-0,22	1,14	0,92	49	0,63	0,00	-3,36	2,70	-0,66
11	0,90	0,00	0,43	2,99	3,42	163	0,28	0,00	1,25	0,79	2,04	95	0,41	0,00	-2,21	1,20	-1,01	50	0,63	0,00	-0,41	2,23	1,82
12	0,89	0,00	1,33	3,25	4,58	164	0,28	0,00	-0,17	0,79	0,62	96	0,40	0,00	-0,22	1,17	0,95	19	0,78	0,00	-2,92	2,82	-0,1
14	0,89	0,00	-0,92	3,78	2,86	165	0,28	0,00	0,32	0,78	1,1	97	0,40	0,00	-2,81	1,26	-1,55	20	0,75	0,00	-2,84	2,40	-0,44
15	0,86	0,00	-1,61	3,33	1,72	167	0,28	0,00	0,32	0,76	1,08	98	0,39	0,00	1,23	1,11	2,34	21	0,75	0,00	1,31	2,47	3,78
16	0,83	0,00	0,28	2,70	2,98	168	0,28	0,00	-0,71	0,74	0,03	100	0,39	0,00	-0,05	1,36	1,31	22	0,75	0,00	-0,24	2,31	2,07
17	0,81	0,00	0,02	2,78	2,8	169	0,28	0,00	0,34	0,81	1,15	101	0,39	0,00	1,84	1,40	3,24	172	0,28	0,00	0,69	0,79	1,48
18	0,78	0,00	-1,73	2,71	0,98	170	0,28	0,00	-0,57	0,76	0,19	102	0,38	0,00	3,26	1,16	4,42	173	0,28	0,00	-0,42	0,79	0,37
107	0,38	0,00	-0,59	1,12	0,53	171	0,28	0,00	-2,10	0,82	-1,28	103	0,38	0,00	0,82	1,11	1,93	174	0,28	0,00	-1,62	0,89	-0,73

Значения ΔH , ΔG и $T\Delta S$ приведены в ккал/моль

Если сайт гидратации имеет ΔG и $\Delta H \gg 0$ ккал/моль, то, как правило, он считается гидрофобным. В таком случае вытеснение одной или нескольких молекул воды из мест связывания метильных групп или галогенов лиганда считается необходимым.

Если $\Delta H \ll 0$ ккал/моль, а $\Delta G \sim 0$ ккал/моль, то это означает, что молекулы воды взаимодействуют с белком, но такое взаимодействие приводит к значительному энтропийному штрафу. Наилучшим решением данной проблемы будет замена полярной группы лиганда, которая не приведет к потере взаимодействия, но позволит уменьшить значение штрафа за счет удаления молекулы воды.

Некоторые же сайты гидратации сложно поддаются вытеснению или замене ($\Delta G \ll 0$ ккал/мол) из-за энергетических или геометрических особенностей. В таких случаях обычно стоит рассматривать воду, как часть белка, формируя при этом солевые мостики.

Алгоритм IFD-MD проводит анализ всех десяти пятисотпикосекундных симуляций с помощью несколько видоизмененной оценочной функции WScore (формула 4)

$$\begin{aligned}
 WScore = & \omega_1 AllMDTrailsPenalized + \\
 & \omega_2 (Num_{WScorePenaltyMDTrails} - 5.0) + \omega_3 \Delta WScore_{BestTrail} + \\
 & \omega_4 WScore_{PhobicEnclosure} + \omega_5 Num_{OHDesolvMDTrails},
 \end{aligned} \tag{4}$$

где $AllMDTrailsPenalized$ – показывает содержат ли все 500 пс МД-симуляции хотя бы один штраф по оценке WScore;

$Num_{WScorePenaltyMDTrails}$ – количество МД-симуляций, имеющих хотя бы один штраф по оценке WScore;

$WScore_{BestTrail}$ – лучшее значение WScore, полученное за все время симуляции для одной позы;

$WScore_{PhobicEnclosure}$ – дополнительная оценка гидрофобного окружения воды (*англ.*, hydrophobic enclosure);

$\text{Num}_{\text{OHDesolvMDTrials}}$ – количество МД-симуляций, в которых гидроксильная группа была десольватирована

$\omega_1, \omega_2, \omega_3, \omega_4, \omega_5$ – весовые коэффициенты.

После чего происходит отбор пяти лучших поз по значению $\text{WScore}_{\text{Func}}$. Значения скоринга для 20 поз, прошедших независимую MD представлены в таблице ниже (таблица 8) – пять отобранных для финальной оценки структур выделены зеленым цветом.

Таблица 8 – Значения оценок WScore для топ-20 поз PROTAC

Поза	$\text{PGP}_{\text{Score}}$, ккал/моль	$\text{WScore}_{\text{Func}}$, ккал/моль	$\text{WScore}_{\text{Pen}}$, ккал/моль
2	347518.7288	-19.822	0.320
11	347543.1091	-14.328	4.600
1	347496.7918	-13.920	4.600
19	347549.2145	-12.352	0.000
10	347539.9452	-10.852	0.640
9	347539.8867	-10.810	1.280
18	347548.0216	-10.215	0.320
5	347530.3223	-7.652	1.280
3	347526.3662	-6.468	4.600
6	347530.7845	-6.364	4.600
20	347550.2064	-6.173	1.280
12	347545.7513	-5.906	4.600
4	347527.2753	-5.505	1.280
8	347538.5263	-5.080	4.600
14	347546.6830	-3.368	4.600
16	347547.1087	-2.459	4.600
15	347547.0853	-1.790	4.600
7	347532.4209	0.271	4.600
13	347546.1730	0,881	4.600
17	347547.4057	4.250	4.600

Для дополнительной оценки позы были подвергнуты метадинамической симуляции с помощью алгоритма Binding Pose Metadynamics (BPMD), используя RMSD лиганда в качестве коллективной переменной. Проводилось десять запусков по 10 нс каждый. Оригинальный алгоритм BPMD предоставляет две оценки, которые связаны со стабильностью лиганда в ходе метадинамического моделирования: PoseScore, определяющий среднее среднеквадратичное отклонение от начальной позы, и PersistenceScore (PersScore), являющийся мерой устойчивости водородных связей, рассчитываемый как доля кадров за последние 2 нс моделирования, которые имеют те же водородные связи, что и входная структура, усредненный по всем 10 повторным моделированиям. Низкий показатель PersScore наблюдается в структурах, сеть контактов в которых ослаблена предвзятостью BPMD. PersScore варьируется в диапазоне от 0 до 1, и если 60% общего числа водородных связей сохраняются во время моделирования ($\text{PersScore} \geq 0.6$) – лиганд образует устойчивый комплекс с белком. Дополнительным признаком стабильности лиганда в сайте мишени считается $\text{PoseScore} < 2.0 \text{ \AA}$ [74] (рисунок 28).

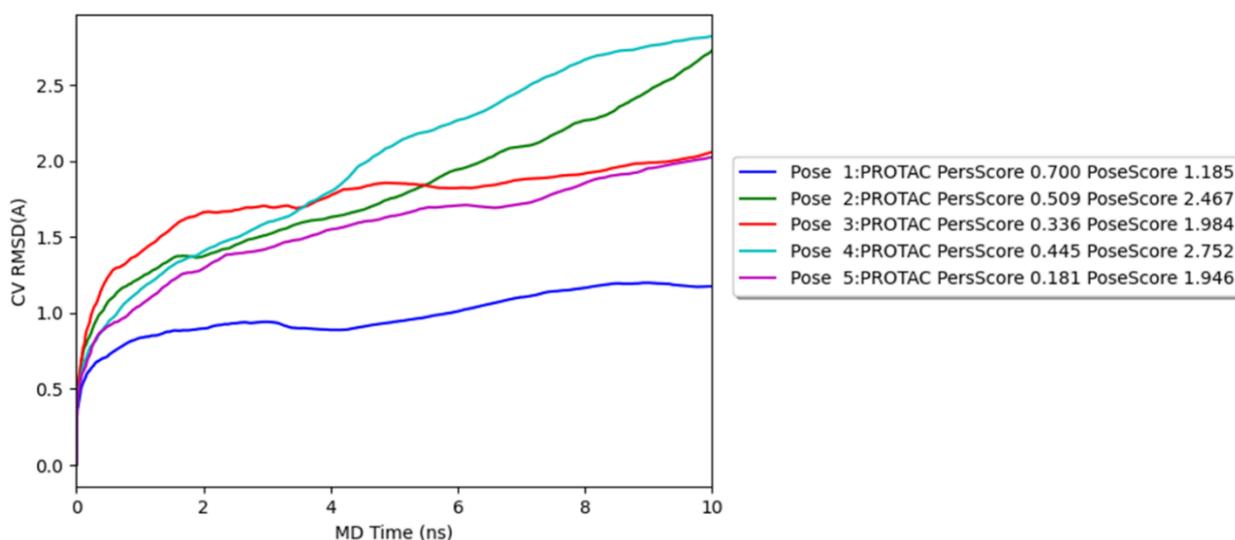


Рисунок 28 – График зависимости CV RMSD от времени симуляции для пяти лучших поз

Можно увидеть, однако, что только одна из поз, находящихся в топ-5, показала удовлетворительные значения PersScore и PoseScore – поза №2.

Мегадинамической симуляции также подвергался нативный комплекс бВОУ. Можно заметить на рисунке ниже, что значения PoseScore=1.290 и PersScore=0.673 сопоставимы со значениями, полученными для бензотриазольного аналога dВЕТ6. Из этого можно сделать вывод, что полученный нами комплекс не уступает кристаллическому по устойчивости водородных связей и по среднеквадратичному отклонению.

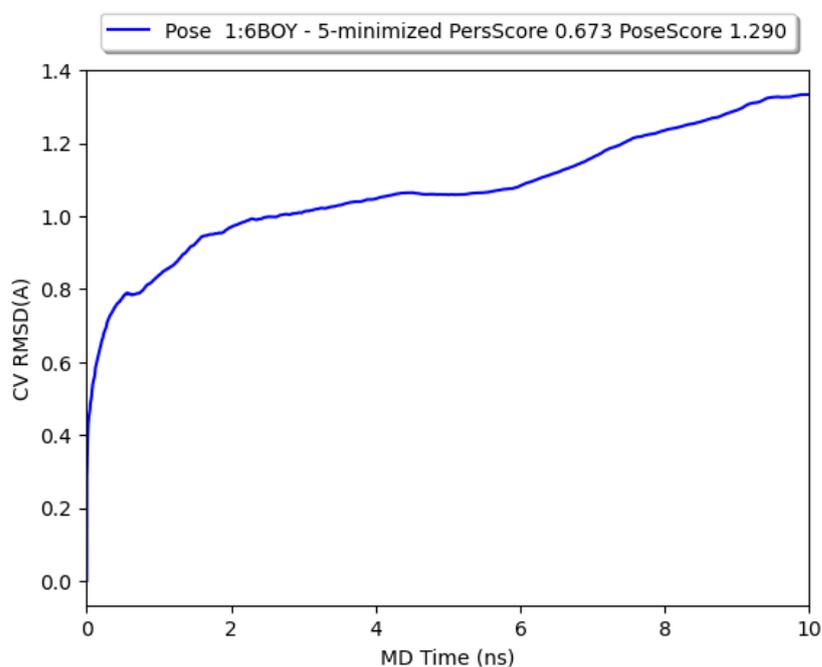


Рисунок 29 – График зависимости CV RMSD от времени симуляции для пяти лучших поз

В качестве параметров для финального отбора поз алгоритм IFD-MD использует такие параметры как PoseStability и HBPersist – то же самое, что и PoseScore/PersScore. Их линейная комбинация – $PoseStability_{func}$ – имеет следующий вид (формула 5):

$$PoseStability_{func} = \omega_6 PoseStability + \omega_7 (PoseStability)^2 + HBPersist_{func}, \quad (5)$$

где $PoseStability_{func}$ – оценка поз после проведения метадинамики. Качественно вычисляется из RMSD лиганда в конце моделирования метадинамики по сравнению с его начальным положением. Также определяется по устойчивости водородных связей в ходе моделирования метадинамики;

ω_6, ω_7 – весовые коэффициенты.

$HBPersist_{func}$ – означает долю устойчивости (в диапазоне от нуля до единицы) всех водородных связей лиганд-рецептор за время моделирования метадинамики. P_1 и P_2 определяют область, в которой функция изменяется линейно, и рассматривались как настраиваемые параметры (формула 6):

$$HBPersist_{func} = 5 \left\{ \frac{hb_{persist}}{P_1 - P_2} + \frac{P_2}{P_2 - P_1} \right\}, \quad (6)$$

где $hb_{persist}$ – доля устойчивости водородных связей между лигандом и рецептором в течение МД;

P_1, P_2 – области, в которых функция изменялась линейно.

Финальная композитная оценка – IFD-MDScore – рассчитываемая в ходе IFD-MD имеет следующий вид (формула 7):

$$\begin{aligned}
 IFD - MD_{Score} = & \omega_8 \Delta Prime_{func} + \omega_9 \Delta Glide_{func} + \\
 & WScore_{func} + PoseStability_{func} + SCMotion_{func} + \\
 & TemplateOverlap_{func},
 \end{aligned}
 \tag{7}$$

где $Prime_{func}$ – значение потенциальной энергии оптимизированного комплекса;

$Glide_{func}$ – композитная оценка $Glide$, включающая в себя изменение $Glide_{docking}$ и $Glide_{Ecoul}$;

$WScore_{func}$ – оценочная функция, которая позволяет оценить способ связывания лиганда в сайте, с учетом молекул воды;

$PoseStability_{func}$ – оценка поз после проведения метадинамики. Качественно вычисляется из RMSD лиганда в конце моделирования метадинамики по сравнению с его начальным положением. Также определяется по устойчивости водородных связей в ходе моделирования метадинамики.

$SCMotion_{func}$ – штраф за подвижность боковой цепи;

$TemplateOverlap_{func}$ – штраф за объемное перекрытие между позами исследуемого и нативного лиганда. Нормирован, может принимать значения от нуля до единицы.

Δ – изменение значения того или иного параметра в каждой из 20 поз, получаемых на стадии PGP по отношению к наибольшему значению;

ω_8, ω_9 – весовые коэффициенты.

Как можно увидеть, на ряду с оценками $Glide$, $Prime$, $WScore$ и параметрами $BPMD$, также учитывается, например, перекрывающийся объем референсного и данного лигандов, и степень подвижности остатков с небольшими значениями $SASA$.

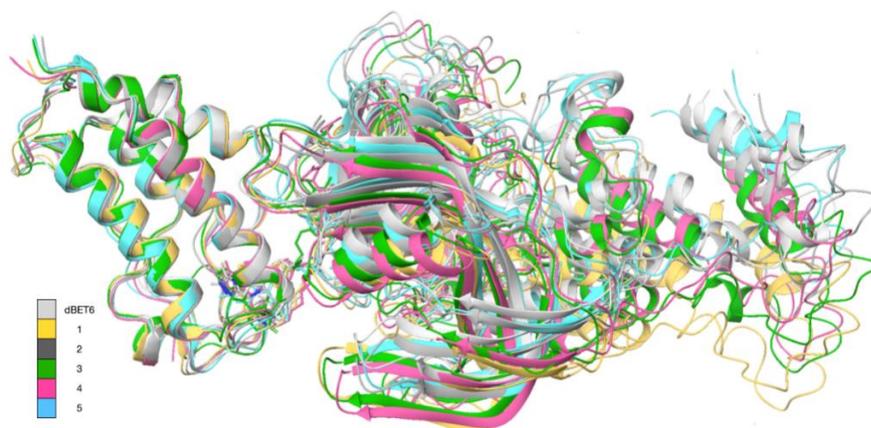
В таблице ниже представлены финальные значения оценок для пяти лучших поз (таблица 9).

Таблица 9 – Финальные значения оценок для пяти лучших поз IFD-MD

Поза	WScore _{Func} , ккал/моль	IFD- MDScore, ккал/моль	HBPres	PoseScore, Å	SCMotion _{Func}	TemplateOverlap _{Func}
1	-13.920	13.951	0.445	2.752	0.000	6.343
2	-19.822	-2.450	0.700	1.185	6.000	4.659
10	-10.852	15.968	0.336	1.984	0.000	7.656
11	-14.328	20.764	0.509	2.467	0.000	6.629
19	-12.352	19.876	0.181	1.946	3.600	5.604

Ранжируемая алгоритмом на первое место поза №2, как можно увидеть исходя из данных, представленных в таблице выше, действительно лидирует по ряду параметров, обладает лучшим значением WScore, наименьшим штрафом за перекрытие объемов с dBET6, а также показывает стабильность лиганда по результатам ВРМД, чего нельзя сказать об остальных.

Наложение структур тройных комплексов для поз №1-5 с кристаллическим бВОУ (выравнивание по BRD4) представлено ниже (рисунок 29).



Наложение позы №1 (жёлтый), №2 (тёмно-серый), №3 (зелёный), №4 (пурпурный), №5 (голубой) и нативного dBET6 (светло-серый)

Рисунок 29 – Наложение структур тройных комплексов для поз №1-5

На первый взгляд, кажется, что алгоритму неплохо удалось воспроизвести геометрию комплекса dBET6, однако более пристальный анализ наложения лигандов не оставляет шанса позам №2-4 (рисунок 30).

Кроме того, не стоит удивляться сходству белок-белковых комплексов – вероятно, серии молекулярно-динамических симуляций длительностью 500 пс просто недостаточно для того чтобы сильно ремоделировать геометрию макромолекул.

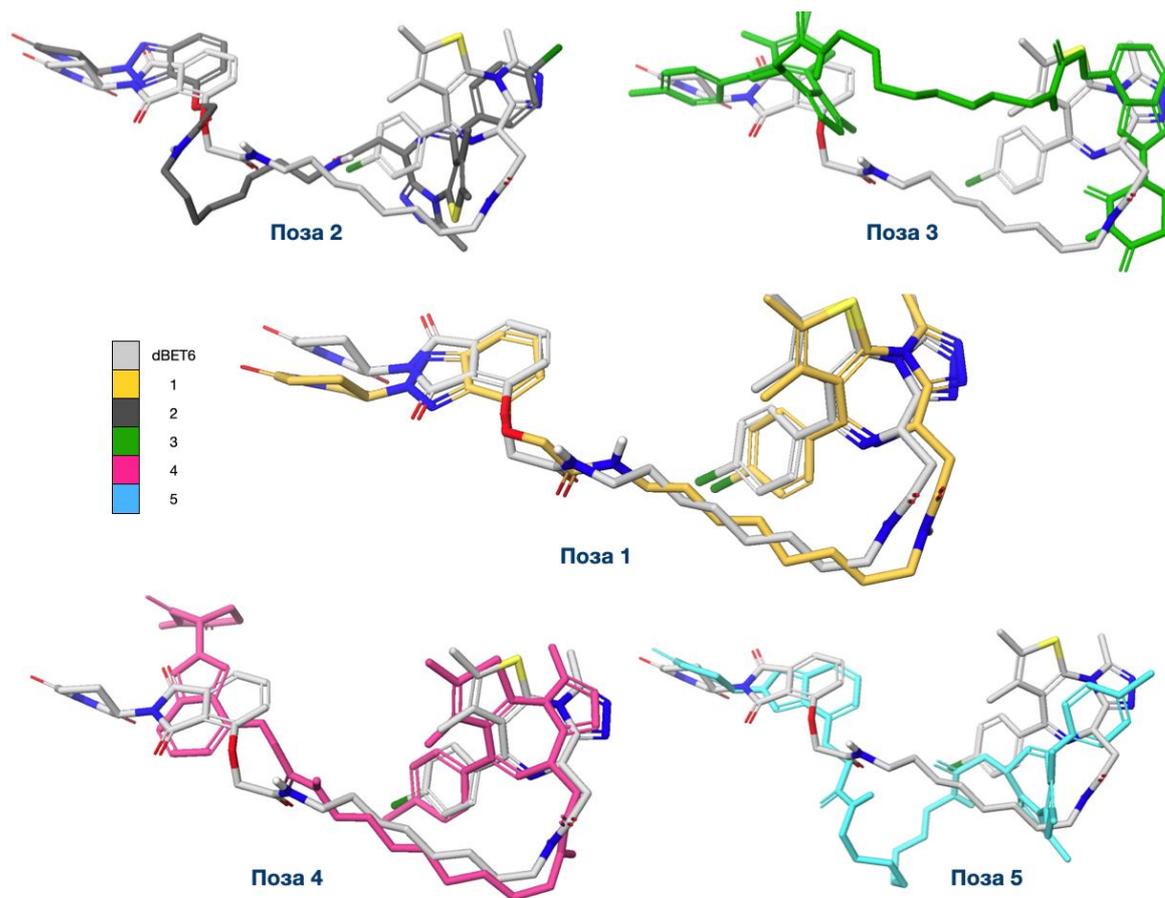


Рисунок 30 – Наложение позы №1 (жёлтый), позы №2 (тёмно-серый), позы №3 (зелёный), позы №4 (пурпурный), позы №5 (голубой) на кристаллический dBET6 (светло-серый)

Только в первой позе VT-PROTAC воспроизводит как геометрию боегловок, так и линкера между ними. Поза №5 демонстрирует тенденцию к циклизации соединительного мотива, а также инверсию лиганда BRD4, равно как и №2. Структура PROTAC в третьей позе оказалась пристыкована наоборот, причем такой способ связывания наблюдался еще на этапе PGR.

Поза №4 имеет близкую к кристаллической геометрию JQ-1 и линкера, однако полную потерю сходства в области CRBN-связывающего мотива.

Дополнительно нами были проанализированы интерфейсы белок-белковых взаимодействий в позах №1 и 5 по сравнению с комплексом dBET6 с помощью алгоритма Protein Interaction Analysis.

Таблица 10 Количественное описание интерфейса белок-белковых взаимодействий

	Взаимодействующие остатки		Тип	Средняя комплементарность поверхности	Средняя buried SASA, CRBN (%)	Средняя buried SASA, BRD4 (%)
	BRD4	CRBN				
6BOY	Asp145	Hip103	hb*	0.61	74.30	75.26
	Trp81	His353	clash**			
	Gln78	Lys156	hb			
	Phe79	Phe150	π -stack***			
	Gln79	Pro352	hb			
pose 1	Asp145	Hip103	hb	0.62	83.40	91.85
	Phe79	Hip353	π -stack			
	Trp81	Hip353	π -stack			
	Hie77	Phe150	π -stack			
	Gln78	Pro352	hb			
	Asp145	Tyr355	hb			
pose 5	Asp145	Hip103	hb	0.62	83.40	91.85
	Phe79	Hip353	π -stack			
	Trp81	Hip353	π -stack			
	Hie77	Phe150	π -stack			
	Gln78	Pro352	hb			
	Asp145	Tyr355	hb			

* – hb (англ, hydrogen bond) водородная связь;

** – clash стерическое столкновение;

*** – π -stack взаимодействие ароматических систем.

По представленным в таблице 10 данным видно, что ключевая водородная связь, образованная между BRD4 и CRBN (Hip103:Asp145), сохраняется как в нативном комплексе, так и в позах №1 и 5. Также стоит отметить, что в полученных позах реализуются новые взаимодействия, а именно водородные связи Pro352:Gln78 и Tyr355:Asp145 и π -стекинг, образование которых повлекло за собой изменение геометрии всего

комплекса. Большинство взаимодействующих остатков отличаются высокой степенью комплементарности. Среднее значение параметра, описывающего долю площади, доступной для растворителя, которая оказывается скрыта из-за взаимодействий (*англ.* buried SASA) в полученных позах одинаковый и незначительно отличается от значений нативного комплекса. Что подтверждает визуально наблюдаемую схожесть геометрий белков в области интерфейса PPI в позах 1,5 и 6BOY.

2.4 Молекулярная динамика комплекса, полученного с помощью алгоритма IFD-MD

После оценки полученных результатов мы решили провести дополнительное молекулярно-динамическое исследование, чтобы проверить подвергнется ли геометрия комплекса ремоделированию в процессе более длительных симуляций, а также для проведения сравнительного анализа пространственных изменений нативного тройного комплекса 6BOY и полученного с помощью IFD-MD комплекса для VT-PROTAC.

С помощью Desmond нами был осуществлен запуск непредвзятого моделирования как для комплекса, соответствующего лучшей позе IFD-MD, так и для кристаллического ТК. Проводилось три симуляции по 100 нс, подготовка структур и сольватных боксов осуществлялась по стандартному протоколу, описанному в руководстве пользователя Desmond [75].

На рисунке ниже представлены графики зависимости RMSD координат C α атомов белков и атомов лигандов от времени симуляции. Одним из критериев стабильности комплексов во время молекулярной динамики и сходимости симуляции является выход колебаний RMSD на плато в последние 10-20 нс. В нашем случае такая ситуация наблюдается только для VT-PROTAC начиная с ~80 нс комплекс колеблется вокруг некоторой средней тепловой структуры, в то время как ТК, связанный с dBET6, не уравнивается.

Также отметим, что у обоих комплексов колебания RMSD $C\alpha$ в среднем не превышают 4.5-5 Å, что свидетельствует о наличии только небольших конформационных перестроек белков (рисунок 31).

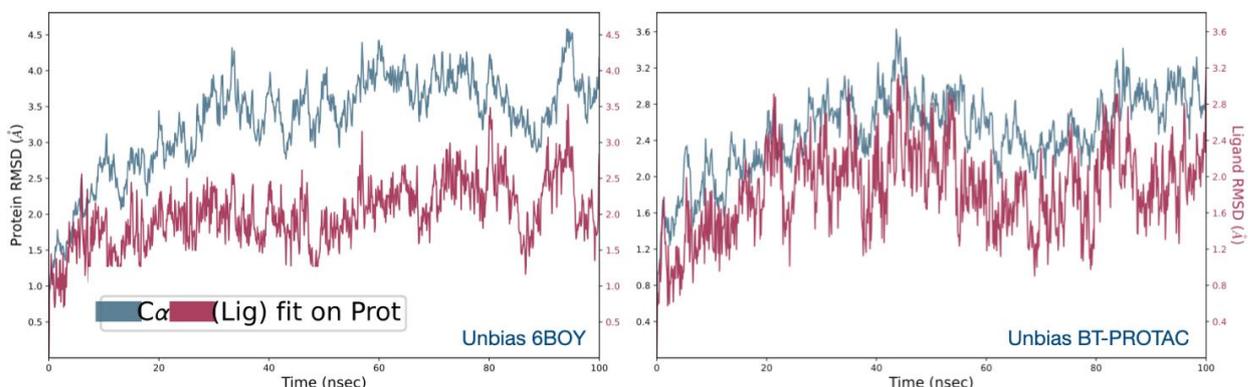


Рисунок 31 – Графики зависимости RMSD атомов белка и лиганда от времени симуляции для комплексов dBET6 и BT-dBET6

На рисунках ниже представлены графики значений среднеквадратичной флуктуации атомов лигандов для dBET6 и его биоизостерного аналога. Как можно увидеть, в случае BT-dBET6, наибольшую подвижность проявляют атомы 5-8, 41-43, 46, 55, соответствующие атомам линкера, что весьма логично и ожидаемо (рисунок 32).

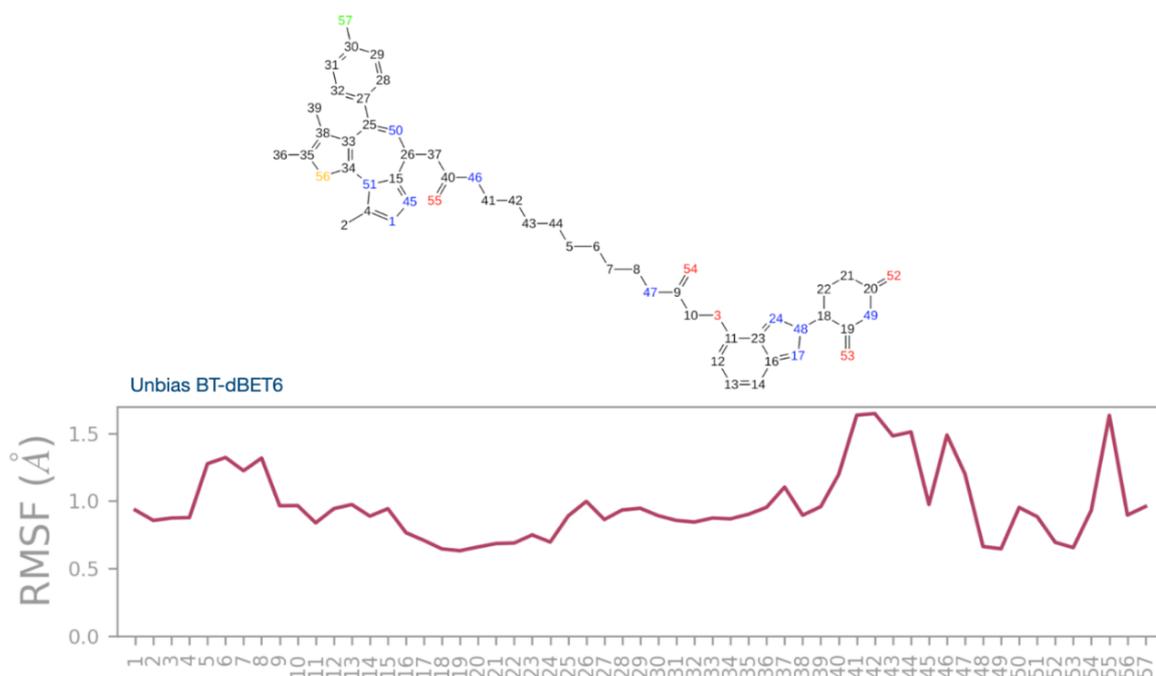


Рисунок 32 – График изменения значений RMSF координат атомов BT-dBET6

В случае самого dBET6 можно наблюдать ту же картину. Также отметим, что в обоих случаях максимальные значения RMSF не превышают 1.7 Å (рисунок 33).

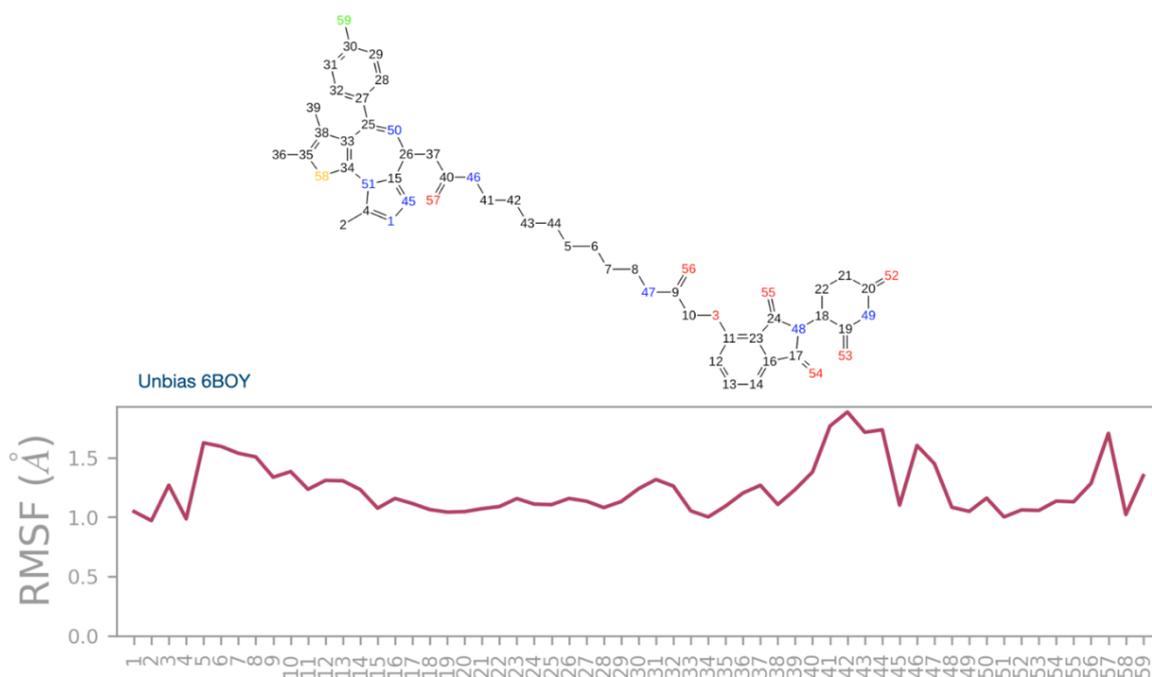


Рисунок 33 – График изменения значений RMSF координат атомов dBET6

Анализ конформационного профиля лигандов, подтверждает выводы о подвижности линкерных мотивов. В целом можно отметить, что все демонстрируемые значения двугранных углов в исследуемых молекулах соответствуют только минимумам на кривых потенциальной энергии. Различия наблюдаются для вращения по осям 1 (красный), 2 (светло-розовый), 4 (темно-зеленый), 5 (светло-голубой, сет №2) соответствующие связям линкера со стороны, ближней к лиганду CRBN, а также 3 (светло-голубой, сет №1) соответствующей связи с *para*-хлорфенильным фрагментом в лиганде BRD (рисунок 34, 35).

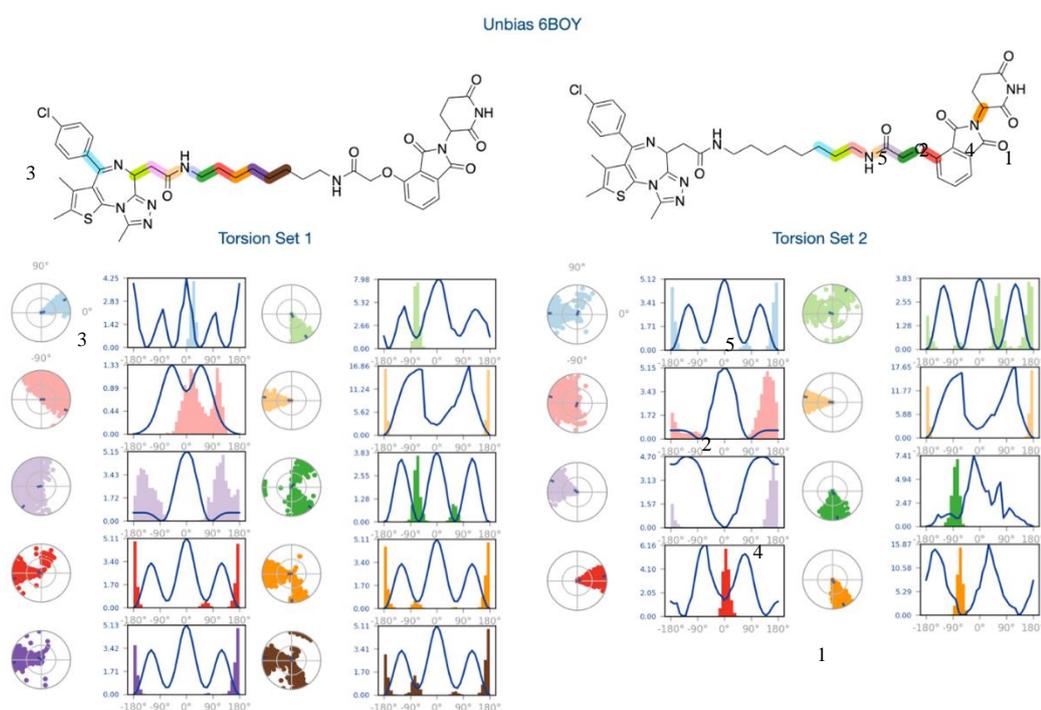


Рисунок 34 – Конформационный профиль лигандов dVET6

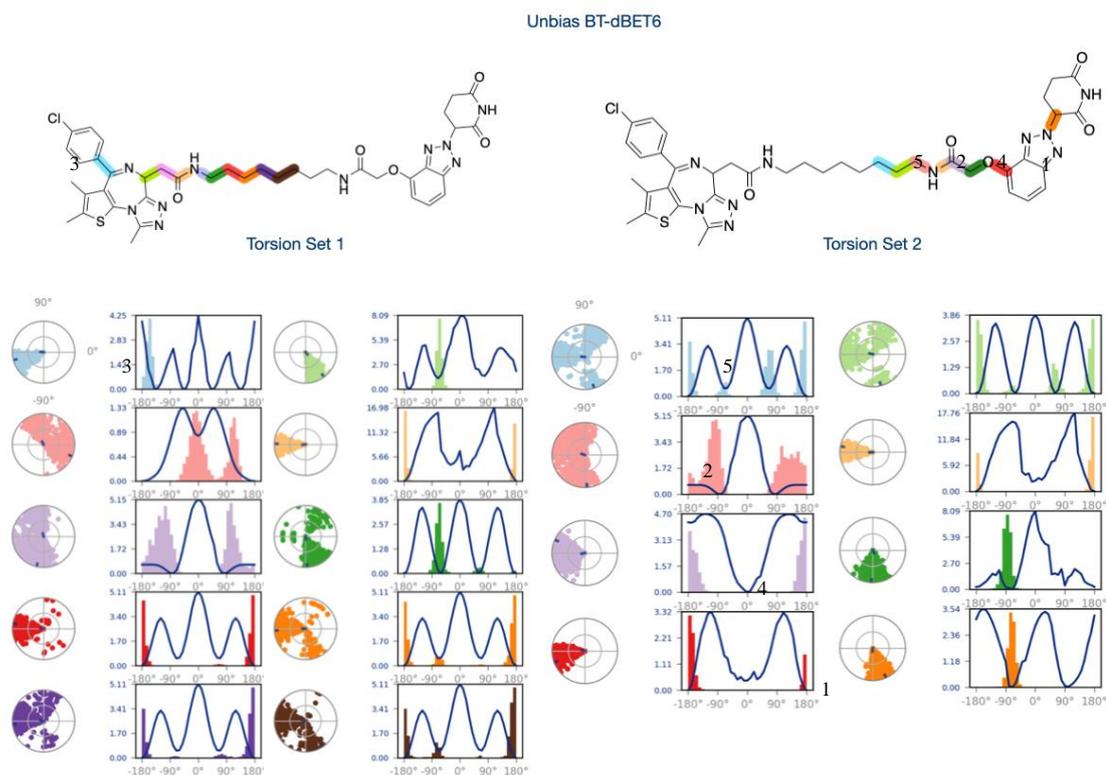


Рисунок 35 – Конформационный профиль лигандов BT-dBET6

Анализируя гистограммы белково-лигандных взаимодействий можно сказать, что характер и степень сохранности контактов не изменилась, однако гидрофобные взаимодействия с остатком Asp351 и водородные связи, транслируемые через молекулы воды, с His357 существует практически на 30% меньше времени для BT-dBET6 по сравнению с референсным PROTAC. Напомним, что эти остатки располагаются в сайте связывания CRBN. Параметр Interaction Fraction на графиках ниже показывает процент времени моделирования, на протяжении которого существует контакт (рисунок 36).

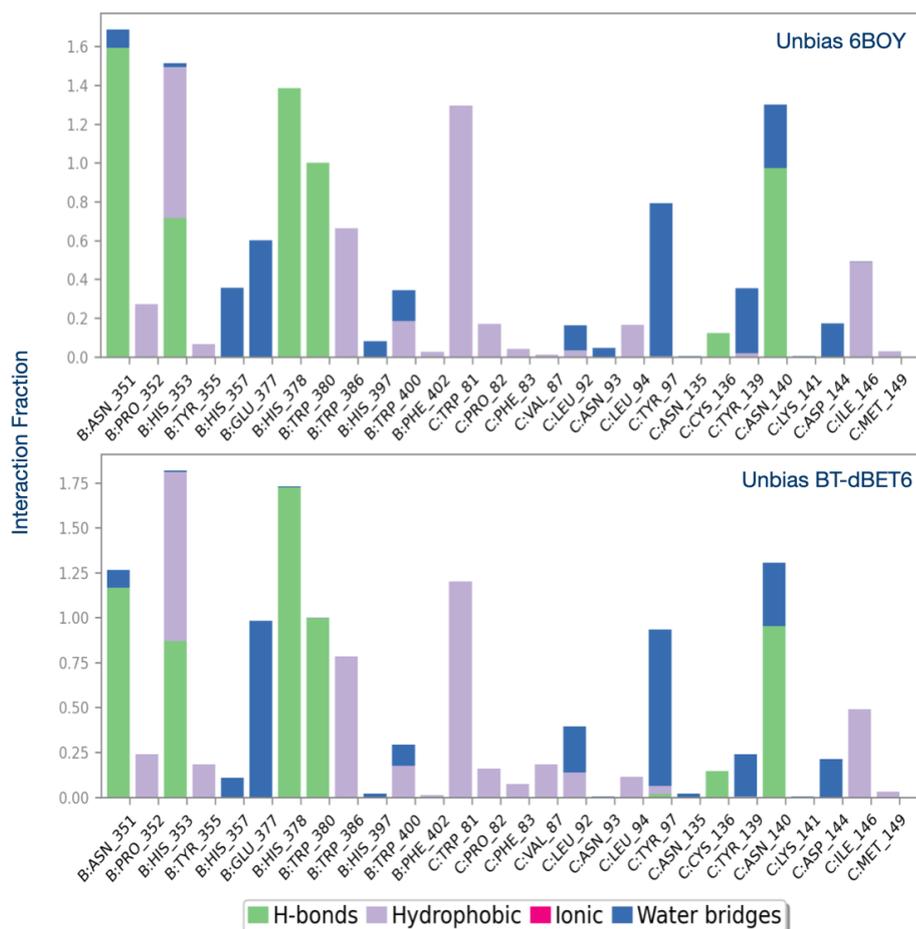


Рисунок 36 – Белково-лигандные контакты

Более пристальный анализ диаграммы взаимодействия деградеров с белками позволяет сделать главный вывод в ходе симуляции способ связывания молекул-химер схож и не изменяется существенным образом. Если говорить более подробно, то:

- взаимодействия, образуемые ацепторно-донорно-акцепторной триадой глутаримидного кольца с остатками Trp и His сохраняются, и демонстрируют даже большую сохранность в случае BT-dBET6;

- биоизостерная замена карбонильной группы во фталимиде не приводит к потере связи с Glu, образуемой через молекулу воды, а напротив, приводит к стабилизации контакта, чего нельзя сказать про взаимодействие с Asn 351;

- ключевые взаимодействия лиганда BRD4 сохраняются (рисунок 37).

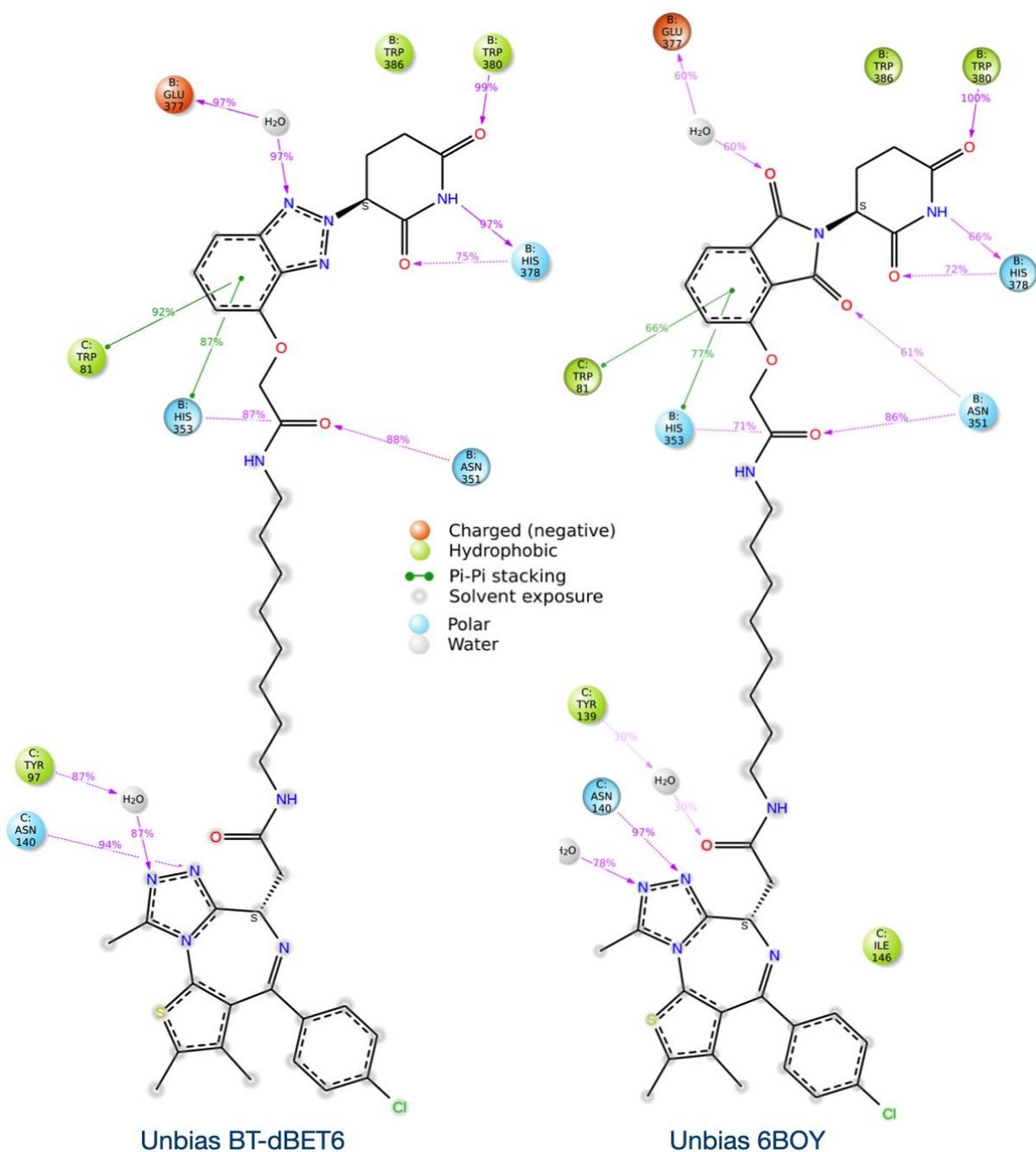


Рисунок 37 – Диаграммы взаимодействий деградеров с белками

Дополнительно была проведена стандартная симуляция с помощью все того же Desmond длительностью в 100 нс. Отметим, что существенной разницы в динамике белков и лигандов обнаружено не было.

В заключение к двум подразделам, посвященным запуску IFD-MD и молекулярной динамике полученных результатов, еще раз резюмируем, что полученный ТК бензотриазольного аналога dBET6, соответствующий лучшей позе IFD-MD можно назвать стабильной системой. Все ключевые взаимодействия сохраняются в полном объеме. Конформационный профиль биоизостерного аналога dBET6 практически полностью воспроизводит таковой для референсного PROTAC. На практическом опыте мы убедились, что исследуемый алгоритм IFD способен генерировать комплекс, который является близким к кристаллическому. Из чего можно сделать предположение о том, что результаты биологических испытаний, вероятно, будут приемлемыми, а именно, BT-dBET6 будет обладать схожей деградирующей способностью.

2.5 Метод возмущения свободной энергии

Метод FEP (Free Energy Perturbation, возмущение свободной энергии) является одним из основных методов молекулярной динамики для вычисления энергетических характеристик таких процессов, как, например, связывание лигандов с белками, конформационные изменения молекул и химические реакции. Основная идея метода FEP заключается в вычислении разницы свободной энергии между двумя состояниями системы путем введения промежуточных состояний и оценки изменения свободной энергии при переходе от одного состояния к другому.

Расчет свободной энергии связывания лигандов с макромолекулами с помощью метода возмущения свободной энергии (FEP) может производиться двумя разными путями с использованием протоколов Absolute Binding FEP (абсолютное связывание) и Relative Binding FEP (относительное связывание).

Absolute Binding FEP (AB-FEP) используется для оценки свободной энергии связывания одного лиганда с белком. Он требует учета всех межмолекулярных взаимодействий и комплексного моделирования системы в

растворе, что делает его вычислительно затратным, но обеспечивает высокую точность. В методе Absolute Binding FEP начальным состоянием является система без связанного лиганда, а конечным состоянием система со связанным лигандом.

Relative Binding FEP применяется для оценки разницы в свободной энергии связывания между двумя или более лигандами. Он особенно полезен для сравнения относительной аффинности и эффективности двух подобных соединений, что позволяет эффективно анализировать вариации в их связывании с целевым белком. В RB-FEP используются так называемые "алхимические превращения" для оценки разницы в свободной энергии связывания между двумя лигандами один лиганд постепенно превращается в другой в процессе симуляции.

В данном исследовании использовались оба подхода FEP с помощью алгоритма FEP+, предоставляемого Schrodinger.

Значения ΔG_{ABFEP} , рассчитанные в режиме абсолютного связывания, для dBET6 и VT-dBET6 составили -27.3 ± 0.17 и -27.4 ± 0.15 ккал/моль соответственно, из чего можно сделать вывод, что биоизостерный аналог связывается с белками чуть более эффективно, чем нативный PROTAC.

Моделирование проводилось на основе кристаллического комплекса бВОУ для dBET6 и на основе ТК полученного с помощью IFD-MD для VT-dBET6. Время симуляций составляло 20 нс.

Результаты запуска RB FEP представлены на рисунке ниже (рисунок 38).

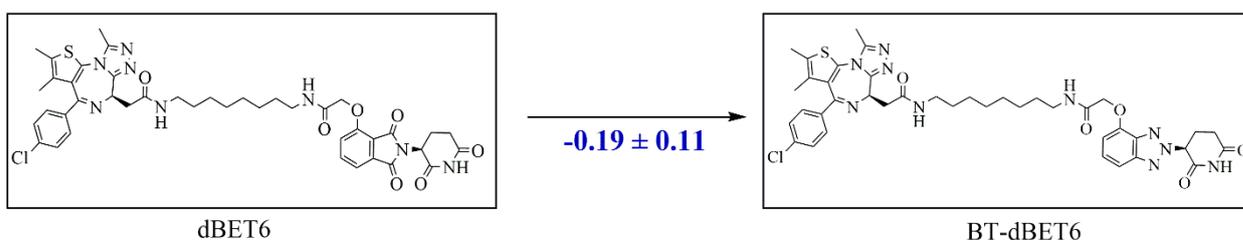


Рисунок 38 – Данные расчета относительной энергии связывания

Как можно увидеть, разница в изменении свободной энергии связывания составила 0.19 ± 0.11 . Полученное значение говорит о том, что при переходе от нативного dBET6 к BT-dBET6 энергия связывания убывает, а значит биоизостерный аналог имеет большую аффинность связывания.

2.6 Генерация тройных комплексов с помощью Protein Degradation Sampler

Для прогнозирования структур тройных комплексов в данной работе так же был применен альтернативный вычислительный алгоритм, который будет описан в этом разделе.

Относительно недавно, в 2022 году, коллегами из Schrodinger был предложен и описан алгоритм, который позволяет генерировать и оценивать структуры тройных комплексов POI-PROTAC-E3. Входными данными для алгоритма являются оптимизированная 3D-структура PROTAC, структуры E3-лигазы и белка интереса, включающие в себя соответствующие лиганды в своих сайтах связывания. Последние обычно получают путем проведения классического докинга или же из соответствующих кристаллических структур.

На первом этапе алгоритм проводит исследование фазового пространства деградера. Пользователю доступны следующие способы осуществления данного конформационного поиска:

- MCMM – моделирование классическим стохастическим методом Монте-Карло (англ. Monte-Carlo Molecular Modeling) [76];

- LMCS - конформационный поиск методом низкочастотных нормальных колебаний (англ. Low Mode Conformational Search) [77];

- ConfGen – оригинальный алгоритм Schrodinger, который подразумевает собой разрыв молекулы по всем свободновращающимся связям, исследование конформационного пространства полученных фрагментов и, на завершающем этапе, их соединение в единую молекулу [78];

– RDKit – подход, в котором для поиска нужной структуры реализуется создание матрицы, состоящей из расстояний между атомами, которые впоследствии изменяются стохастическим образом [79];

– Desmond – метод исследования конформационного пространства с помощью проведения симуляции молекулярной динамики.

После завершения первого этапа и получения набора конформаций выполняется наложение структур POI и E3 в соответствии с их лигандами. На данном шаге также происходит фильтрация поз, которые приводят к стерическим столкновениям. Пользователю разрешено самостоятельно выбирать дистанцию и тип атомов, между которыми будут рассчитываться столкновения (атомы главных цепей белков, только тяжелые атомы, тяжелые атомы главных цепей или вообще все атомы).

На последнем этапе обычно происходит оценка и уточнение структур полученных тройных комплексов, опосредованные алгоритмом белок-белкового докинга PIPER, но в данном исследовании было принято решение исключить данный этап, т.к скоринговые функции PIPER не могут адекватно оценивать взаимодействия таких больших комплексов.

Детальная схема полного рабочего процесса Protein Degradation Sampler представлена на рисунке ниже (рисунок 39).

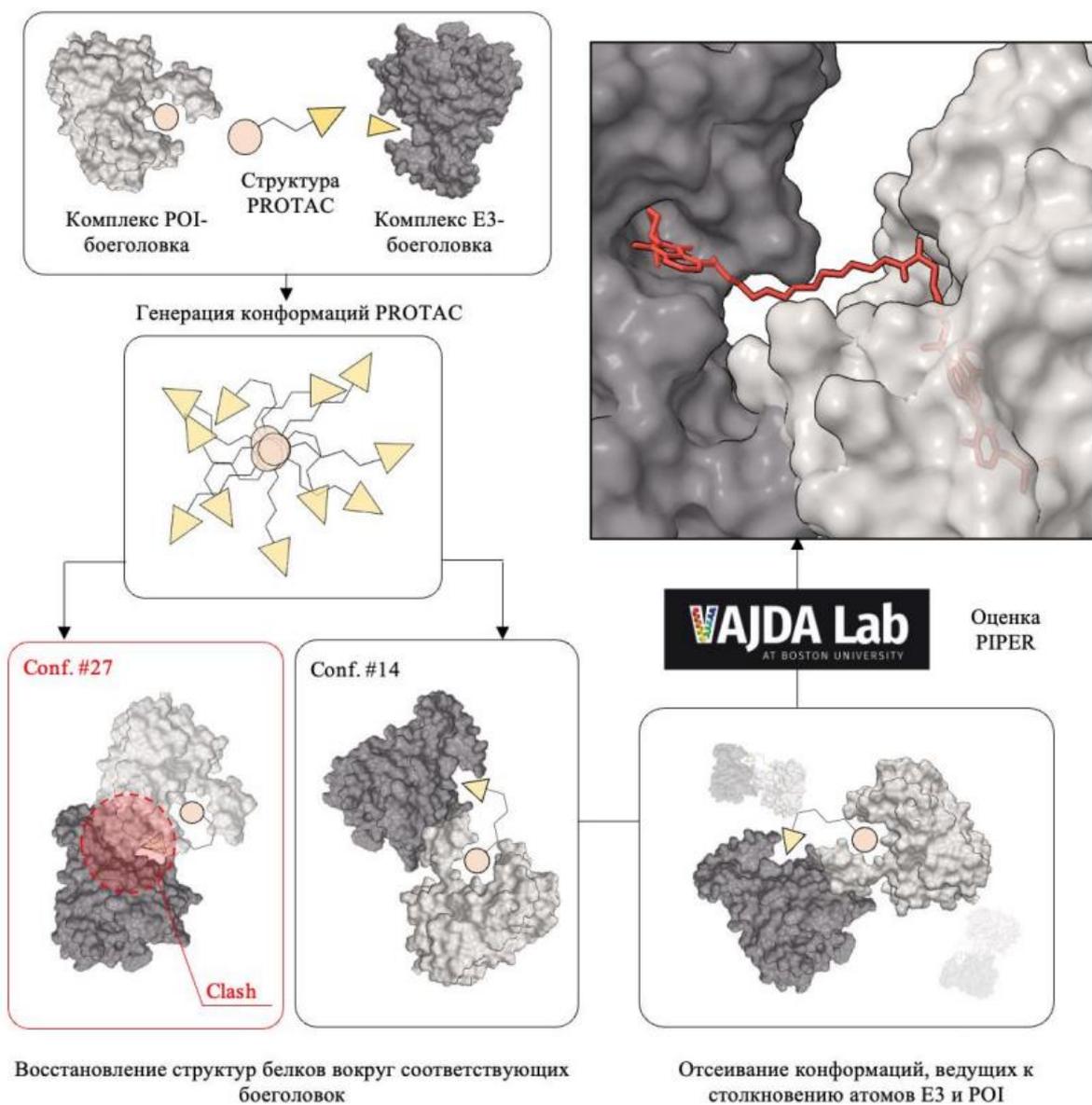


Рисунок 39 – Полный рабочий процесс алгоритма Protein Degradation Sampler

Задача состояла в том, чтобы выявить те параметры запуска, которые давали бы максимально обогащенные кристаллоподобными ТК результаты минимальном общем числе генерируемых комплексов. После выполнения расчетов мы выравнивали конформеры по BRD4 с помощью модуля Protein Structure Alignment [80] и рассчитывали RMSD лиганда, сравнивая его с входными референсными структурами, а именно с кристаллическим бВОУ для dBET6 и с полученным ранее с помощью IFD-MD комплексом для VT-PROTAC.

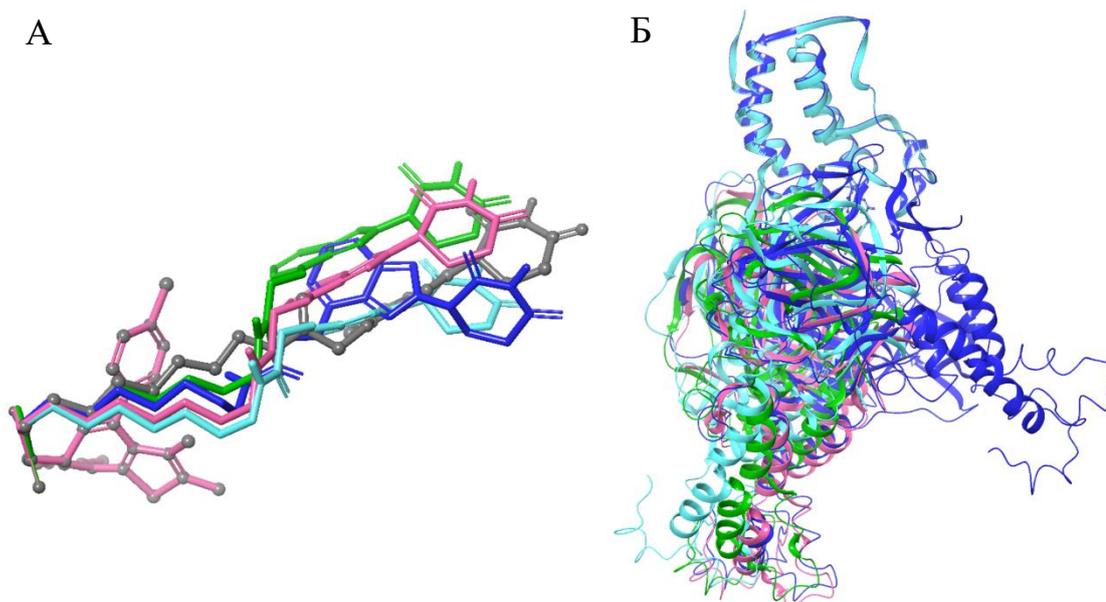
Результаты расчетов с различными параметрами представлены в таблице ниже (таблица 11).

Таблица 11 Количественное описание результатов работы алгоритма

PROTAC	Дистанция, Å	Типы атомов	Число комплексов	Число комплексов с RMSD лиганда		
				<3Å	3Å<X<4Å	4Å<X<5Å
dBET6	0.7	Ca	1298	7	9	6
		тяжелые атомы Ca	1517	8	7	5
		все тяжелые	362	0	0	0
		все	176	0	0	0
	1.4	Ca	500	2	1	0
		тяжелые атомы Ca	564	3	0	1
		все тяжелые	0	0	0	0
		все	0	0	0	0
BT-dBET6t	0.7	Ca	1368	12	13	14
		тяжелые атомы Ca	1668	19	14	18
		все тяжелые	218	2	1	3
		все	113	1	0	1
	1.4	Ca	292	2	3	6
		тяжелые атомы Ca	352	2	4	9
		все тяжелые	0	0	0	0
		все	0	0	0	0

Как видно из таблицы при очень жестких ограничениях не было сгенерировано ни одного комплекса, что вполне ожидаемо.

В любом случае можно заметить, что алгоритм способен генерировать некоторый набор комплексов, обогащённый конформерами, близкими к кристаллическому (рисунок 40).



А – наложение структур PROTACs; Б – наложение структур тройных комплексов

Рисунок 40 – Наложение конформеров, найденных алгоритмом и полученного комплекса для ВТ-dBET6 (clash fr 0.7, mddist 10, clash atoms C α)

Еще раз подчеркнем, что алгоритм способен предоставлять некоторый набор комплексов, геометрия которых близка к кристаллической.

Дальнейшей нашей задачей стала проверка возможности нахождения среди всех результатов наиболее кристаллоподобных, опираясь на энергетические характеристики полученных ТК. Основываясь на том факте, что в большинстве литературных кейсов исследователи используют в качестве средства для ранжирования значения полных потенциальных энергий и свободных энергий связывания тройных комплексов, мы также решили рассчитать эти параметры с помощью энергетической модели Prime, предположив, что наиболее кристаллоподобный ТК будет обладать либо наилучшими энергетическими характеристиками, либо схожими с кристаллическим или IFD-MD референсом. Данная процедура позволила нам получить два значения – Prime Energy полная потенциальная энергия лиганд-белкового комплекса, и ΔG_{bind} MM GBSA свободная энергия связывания, значения которых предоставлены в таблице ниже (таблица 12).

Таблица 12 Энергетическая оценка полученных результатов (clash fr 0.7, mddist 10, clash atoms C α)

Conformer	Prime energy	ΔG_{bind} MM GBSA	RMSD C α	RMSDlig	Ранг	
					Prime	ΔG_{bind}
dBET6	-20327.80	-118.69	0	0	-	-
1426	-19942.64	-114.39	2.584	5.79	1	18
7976	-19940.65	-109.40	6.314	20.83	2	26
9912	-19937.52	-112.58	15.200	41.55	3	20
5453	-19934.61	-97.59	4.369	31.21	4	64
7610	-19933.81	-93.54	20.110	57.49	5	79
4551	-10544.78	9286.93	17.198	43.18	1294	1293
4653	-10471.45	9371.63	5.024	7.51	1295	1295
4905	-9802.15	10069.94	15.138	39.51	1296	1296
6243	-8415.89	11471.32	15.224	35.61	1297	1297
8224	-6661.71	13149.79	19.989	57.80	1298	1298
BT-dBET6	-19142.32	1.16	0	0	-	-
5850	-19165.48	16.73	17.24	57.38	1	5
1753	-19162.73	20.73	2.16	5.78	2	14
7046	-19159.39	33.29	19.50	59.05	3	42
9088	-19156.47	18.49	19.01	58.35	4	8
2207	-19154.49	20.29	15.10	27.81	5	11
6483	≥ 100000	42.77	6.65	19.34	1364	99
7522	≥ 100000	71.56	13.01	44.04	1365	522
4624	≥ 100000	42.38	14.45	37.26	1366	94
3246	≥ 100000	53.47	15.65	38.93	1367	222
6723	≥ 100000	≥ 100000	18.00	≥ 100	1368	1368

Анализируя полученные результаты, можно сделать вывод о том, что значения ΔG_{bind} MM GBSA не позволяют ранжировать кристаллоподобные комплексы в верхние строчки рейтинга, чего нельзя сказать про скоринг по Prime energy. Например, в случае dBET6, поза, обладающая минимальным RMSD, ранжируется на первое место.

Таким образом, мы убедились, что алгоритм Protein Degradер Sampler способен генерировать решения, близкие к кристаллическим. В первом приближении, ранжирование на основе Prime energy позволяет однозначно идентифицировать как минимум один кристаллоподобный результат.

3 Экспериментальная часть

3.1 Компьютерное моделирование

Все расчеты в рамках выполненной выпускной квалификационной работы проводили с применением программного пакета Schrodinger Suite 2022-4 на рабочей станции на рабочей станции, снабженной 64 процессорами AMD Ерус 7513 (тактовая частота 2.6 ГГц) и двумя графическими картами NVIDIA GA102GL (RTX A5000), под управлением ОС Ubuntu 22.04.

3.1.1. Подготовка белков

Кристаллическая структура CRBN-dBET6-BRD(BD1) (PDB ID: 6BOY) была загружена из Protein Data Bank [57].

Подготовка белков проводилась с помощью модуля Protein Preparation Workflow [58]. Были добавлены отсутствующие атомы в боковых цепях, уточнены некоторые порядки связей, предсказаны состояния протонирования для гетероатомов с помощью алгоритма Epik [81] $pH = 7.4 \pm 0.2$, а также были добавлены атомы водорода. Для снижения вычислительных затрат цепь, содержащая DDB1 в тройном комплексе 6BOY, была удалена. Также комплекс был подвергнут этапу минимизации.

3.1.2 Подготовка лигандов

Структура VT-PROTAC была создана с помощью модуля 2D-Sketcher, подготовлена с помощью модуля LigPrep [59], хиральность определялась на основании 3D-представления, состояния протонирования гетероатомов определялись с помощью алгоритма Epik [81] $pH = 7.4 \pm 0.2$.

3.1.3 Индуцированный докинг

Рабочий процесс Induced Fit Docking-Molecular Dynamics [56] запускался с использованием графического интерфейса модуля IFD-MD, представленного в Maestro.

В качестве шаблона был выбран подготовленный комплекс dBET6 бBOY, а бензотриазольный аналог в качестве лиганда для стыковки. Сам расчет производился в режиме нековалентной (noncovalent) стыковки, дополнительно было произведено уточнение силового поля OPLS [83]. Значения остальных параметров запуска были оставлены по умолчанию и не изменялись.

3.1.4 Симуляция молекулярной динамики

Все запуски молекулярной динамики выполнялись с помощью Desmond [75]. Заряды белки были нейтрализованы добавлением соответствующего количества ионов Na⁺ или Cl⁻. Сольватные боксы были приготовлены так, как описано в руководстве пользователя Desmond. TIP3P [82] использовалась в качестве модели сольватации, сама симуляция выполнялась в течение 100 нс при T = 300 K и P = 1 атм. NVP был выбран в качестве канонического ансамбля. Значения остальных параметров и настроек запуска были оставлены по умолчанию и не изменялись.

Описанные выше действия являются справедливыми и для осуществления непредвзятого моделирования, за тем лишь исключением, что параметр «Seed» устанавливался как «Random».

3.1.5 Binding Pose Metadynamics

Для сравнения с результатами, полученными для бензотриазольного PROTAC в ходе расчета IFD-MD, нами был дополнительно выполнен расчет метадинамики кристаллического тройного комплекса бBOY с помощью Binding Pose Metadynamics [74].

Запуск производился с использованием оболочки командной строки Linux, команда имела следующее содержание:

```
$$SCHRODINGER/run -FROM psp binding_pose_metadynamics.py бBOY  
maegz - ligand_asl 'mol. Num 3' -ntrial 10 - jobname  
metadynamics_binding_бBOY -SUBHOST localhost:1 -HOST localhost:1
```

В качестве лиганда был выбран dBET6, производилось 10 независимых симуляций по 10 нс каждая. Значения остальных параметров запуска были оставлены по умолчанию и не изменялись.

3.1.6 FEP+

Дополнительно производился расчет абсолютной энергии связывания как для dBET6, так и для бензотриазольного аналога методом возмущения свободной энергии (англ. Free Energy Perturbation) с помощью модуля FEP+ Schrodinger [84]. Все симуляции выполнялись в течение 20 нс, для dBET6 использовалась кристаллическая структура 6BOY, для бензотриазольного PROTAC комплекс, полученный в ходе IFD-MD.

Расчет относительной энергии связывания был запущен с аналогичными параметрами для времени симуляций, использовалась структура 6BOY, соответствующая структура BT-PROTAC в связывающей конформации была получена модификацией dBET6 с помощью модуля 3D-Builder.

3.1.7 Protein Degradation Sampler

Рабочий процесс запускался с помощью платформы автоматизации, включающий оболочку командной строки Schrodinger PowerShell.

Входные данные (белковые комплексы и структуры PROTAC) предоставлялись в формате .mae. В качестве типа алгоритма для генерации структур PROTAC выбирался модуль молекулярной динамики Desmond. Устанавливались дистанционные ограничения (0.7 или 1.4 Å) на позиционирование всех атомов, Ca атомов, тяжелых атомов Ca или всех тяжелых атомов. Время симуляции МД составляло 10000 пс. Минимальное расстояние, на которое атомам боеголовок PROTAC разрешено сближаться, устанавливалось 10 Å. Максимальное число отбираемых комплексов составляло 5000. Также пропускался этап оценки PIPER.

Команда для запуска имела следующее содержание:

```
$$SCHRODINGER/run -FROM psp protein_degrader_sampler.py  
BRD4_IFD-MD_p1.mae 'res.num 1000' CRBN_IFD-MD_p1.mae 'res.num 2000'  
triazole_PROTAC.mae -poses 5000 -clash_fraction 0.7 (1.4) -skip_piper -
```

```
clash_atoms backbone_heavy (heavy or backbone_all or all) -sampling_protocol  
DESMOND -sampling_steps 10000 -min_md_distance 10 -jobname -NJOBS 1 -  
HOST localhost
```

3.1.8 Энергетическая оценка Prime

Энергетическая оценка происходила с помощью модуля Prime с использованием оболочки командной строки. В качестве «лиганда» был выбран BRD4. Режим расчета устанавливался как ограниченная минимизация геометрии.

Запуск производился с использованием оболочки командной строки Linux, команда имела следующее содержание:

```
$SCHRODINGER/prime_mmgbsa input.maegz -job_type REAL_MIN -  
ligand 'chain.name B' -jobname -HOST localhost:5 -NJOBS 1000
```

Заключение

По результатам проделанной работы можно сделать следующие выводы:

– протестирована работа алгоритма Schrodinger IFD-MD на структурах белок-белковых тройных комплексов PROTAC. Доказано, что алгоритм может быть использован для моделирования сложных структур и предсказания способов связывания в сайтах, образованных белок-белковым интерфейсом. Был получен ряд тройных комплексов бензотриазольного аналога dBET6. Общая оценка $IFD-MDScore = -2.450$ ккал/моль, $RMSD_{Lig} = 5.328$ Å относительно кристаллического референса;

– по результатам проведения МД-симуляции установлено, что комплекс VT-dBET6 достигает состояния равновесия к концу эксперимента, чего нельзя сказать о комплексе нативного dBET6, а также что способ связывания молекул-химер в ходе симуляции оставался схожим и существенно не изменялся – большинство взаимодействий как с BRD4, так и CRBN сохранились;

– протестирован альтернативный подход к генерации тройных комплексов — Protein Degradar Sampler — и оптимизированы параметры расчета. Для наибольшего обогащения результатов комплексами близкими к кристаллическим дистанционные ограничения на столкновения должны составлять 0.7 Å при учете только C α атомов;

– в ходе оценки с помощью энергетической модели MM-GBSA Prime установлено, что значение Prime Energy можно использовать для ранжирования наиболее кристаллоподобных комплексов из всех сгенерированных на первое или второе место. Полученные результаты имеют важную практическую значимость для разработки новых химерных молекул.

Список используемой литературы

1. The ubiquitin–proteasome system in immune cells / G. Cetin, S. Klafack, M., Studencka-Turski et al. // *Biomolecules*. – 2021. – Vol. 11, № 1. – P. 60.
2. Baloghova N. Ubiquitin ligases involved in the regulation of Wnt, TGF- β , and notch signaling pathways and their roles in mouse development and homeostasis / N. Baloghova, T. Lidak, L. Cermak // *Genes*. – 2019. – Vol. 10, № 10. – P. 815.
3. Kleiger G. Perilous journey: a tour of the ubiquitin–proteasome system / G. Kleiger, T. Mayor // *Trends in cell biology*. – 2014. – Vol. 24, № 6. – P. 352-359.
4. Kravtsova-Ivantsiv Y. Non-canonical ubiquitin-based signals for proteasomal degradation / Y. Kravtsova-Ivantsiv, A. Ciechanover // *Journal of cell science*. – 2012. – Vol. 125, № 3. – P. 539-548.
5. Collins G. A. The logic of the 26S proteasome / A. Collins, A. L. Goldberg // *Cell*. – 2017. – Vol. 169, № 5. – P. 792-806.
6. Chamberlain P. P. Development of targeted protein degradation therapeutics / P. P. Chamberlain, L. G. Hamann // *Nature chemical biology*. – 2019. – Vol. 15, № 10. – P. 937-944.
7. Sahu I. Proteasome in action: substrate degradation by the 26S proteasome / I. Sahu, M. H. Glickman // *Biochemical Society Transactions*. – 2021. – Vol. 49, № 2. – P. 629-644.
8. Lessons in PROTAC design from selective degradation with a promiscuous warhead / D. P. Bondeson, B. E. Smith, G. M. Burslem et al. // *Cell chemical biology*. – 2018. – Vol. 25, № 1. – P. 78-87. e5.
9. Wu J. Computational strategies for PROTAC drug discovery / J. Wu, W. Wang, C. H. Leung // *Acta Materia Medica*. – 2023. – Vol. 2, № 1. – P. 42-53.
10. Wu P. FDA-approved small-molecule kinase inhibitors / P. Wu, T. E. Nielsen, M. H. Clausen // *Trends in pharmacological sciences*. – 2015. – Vol. 36, № 7. – P. 422-439.
11. Androgen receptor degradation by the proteolysis-targeting chimera ARCC-4 outperforms enzalutamide in cellular models of prostate cancer drug

resistance / J. Salami, S. Alabi, R. R. Willard et al. // *Communications biology*. – 2018. – Vol. 1, № 1. – P. 1-9.

12. Degradation of Bruton's tyrosine kinase mutants by PROTACs for potential treatment of ibrutinib-resistant non-Hodgkin lymphomas / Y. Sun, N. Ding, Y. Song et al. // *Leukemia*. – 2019. – Vol. 33, № 8. – P. 2105-2110.

13. Cromm P. M. Addressing kinase-independent functions of Fak via PROTAC-mediated degradation / P. M. Cromm, K. T. G. Samarasinghe, J. Hines et al. // *Journal of the American Chemical Society*. – 2018. – Vol. 140, № 49. – P. 17019-17026.

14. Kim J. Privileged structures: efficient chemical “navigators” toward unexplored biologically relevant chemical spaces / J. Kim, H. Kim, S. B. Park // *Journal of the American Chemical Society*. – 2014. – Vol. 136, № 42. – P. 14629-14638.

15. A potent and selective small-molecule degrader of STAT3 achieves complete tumor regression in vivo / L. Bai, H. Zhou, R. Hu et al. // *Cancer cell*. – 2019. – Vol. 36, № 5. – P. 498-511.

16. Amide-to-ester substitution as a strategy for optimizing PROTAC permeability and cellular activity / V. G. Klein, A. G. Bond, C. Craigon et al. // *Journal of medicinal chemistry*. – 2021. – Vol. 64, № 24. – P. 18082-18101.

17. Catalytic in vivo protein knockdown by small- molecule PROTACs / D. P. Bondeson, A. Mares, I. E. D. Smith et al. // *Nature chemical biology*. – 2015. – Vol. 11, № 8. – P. 611- 617.

18. Casement R. Mechanistic and structural features of PROTAC ternary complexes / R. Casement, A. Bond, C. Craigon et al. // *Targeted Protein Degradation: Methods and Protocols*. – 2021. – P. 79-113.

19. The roles of E3 ubiquitin ligases in cancer progression and targeted therapy / C. Sampson, Q. Wang, W. Otkur et al. // *Clinical and Translational Medicine*. – 2023. – Vol. 13, № 3. – P. e1204.

20. Molecular architecture and assembly of the DDB1–CUL4A ubiquitin ligase machinery / S. Angers, T. Li, X. Yi et al. // *Nature*. – 2006. – Vol. 443, № 7111. – P. 590-593.

21. Bosu D. R. Cullin-RING ubiquitin ligases: global regulation and activation cycles / D. R. Bosu, E. T. Kipreos // *Cell division*. – 2008. – Vol. 3. – P. 1-13.

22. Structure of the DDB1–CRBN E3 ubiquitin ligase in complex with thalidomide / E. S. Fischer, K. Bohm, J. R. Lydeard et al. // *Nature*. – 2014. – Vol. 512, № 7512. – P. 49-53.

23. Mellin G. W. The saga of thalidomide: neuropathy to embryopathy, with case reports of congenital anomalies / G. W. Mellin, M. Katzenstein // *New England Journal of Medicine*. – 1962. – Vol. 267, № 23. – P. 1184-1193.

24. Raedler L. A. Revlimid (lenalidomide) now FDA approved as first-line therapy for patients with multiple myeloma / L. A. Raedler // *American health & drug benefits*. – 2016. – Vol. 9, № Spec Feature. – P. 140.

25. Pomalidomide alone or in combination with low-dose dexamethasone in relapsed and refractory multiple myeloma: a randomized phase 2 study / P. G. Richardson, D. S. Siegel, R. Vij et al. // *Blood, The Journal of the American Society of Hematology*. – 2014. – Vol. 123, № 12. – P. 1826-1832.

26. Design, synthesis, and biological evaluation of IRAK4-targeting PROTACs / Y. Chen, Y. Ning, G. Bai et al. // *ACS Medicinal Chemistry Letters*. – 2020. – Vol. 12, № 1. – P. 82-87.

27. Discovery of SIAIS178 as an effective BCR-ABL degrader by recruiting von Hippel–Lindau (VHL) E3 ubiquitin ligase / Q. Zhao, C. Ren, L. Liu et al. // *Journal of medicinal chemistry*. – 2019. – Vol. 62, № 20. – P. 9281-9298.

28. Discovery of potent epidermal growth factor receptor (EGFR) degraders by proteolysis targeting chimera (PROTAC) / H. Zhang, H. Y. Zhao, X. X. Xi et al. // *European Journal of Medicinal Chemistry*. – 2020. – Vol. 189. – P. 112061.

29. Amide-to-ester substitution as a strategy for optimizing PROTAC permeability and cellular activity / V. G. Klein, A. G. Bond, C. Craigon et al. // *Journal of medicinal chemistry*. – 2021. – Vol. 64, № 24. – P. 18082-18101
30. Cecchini C. Linkers as game-changers in protac technology: emphasizing general trends in PROTAC pharmacokinetics for their rational design / C. Cecchini, S. Tardy, L. Scapozza // *Chimia*. – 2022. – Vol. 76, № 4. – P. 341-341.
31. Paiva S. L. Targeted protein degradation: elements of PROTAC design / S. L. Paiva, C. M. Crews // *Current opinion in chemical biology*. – 2019. – Vol. 50. – P. 111-119.
32. Structure-based discovery of SIAIS001 as an oral bioavailability ALK degrader constructed from Alectinib / C. Ren, N. Sun, Y. Kong et al. // *European Journal of Medicinal Chemistry*. – 2021. – Vol. 217. – P. 113335.
33. Bemis T. A. Unraveling the role of linker design in proteolysis targeting chimeras: Miniperspective / T. A. Bemis, J. J. La Clair, M. D. Burkart // *Journal of Medicinal Chemistry*. – 2021. – Vol. 64, № 12. – P. 8042-8052.
34. Impact of linker length on the activity of PROTACs / K. Cyrus, M. Wehenkel, E. Y. Choi et al. // *Molecular BioSystems*. – 2011. – Vol. 7, № 2. – P. 359-364.
35. BAF complex vulnerabilities in cancer demonstrated via structure-based PROTAC design / W. Farnaby, M. Koegl, M. J. Roy et al. // *Nature chemical biology*. – 2019. – Vol. 15, № 7. – P. 672-680.
36. Discovery of ARD-69 as a highly potent proteolysis targeting chimera (PROTAC) degrader of androgen receptor (AR) for the treatment of prostate cancer / X. Han, C. Wang, C. Qin et al. // *Journal of medicinal chemistry*. – 2019. – Vol. 62, № 2. – P. 941-964.
37. Neeser R. M. Reinforcement Learning-Driven Linker Design via Fast Attention-based Point Cloud Alignment / R. M. Neeser, M. Akdel, D. Kovtun et al. // *arXiv preprint arXiv:2306.08166*. – 2023.

38. Link-INVENT: generative linker design with reinforcement learning / J. Guo, F. Knuth, C. Margreitter et al. // *Digital Discovery*. – 2023. – Vol. 2, № 2. – P. 392-408.
39. Kao C. T. Fragment Linker Prediction Using the Deep Encoder-Decoder Network for PROTACs Drug Design / C. T. Kao, C. T. Lin, C. L. Chou et al. // *Journal of Chemical Information and Modeling*. – 2023. – Vol. 63, № 10. – P. 2918-2927.
40. Imrie F. Deep generative models for 3D linker design / F. Imrie, A. R. Bradley, M. van der Schaar et al. // *Journal of chemical information and modeling*. – 2020. – Vol. 60, № 4. – P. 1983-1995.
41. DRlinker: Deep Reinforcement Learning for Optimization in Fragment Linking Design / Y. Tan, L. Dai, W. Huang et al. // *Journal of Chemical Information and Modeling*. – 2022. – Vol. 62, № 23. – P. 5907-5917.
42. Neeser R. M. Reinforcement Learning-Driven Linker Design via Fast Attention-based Point Cloud Alignment / R. M. Neeser, M. Akdel, D. Kovtun et al. // *arXiv preprint arXiv:2306.08166*. – 2023.
43. Li B. 3D based generative PROTAC linker design with reinforcement learning / B. Li, T. Ran, H. Chen // *Briefings in Bioinformatics*. – 2023. – Vol. 24, № 5. – P. 323.
44. Rational Prediction of PROTAC-Compatible Protein–Protein Interfaces by Molecular Docking / G. P. Pereira, B. Jimenez-Garcia, R. Pellarin et al. // *Journal of Chemical Information and Modeling*. – 2023. – Vol. 63, № 21. – P. 6823-6833.
45. Drummond M. L. In silico modeling of PROTAC-mediated ternary complexes: validation and application / M. L. Drummond, C. I. Williams // *Journal of chemical information and modeling*. – 2019. – Vol. 59, № 4. – P. 1634-1644.
46. Althagafi I. New series of thiazole derivatives: synthesis, structural elucidation, antimicrobial activity, molecular modeling and MOE docking / I. Althagafi, N. El-Metwaly, T. A. Farghaly // *Molecules*. – 2019. – Vol. 24. – №. 9. – P. 1741.

47. Zaidman D. PROsettaC: Rosetta based modeling of PROTAC mediated ternary complexes / D. Zaidman, J. Prilusky, N. London // Journal of chemical information and modeling. – 2020. – Vol. 60, № 10. – P. 4894-490

48. ICM Protein Modeling User Manual. URL: https://www.google.com/url?sa=t&source=web&rct=j&opi=89978449&url=https://www.molsoft.com/docs/ICM_Modeling.pdf&ved=2ahUKEwjE0q3AvPiFAxXFCRAIHYEaCv4QFnoECBwQAQ&usg=AOvVaw0e4-p7vYVJc3amHsqDLyBX (дата обращения 01.03.2024).

49. High Accuracy Prediction of PROTAC Complex Structures / M. Ignatov, A. Jindal, S. Kotelnikov et al. // Journal of the American Chemical Society. – 2023. – Vol. 145, № 13. – P. 7123-7135.

50. Structure-based design of a macrocyclic PROTAC / A. Testa, S. J. Hughes, X. Lucas et al. // Angewandte Chemie. – 2020. – Vol. 132, № 4. – P. 1744-1751.

51. In silico modeling and scoring of PROTAC-mediated ternary complex poses / J. Liao, X. Nie, I. C. Unarta et al. // Journal of Medicinal Chemistry. – 2022. – Vol. 65, № 8. – P. 6116-6132.

52. Drummond M. L. Improved accuracy for modeling PROTAC-mediated ternary complex formation and targeted protein degradation via new in silico methodologies / M. L. Drummond, A. Henry, H. Li et al. // Journal of Chemical Information and Modeling. – 2020. – Vol. 60, № 10. – P. 5234-5254.

53. Li W. Importance of three-body problems and protein–protein interactions in proteolysis-targeting chimera modeling: insights from molecular dynamics simulations / W. Li, J. Zhang, L. Guo et al. // Journal of Chemical Information and Modeling. – 2022. – Vol. 62, № 3. – P. 523-532.

54. Nabuurs S. B. A flexible approach to induced fit docking / S. B. Nabuurs, M. Wagener, J. de Vlieg // Journal of medicinal chemistry. – 2007. – Vol. 50, № 26. – P. 6507-6518.

55. The maximal and current accuracy of rigorous protein-ligand binding free energy calculations / G. A. Ross, C. Lu, G. Scarabelli et al. // *Communications Chemistry*. – 2023. – Vol. 6, № 1. – P. 222.

56. Reliable and accurate solution to the induced fit docking problem for protein–ligand binding / E. B. Miller, R. B. Murphy, D. Sindhikara et al. // *Journal of Chemical Theory and Computation*. – 2021. – Vol. 17, № 4. – P. 2630-2639.

57. RCSB Protein Data Bank (PDB). URL: <https://www.rcsb.org/> (дата обращения 15.04.2024).

58. Protein Preparation Guide Schrödinger Suite 2012 Update 2, Glide version 6.4, Prime version 3.7, Schrodinger, LCC, New York, 2012.

59. LigPrep 3.4 User Manual, Glide version 6.4, Prime version 3.7, Schrodinger, LCC, New York, 2015.

60. Cryo-EM in drug discovery: achievements, limitations and prospects / J. P. Renaud, A. Chari, C. Ciferri et al. // *Nature reviews Drug discovery*. – 2018. – Vol. 17, № 7. – P. 471-492.

61. Maveyraud L. Protein X-ray crystallography and drug discovery / L. Maveyraud, L. Mourey // *Molecules*. – 2020. – Vol. 25, № 5. – P. 1030.

62. Kaptein R. Protein structures from NMR / R. Kaptein, R. Boelens, R. M. Scheek et al. // *Biochemistry*. – 1988. – Vol. 27, № 15. – P. 5389-5395.

63. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening / T. A. Halgren, R. B. Murphy, R. A. Friesner et al. // *Journal of medicinal chemistry*. – 2004. – Vol. 47, № 7. – P. 1750-1759.

64. Bhachoo J. Investigating protein–peptide interactions using the Schrödinger computational suite / J. Bhachoo, T. Beuming // *Modeling peptide-protein interactions: methods and protocols*. – 2017. – P. 235-254.

65. Dixon S. L. PHASE: a novel approach to pharmacophore modeling and 3D database searching / S. L. Dixon, A. M. Smondyrev, S. N. Rao // *Chemical biology & drug design*. – 2006. – Vol. 67, № 5. – P. 370-372.

66. WScore: a flexible and accurate treatment of explicit water molecules in ligand–receptor docking / R. B. Murphy, M. P. Repasky, J. R. Greenwood et al. // *Journal of medicinal chemistry*. – 2016. – Vol. 59, № 9. – P. 4364-4384.

67. Prediction of protein–ligand binding poses via a combination of induced fit docking and metadynamics simulations / A. J. Clark, P. Tiwary, K. Borrelli et al. // *Journal of chemical theory and computation*. – 2016. – Vol. 12, № 6. – P. 2990-2998.

68. Phase 4.3 User Manual, Glide version 6.4, Prime version 3.7, Schrodinger, LCC, New York, 2015.

69. Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein–ligand complexes / R. A. Friesner, R. B. Murphy, M. P. Repasky et al. // *Journal of medicinal chemistry*. – 2006. – Vol. 49, № 21. – P. 6177-6196.

70. Cappel D. Calculating water thermodynamics in the binding site of proteins–applications of WaterMap to drug discovery / D. Cappel, W. Sherman, T. Beuming // *Current Topics in Medicinal Chemistry*. – 2017. – Vol. 17, № 23. – P. 2586-2598.

71. Testing inhomogeneous solvation theory in structure-based ligand discovery / T. E. Balius, M. Fischer, R. M. Stein et al. // *Proceedings of the National Academy of Sciences*. – 2017. – Vol. 114, № 33. – P. E6839-E6846.

72. Motifs for molecular recognition exploiting hydrophobic enclosure in protein–ligand binding / T. Young, R. Abel, B. Kim et al. // *Proceedings of the National Academy of Sciences*. – 2007. – Vol. 104, № 3. – P. 808-813.

73. Contribution of explicit solvent effects to the binding affinity of small-molecule inhibitors in blood coagulation factor serine proteases / R. Abel, N. K. Salam, J. Shelley et al. // *ChemMedChem*. – 2011. – Vol. 6, № 6. – P. 1049-1066.

74. Fusani L. Exploring ligand stability in protein crystal structures using binding pose metadynamics / L. Fusani, D. S. Palmer, D. O. Somers et al. // *Journal of Chemical Information and Modeling*. – 2020. – Vol. 60, № 3. – P. 1528-1539.

75. Desmond 4.2 User Manual, Schrödinger, LLC, New York, NY, 2015. – 166 p.
76. Ustinov E. Kinetic Monte Carlo approach for molecular modeling of adsorption // Current Opinion in Chemical Engineering. – 2019. – Vol. 24. – P. 1-11.
77. Kolossváry I. Low mode search. An efficient, automated computational method for conformational analysis: application to cyclic and acyclic alkanes and cyclic peptides // Journal of the American Chemical Society. – 1996. – Vol. 118, №. 21. – P. 5011-5019.
78. Watts K. S ConfGen: a conformational search method for efficient generation of bioactive conformers // Journal of chemical information and modeling. – 2010. – Vol. 50, №. 4. – P. 534-546.
79. RDKit User Manual URL: <https://rdkit.org/docs/source/rdkit.Chem.html> (дата обращения 20.04.2024).
80. Protein Structure Alignment User Manual URL: https://gohom.win/ManualHom/Schrodinger/Schrodinger_2015_2_docs/maestro/help_Maestro/tools_menu/protein_structure_alignment.html (дата обращения 25.04.2024).
81. Epik 3.2 User Manual, Schrödinger, LLC, New York, NY, 2015. – 64 p.
82. MacKerell A. D. All-atom empirical potential for molecular modeling and dynamics studies of proteins / A. D. MacKerell, Jr., D. Bashford, M. Bellott et al. // J. Phys. Chem. B. – 1998. – Vol. 102, № 18. – P. 3586-3616.
83. Lu C. OPLS4: Improving force field accuracy on challenging regimes of chemical space // Journal of chemical theory and computation. – 2021. – Vol. 17. – №. 7. – P. 4291-4300.
84. Chen W. et al. Enhancing hit discovery in virtual screening through accurate calculation of absolute protein-ligand binding free energies. – 2022.