

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
имени М.В. ЛОМОНОСОВА  
ХИМИЧЕСКИЙ ФАКУЛЬТЕТ  
ФЕДЕРАЛЬНЫЙ НАУЧНЫЙ ЦЕНТР ИССЛЕДОВАНИЙ И РАЗРАБОТКИ  
ИММУНОБИОЛОГИЧЕСКИХ ПРЕПАРАТОВ ИМ. М.П. ЧУМАКОВА РАН  
(Институт полиомиелита)

*На правах рукописи*

**Фомина Анастасия Дмитриевна**

**Систематический ансамблевый докинг потенциальных  
лигандов главной протеазы SARS-CoV-2 и белка NS1 flavivирусов**

1.4.16. Медицинская химия

ДИССЕРТАЦИЯ  
на соискание ученой степени  
кандидата химических наук

Научные руководители:  
кандидат химических наук,  
старший научный сотрудник  
Палюлин Владимир Александрович  
кандидат химических наук  
Осолодкин Дмитрий Иванович

Москва – 2025

## Оглавление

Введение.....	4
Глава 1. Обзор литературы.....	9
1.1. Ансамблевый докинг.....	9
1.1.1. Структурно-обоснованный виртуальный скрининг.....	9
1.1.2. Составление ансамблей белковых структур.....	10
1.1.3. Консенсусная оценка результатов виртуального скрининга.....	14
1.2. Главная протеаза коронавирусов как терапевтическая мишень.....	16
1.2.1. Репликативный цикл коронавирусов.....	17
1.2.2. Структура главной протеазы коронавирусов.....	20
1.2.3. Ингибиторы главной протеазы SARS-CoV-2.....	24
1.2.3.1. Ковалентные пептидомиметики.....	25
1.2.3.2. Ковалентные ингибиторы непептидной природы.....	27
1.2.3.3. Нековалентные ингибиторы.....	29
1.3. Белок NS1 ортофлавивирусов как терапевтическая мишень.....	32
Глава 2. Разработка метода составления ансамбля разнообразных структур.....	40
2.1. Подготовка белковых структур.....	40
2.2. Составление ансамблей.....	43
2.2.1. Выбор структур по среднеквадратичному отклонению.....	43
2.2.2. Выбор структур методом анализа главных компонент.....	46
2.2.3. Выбор состояний из траекторий молекулярной динамики.....	48
2.2.4. Конформационное разнообразие ансамблей.....	48
Глава 3. Подготовка библиотек низкомолекулярных соединений.....	51
3.1. Библиотека соединений с аннотацией по противокоронавирусной активности.....	51
3.2. Библиотека разнообразных соединений.....	54
3.3. Химическое пространство ингибиторов главной протеазы SARS-CoV-2.....	55
Глава 4. Разработка метода систематического ансамблевого докинга.....	61
4.1. Виртуальный скрининг методом систематического ансамблевого докинга.....	61
4.2. Применение методики систематического ансамблевого докинга к CDK2.....	66
4.3. Докинг со структурными ограничениями.....	68
4.3.1. Анализ взаимодействий.....	68

4.3.2. Влияние структурных ограничений на результаты докинга.....	70
Глава 5. Виртуальный скрининг ингибиторов главной протеазы SARS-CoV-2.....	74
5.1. Поиск новых хемотипов ингибиторов главной протеазы.....	74
5.2. Расширение серии тиенопирамидинонов.....	79
Глава 6. Прогнозирование класса активности ингибиторов Mpro SARS-CoV-2.....	88
6.1. Обучающая выборка и декрипторы.....	88
6.2. Подбор параметров моделей.....	90
6.3. Анализ вкладов дескрипторов в прогноз моделей машинного обучения.....	98
6.4. Влияние релевантности выборки на качество прогнозирования.....	100
Глава 7. Виртуальный скрининг лигандов белка NS1 ортофлавивирусов.....	109
7.1. Сравнительный анализ структур белка NS1 ортофлавивирусов.....	109
7.2. Поиск потенциальных сайтов связывания малых молекул.....	114
7.2.1. Поиск методом FTSite.....	114
7.2.2. Поиск методом DoGSiteScorer.....	116
7.2.3. Выбор потенциального сайта связывания.....	119
7.3. Моделирование ансамбля.....	121
7.3.1. Построение моделей структур белка NS1 ортофлавивирусов по гомологии.....	122
7.3.2. Выбор моделей по среднеквадратичному отклонению.....	123
7.3.3. Составление ансамбля моделей по результатам докинга.....	127
7.4. Виртуальный скрининг лигандов белка NS1 ортофлавивирусов.....	133
Заключение.....	136
Список сокращений.....	138
Список литературы.....	139
Приложение А.....	160
Приложение Б.....	171
Приложение В.....	173
Приложение Г.....	176
Приложение Д.....	179
Приложение Е.....	182

## Введение

**Актуальность темы исследования.** Метод молекулярного докинга – один из основных инструментов медицинской химии. В частности, он применяется для структурно-обоснованного виртуального скрининга – вычислительной процедуры отбора и приоритизации потенциальных лекарственных молекул по целевым свойствам для дальнейших экспериментальных исследований. Благодаря развитию кристаллографического скрининга, моделирования по гомологии [1] и генерации при помощи нейронных сетей [2] стали доступны пространственные структуры для большинства белков, закодированных в геномах человека и различных патогенов. Одним из методов структурно-обоснованного виртуального скрининга, который позволяет наиболее полно использовать доступность значительного количества структурных данных о молекулярных мишениях, является ансамблевый докинг. Этот метод позволяет учитывать конформационное пространство белка за счет докинга в ансамбль структур и обеспечивает более высокое обогащение хитов виртуального скрининга активными соединениями, чем докинг в индивидуальные структуры [3]. Кроме того, этот подход требует значительно меньших вычислительных затрат по сравнению с другими способами описания конформационной подвижности мишени, что позволяет использовать его в задачах высокопроизводительного виртуального скрининга. В частности, метод ансамблевого докинга обладает особой значимостью для поиска противовирусных лекарственных средств, так как белки-мишени родственных вирусов имеют высокую гомологию, и хиты такого виртуального скрининга могут обладать широким спектром действия.

Заболевания, вызванные коронавирусами и ортофлавивирусами, представляют значительную угрозу для жизни и здоровья людей по всему миру. Вакцинация позволяет снизить риск развития тяжелого заболевания, однако не для всех вирусов разработаны вакцины, а их использование не всегда возможно. В частности, для некоторых ортофлавивирусов наблюдается антитело-зависимое усиление инфекции [4], а лечение носит только поддерживающий характер, из-за чего высока смертность и вероятность тяжелых долгосрочных последствий при остром протекании заболевания. Поэтому важной задачей является разработка лекарственных средств прямого противовирусного действия, которые действовали бы уже после заражения. В ходе пандемии COVID-19 также возникла острая потребность в быстрой разработке и репозиционировании противовирусных лекарственных средств, и в результате многочисленных и разноплановых исследований было накоплено значительное количество структурной и биологической информации. Такое количество данных позволяет максимально эффективно применить метод ансамблевого докинга. Однако этот метод недостаточно автоматизирован, отличается трудоемкостью и требует постоянного контроля со стороны оператора, что ограничивает его применение. Кроме того, возникает проблема выбора из

значительного количества разнообразных структур тех, которые будут репрезентативны относительно конформационного пространства и надежны в качестве мишеней для виртуального скрининга. Подобный отбор зачастую основан на новизне структуры, литературных данных об успешном применении отдельных структур для виртуального скрининга или случайности, и не является систематическим. Область применимости ансамблевого докинга для поиска противовирусных препаратов широкого спектра действия также остается вопросом для изучения.

**Целью** работы была разработка автоматизированной процедуры виртуального скрининга методом ансамблевого докинга, основанной на стандартизованных подходах к отбору структур белковой мишени, генерации и обработке результатов докинга и ретроспективной валидации прогностической способности модели для применения в поиске потенциальных лигандов вирусных белков — главной протеазы SARS-CoV-2 и белка NS1 ортофлавивирусов. Для достижения этой цели в работе были поставлены и решены следующие **задачи**:

- Разработка систематического подхода к составлению ансамбля структур биологических мишеней;
- Формирование и аннотирование библиотеки экспериментально исследованных низкомолекулярных соединений на основании литературных данных и библиотеки разнообразных соединений, подобных соединениям-лидерам;
- Разработка протоколов генерации и обработки результатов виртуального скрининга методом докинга в ансамбль структур биологических мишеней, и анализа результатов на основе консенсусных метрик;
- Анализ применимости информации о способах связывания лигандов с мишенями, полученной по результатам виртуального скрининга, для прогнозирования класса активности малых молекул при помощи машинного обучения;
- Виртуальный скрининг потенциальных ингибиторов главной протеазы SARS-CoV-2 методом систематического ансамблевого докинга на основе однородных кристаллографических данных и формирование фокусированной библиотеки для экспериментальных исследований;
- Виртуальный скрининг потенциальных лигандов широкого спектра действия для белка NS1 ортофлавивирусов методом систематического ансамблевого докинга и формирование фокусированной библиотеки для экспериментальных исследований.

**Научная новизна.** Впервые разработана и реализована стратегия систематического отбора разнообразных ансамблей белковых структур на основании среднеквадратичного отклонения атомов аминокислотных остатков, формирующих сайт связывания.

Систематизирована и оформлена в виде базы данных информация об экспериментальном

исследовании активности низкомолекулярных соединений в отношении главной протеазы SARS-CoV-2 и других бетакоронавирусов, опубликованная в научной литературе.

На основе виртуального скрининга методом систематического ансамблевого докинга выявлен новый класс ингибиторов главной протеазы SARS-CoV-2 — замещенные тиено[3,2-*d*]пиrimидиноны, активность которых подтверждена экспериментально.

Показано обогащение прогнозов при построении моделей случайного леса истинноположительными результатами при использовании фингерпринтов взаимодействия относительно аналогичных моделей, основанных на молекулярных фингерпринтах.

Впервые проведен систематический поиск пригодных для докинга мест связывания в разных структурах белка NS1 ортофлавивирусов и в рамках ансамблевого докинга показано преимущество группировки моделей структуры белка NS1 по переносчикам вируса.

**Практическая значимость.** Разработана программа для ЭВМ, реализующая метод систематического ансамблевого докинга для мишень-ориентированного виртуального скрининга потенциальных противовирусных препаратов. Получено свидетельство о государственной регистрации в Реестре программ для ЭВМ.

Разработана автоматизированная методика составления ансамблей белковых структур и ансамблевого докинга, которая может применяться для виртуального скрининга лигандов любой биологической мишени, для которой существует большое количество структурной информации.

На основании литературных данных сформирована и аннотирована библиотека низкомолекулярных органических соединений, активность которых в отношении главной протеазы SARS-CoV-2 исследована экспериментально. Показана эффективность применения этой библиотеки при валидации схем виртуального скрининга потенциальных ингибиторов Mpro SARS-CoV-2. Получено свидетельство о государственной регистрации в Реестре баз данных.

Изучены возможные способы связывания низкомолекулярных соединений с белком NS1 ортофлавивирусов и выявлены структурные фрагменты, потенциально способные обеспечивать специфическое связывание с белком NS1. На основании этой информации сформирована фокусированная библиотека потенциальных лигандов белка NS1 для экспериментальных исследований.

#### **Положения, выносимые на защиту:**

1) Выбор разнообразных белковых структур в ансамбли при помощи среднеквадратичного отклонения между координатами атомов аминокислотных остатков, формирующих сайт связывания, позволяет достаточно репрезентативно описать конформационное пространство белка и нивелировать фактор случайности при выборе структуры мишени для виртуального

скрининга;

2) Разработан программный комплекс для систематического ансамблевого докинга, позволяющий составить разнообразный репрезентативный ансамбль из любого предложенного набора структур белка и рассчитать консенсусный результат виртуального скрининга;

3) Собрана аннотированная база данных соединений с экспериментально определенной активностью против главной протеазы бетакоронавирусов. В ней содержится количественная информация об эффективности ингибиования протеазы и противовирусной активности, наличии в соединении реакционноспособных групп для ковалентного связывания, а также цифровые идентификаторы источников данных;

4) Добавление фингерпринтов лиганд-белкового взаимодействия в модели случайного леса для прогнозирования класса активности низкомолекулярных ингибиторов главной протеазы SARS-CoV-2 позволяет увеличить количество истинноположительных результатов;

5) В ходе отбора моделей структур белка NS1 ортофлавивирусов разделение ансамблей по переносчикам вируса является эффективной стратегией при виртуальном скрининге потенциальных лигандов.

**Личный вклад автора** состоит в анализе литературы, проведении расчетов и моделирования, интерпретации и обобщении полученных результатов, подготовке материалов к публикации, представлении полученных результатов на конференциях. Автором проведены отбор и аннотация библиотеки соединений с экспериментально определенной активностью в отношении главной протеазы коронавирусов, разработана методика отбора ансамбля белковых структур, написаны программы для автоматизации отбора ансамблей, ансамблевого докинга, ранжирования молекул на основе консенсусной оценки и расчетов фингерпринтов взаимодействия, разработаны модели машинного обучения для прогнозирования активности ингибиторов главной протеазы SARS-CoV-2, проведен виртуальный скрининг различных библиотек малых молекул, дизайн серий аналогов соединений-лидеров и их отбор для экспериментальной валидации на основании результатов ансамблевого докинга. Во всех опубликованных в соавторстве работах по теме диссертационной работы вклад автора (Фоминой А.Д) является основополагающим, в том числе в работе [5], где автором проделана работа по виртуальному скринингу новых хемотипов потенциальных противокоронавирусных соединений, которая легла в основу экспериментального исследования.

**Достоверность полученных результатов** обеспечивается использованием для расчетов стандартных, широко используемых, статистически обоснованных алгоритмов и программного обеспечения, публикациями в рецензируемых научных журналах и апробацией на конференциях различного уровня.

**Апробация результатов.** Основные результаты работы были представлены в виде устных и стеновых докладов на российских и международных конференциях: 5th Advanced in silico Drug Design workshop/challenge (3-7 февраля 2020, Оломоуц, Чехия), XXVII Симпозиум "Биоинформатика и компьютерное конструирование лекарств" (5–8 апреля 2021, Москва, Россия), 5-я Российская конференция по медицинской химии с международным участием «МедХим – Россия 2021» MedChem Russia 2021 (5-8 октября 2021, Волгоград, Россия), Конференция молодых учёных и специалистов ФНЦИРИП им. М.П. Чумакова РАН (Институт полиомиелита) (2 сентября 2022, Москва, Россия), XXVIII Symposium "Bioinformatics and computer-aided drug discovery" (24-26 мая 2022, Москва, Россия), XXIX International Symposium "Bioinformatics and Computer-Aided Drug Discovery" (18-20 сентября 2023, Москва, Россия), XXX Symposium "Bioinformatics and computer-aided drug discovery" (16-18 сентября 2024, Москва, Россия, 2024).

Исследования выполнялись при поддержке государственного задания ФГАНУ «ФНЦИРИП им. М.П. Чумакова РАН» (Институт полиомиелита) и Некоммерческого фонда содействия развитию науки и образования «ИНТЕЛЛЕКТ».

**Публикации.** По результатам работы опубликованы 3 статьи в рецензируемых научных журналах, индексируемых в базе ядра РИНЦ «eLibrary Science Index», международными базами данных (Web of Science, Scopus, RSCI) и рекомендованных для защиты в диссертационном совете МГУ для публикации результатов диссертационных работ по специальности 1.4.16. Медицинская химия (химические науки), 1 свидетельство о регистрации в государственном реестре программ для ЭВМ РФ, 1 свидетельство о регистрации в государственном реестре баз данных РФ и 8 тезисов докладов, представленных на вышеуказанных российских и международных конференциях.

**Структура и объем работы.** Работа состоит из введения, обзора литературы, 6 глав, каждая из которых включает в себя описание использованных методов исследования, полученных результатов и их обсуждение, заключения, списка цитируемой литературы и приложений. Работа изложена на 200 страницах и содержит в основной части 20 таблиц и 86 рисунков. Дополнительная информация представлена в 6 приложениях (6 таблиц и 17 рисунков). Библиографический список включает в себя 275 источников литературы.

## Глава 1. Обзор литературы<sup>1</sup>

### 1.1. Ансамблевый докинг

#### 1.1.1. Структурно-обоснованный виртуальный скрининг

Метод молекулярного докинга – это стандартный инструмент ранних стадий разработки лекарственных средств. Он позволяет спрогнозировать, насколько хорошо малая молекула (лиганд) может связываться со структурой белка (мишень). Процесс виртуального скрининга методом докинга включает следующие задачи: подбор конформации лиганда в заданном сайте связывания мишени; оценку эффективности связывания на основании физико-химической, эмпирической, статистической или иной оценочной функции; ранжирование лигандов на основании значений оценочной функции [6].

Докинг может проводиться без учета конформационной подвижности лиганда и мишени (“жесткий” докинг) или учитывать подвижность лиганда и/или мишени (“гибкий” докинг). “Жесткий” докинг наименее затратен с вычислительной точки зрения и может быть использован для виртуального скрининга очень больших библиотек низкомолекулярных органических соединений, однако требует предварительной генерации конформаций лигандов, которые затем размещаются в жестком сайте связывания мишени. Успешный пример применения “жесткого” докинга для разработки ингибиторов главной протеазы SARS-CoV-2 продемонстрировали авторы [7]. Они объединили виртуальный скрининг библиотеки более 200 миллионов разнообразных соединений из базы данных ZINC15 и виртуальный скрининг фокусированной библиотеки 10 миллиардов соединений, построенной на результатах фрагментного кристаллографического скрининга. Скаффолды для дальнейшей структурно-обоснованной оптимизации выбирали на основании виртуального скрининга при помощи программы DOCK 3.7, причем соединение-лидер имело  $EC_{50} 0,11 \pm 0,03$  мкМ, что сравнимо с активностью первого зарегистрированного ингибитора Mpro SARS-CoV-2, – нирматрелвира ( $EC_{50} 0,08 \pm 0,03$  мкМ).

В ходе “полугибкого” докинга конформация лиганда подбирается в сайте связывания так, чтобы обеспечить наилучшее значение оценки докинга. Подбор конформации может осуществляться при помощи различных алгоритмов: “сборки” лиганда из фрагментов (DOCK

1 При работе над данным разделом диссертации использованы материалы следующих публикаций автора, в которых согласно Положению о присуждении ученых степеней в МГУ отражены основные результаты, положения и выводы исследования:

1. Фомина А.Д., Палюлин В.А., Осолодкин Д.И. Моделирование структуры белка NS1 ортофлавивирусов по гомологии для виртуального скрининга потенциальных лигандов // Биомедицинская химия – 2024 – т. 70 (6) – с. 456–468
2. Fomina A. D, Uvarova V. I., Kozlovskaya L. I., Palyulin V. A., Osolodkin D.I., Ishmukhametov A. A. Ensemble Docking Based Virtual Screening of SARS-CoV-2 Main Protease Inhibitors // Molecular Informatics – 2024 – V.43. – e202300279.
3. Zakharova M.Yu., Kuznetsova A.A., Uvarova V.I., Fomina A.D., Kozlovskaya L.I., Kaliberda E.N., Kurbatskaia I.N., Smirnov I.V., Bulygin A.A., Knorre V.D., Fedorova O.S., Varnek A., Osolodkin D.I., Ishmukhametov A.A., Egorov A.M., Gabibov A.G., Kuznetsov N.A. Pre-Steady-State Kinetics of the SARS-CoV-2 Main Protease as a Powerful Tool for Antiviral Drug Discovery // Frontiers in Pharmacology – 2021 – V. 12. – P. 773198.

[8], FlexX [9]), генетического алгоритма (AutoDock [10], GOLD [11]), метода Монте-Карло (Glide [12], ICM [13]) и др. [14]. Такие методы позволяют более точно прогнозировать способ связывания лиганда и белка по сравнению с “жестким” докингом, однако занимают как минимум в два раза больше расчетного времени [15].

Полностью “гибкий” докинг требует наибольших вычислительных затрат, и поэтому возможности его использования для виртуального скрининга ограничены [16]. В этом варианте конформационное пространство белка моделируется при помощи молекулярной динамики, библиотек ротамеров [17] или набора структур белка, служащих источником информации о разнообразии конформаций аминокислотных остатков, которые используют для локальной оптимизации способа связывания лиганда [18].

Использование нескольких конформаций белка для моделирования его подвижности было предложено в 1999 году на примере фармакофорного виртуального скрининга ингибиторов интегразы ВИЧ-1 [19]. Разнообразные конформации активного центра белка были извлечены из траектории молекулярной динамики, и на их основе были созданы фармакофорные модели. Консенсусные фармакофорные модели позволяли точнее прогнозировать связывание ингибиторов с белком, чем модели, полученные на основе одиночных конформаций белка. Также наблюдалось обогащение результатов виртуального скрининга активными соединениями при использовании ансамбля структур в ходе докинга по сравнению с индивидуальными структурами [3]. Было показано, что использования ансамбля имеет преимущество не для любого белка, однако конкретных факторов, которые влияют на это, не было выявлено, [20].

### **1.1.2. Составление ансамблей белковых структур**

При наличии большого объема экспериментальных кристаллографических данных для целевого белка-мишени и его близких гомологов можно использовать их в качестве источника информации о конформационном пространстве мишени. Связывание малых молекул с белком вызывает конформационные изменения в центре связывания [21]. Как следствие, докинг в структуру комплекса с лигандром зачастую обладает более высокой надежностью, чем докинг в апо-структур<sup>2</sup> как в экспериментах по воспроизведению кристаллографической конформации лиганда [22], [23], так и при виртуальном скрининге [24], [25]. Использование ансамблей холо-структур значительно увеличивает обогащение результатов виртуального скрининга по сравнению с ансамблями апо-структур (Рисунок 1), хотя в отдельных случаях докинг в индивидуальную структуру может обладать более высокой точностью [24].

---

<sup>2</sup> Апо-структура здесь – структура белка, не связанного с лигандром; холо-структура – структура комплекса белка с низкомолекулярным лигандром.

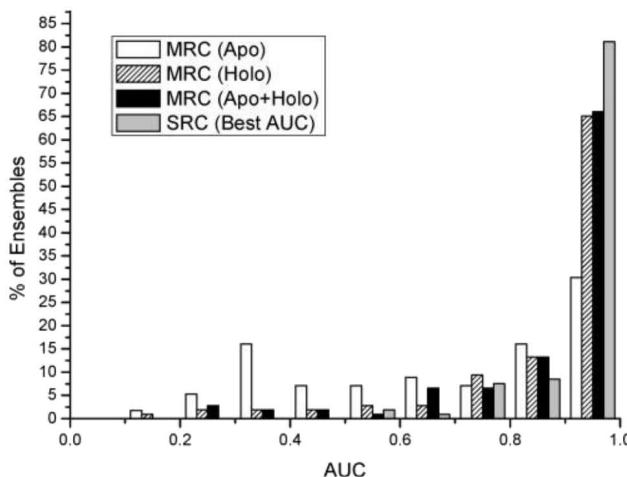


Рисунок 1. Значения площади под ROC-кривой (ROC AUC), полученные при докинге в 106 белковых ансамблей, собранных из апо-, холо-структур или структур обоих типов. Для сравнения приведены лучшие значения AUC, полученные для докинга в одну, лучшую структуру [24].

Зачастую при составлении ансамбля кристаллографических структур не применяется системный подход к отбору конформаций. Например при оптимизации ингибиторов главной протеазы SARS-CoV-2, выявленных по результатам кристаллографического скрининга соединений из базы данных ZINC, применяли ансамблевый докинг для оценки серий производных изначальных хитов [26]. Ансамбль состоял из двух структур (PDB ID: 6LU7 и 5R7Z), выбранных на основании визуального сравнения конформаций остатков, формирующих активный центр. Авторы подчеркивали важность конформационного разнообразия остатков а/к, формирующих “карман” S2<sup>3</sup> активного центра белка, в частности, Met49. В рассмотренных структурах изменение ориентации боковой цепи этого остатка уменьшает глубину кармана S2 на 3,1 Å.

Аналогичным образом поступили в работе [27] по виртуальному скринингу миллиарда молекул из базы данных Enamine REAL при помощи программы AutoDock-GPU [28]. Для докинга были использованы две кристаллические структуры (PDB ID: 5R84 и 6WQF). После докинга для 1% наиболее высоко оцененных молекул проводили переоценку генерированных способов связывания при помощи программы RF-Score-3 [29], а топ-50% соединений, ранжированных при помощи новых оценок, докировали в десять конформеров, отобранных из траектории молекулярной динамики. Результаты докинга в каждую индивидуальную конформацию оценивали при помощи 5 моделей машинного обучения (МО). Список хитов формировали из молекул, которые попали в топ-500 для 25 из 50 списков, ранжированных при помощи возможных комбинаций конформеров белка и модели МО.

3 Описание структуры активного центра Mpro SARS-CoV-2 см. в разделе 1.3.2.

Самым распространенным источником структур для ансамблевого докинга является моделирование методом молекулярной динамики (МД). Оно позволяет получить большой набор конформеров, однако большинство из них не являются биологически значимыми (то есть реализующимися при связывания белка с каким-либо лигандом) [24],[30]. Существуют различные подходы к отбору конформеров белка из траекторий МД. Самый простой заключается в выборе конформеров, соответствующих равноудаленным временными точкам траектории. Такой подход в сочетании с ансамблем кристаллографических холо-структур был успешно использован в работе по виртуальному скринингу методом ансамблевого докинга библиотеки белковых миметиков Enamine для поиска ингибиторов главной протеазы SARS-CoV-2 [31]. Однако он не обеспечивает ни максимального разнообразия конформеров, ни высокой точности виртуального скрининга.

Для максимизации покрытия конформационного пространства структурами ансамбля применяются различные методы кластеризации. Одним из таких методов является кластеризация по радиальному порогу среднеквадратичных отклонений (СКО) координат атомов [32], [33], [34]. Алгоритм радиальной пороговой кластеризации (Radial Threshold Clusterization, RTC) [35] включает несколько стадий. Сначала рассчитывают попарные СКО координат тяжелых атомов всех структур. Затем для каждой структуры назначают соседей – структуры, имеющие СКО ниже назначенного порогового значения. Структура с наибольшим числом соседей выбирается в качестве центра кластера, в который включаются ее соседи. Структуры, включенные в кластер, исключаются из рассмотрения, и процедура повторяется, пока всем структурам не будет назначен кластер.

Другой широко применяемый подход к кластеризации структур – анализ главных компонент (Principal Component Analysis, PCA) координат атомов с последующей кластеризацией (например, методом  $k$ -средних) [36]. PCA позволяет при помощи ортогонального преобразования (линейного преобразования, в котором сохраняются скалярные произведения векторов) уменьшить размерность векторного пространства (обычно до двумерного). При этом оси новой системы координат (главные компоненты) выбираются так, чтобы дисперсия данных вдоль них была максимальна, так что первая главная компонента описывает наибольший процент дисперсии данных, вторая – второй по величине, и так далее [37]. В качестве входных данных можно использовать любой набор динамических переменных, описывающих структуру белка (например, значения двугранного угла или межатомные расстояния, что устраняет необходимость выравнивать структуры). Кластеризация методом  $k$ -средних начинается со назначения  $k$  центров масс. Выбор числа  $k$  осуществляется произвольно или так, чтобы оптимизировать число кластеров. Каждая точка присваивается одному из  $k$

кластеров так, чтобы минимизировать среднеквадратичное евклидово расстояние (или другую выбранную метрику) между точками внутри кластера в координатах главных компонент. Затем центр масс для каждого кластера вычисляется заново, и процедура повторяется до тех пор, пока целевая метрика не прекратит изменяться [38].

PCA с кластеризацией методом  $k$ -средних был применен для генерации ансамбля в веб-сервере DINC-COVID, разработанном во время пандемии COVID-19 для виртуального скрининга противокоронавирусных лекарственных средств [39]. Сервер позволял проводить докинг в сформированные заранее ансамбли структур главной протеазы, папаин-подобной протеазы и РНК-зависимой РНК-полимеразы. Для каждой структуры предлагали три ансамбля: два из конформаций белка, полученных из траекторий молекулярной динамики с двумя разными силовыми полями (CHARMM36 и GROMOS53a6), а третий – из кристаллографических структур, размещенных в базе данных RCSB PDB. Структуры кластеризовали методом  $k$ -средних для результатов PCA расстояний между ближайшими парами тяжелых атомов аминокислотных остатков активного центра. Структуры, наиболее близкие центроидам кластеров, помещали в ансамбль. Для библиотеки из 29 молекул с известной активностью авторы провели анализ ROC-кривых и сравнили площади под ними (AUC). Наибольшая ROC AUC и ROC-кривые, характеризующиеся ранним обогащением, были получены для ансамбля, сформированного из кристаллографических структур.

В качестве источника данных для анализа главных компонент могут служить и сами координаты атомов. Например, координаты тяжелых атомов остатков, формирующих активный центр, были использованы для сравнения конформаций, полученных из траекторий МД, с конформациями кристаллографических структур, содержащих лиганд в активном центре [33].

При отборе структур в ансамбль необходимо найти компромисс между степенью подвижности белка, учитываемой в процессе докинга, и размером ансамбля. Увеличение размера ансамбля увеличивает вычислительную сложность виртуального скрининга. Было показано, что разрешающая способность и обогащение результатов виртуального скрининга могут уменьшаться с ростом размера ансамбля [25].

Для отбора ансамбля оптимального размера в ряде работ проводят предварительный виртуальный скрининг небольшой библиотеки соединений во все рассматриваемые структуры, рассчитывают показатели обогащения (ROC AUC [40], разность средних оценок для активных и неактивных соединений [41], корреляцию между экспериментальной и спрогнозированной аффинностью связывания [42]) и выбирают структуры, дающие наибольшее обогащение результатов виртуального скрининга.

Однако отбор структур с лучшими метриками обогащения результатов виртуального

скрининга не гарантирует синергического эффекта ансамбля. Для составления ансамбля, который демонстрировал бы более высокие значения метрик обогащения, чем составляющие его структуры, можно использовать комбинаторный перебор [43]. Этот подход вычислительно затратен, так как если  $N$  – общее количество рассматриваемых конформаций белка, то необходимо оценить  $2N - 1$  ансамблей. Там же был предложен “медленный эвристический” метод, в котором конформации добавляют в ансамбль итеративно, начиная со структуры, для которой наблюдается самая высокая разрешающая способность (качество ранжирования) при докинге в одну структуру. На каждой итерации все конформации добавляются в ансамбль по очереди и вычисляются целевые метрики. Конформацию, при которой была получена самая высокая доля истинноположительных результатов ВС в ансамбле, оставляют в нем, и продолжают процесс, пока не будет определен ансамбль с оптимальными значениями целевых метрик. В этом случае необходимо оценить  $N(N+1)/2$  ансамблей. Для отбора ансамбля “быстрым эвристическим” методом итеративно добавляют структуры в порядке уменьшения обогащения результатов виртуального скрининга, и на каждой итерации рассчитывают метрики обогащения по ансамблю. Это наиболее быстрый из рассмотренных авторами методов – он требует  $2N - 1$  итераций. На примере небольших наборов структур (11 и 6 конформаций из траекторий молекулярной динамики андрогенового рецептора и циклинзависимой киназы-2, соответственно, и 12 кристаллических структур рецептора, активируемого пролифератором пероксисом  $\delta$  (PPAR- $\delta$ ), было показано, что ансамбли, собранные “медленным эвристическим” методом, имеют такие же высокие показатели обогащения, как ансамбли, отобранные полным перебором. При этом ансамбли имеют небольшой размер – 2-4 структуры, и отбор “быстрым эвристическим” методом приводит к ансамблям меньшего размера, чем отбор “медленным эвристическим” методом или полным перебором.

### **1.1.3. Консенсусная оценка результатов виртуального скрининга**

Консенсусная оценка может быть основана на консенсусных метриках (в том числе использующих машинное обучение), консенсусном голосовании и “композитных” оценочных сеток для докинга.

Минимальная оценка по ансамблю является простой и легко интерпретируемой консенсусной метрикой. Она была использована, например, в упомянутых ранее работах по поиску потенциальных лекарств против COVID-19 [26], [34], [39]. Также могут использоваться схожие метрики: максимальная оценка, средняя оценка, медиана, корень из суммы квадратов оценок, среднее гармоническое или геометрическое среднее по ансамблю [44]. В результате сравнения метрик обогащения (ROC AUC, EF – Enrichment Factor, BEDROC [45]) для пяти

мишеней (янус-киназы 1 и 2, серотониновый рецептор 5-HT<sub>6</sub>, альдозоредуктаза 2 и эстрогеновый рецептор) было показано, что наибольшее обогащение результатов виртуального скрининга для всех мишеней достигается при использовании гармонического или геометрического среднего для ранжирования молекул, в то время как самые низкие значения ROC AUC и фактора обогащения (EF) достигались при использовании максимальной оценки по ансамблю. Минимальная оценка оказалась эффективной не для всех мишеней – так, для серотонинового рецептора 5-HT<sub>6</sub> результаты ВС оказались наименее обогащены при использовании минимальной оценки, в то время как для янус-киназы 1 и альдозоредуктазы 2 результаты виртуального скрининга с использованием минимальной оценки были сравнимы по обогащению с результатами, полученными с использованием гармонического и геометрического среднего.

Развитием этих оценок можно считать взвешенное среднее и метрики, основанные на машинном обучении. На примере 21 киназы было показано, что машинное обучение может улучшить обогащение результатов ансамблевого докинга [46]. Соединения из базы данных DUD-E [47] докирировали в ансамбли кристаллографических структур, и из оценок докинга в каждую индивидуальную структуру формировали векторы, на которых проводили обучение четырех отдельных классификационных моделей: *k*-ближайших соседей, логистической регрессии, случайного леса и SVM (метод опорных векторов, support vector machine). Авторы показали, что значения ROC AUC увеличиваются при использовании прогнозов классификационных моделей, обученных на оценках докинга, для результатов виртуального скрининга по сравнению с индивидуальными оценками докинга или наивного байесовского классификатора. Кроме того, была обнаружена зависимость разрешающей способности классификации от количества структур мишени в ансамбле: для *k*-ближайших соседей, случайного леса и SVM значения ROC AUC росли с увеличением количества структур.

Отдельно стоит отметить подход к оценке результатов докинга, основанный на сумме рангов. При этом результаты виртуального скрининга вначале ранжируют на основании индивидуальных оценок, а затем для каждой молекулы вычисляют сумму рангов, на основании которой повторно проводят ранжирование [48].

В подходе консенсусного голосования хитом ВС назначается та молекула, для которой для всех элементов ансамбля или их заранее определенной доли оценка превышает установленный порог. Например, такой подход был применен на финальной стадии ВС в работах [49], [27] (см. далее).

Другим способом учета влияния различных конформаций являются “композитные” сеточные поля для докинга. С их помощью общая оценка по ансамблю не является функцией от

индивидуальных оценок докинга, вместо этого для докинга малых молекул используется взвешенная усредненная оценочная функция, построенная на ансамбле структур. Такая методика была реализована с использованием программы DOCK3.5 [50]. Были предложены два метода построения “композитных” сеточных полей: “энергетически-взвешенный” и “геометрически-взвешенный”. “Энергетический” метод предполагает расчет энергии Ван-дер-Ваальсовых и кулоновских взаимодействий для каждого атома макромолекулы в каждом комплексе белок-лиганд с последующим вычислением взвешенного среднего. Веса энергий назначаются так, чтобы при наличии разнонаправленных взаимодействий те из них, что приводят к усилению связывания, вносили больший вклад, чем взаимодействия, приводящие к его ослаблению. “Геометрический” метод основан на вычислении среднего положения для каждого атома. Более высокая точность методики по сравнению с докингом в одну структуру в задачах воспроизведения кристаллографической конформации и виртуального скрининга была показана на примере протеазы ВИЧ и белка p21 (ингибитора циклин-зависимой киназы A1). Точность воспроизведения способа связывания лиганда при докинге с использованием “композитного” сеточного поля не превосходит точность при докинге в индивидуальную “нативную” структуру (из которой был извлечен лиганд). Однако точность ансамблевого докинга с “композитным” сеточным полем выше точности докинга в индивидуальную не-“нативную” структуру. Кроме того, скорость докинга с использованием “композитного” сеточного поля значительно превосходит скорость докинга в набор индивидуальных структур.

Аналогичный метод был применен в программе AutoDock 3 [51] и протестирован на примере протеазы ВИЧ-1. Авторы развили предложенный ранее “энергетически-взвешенный” метод, в котором веса зависели от значения энергии (0,0001, если энергия больше порогового значения, и 1, если энергия меньше порогового значения). Они предложили назначать веса в соответствии с формулой:

$$w = e^{\frac{\Delta E}{RT}}, \quad (1)$$

где  $\Delta E$  – изменение энергии,  $T$  – температура,  $R$  – универсальная газовая постоянная. Докинг с использованием “энергетически-взвешенных” сеточных полей оказался более точным в задаче воспроизведения конформации и энергии связывания, чем докинг с использованием усредненных сеточных полей и полей, в которых для каждого узла выбирали минимальное значение энергии.

## 1.2. Главная протеаза коронавирусов как терапевтическая мишень

Вирус SARS-CoV-2 относится к роду *Betacoronavirus* подсемейства *Orthocoronavirinae* семейства *Coronaviridae*. Своё название семейство получило благодаря характерным

шиповидным гликопротеинам (spike glycoprotein, белок S) на поверхности оболочки вирионов [52]. Коронавирусы вызывают респираторные и кишечные заболевания у млекопитающих и птиц [53] и могут передаваться от животного к человеку, причем самым многочисленным природным резервуаром коронавирусов являются летучие мыши [54].

SARS-CoV-2 передается воздушно-капельным путем, средний инкубационный период заболевания составляет 4–5 дней. В некоторых случаях инфекция протекает бессимптомно, но у большинства пациентов наблюдаются легкие или умеренные симптомы респираторного заболевания: кашель, лихорадка, головная боль, миалгия и диарея. Очень частым симптомом инфекции COVID-19 является аносмия – потеря обоняния, связанная с поражением поддерживающих клеток обонятельной слизистой оболочки носа [55], [56].

Наиболее частым симптомом тяжелого заболевания является одышка, вызванная пониженным содержанием кислорода в крови [57]. Тяжелое протекание болезни приводит к острому респираторному дистресс-синдрому, характеризующемуся воспалением, нарушением целостности сосудистых стенок в легких и, как следствие, потерей аэрированной легочной ткани. Тяжелая форма COVID-19 может также привести к поражению желудочно-кишечного тракта, острому поражению сердца, почек и печени, сердечным аритмиям, разрушению скелетной мускулатуры, нарушению свертываемости крови и шоку [58]. Риски развития тяжелой формы заболевания зависят от варианта вируса, возраста пациента и сопутствующих заболеваний: смертность от альфа-варианта SARS-CoV-2 в 2,7 раз выше, чем от дельта-варианта [59]. С увеличением возраста пациента на год риск госпитализации увеличивается на 3,4%, а риск смертельного исхода после госпитализации – на 7,4% [60].

### **1.2.1. Репликативный цикл коронавирусов**

На начальном этапе инфекции (Рисунок 2) происходит специфическое связывание белка S коронавируса с рецепторами клетки. В случае SARS-CoV-2 рецептором является ангиотензинпревращающий фермент 2 (ACE2), однако другие коронавирусы используют другие рецепторы – дипептидилпептидазу 4 (MERS-CoV) или аминопептидазу N (HCoV-229E) [61]. Вирус связывается с ACE2 на поверхности клетки-мишени при помощи домена S1 белка S. Затем белок S расщепляется трансмембранный сериновой протеазой TMPRSS2 [62]. Это активирует субъединицу S2, обеспечивающую конформационные перестройки клеточной мембраны и слияние липидных бислоев вириона и клетки хозяина, в результате чего вирусный рибонуклеокапсид высвобождается в цитоплазму.

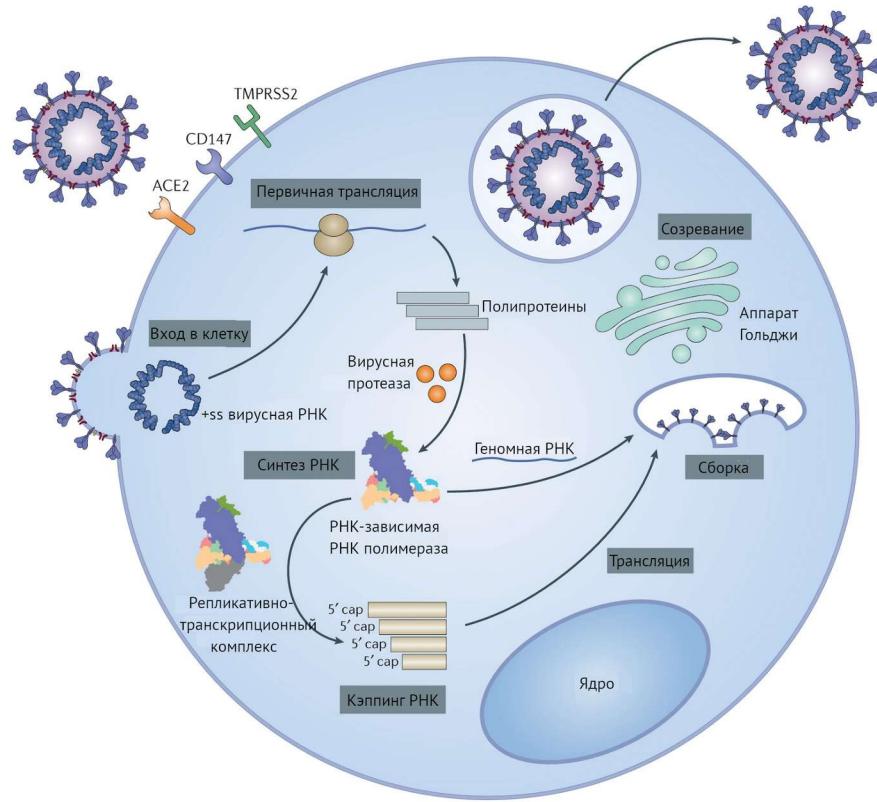


Рисунок 2. Репликативный цикл коронавирусов [63].

Трансляция генома коронавирусов происходит в цитоплазме [64]. Геном коронавирусов представлен одноцепочечной РНК положительной полярности длиной 26–32 kb [53; 65] (Рисунок 3). 5'-Конец содержит кэп-структуру, облегчающую трансляцию вирусных белков, обеспечивающую защиту от экзонуклеаз и затрудняющую обнаружение вирусной РНК иммунной системой организма хозяина [66]. В геноме закодированы вирусные полипротеины и несколько открытых рамок считывания, из которых в ходе процессинга получаются неструктурные белки [61]. Ген вирусного полипротеина имеет две открытые рамки считывания: 1а и 1b. При трансляции рибосомами зараженной клетки на последовательности UUUAAAC может происходить сдвиг открытых рамок считывания, таким образом, транслируются полипротеины pp1a и pp1ab длиной 4405 и 7096 аминокислотных остатков соответственно (Рисунок 3). Полипротеин pp1a транслируется примерно в 1,5-2 раза больше, чем pp1ab [67]. Также в геноме вируса в других открытых рамках считывания закодированы 4 структурных белка (S, E (Envelope), M (Membrane), N (Nucleocapsid)) и некоторое число, в зависимости от вида, вспомогательных белков в отдельных открытых рамках считывания.

Процессинг полипротеинов коронавирусов осуществляется двумя вирусными протеазами. Папаин-подобная протеаза (PLpro, nsp3) катализирует протеолиз в трех сайтах, в результате чего высвобождаются неструктурные белки (nsp) 1 и 2, а также сама PLpro [68]. Главная протеаза (3С-подобная протеаза, Mpro, 3CLpro, nsp5) тоже катализирует свой автопроцессинг, а

также процессинг всех остальных неструктурных белков (Рисунок 3). Неструктурные белки коронавирусов (Таблица 1) формируют репликативно-транскрипционный комплекс (replication and transcription complex, RTC).

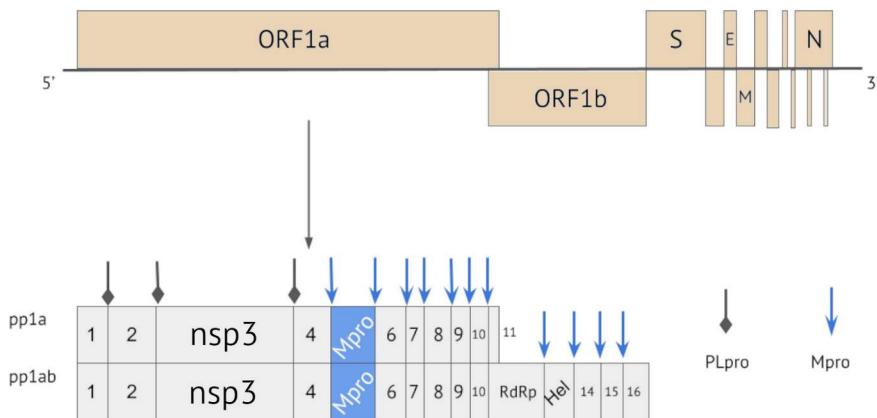


Рисунок 3. Схема генома, трансляции и процессинга вирусными протеазами PLpro и Mpro коронавирусных полипротеинов.

Таблица 1. Неструктурные белки коронавирусов и их роли

Название	Функция	Размер, а\к	Ссылка
nsp1	Связывается с 40S-субъединицей рибосом, ингибируя трансляцию белков инфицированной клетки, способствует деградации мРНК зараженной клетки	180	[69]
nsp2	Функция неизвестна	638	
nsp3/PLpro	Протеолиз вирусных полипротеинов в 3 сайтах, связывается с вирусной РНК. За счет де-моно-АДФ-рибозилирования и де-поли-АДФ-рибозилирования блокирует иммунный ответ	1922	[68]
nsp4	Реструктуризация мембран клетки, стабилизация репликативных комплексов (связывается с PLpro)	500	[70]
nsp5/Mpro/3CLpro	Протеолиз вирусных полипротеинов	306	
nsp6	Реструктуризация мембран клетки, стабилизация репликативных комплексов	290	[70]
nsp7	Фактор процессивности РНК-зависимой РНК-полимеразы, компонент репликативного комплекса	83	[71]
nsp8	Кофактор РНК-зависимой РНК-полимеразы, катализ синтеза РНК-праймеров, компонент репликативного комплекса	198	[71]
nsp9	Ингибитор уридинмонофосфат-трансферазного домена nsp12, компонент репликативного комплекса	113	[72]
nsp10	Кофактор экзонуклеазы nsp14 и метилтрансферазы nsp16; регулирует сдвиг рамки считываивания в рибосомах	139	[73],[74], [75]

Продолжение Таблицы 1.

Название	Функция	Размер, а\к	Ссылка
nsp11	Функция неизвестна	13	
nsp12/RdRp	РНК-зависимая РНК-полимераза	932	[71]
nsp13/HeI	РНК-хеликаза, содержащая цинк-связывающий домен, РНК-5'-фосфатаза	601	[76], [77]
nsp14	3'-5'-экзонуклеаза, N7-метилтрансфераза	527	[78],[79]
nsp15	Уридилат-специфичная эндорибонуклеаза, укорачивает полиуридиновые последовательности на 5'-конце антисмысловой последовательности вирусной РНК	346	[80], [81]
nsp16	2'-О-метилтрансфераза	298	[82]

В ходе репликации вирусного генома происходит синтез полноразмерных антисмысловых геномных копий, которые функционируют как матрицы для создания новой геномной РНК. Синтез РНК осуществляется полимеразным комплексом: РНК-зависимой РНК-полимеразой nsp12 (RdRp) в комплексе с двумя ее кофакторами nsp7 и nsp8 [71]. Также в результате прерывистой транскрипции синтезируются фрагменты антисмысловой РНК, выступающие в качестве матриц для синтеза субгеномных мРНК, которые транслируются в структурные и вспомогательные белки [83]. Проверка корректности репликации генома производится экзонуклеазой nsp14 [78]: в случае обнаружения некомplementарной пары оснований хеликаза nsp13 стимулирует обратное прокручивание синтезированной цепи РНК для исправления ошибки [77].

После синтеза вирусной РНК происходит ее кэпирование при помощи последовательного действия ферментов nsp13 [76], nsp12 [84], nsp14 [79] и nsp16 [85],[82]. Структурный белок N связывается с новыми полноразмерными кэпированными РНК, формируя рибонуклеопротеин. Связываясь со структурными белками E, M и S, они образуют нуклеокапсид. Созревание новых вирионов происходит по механизму почкования в промежуточный компартмент эндоплазматического ретикулума и аппарата Гольджи (endoplasmic reticulum – Golgi intermediate compartment, ERGIC) [86]. Вирионы выводятся из инфицированных клеток путем транспорта к плазматической мембране в гладкостенных везикулах и экзоцитоза [87] или, согласно другим исследованиям [88], посредством лизосомного транспорта (Рисунок 2).

### 1.2.2. Структура главной протеазы коронавирусов

Главная протеаза коронавирусов активна в виде гомодимера (Рисунок 4), так как амидный атом азота Ser1 N-конца каждого из двух протомеров образует водородные связи с карбоксильной группой остатка Glu166 и карбонильным атомом кислорода основной цепи остатка Phe140 другого протомера и тем самым участвует в формировании области связывания

субстрата [89]. Phe140 образует стекинг-взаимодействие с His163, из-за чего, предположительно, остаток гистидина остается незаряженным при физиологических значениях рН и может формировать водородные связи с субстратом через  $\epsilon$ -NH [90]. Каждый протомер состоит из 306 аминокислотных остатков, имеет молекулярный вес 33,8 кДа. В структуре протомера выделяют три домена (Рисунок 4): домен I (химотрипсин-подобный, остатки 8–101) и домен II (подобный 3C-протеазе пикорнавирусов [91], остатки 102–184) имеют структуру антипараллельной  $\beta$ -бочки, между ними расположен активный центр; домен III (остатки 201–303) содержит пять  $\alpha$ -спиралей, объединенных в антипараллельные глобулы, и связан с доменом II длинной петлей (остатки 185–200) [92]. Домен III обеспечивает димеризацию Mpro за счет образования солевого мостика между Glu290 одной субъединицы и Arg4 другой, а также взаимодействия Glu288, Asp289 и Glu290 с пятью первыми остатками N-конца другого протомера [93].

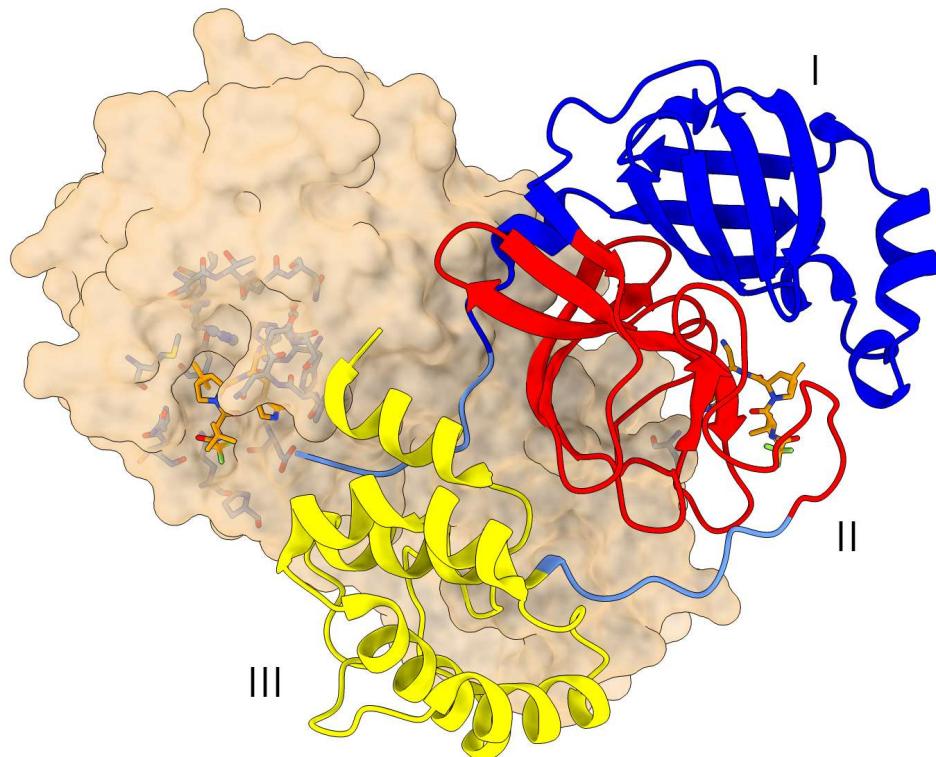


Рисунок 4. Структура главной протеазы SARS-CoV-2 с молекулой ковалентного ингибитора нирматрелвира в активном центре (PDB ID: 7TLL [94]). Окраска выполнена по доменам: I – синий, II – красный, III – желтый, линкеры – голубой [95]. Для второго протомера показана поверхность.

Mpro достаточно консервативна у разных коронавирусов. Mpro SARS-CoV-2 имеет 96% идентичности аминокислотных последовательностей с Mpro SARS-CoV, 51% – MERS-CoV и одинаковую с ними третичную структуру [96]. Методом ConSurf было показано, что у главных протеаз SARS-CoV-2, SARS-CoV, MERS-CoV, вируса инфекционного бронхита птиц (Infectious bronchitis virus, IBV), вируса эпидемической диареи свиней (Porcine epidemic diarrhea virus,

PEDV) и коронавируса летучих мешей (Bat-CoV) наиболее консервативны остатки, образующие активный центр и сайт димеризации (Рисунок 5) [97]. Для Mpro SARS-CoV-2 была описана только одна закрепившаяся мутация в варианте Омикрон (Pro132His), не оказавшая влияние на каталитические свойства белка и его чувствительность к известным ингибиторам, однако понизившая термическую стабильность фермента [98].

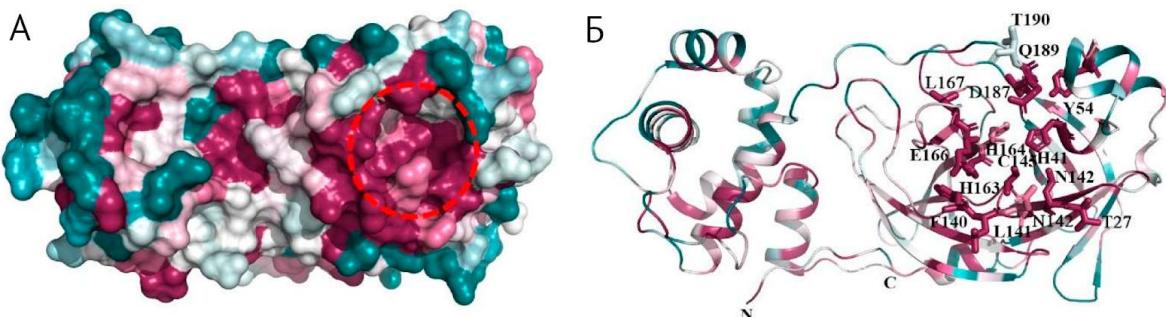


Рисунок 5. Анализ консервативности протеазы Mpro у разных коронавирусов методом ConSurf. Наиболее консервативные остатки окрашены малиновым, наименее консервативные – бирюзовым [97]. А – отображена поверхность мономера, Б – структура показана в виде ленты, отображены а\к остатки, формирующие активный центр.

Главная протеаза коронавирусов катализирует протеолиз полипротеина в 11 сайтах. Субстратная специфичность определяется последовательностью X-X-Leu-Gln↓(Ser/Ala/Gly) (положения остатков обозначаются P4-P3-P2-P1↓P1’), где ↓ соответствует расщепляемой пептидной связи (Рисунок 6А) [91]. Карман S1' – неглубокий карман, в котором в естественном субстрате находятся небольшие аминокислотные остатки – серин, аланин или глицин в положении P1'. Он образован боковыми цепями Thr25, Leu27, His41 и Cys145, а также атомами основной цепи Thr26 и Cys145. Карман S1 специфично принимает остаток глутамина в положении P1 и состоит из боковых цепей остатков Phe140, Asn142, Ser144, His163, Glu166 и His172, а также атомов основной цепи Phe140, Leu141, Asn142, Met165 и Ser1 соседнего протомера. Глубокий гидрофобный карман S2 специчен для лейцина в положении P2 и состоит из боковых цепей His41, Met49, Тир54 и Met165 и алкильной части боковой цепи Asp187. Карманы S3 и S4 доступны для растворителя и образованы боковыми цепями остатков Met165, Leu167, Phe185 и Gln192 и атомами основной цепи Gln189 [99].

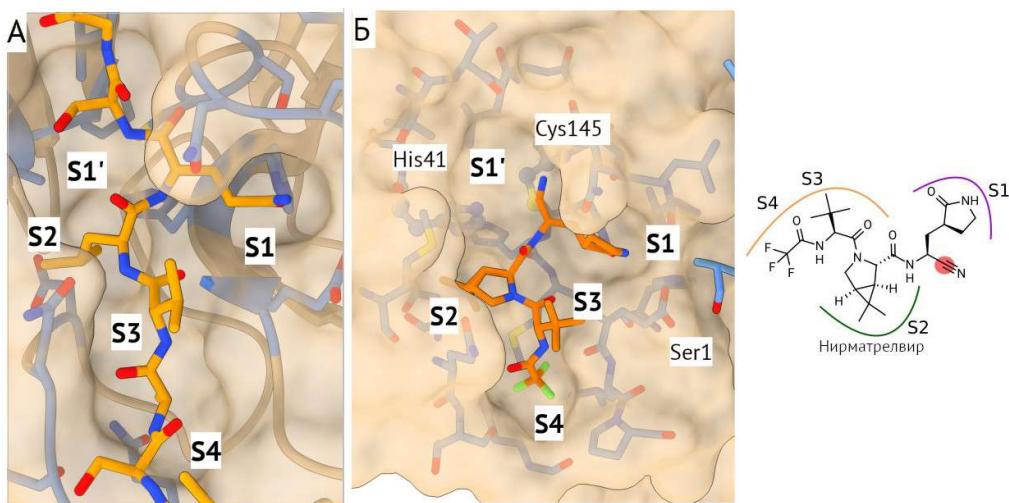


Рисунок 6. Структура активного центра Mpro SARS-CoV-2 в комплексе с пептидным субстратом (А, PDB ID: 7DVP [99]) и нирматрелвиром (Б, PDB ID: 7TLL [94]) [95].

Каталитическая диада протеазы представлена аминокислотными остатками Cys145 и His41 (Рисунок 7, I) [91]. На первой стадии катализа происходит перенос иона водорода с Cys145 на His41 с образованием пары ионов тиолата и имидазолия (Рисунок 7, II). Затем депротонированный Cys145 нуклеофильно атакует карбонильный атом углерода пептидной связи субстрата (Рисунок 7, II). Протон от протонированного His41 переносится на атом азота расщепляемой пептидной связи, высвобождая фрагмент субстрата R2 (Рисунок 7, III-IV). На финальном этапе происходит деацетилирование Cys145 под действием молекулы воды, активированной His41 (Рисунок 7, V), для восстановления изначального состояния фермента (Рисунок 7, I) и высвобождения фрагмента субстрата R1 [100]. Для инициирования каталитической реакции Cys145 и His41 должны находиться на расстоянии менее 4 Å и в конфигурации, в которой донор и акцептор водородной связи обращены друг к другу [101].

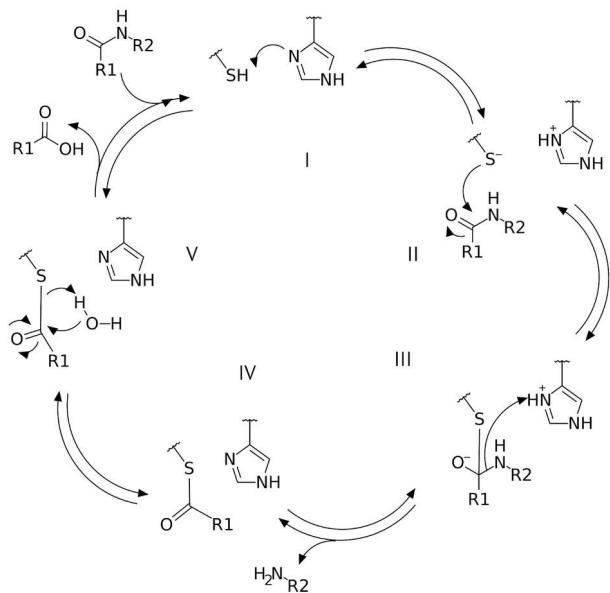


Рисунок 7. Каталитический цикл протеазы Mpro коронавирусов.

### 1.2.3. Ингибиторы главной протеазы SARS-CoV-2

Главная протеаза коронавирусов является одной из наиболее хорошо изученных мишеней для разработки лекарств против COVID-19. Более половины исследований направлены на разработку ингибиторов, ковалентно связывающихся с Cys145 [102]. Ковалентные ингибиторы обычно имеют более высокую эффективность по сравнению с нековалентными ингибиторами, действующими на ту же мишень, что позволяет снизить дозировку лекарства и увеличить промежутки между введениями [103].

Ковалентные ингибиторы главной протеазы коронавирусов должны соответствовать основному структурному требованию: для ковалентного связывания с Cys145 соединение должно содержать в положении P1' электрофильную активную группу – альдегид, активированный кетон,  $\alpha$ -кетоамид, сложный эфир, акцептор Михаэля, электрофильный атом серы или селена, или нитрил (Рисунок 8) [103].

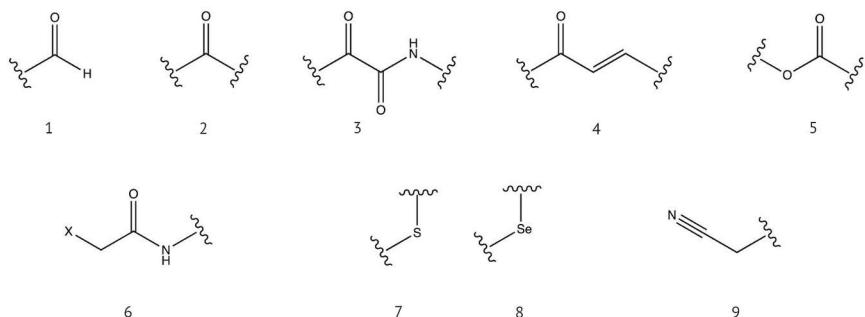


Рисунок 8. Примеры электрофильных активных групп. 1 – альдегид, 2 – кетон, 3 –  $\alpha$ -кетоамид, 4 – акцептор Михаэля, 5 – сложный эфир, 6 –  $\alpha$ -галогенамид, 7 – сера, активированная акцепторными группами, 8 – активированный селен, 9 – нитрил.

Фрагмент P1, обычно представленный  $\gamma$ -лактамным циклом – аналогом глутамина, – должен быть способен формировать водородные связи и гидрофобные взаимодействия с аминокислотными остатками кармана S1 – His163, Glu166 и His172; фрагменты P2 и P3 должны быть гидрофобными и связываться в карманах S2 и S3/S4 [104]. Сам остаток глутамина не может быть использован в качестве заместителя, так как амидная группа боковой цепи может реагировать с ковалентной активной группой ингибитора с образованием циклического продукта (Рисунок 9).  $\gamma$ -Лактамный цикл повышает жесткость фрагмента по сравнению с гибким остатком глутамина, что, вероятно, обеспечивает снижение потери энтропии при связывании с мишенью [105].

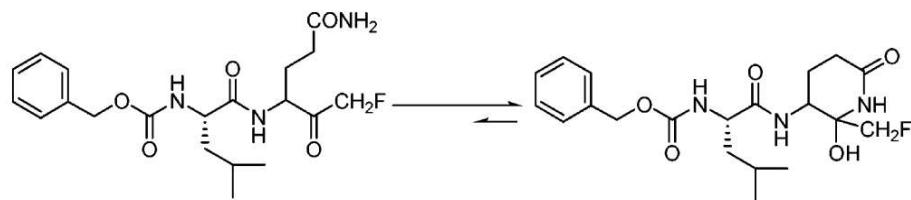


Рисунок 9. Циклизация остатка глутамина с активированной карбонильной группой [105]

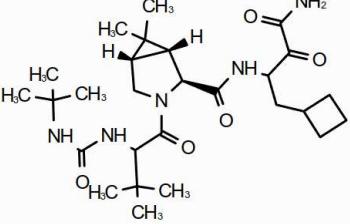
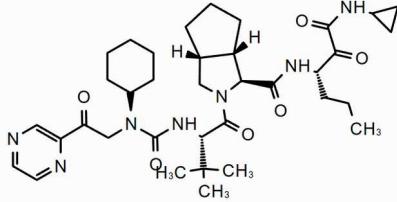
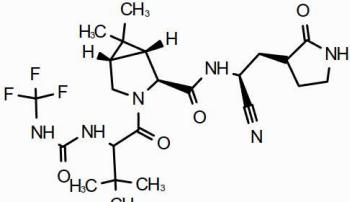
### 1.2.3.1. Ковалентные пептидомиметики

Вследствие высокой консервативности Mpro [106] первыми клиническими кандидатами в ходе пандемии COVID-19 стали ингибиторы Mpro других коронавирусов. GC-376 (1, Таблица 2) – ветеринарное лекарство от инфекционного перитонита кошачьих [107]. Это ковалентный пептидомиметик с сульфитной активной группой. В положении Р1 находится  $\gamma$ -лактамный цикл, в положении Р2 – остаток лейцина и в положении Р3 – бензильная группа [108].

Таблица 2. Ковалентные пептидомиметики – ингибиторы главной протеазы SARS-CoV-2

№	Название	Структура	$K_i$ , нМ	$IC_{50}$ Mpro SARS-CoV-2, нМ	$EC_{50}$ SARS-CoV-2, нМ	Ссылка
1	GC-376		59,9	30	3370 (штамм USA-WA1/2 020)	[108]
2	PF-00835231		0,271	8	40000	[109]
3	PF-07304814		174	-	86700	[109]

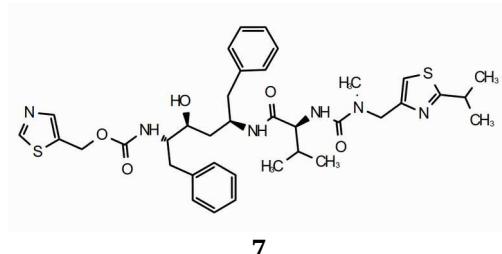
Продолжение Таблицы 2

№	Название	Структура	$K_i, \text{ нМ}$	$IC_{50} \text{ Mpro SARS-CoV-2, нМ}$	$EC_{50} \text{ SARS-CoV-2, нМ}$	Ссылка
4	Боцепревир		1180	1590	1900	[108]
5	Телапревир		-	18700	>50000	[110]
6	Нирматрелвир		3,11	23	4480	[111], [112]

PF-00835231 (2) [113] был разработан как ингибитор Mpro SARS-CoV с внутривенной формой введения, однако из-за окончания вспышки в 2005 году для него в тот момент не были проведены клинические испытания. Молекула содержит  $\alpha$ -гидроксиметилкетонную активную группу,  $\gamma$ -лактамный аналог глутамина в положении P1, остаток лейцина в положении P2 и замещенный индол-2-ил в положении P3. Для увеличения растворимости гидроксильную группу фосфорилируют, получая пролекарство PF-07304814 (3) [109]. Это соединение имеет гораздо более низкую по сравнению с другими ковалентными ингибиторами Mpro константу ингибирования, однако недостаточная растворимость и биодоступность препятствуют его широкому применению.

В ходе работ по репозиционированию лекарств для COVID-19 было обнаружено, что ингибиторы протеазы вируса гепатита С – боцепревир (4) и телапревир (5) – также ингибируют Mpro SARS-CoV-2 [108]. В кристаллографических структурах Mpro с этими молекулами в активном центре 6,6-диметил-3-азабицикло[3.1.0]гексановый фрагмент боцепревира и (3aR,6aS)-октагидроцикlopента[с]пиррольный фрагмент телапревира занимают карман S2. Это показало возможность использования таких структурных фрагментов для увеличения жесткости потенциальных ингибиторов главной протеазы SARS-CoV-2.

Первый ингибитор главной протеазы коронавирусов, разрешенный для экстренного медицинского применения Управлением по контролю за продуктами и лекарствами США (FDA), нирматрелвир (**6**), также относится к классу ковалентных пептидомиметиков. Он имеет нитрильную активную группу,  $\gamma$ -лактамный аналог глутамина в положении P1, жесткий циклический аналог лейцина в положении P2, трет-бутильную группу в положении P3 и трифторметил в положении P4 (Рисунок 6Б). Важным преимуществом нирматрелвира является его пероральный способ введения. Пероральную биодоступность удалось повысить по сравнению с PF-00835231 благодаря замене реакционноспособной группы –  $\alpha$ -гидроксиметилкетона – на нитрил [111]. Комбинация нирматрелвира с ритонавиром (**7**) зарегистрирована FDA под торговой маркой Paxlovid. Ритонавир ингибитирует цитохром CYP3A4, препятствуя быстрому метаболизму пептидомиметиков – ингибиторов протеаз [114]. В рандомизированном двойном слепом клиническом испытании с участием 2246 пациентов терапия паксловидом в течение 5 дней после проявления симптомов COVID-19 уменьшила риск развития тяжелой формы COVID-19 на 89% по сравнению с плацебо [115].



7

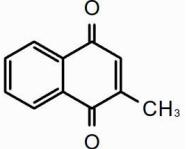
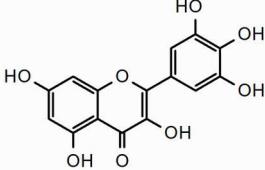
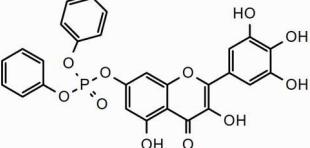
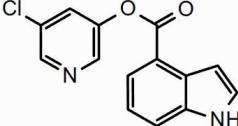
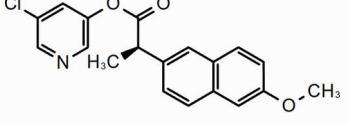
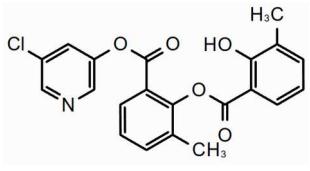
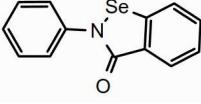
### 1.2.3.2. Ковалентные ингибиторы непептидной природы

В ходе исследований по репозиционированию соединений растительного происхождения было показано, что витамин К3 (**8**, Таблица 3) способен ковалентно ингибировать Mpro, присоединясь по Михаэлю к Cys145. Оптимизация структуры витамина позволила десятикратно увеличить активность. При этом при ковалентном докинге витамина К3, связанного с Cys145 по положению 3, бициклический центральный фрагмент располагается в кармане S1.

Многие исследования *in silico* и *in vitro* были посвящены флавоноидам как потенциальным ковалентным ингибиторам Mpro из-за наличия акцептора Михаэля в их структуре. Например, было обнаружено, что мирицетин (**9**) и его производные (**10**) ингибируют Mpro SARS-CoV-2 и подавляют репродукцию вируса, а также снижают воспаление [121]. Связывание мирицетина с цистеином в активном центре главной протеазы было подтверждено кристаллографическими данными (PDB ID: 7B3E) [122]. Молекула занимает карманы S1' и S2, при этом атом в свободном орто-положении тригидроксифенильной группы мирицетина ковалентно связан с Cys145 активного центра. В отличие от большинства ковалентных ингибиторов Mpro

мирицетин связывается с ферментом необратимо.

Таблица 3. Ковалентные ингибиторы главной протеазы SARS-CoV-2 непептидной природы

№	Название	Структура	IC <sub>50</sub> Mpro SARS-CoV-2, нМ	EC <sub>50</sub> SARS- CoV-2, нМ	Ссылка
8	Витамин К3		7960	-	[116]
9	Мирицетин		630	8000	[117]
10	Производное мирицетина		3130	3150	[117]
11	GRL-0920		250	2800	[118]
12	Производное R- напроксена		160	>100	[119]
13	Производное салициловой кислоты		4900	24000	[119]
14	Эбселен		670	4670	[120]

Другая большая группа ингибиторов Mpro – сложные эфиры с хорошими уходящими группами, которые необратимо ацилируют Cys145, например, хлорпиридиниловые эфиры индолов или индолинов (11) [118],[123],[104], а также производные нестериоидных

противовоспалительных препаратов: салициловой кислоты, ибuproфена и др., и их аналогов [119] (12, 13, Таблица 3).

Эбселен (14) и его изостерический аналог эбсульфур разрабатывались как антиоксиданты и противовоспалительные препараты [124]. В ходе репозиционирования этих соединений было обнаружено, что они и их производные проявляют активность против Mpro SARS-CoV-2, связываются с Cys145 в активном центре и ингибируют репродукцию вируса [120],[125] (Рисунок 10). Кристаллическая структура главной протеазы с эбселеном [PDB ID: 7BFB] демонстрирует неселективное связывание трех молекул лиганда с Cys145 в активном центре, Cys156 в домене II с противоположной активному сайту стороны и с Cys300 между доменами II и III. Также была получена кристаллическая структура главной протеазы с эбселеном, содержащая остатки цистеина, связанные с атомами селена [PDB ID: 7BAK], что позволяет предположить дальнейший гидролиз продукта связывания эбселена с остатком цистеина [120]. Были получены более активные производные эбселена с наномолярными концентрациями полуингибиования, обладающие большей селективностью, чем немодифицированный эбселен [126], [125].

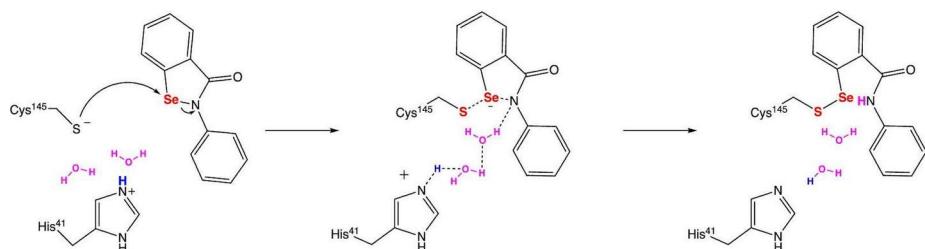


Рисунок 10. Возможный механизм присоединения эбселена к Cys145 в активном центре Mpro [104]

### 1.2.3.3. Нековалентные ингибиторы

Несмотря на высокую эффективность ковалентных ингибиторов, они могут иметь побочные эффекты из-за взаимодействия с нецелевыми белками, в частности, вызывать аутоиммунные реакции [127],[128]. Поэтому существует интерес к разработке нековалентных ингибиторов Mpro SARS-CoV-2. Их способы связывания обычно отличаются от способов связывания ковалентных ингибиторов – чаще всего они занимают карманы S1, S1' и S2 (Рисунок 11). Первым нековалентным ингибитором главной протеазы SARS-CoV-2, получившим одобрение от регулятора [129], является энситрелвир (S-217622, 15, Таблица 4). Разработка включала несколько стадий. Сначала провели крупномасштабный виртуальный скрининг методом молекулярного докинга корпоративной библиотеки соединений, хиты этого скрининга отфильтровали в соответствии с фармакофорной гипотезой и протестировали *in vitro*. Соединения с наибольшей активностью служили стартовой точкой для структурной оптимизации [130]. Метилтриазольный, трифтобензильный и хлор-метил-индазольный

фрагменты энситрелвира занимают карманы S1, S2 и S1', соответственно, и образуют специфические взаимодействия с белком (Рисунок 11А). Атом азота в положении 4 метилтриазольного цикла образует водородную связь с  $\epsilon$ -NH боковой цепи His163. Положение трифтобензильного фрагмента в кармане S2 стабилизируется за счет  $\pi$ -стекинга ароматического цикла с His41. Иминный линкер между центральным и 6-хлор-2-метил-2Н-индазольным фрагментом образует систему водородных связей с боковой цепью His41 через молекулу воды. Центральный триазиновый цикл связывается между карманами S1, S2 и S1' за счет водородной связи между карбонильным атомом кислорода в положении 2 и NH основной цепи Glu166. Активность энситрелвира *in vivo* сравнима с активностью ковалентных ингибиторов, при этом соединение метаболически стабильно и обладает высокой пероральной доступностью [131].

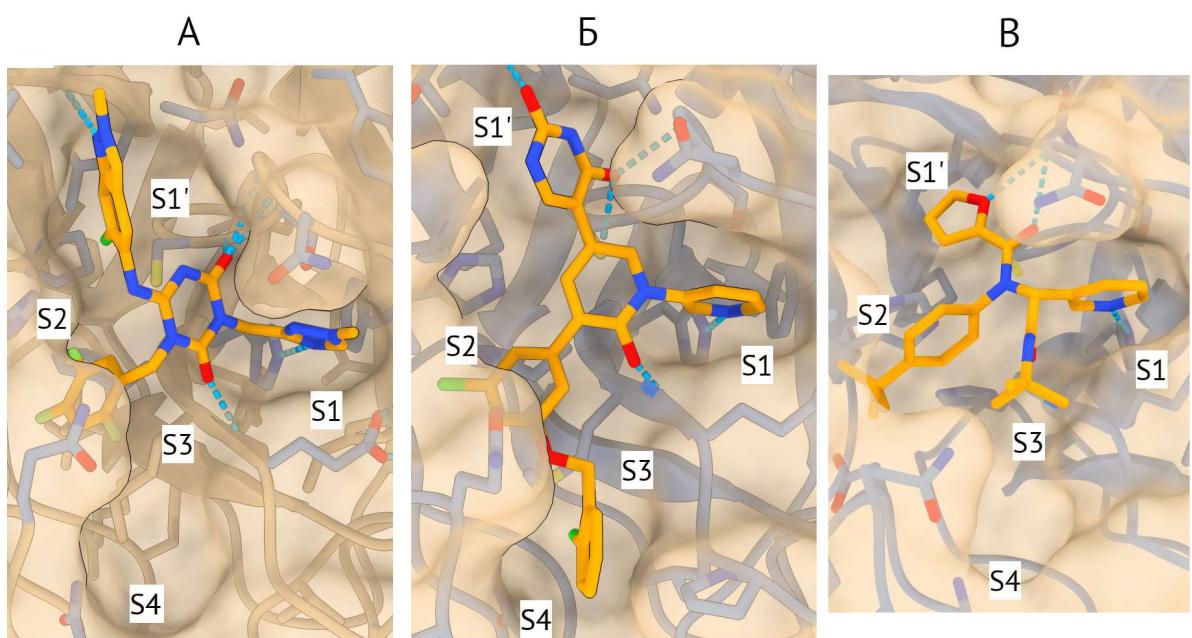


Рисунок 11. Положение энситрелвира (15) [PDB ID: 8DZ0] (А), производного перампанела (16) [PDB ID: 7L13] (Б) и ML-188 (17) [PDB ID: 7L0D] (В) в активном центре Mpro SARS-CoV-2. Водородные связи выделены пунктиром.

Таблица 4. Нековалентные ингибиторы главной протеазы SARS-CoV-2

№	Название	Структура	IC <sub>50</sub> SARS- CoV-2, нМ	Mpro SARS- CoV-2, нМ	EC <sub>50</sub> SARS- CoV-2, нМ	Ссылка
15	Энситрелвир		13	370 (штамм WK- 521)	[130]	
16	Производное перампанела		18	980	[132]	
17	ML-188		2500	-	[13 3, p. 188 ]	

Ингибиторная активность противоэпилептического препарата перампанела в отношении Mpro SARS-CoV-2 была обнаружена в ходе репозиционирования [134]. Работу по его оптимизации начали с докинга молекулы в жесткую структуру Mpro. С помощью метода возмущения свободной энергии (FEP) были разработаны производные с улучшенной энергией связывания (например, **16**). Оптимизированные молекулы имели IC<sub>50</sub> против Mpro SARS-CoV-2 в ферментативном анализе в диапазоне от 0,018 до 10 мкМ [132]. Подобно энситрелвиру, производные перампанела имеют клевероподобную структуру с центральным фрагментом, располагающимся в области между карманами вблизи от Cys145, и заместителями, заполняющими карманы S1, S1', S2 и S3 (Рисунок 11Б).

Нековалентный ингибитор ML-188 (**17**) был обнаружен в ходе высокопроизводительного скрининга ингибиторов главной протеазы SARS-CoV [135]. Молекула представляет собой третичный амин, заместители которого располагаются в карманах S1, S1' и S2 (Рисунок 11Б).

Он проявляет активность против главной протеазы SARS-CoV-2 (Таблица 4) и послужил отправной точкой для разработки других ингибиторов, в том числе, ковалентных [136], в которых вместо фурана в положении P1' находится активная электрофильная группа, а также добавлен заместитель в положении P4, способный образовывать специфические взаимодействия в карманах S3 и S4.

### **1.3. Белок NS1 ортофлавивирусов как терапевтическая мишень**

Вирусы рода *Orthoflavivirus* (Рисунок 12) вызывают такие опасные заболевания, как лихорадка денге, лихорадка Зика, клещевой энцефалит и желтая лихорадка. Распространение ортофлавивирусов зависит от ареала обитания их переносчиков (комары и клещи) и природных резервуаров (чаще всего мелких грызунов и птиц) [137]. Расширение ареалов фиксируется в настоящий момент в связи с изменением климата [138], [139]. Ортофлавивирусные инфекции могут протекать бессимптомно или вызывать лихорадки с головной болью, миалгией, артритом и сыпью. В некоторых случаях инфекции могут прогрессировать в серьезные заболевания, сопровождающиеся геморрагической лихорадкой, менингоэнцефалитом и различными неврологическими осложнениями, приводя к снижению качества жизни, инвалидизации и смерти [140], [137]. Ортофлавивирусные инфекции можно в целом разделить на две категории: одни вирусы (вирус денге (ДЕНВ), вирус желтой лихорадки (ВЖЛ)) обычно вызывают системную инфекцию, геморрагическую лихорадку и поражение эндотелия капилляров, а другие (вирус Зика (ЗИКВ), вирус лихорадки Западного Нила (ВЛЗН), вирус клещевого энцефалита (ВКЭ), вирус японского энцефалита (ВЯЭ)) могут приводить к неврологическим осложнениям – менингитам и энцефалитам [141], [142]. Патогенез вирусов определяется не только рецепторами, с которыми взаимодействует вирус при проникновении в клетку [143], но и тем, с какими клетками связывается белок NS1: NS1 ДЕНВ преимущественно связывается с эндотелиальными клетками легочных сосудов, NS1 ВЖЛ связывается в сосудах легких и печени, NS1 ЗИКВ связывается с клетками сосудов мозга [144] и пупочной вены, NS1 ВЛЗН и ВЯЭ вызывают поражения в эндотелиальных клетках сосудов мозга [145]. Связывание белка NS1 с клетками эндотелия сосудов провоцирует запуск воспалительных процессов и разрушение гликопротеинов, протеогликанов и гликозаминогликанов на поверхности клеток [146]. Одновременно с этим на поверхности зараженной клетки белок NS1 ингибирует классический и лектиновый пути комплемента [147], [148], не давая иммунной системе распознать и уничтожить инфицированную клетку.

В настоящее время не существует зарегистрированных препаратов прямого действия против инфекций, вызванных ортофлавивирусами, несмотря на продолжающиеся работы по

поиску таких соединений [149]. Для ряда вирусов (ВКЭ [150], ДЕНВ [151], ВЖЛ [152], ВЯЭ [153]) разработаны вакцины, однако их разработка и широкое применение осложнены тем, что для некоторых ортофлавивирусов (ДЕНВ, ЗИКВ) наблюдается антитело-зависимое усиление инфекции [154], [4]. Кросс-реактивные антитела связывают вирусные частицы в недостаточной концентрации или с недостаточной avidностью, чтобы нейтрализовать их, но облегчают проникновение вируса в клетки, что приводит к таким острым состояниям, как геморрагическая лихорадка или шоковый синдром [155].

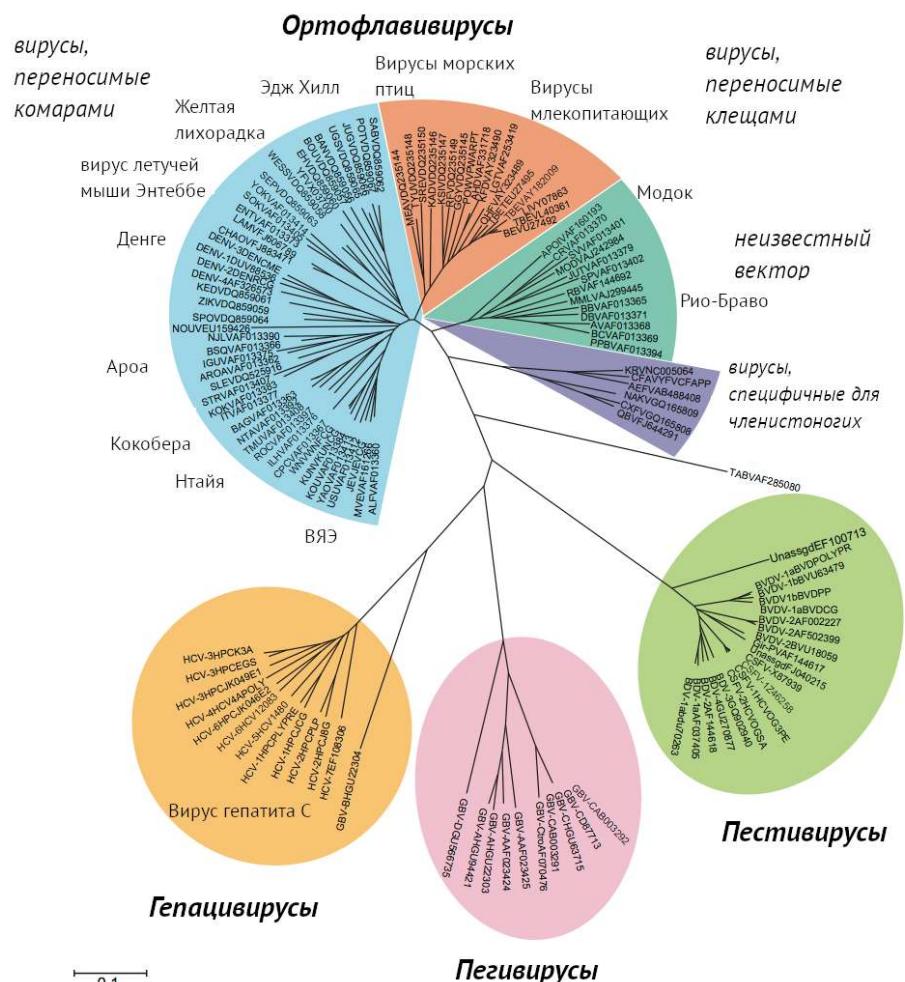


Рисунок 12. Филогенетическое дерево семейства *Flaviviridae* [156]

Ортофлавивирусы – это оболочечные вирусы, геном которых представлен одноцепочечной РНК положительной полярности длиной 10–11 тыс. осн. (Рисунок 13) [157]. Геном состоит из длинной открытой рамки считывания, окруженной 5'- и 3'-нетранслируемыми областями с кэпом на 5'-конце.

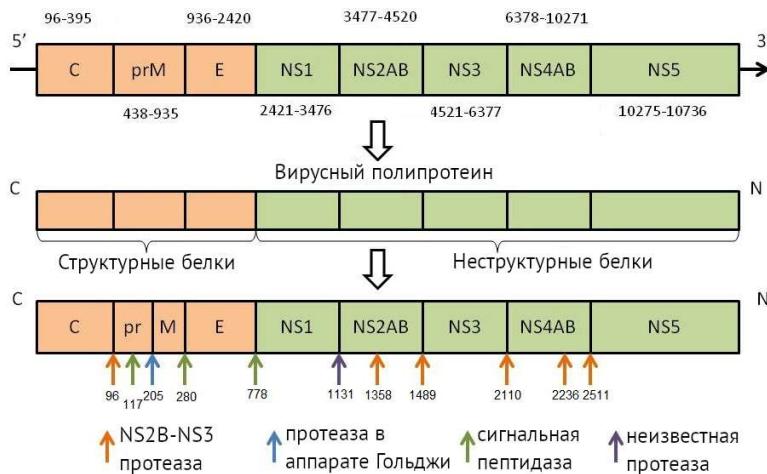


Рисунок 13. Схема генома и полипротеина ортофлавивирусов [158]. Нумерация приведена по аннотированному геному вируса ДЕНВ1 (GenBank: EU848545.1).

Вирусные частицы проникают в клетку путем рецептор-опосредованного эндоцитоза [159]. Под влиянием низкого рН эндосомы происходит изменение конформации белка Е оболочки вируса и высвобождение вирусной РНК в цитоплазму клетки [160]. РНК транслируется в вирусный полипротеин, процессинг которого в структурные и неструктурные белки осуществляется вирусной протеазой NS2B/NS3 и сигнальной пептидазой зараженной клетки (Таблица 5, Рисунок 13) [161]. Трансмембранные вирусные белки NS1, NS2A, NS4A и NS4B формируют карманы в ЭР, в которых концентрируются неструктурные белки, составляющие репликативный комплекс [162]. Вновь синтезированный геном упаковывается вместе с вирусным капсидным белком (С) и окружается липидным бислоем, в который заякорены белки Е и prM, с образованием незрелой вирусной частицы [163]. Созревание вирусной частицы происходит в транс-отделе аппарата Гольджи, где под влиянием низкого рН комплекс prM-Е меняет конформацию и белок prM подвергается протеолизу с высвобождением пептида pr [164]. Зрелые частицы покидают клетку путем эндоцитоза [165].

Белок NS1 ортофлавивирусов – гликопротеин массой 48-50 кДа, состоящий из 352 аминокислотных остатков [158]. В его структуре (Рисунок 14) можно выделить четыре домена:  $\beta$ -бочку (а\к 1-29), связующий домен (а\к 30-37, 152-182),  $\alpha$ -спиральное крыло (а\к 38-151) и  $\beta$ -лестница (а\к 183-352). Третичная структура поддерживается шестью дисульфидными связями [177]. Белок NS1 присутствует в клетке и на ее поверхности в виде димера, а в кровотоке – в виде тетра- и гексамеров (Рисунок 15), в которых  $\beta$ -бочки ориентированы внутрь, а “спагетти-петля” (часть  $\beta$ -лестницы, а\к 219-272) – наружу [168]. Поверхность  $\beta$ -бочки и внутренняя поверхность  $\alpha$ -спирального крыла гидрофобны, вследствие чего центр секретированного гексамиера заполнен липидами [178]. Белок NS1 гликозилирован в двух (в случае ВЖЛ – в трех) сайтах: Asn130, Asn207 (также Asn175 у NS1 ВЖЛ и утиного вируса Тембусу [168], [179]). Гликозилирование важно для увеличения стабильности димерной и гексамерной форм белка и

для взаимодействия с компонентами системы комплемента [180], [181].

Таблица 5. Белки ортофлавивирусов и их функции

Название	Функция	Ссылка
C	Белок нуклеокапсида	[166]
prM	Участвует в формировании вирусных частиц, защищает белок E от рН-индуцированных конформационных изменений в незрелом вирионе	[167]
E	Участвует в формировании вирусных частиц, инициирует вход вируса в клетку, формирует внешнюю оболочку вириона, взаимодействует с антителами	[160]
NS1	Стабилизирует репликативный комплекс, связываясь с белками NS4A и NS4B, участвует в подавлении иммунного ответа, увеличивает вирулентность	[168], [147], [169]
NS2A	Участвует в сборке вириона, компонент репликативного комплекса	[170]
NS2B	Кофактор протеазы NS3	[171]
NS3	Включает два домена. Протеазный домен катализирует процессинг полипротеина, хеликазный домен участвует в работе репликативного комплекса благодаря хеликазной и РНК-трифосфатазной активности	[172], [173], [174]
NS4A	Стабилизирует репликативный комплекс	[162], [175]
NS4B	Стабилизирует репликативный комплекс, ингибит сигнальный путь $\alpha/\beta$ интерферона	[162], [176]
NS5	Включает два домена. NS5A обладает метилтрансферазной активностью, NS5B – РНК-зависимой РНК-полимеразной активностью	[171], [157]

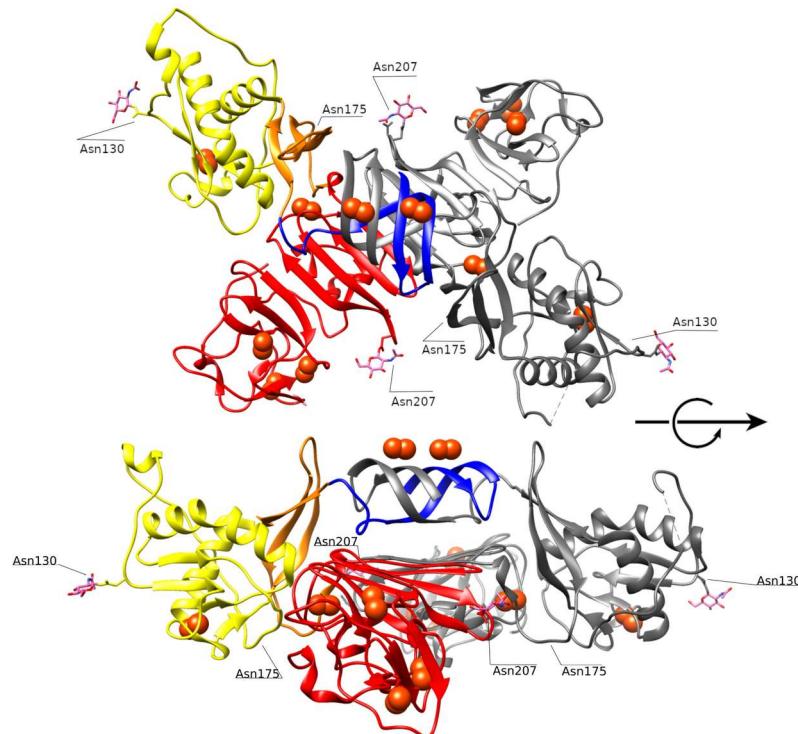


Рисунок 14. Структура NS1 ЗИКВ (PDB ID: 5K6K) [158]. Цветами выделены домены одного из мономеров: синим –  $\beta$ -бочка (а\к 1-29), оранжевым – связующий домен (а\к 30-37, 152-182), желтым –  $\alpha$ -спиральное крыло (а\к 38-151), красным –  $\beta$ -лестница (а\к 183-352). Отмечены сайты гликозилирования – Asn130 и Asn207 (Asn175 гликозилирован у вируса желтой лихорадки) и дисульфидные связи (темно-оранжевые сферы).

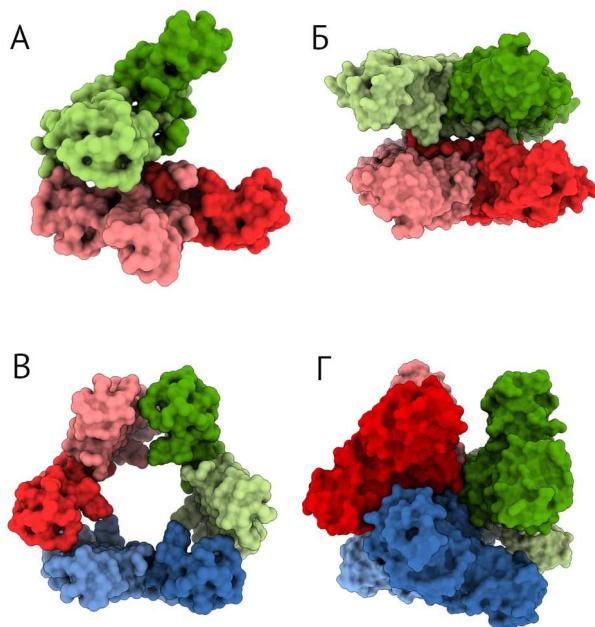


Рисунок 15. Четвертичная структура мультимера белка NS1 ортофлавивирусов [158]. (А) — “мягкий” тетramer (PDB ID 7WUU [182]), (Б) — “стабильный” тетramer (PDB ID: 7WUT [182]), (В) — секретируемый гексамер “голова-к-голове” (PDB ID: 7WUV [182]), (Г) — гексамер “сторона-к-стороне” (PDB ID: 8WBE [183]).

Белок NS1 слабо консервативен у различных ортофлавивирусов, его аминокислотные последовательности имеют 20–40% идентичности и 60–80% сходства [184], [185]. Наименее консервативны аминокислотные остатки на экспонированных в цитоплазму либо во внеклеточное пространство участках белка, где находятся антигенные детерминанты (Рисунок 16) [186].

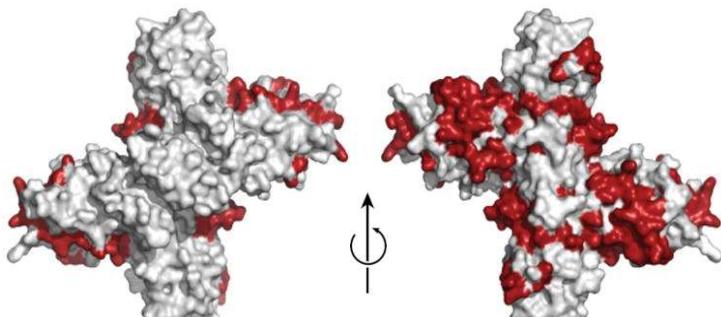


Рисунок 16. Вариабельные области белка NS1 вириона денге (красные), идентифицированные в результате сопоставления аминокислотных последовательностей четырех основных серотипов денге [186]. Проекция слева аналогична проекции на рис. 14 (верхняя панель), повернутой вокруг оси, перпендикулярной плоскости рисунка.

Белок NS1 ортофлавивирусов используется в качестве диагностического маркера во время острой фазы заболевания [187]. Кроме того, было показано, что несмотря на то, что белок NS1 не является частью вириона, антитела против него способны снижать вероятность летального исхода при заболевании благодаря комплемент-опосредованному лизису [188] или фагоцитозу [189] инфицированных клеток, на поверхности которых экспонирован белок NS1 [190].

Важная роль в жизненном цикле вириона и возникновении ряда патологических проявлений заболевания делает белок NS1 перспективной мишенью для разработки противовирусных препаратов. Были предложены следующие стратегии воздействия на него: нарушение димеризации белка, нарушение связывания белка с другими компонентами репликативного комплекса и нарушение связывания белка с белками иммунной системы зараженного организма [191]. Отдельно можно выделить разработку малых интерферирующих РНК, препятствующих экспрессии белка NS1 [192].

В качестве лигандов белка NS1 ДЕНВ были предложены пептиды длиной 12 а/к остатков, специфически связывающиеся с белком и препятствующие его димеризации и связыванию с мембраной ЭПР [193]. Их скрининг проводили при помощи фагового дисплея. Для пептидов, обнаруженных в ходе скрининга, проводили докинг и моделирование молекулярной динамики и выявили потенциальный сайт связывания – между  $\beta$ -бочкой,  $\beta$ -лестницей и связующим доменом (Рисунок 17). Все четыре пептида в концентрации 10 мкМ снижали урожай вирусных

частиц ДЕНВ-2 на 25–45%, 42–57% и 12–26% при обработке через 0, 4 и 8 ч после заражения соответственно.

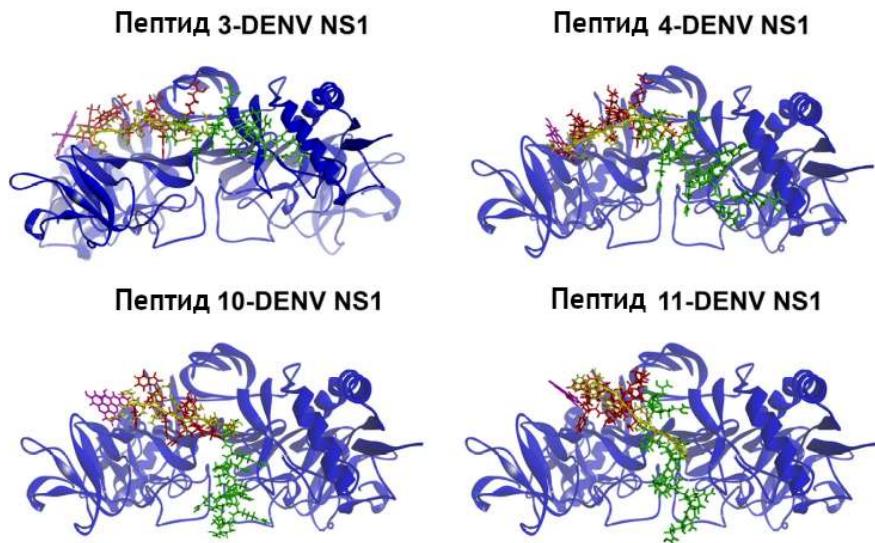
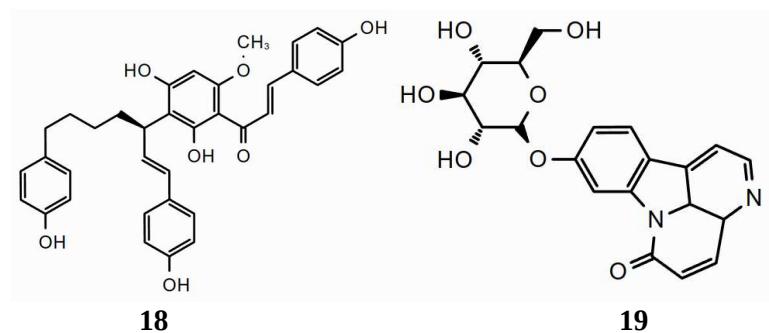


Рисунок 17. Результаты моделирования связывания пептидов с мономером белка NS1 [193].

Консенсусный молекулярный докинг был использован в процедуре виртуального скрининга малых молекул, способных препятствовать димеризации белка NS1 [194]. По литературным данным была собрана библиотека соединений, проявивших активность против различных ортофлавивирусов (ЗИКВ, ДЕНВ, ВЖЛ, ВЯЭ). Молекулы докировали в структуру мономера белка NS1 вируса Зика (PDB ID: 5K6K, цепь В) при помощи двух программ: AutoDock Vina использовали для слепого докинга, чтобы определить сайт связывания большинства молекул; GOLD использовали для ранжирования молекул по оценке эффективности связывания в области в радиусе 10 Å от остатка Ile23, соответствующей сайту димеризации белка NS1 между  $\beta$ -бочкой и  $\beta$ -лестницей. Три соединения, которые вошли в топ-10 по оценкам обеих программ, были выбраны для моделирования МД комплексов с мономером белка NS1, спрогнозированных двумя программами докинга. Одна из молекул, деоксикиаликсин-А (18), была предложена для экспериментальной валидации на основании энергии связывания и анализа траектории МД. Дальнейшее экспериментальное исследование, по всей видимости, не проводилось.



Другой подход к нарушению димеризации белка NS1 – ингибирование гликозилирования [195] – был применен в работе по виртуальному скринингу методом докинга в структуру белка NS1 ДЕНВ собранной авторами библиотеки из 10 тысяч веществ растительного происхождения, обладающих активностью против различных ортофлавивирусов [196]. Для всех результатов докинга вычислялись новые оценки при помощи специальной оценочной функции, и на их основании, а также на основании наличия связывания с сайтом гликозилирования Asn130 проводили ранжирование и отбор хитов виртуального скрининга. Для четырех молекул провели моделирование МД, в ходе которого хит виртуального скрининга – кантин-6-он-9-О- $\beta$ -глюкопиранозид (19) – связывался с белками NS1, NS3/NS2B и NS5 стабильно, но неспецифично. Экспериментальная валидация результатов моделирования не проводилась.

Разработка противовирусных лекарственных средств, способных действовать после заражения, является актуальной задачей медицинской химии. Вспышки новых инфекций, недостаточная доступность вакцин и развитие резистентности вирусов к существующим лекарствам [197] обуславливают непрерывную потребность в новых активных соединениях. Вычислительные методы помогают снизить время и ресурсы, необходимые для их поиска. В ходе пандемии COVID-19 стали хорошо заметны как преимущества и возможности структурно-обоснованного виртуального скрининга, так и его недостатки. Благодаря развитию метода кристаллографического скрининга малых молекул возник большой объем данных о структуре белков, в связи с чем возникла потребность в автоматизации их анализа. Работы по виртуальному скринингу и молекулярному моделированию противокоронавирусных лекарственных средств помогли ускорить их разработку и обнаружить новые типы активных соединений. Вместе с тем, множество исследований *in silico* не были валидированы и подтверждены экспериментально [198]. Тем не менее, доступный объем структурной, химической и биологической информации позволяет развить существующие вычислительные методы и систематизировать подходы к виртуальному скринингу не только противокоронавирусных, но и других потенциальных противовирусных препаратов.

## Глава 2. Разработка метода составления ансамбля разнообразных структур<sup>4</sup>

Первой стадией работы была разработка систематической, автоматизированной методики составления ансамблей белковых структур, применимой как к кристаллографическим данным, так и к конформерам из траекторий молекулярной динамики.

### 2.1. Подготовка белковых структур

Поиск структур Mpro SARS-CoV-2 в базе данных PDB [199] осуществляли по запросу: Polymer Entity Description CONTAINS PHRASE "3C-like proteinase" AND Scientific Name of the Source Organism = "Severe acute respiratory syndrome coronavirus 2". Структуры фильтровали для дальнейшего использования в соответствии с критериями, перечисленными в Таблице 6 (отмечены \*). В первый раз структуры были извлечены 18 марта 2021 г.: общее количество структур, депонированных в PDB, составило 283, из них 251 подходили для анализа разнообразия. Во второй итерации (11 января 2022 г.) в PDB было 415 доступных структур, из которых подходили 306. Структуры главной протеазы SARS-CoV-2 стандартизовали для сравнения и составления ансамблей (Рисунок 18). Исходные PDB-файлы, содержащие димерные структуры, использовали без изменений, тогда как для PDB-файлов, содержащих только мономер, структуру второго мономера копировали и поворачивали при помощи операций симметрии, а затем объединяли субъединицы в единую модель с помощью пользовательского сценария для *pychimera* 0.2.1 [200] – библиотеки Python2, позволяющей выполнять команды UCSF Chimera 1.16 [201] из командной строки. Молекулы растворителя, лиганды и альтернативные конформации остатков были удалены из файлов (для единообразия всегда сохраняли конформацию А).

---

4 При работе над данным разделом диссертации использованы материалы следующих публикаций автора, в которых, согласно Положению о присуждении ученых степеней в МГУ, отражены основные результаты, положения и выводы исследования:

1. Фомина А.Д., Палюлин В.А., Осолодкин Д.И. Моделирование структуры белка NS1 ортофлавивирусов по гомологии для виртуального скрининга потенциальных лигандов // Биомедицинская химия – 2024 – т. 70 (6) – с. 456–468
2. Fomina A. D., Uvarova V. I., Kozlovskaya L. I., Palyulin V. A., Osolodkin D. I., Ishmukhametov A. A. Ensemble Docking Based Virtual Screening of SARS-CoV-2 Main Protease Inhibitors // Mol. Inf. – 2024 – V.43. – e202300279.
3. Zakharova M.Yu., Kuznetsova A.A., Uvarova V.I., Fomina A.D., Kozlovskaya L.I., Kaliberda E.N., Kurbatskaia I.N., Smirnov I.V., Bulygin A.A., Knorre V.D., Fedorova O.S., Varnek A., Osolodkin D.I., Ishmukhametov A.A., Egorov A.M., Gabibov A.G., Kuznetsov N.A. Pre-Steady-State Kinetics of the SARS-CoV-2 Main Protease as a Powerful Tool for Antiviral Drug Discovery // Front. Pharmacol. – 2021 – V. 12. – P. 773198.
4. Фомина А.Д., Осолодкин Д.И. Программный комплекс для систематического ансамблевого докинга // Свидетельство о государственной регистрации программы для ЭВМ № 2024684465. – Номер заявки 2024684027. – Дата поступления 17.10.24. – Дата государственной регистрации в Реестре программ для ЭВМ 17.10.24.
5. Фомина А.Д., Осолодкин Д.И. База данных соединений с экспериментально определенной активностью против главной протеазы β-коронавирусов // Свидетельство о государственной регистрации базы данных № 2024625235. – Номер заявки 2024624786. – Дата поступления 29.10.2024 – Дата государственной регистрации в Реестре баз данных 15.11.2024 Бюл. № 11

Таблица 6. Характеристики структур главной протеазы SARS-CoV-2 из базы данных PDB

	03.2021	01.2022
<b>Общее число структур</b>	<b>283</b>	<b>415</b>
Структуры с мутациями в области активного центра*	3	6
Структуры с окисленным остатком Cys145*	4	8
Незрелые структуры*	6	8
Некорректный способ димеризации*	2	8
Структуры с неупорядоченными остатками в активном центре*	17	78
<b>Число структур, прошедших все фильтры</b>	<b>251</b>	<b>306</b>
Апо-структуры	36	69
Структуры с аллостерическими лигандами	39	46

\* характеристики, использованные для фильтрации структур

Был проведен визуальный анализ способа связывания ковалентного пептидомиметика N3 в активном центре Mpro (PDB ID 6LU7 [202]) и выделены аминокислотные остатки, формирующие сайт связывания: Thr24, Thr25, Thr26, Leu27, Asn28, Pro39, Arg40, His41, Val42, Ile43, Cys44, Thr45, Asp48, Met49, Leu50, Pro52, Tyr54, Tyr118, Asn119, Phe140, Leu141, Asn142, Ser144, Cys145, His163, His164, Glu166, Asp187, Gln189, Thr190, Gln192.

Структуры окисленных форм белка и структуры с неупорядоченными аминокислотными остатками в области активного центра были выявлены в VMD 1.9.3 [203] по количеству неводородных атомов в остатках, формирующих активный центр, и исключены из дальнейшего рассмотрения.



Рисунок 18. Порядок стандартизации и отбора структур для составления ансамбля

Аминокислотные последовательности белка извлекали из файлов PDB с помощью сценария `pdb_seq` из библиотеки `pdbscripts 0.2` [204]. Выравнивание последовательностей для поиска и исключения структур с мутациями в активном сайте проводили с помощью веб-интерфейса MUSCLE [205], [206]. При анализе выравнивания аминокислотных последовательностей была обнаружена некорректная аннотация структуры 6W79 – она принадлежит главной протеазе SARS-CoV, а не SARS-CoV-2. Однако во время первой итерации построения ансамбля было решено оставить эту структуру в наборе, чтобы обеспечить структурное разнообразие, основываясь на высоком структурном сходстве и фактическом отсутствии селективности ингибиторов между протеазами этих вирусов. Во второй итерации структура была удалена.

Пространственное выравнивание белковых структур проводили в VMD 1.9.3 по методу Кабша [207] для набора атомов Сα цепи А, присутствующих в каждой структуре. В качестве

эталона использовали структуру с идентификатором PDB 5R7Y [208]. Структуры с некорректным способом димеризации исключали по порогу среднего попарного СКО, равному 3, так как на основании визуальной оценки структур с  $\text{СКО} > 3$  было установлено, что такие структуры характеризуются некорректной укладкой мономеров в димере (Таблица 6). Все дальнейшие процедуры проводили с набором выровненных структур.

## 2.2. Составление ансамблей

### 2.2.1. Выбор структур по среднеквадратичному отклонению

Принцип подбора структур для составления ансамбля был основан на среднеквадратичном отклонении (СКО) координат тяжелых атомов остатков активного центра (Рисунок 27). Попарные значения СКО рассчитывали в VMD 1.9.3. Структуры с наибольшим средним попарным СКО оказались незрелыми и были исключены из рассмотрения. Значения СКО координат атомов были визуализированы на тепловых картах, реализованных [209] на языке программирования Python 3.9 с помощью Pandas 1.4.1 [210], NumPy 1.21.6 [211], библиотеки Matplotlib 3.5.1 [212] и seaborn 0.11.2 [213]. На основании распределения значений СКО мы отобрали 1% самых разнообразных структурных пар. Многие из них включали в себя одну и ту же («несходную») структуру в одном положении и разные похожие между собой структуры в другом, таким образом, являясь вырожденными. Такие вырожденные пары на основании распределения попарных СКО были отфильтрованы по порогу подобия так, чтобы отсекать от 55 до 95 % наименее разнообразных структур и сохранить одну репрезентативную пару из набора сходных пар [209]. Далее в качестве порога подобия использовали 75% пар структур (Рисунок 19), так как в результате получался ансамбль оптимального размера – шесть или семь наименее похожих между собой структур. Ансамбль 1 был выбран из структур, полученных из PDB 18 марта 2021 г., а Ансамбль 2 – из структур, полученных 11 января 2022 г.

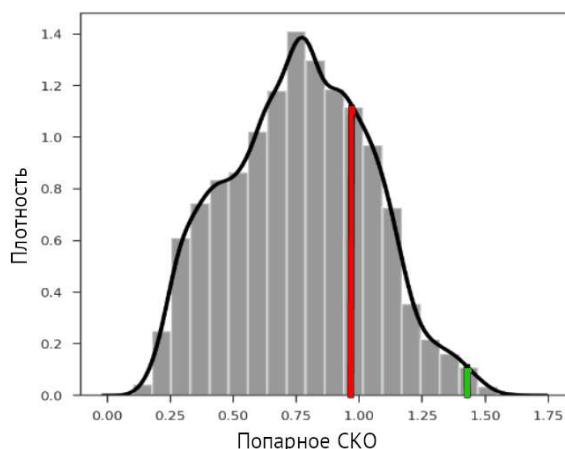


Рисунок 19. Распределение СКО координат атомов остатков активных центров для структур, извлеченных из PDB 11 января 2022 г. Вертикальные линии отображают порог подобия (красный) и порог разнообразия (зеленый).

В *Ансамбль 1* вошли семь структур: 7AKU, 6W79, 6XHM, 7NEV, 7L13, 7BFB, 7B83 (Таблица 7). В *Ансамбль 2* вошли шесть структур: 7JKV, 7CB7, 7MPB, 7MHP, 7VH8, 7RC0 (Таблица 7). Только одна апо-структура вошла в *Ансамбль 2* (PDB ID: 7MHP). Остальные структуры обоих ансамблей имели лиганды в своих активных центрах, как ковалентные, так и нековалентные. В ансамбли не вошли структуры комплексов Mpro с аллостерическими ингибиторами.

Таблица 7. Информация о структурах *Ансамбля 1* и *Ансамбля 2* в базе данных PDB

ID	Разрешение, Å	Лиганд	Дата публикации	Ссылка
Ансамбль 1				
7AKU	2,49	калпептин	2020-10-28	[214]
7NEV	1,69	леупептин	2021-03-03	[214]
7L13	2,17	аналог перампанела 21	2021-03-03	[132]
6XHM	1,41	PF-00835231	2020-07-08	[113]
7B83	1,8	пиритион цинк	2021-01-13	[214]
7BFB	2,05	эбселен	2021-03-03	Не опубликована в статье
6W79	1,46	нековалентный ингибитор X77	2020-08-26	Не опубликована в статье
Ансамбль 2				
7MHP	2	нет	2021-05-12	[215]
7MPB	2,3	аскорбиновая кислота	2021-05-12	[216]
7CB7	1,69	GC-376	2021-05-05	[217]
7RC0	1,65	GRL-091-20	2021-09-29	[123]
7VH8	1,59	PF-07321332	2021-11-03	[218]
7JKV	1,25	GRL-2420	2020-09-23	[219]
7C7P*	1,74	телапревир	2020-07-29	[220]

\* только в ens2\_3

Зависимость размера ансамбля от порогового значения подобия была проанализирована для более крупного набора данных, полученного в январе 2022 г. (Рисунок 20, Таблица 8). Пороговое значение варьировали для сравнения ансамблей размеров 3 (ens2\_3: 7C7P, 7CB7, 7VH8) и 4 (ens2\_4: 7CB7, 7MPB, 7RC0, 7VH8) с полноразмерным *Ансамблем 2* (ens2\_6).

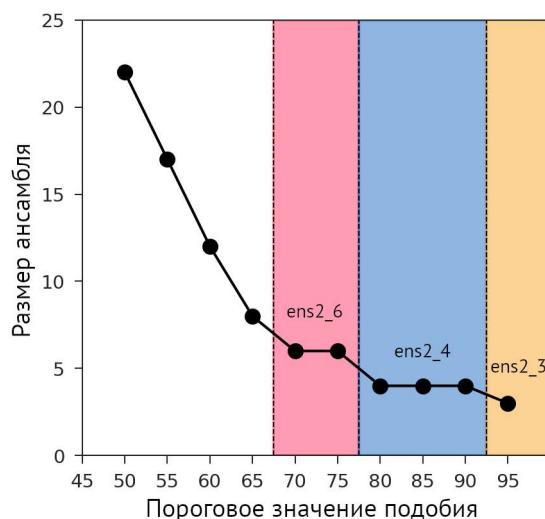


Рисунок 20. Зависимость размера ансамбля от порогового значения подобия

Таблица 8. Ансамбли структур Mrgo, формирующиеся при различных пороговых значениях подобия

Порог подобия, %	N	Структуры, вошедшие в ансамбль
50	22	6WTT, 6XHM, 6XOA, 6XR3, 7AKU, 7BUY, 7C6U, 7CB7, 7DHJ, 7FAZ, 7JKV, 7MBI, 7MHN, 7MHO, 7MHP, 7MHQ, 7MPB, 7NG6, 7RBZ, 7RC0, 7RN0, 7VH8
55	17	6WTT, 6XHM, 6XOA, 7BUY, 7C6U, 7CB7, 7FAZ, 7JKV, 7MBI, 7MHO, 7MHP, 7MHQ, 7MPB, 7NG6, 7RC0, 7RN0, 7VH8
60	12	6XHM, 6XOA, 6XR3, 7C7P, 7CB7, 7JKV, 7MBI, 7MHO, 7MHP, 7MPB, 7RC0, 7VH8
65	8	6XHM, 7C7P, 7CB7, 7JKV, 7MHP, 7MPB, 7RC0, 7VH8
70	6	7CB7, 7JKV, 7MHP, 7MPB, 7RC0, 7VH8
75	6	7CB7, 7JKV, 7MHP, 7MPB, 7RC0, 7VH8
80	4	7CB7, 7MPB, 7RC0, 7VH8
85	4	7CB7, 7MPB, 7RC0, 7VH8
90	4	7CB7, 7MPB, 7RC0, 7VH8
95	3	7C7P, 7CB7, 7VH8

## 2.2.2. Выбор структур методом анализа главных компонент

Анализ главных компонент (PCA, principal component analysis) — распространенный способ отбора репрезентативных белковых структур, и мы использовали его как альтернативу попарному СКО координат атомов. PCA проводили при помощи sklearn 0.24.2 для координат неводородных атомов того же набора остатков активного сайта, который использовали для расчета СКО. Координаты были последовательно считаны из файлов PDB, преобразованы в одномерные векторы, нормированы с помощью *MinMaxScaler* и использованы в качестве входных данных для PCA [209].

Точки, соответствующие структурам, были кластеризованы в координатах двух первых главных компонент с использованием метода  $k$ -средних. Оптимальное количество кластеров и, соответственно, размер ансамбля определяли методом локтя, который заключается в построении зависимости внутрикластерной суммы квадратов отклонений (WCSS, within cluster sum of squares) от числа кластеров (Рисунок 21). Перелом на графике этой зависимости соответствует оптимальному числу кластеров, так как при увеличении числа кластеров более не наблюдается значительного снижения суммы квадратов отклонений. На основании визуальной оценки зависимости (Рисунок 21) количество кластеров было выбрано равным четырем, хотя зависимость является плавной. При визуальной оценке результатов PCA в координатах первых двух главных компонент (Рисунок 22) можно наблюдать только два отдельных кластера. Однако было решено выделить четыре кластера для максимальной репрезентации конформационного пространства.

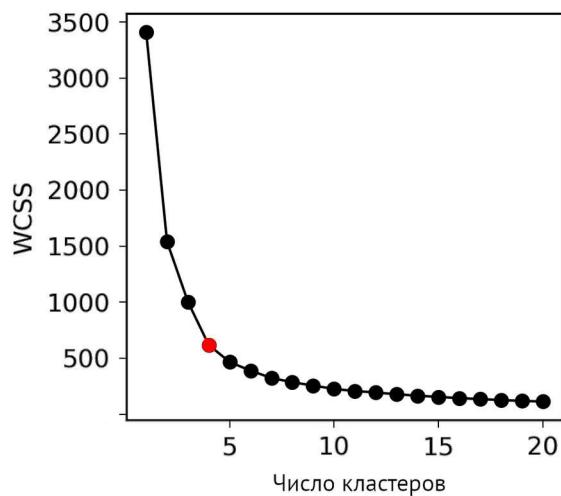


Рисунок 21. Зависимость внутрикластерной суммы квадратов отклонений (WCSS) от числа кластеров в результатах PCA координат атомов аминокислотных остатков активного центра.

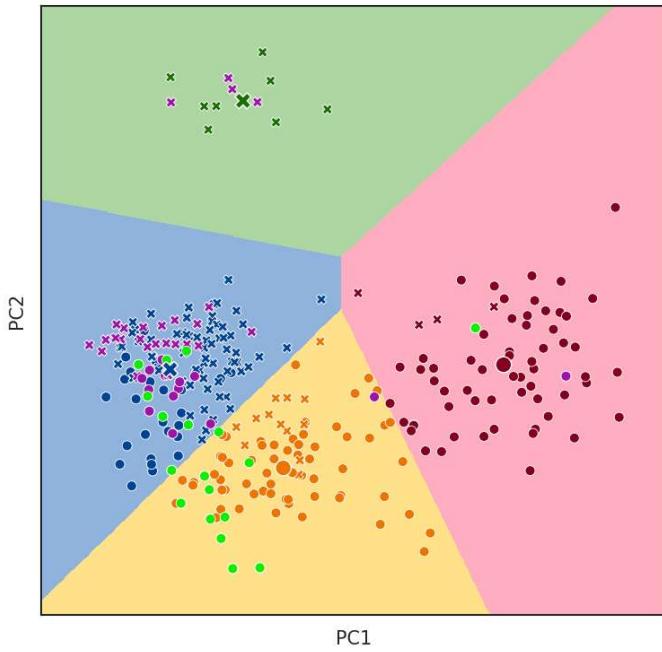


Рисунок 22. Кластеризация структур Mrго, доступных в январе 2022 г., методом  $k$ -средних ( $k = 4$ ) на графике двух главных компонент. Структуры, полученные методом PanDDA, отмечены крестиками. Точки, ближайшие к центроидам каждого кластера, увеличены в размере и окрашены в цвет, соответствующий цвету кластера. Фиолетовым выделены структуры с аллостерическими ингибиторами, а ярко-зеленым – апо-структуры.

Кластеризация  $k$ -средних была выполнена с помощью функции *KMeans* из *sklearn 0.24.2* с алгоритмом инициации  $k$ -means++ [221] и зерном генератора псевдослучайных чисел, равным 42. Кластер 1 включает только структуры, решенные с помощью метода PanDDA [208], и эти структуры, как правило, имеют более консервативные конформации аминокислотного остатка Ser1 по сравнению с другими структурами, в то время как другие остатки активного центра демонстрируют такое же конформационное разнообразие. При этом все структуры в кластере 1 являются либо апо-структурами, либо структурами с аллостерическими ингибиторами. Структуры PanDDA в более крупном кластере собраны в верхней левой плотной части, хотя общих конформационных признаков, как в другом обособленном кластере, не наблюдается. Большинство структур, содержащих аллостерические ингибиторы, скапливаются в плотном кластере 2 (рис. 22, оранжевый). Субкластеры 3 и 4 более разрежены и содержат структуры с разными лигандами в активном центре.

Ансамбль PCA был составлен из структур, соответствующих точкам, ближайшим к центроидам кластеров: 5REZ, 5RGM, 7JQ2, 7L14 (Таблица 9).

Таблица 9. Информация о структурах *Ансамбля PCA* в базе данных PDB

ID	Разрешение, Å	Лиганд	Дата публикации	Ссылка
Ансамбль PCA				
5REZ	1,79	POB0129	2020-03-25	[208]
5RGM	2,04	PCM-0102142	2020-04-15	[208]
7JQ2	1,4	MPI5	2020-12-23	[222]
7L14	1,66	Аналог перампанела 26	2020-03-25	[132]

### 2.2.3. Выбор состояний из траекторий молекулярной динамики

Моделирование молекулярной динамики (МД) – наиболее распространенный источник белковых структур для ансамблевого докинга. Чтобы составить *Ансамбль МД*, мы использовали пять конформаций, соответствующих временным отсечкам 2, 4, 6, 8 и 10 мкс, извлеченных из 10-микросекундной траектории МД [223] структуры 6LU7 с помощью функции GROMACS trjconv[224]:

```
gmx trjconv -s protein_conf.gro -f trajectory.xtc -tu us -dump
10 -o protein_snap_10us.pdb
```

Конформации мономеров из тех же временных точек траектории были успешно использованы в виртуальном скрининге, основанном на структуре мишени [31]. Однако мы извлекли конформации димера Мрго и выровняли их со структурой 5R7Y, как описано ранее. Остатки были перенумерованы, а цепи разделены и названы с помощью инструментов USCF Chimera Rerumber Residues и Change Chain ID.

### 2.2.4. Конформационное разнообразие ансамблей

Из анализа тепловой карты попарных СКО между неводородными атомами остатков активного центра структур всех ансамблей (Рисунок 23) следует, что конформации из траектории МД значительно разнообразнее, чем кристаллографические структуры, извлеченные из PDB. Хотя такое разнообразие может быть полезным для обеспечения разнообразия результатов виртуального скрининга, биохимическое значение таких конформаций может быть ограничено. С другой стороны, визуальный анализ ансамблей PCA и СКО позволяет отметить, что одни остатки в основном регистрируются в одних и тех же конформациях, а другие более гибки (Рисунок 24). Примечательно, что процедуры на основе СКО и PCA привели к различному разнообразию выбранных структур, при этом структуры, отобранные при помощи PCA, были менее разнообразны, чем структуры, отобранные на основании попарных СКО.

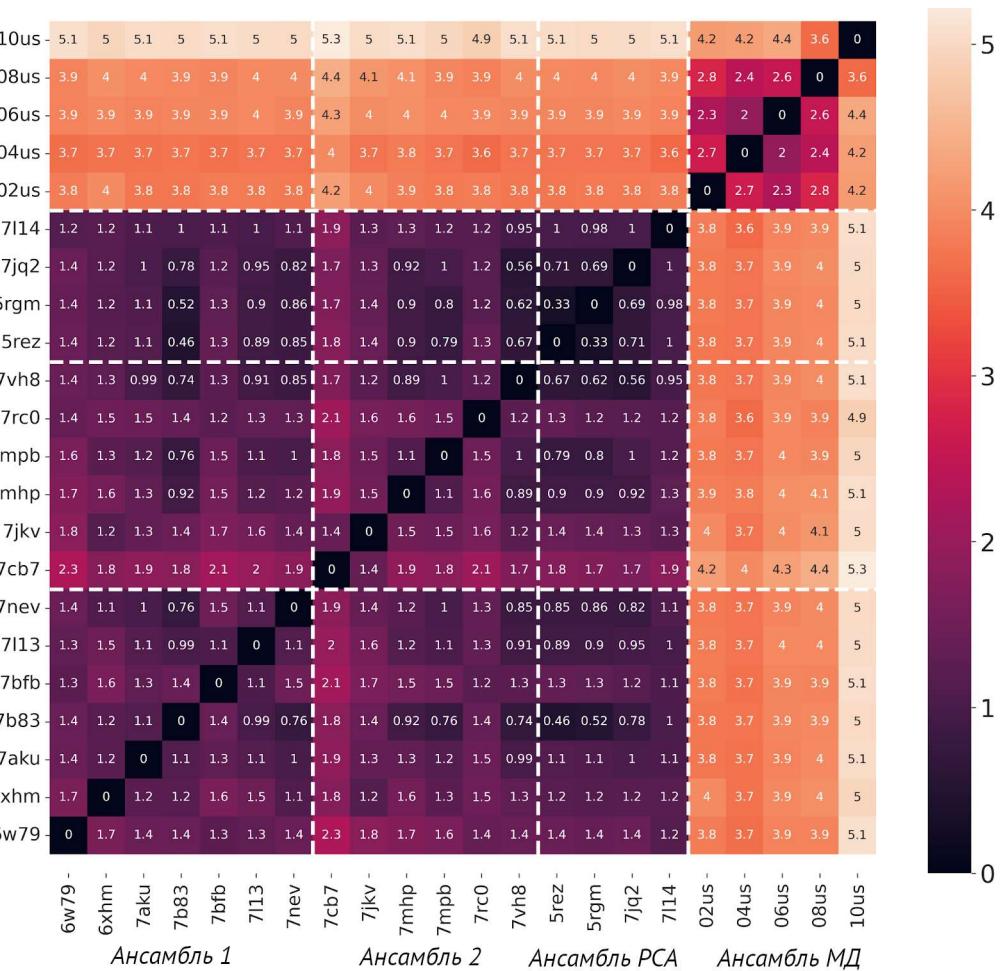


Рисунок 23. Тепловая карта попарных СКО (Å) между координатами тяжелых атомов остатков активного центра структур Ансамбля 1, Ансамбля 2, Ансамбля РСА и Ансамбля МД.

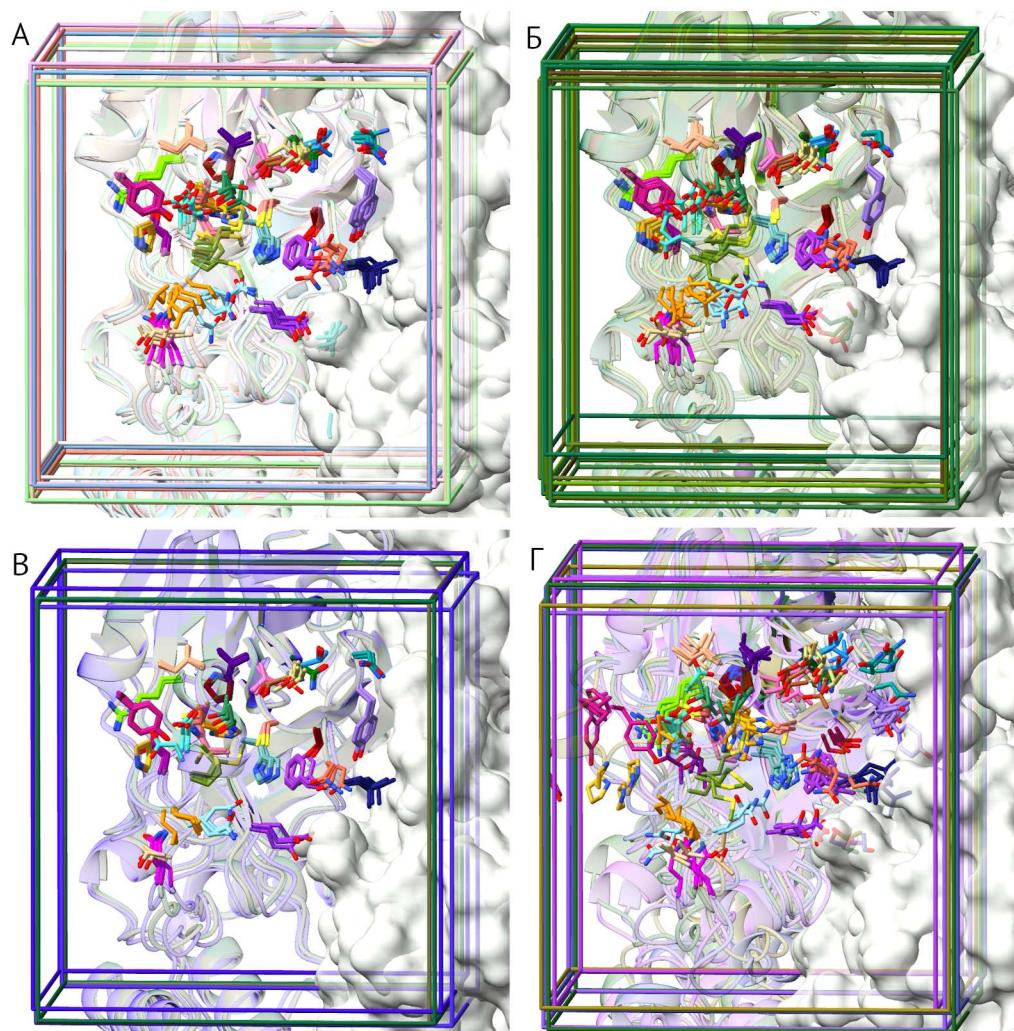


Рисунок 24. Наложение структур Ансамбля 1 (А), Ансамбля 2 (Б), Ансамбля PCA (В) и Ансамбля МД (Г) и соответствующих боксов для докинга. Для цепи А показаны а/к остатки активного центра, окрашенные в соответствии с номером остатка. Для цепи В показана поверхность.

Разработанная процедура отбора структур по разнообразию конформаций остатков активного центра автоматизирована, и получаемые таким образом ансамбли всегда детерминированы начальным набором данных. Таким образом процедура систематического отбора исключает влияние случайности при формировании ансамбля. Однако для того, чтобы определить надежность виртуального скрининга при помощи таких ансамблей и сравнить их с ансамблями, отобранными при помощи моделирования молекулярной динамики, необходим внешний контроль при помощи докинга аннотированной библиотеки.

## Глава 3. Подготовка библиотек низкомолекулярных соединений<sup>5</sup>

Для разработки систематического метода составления ансамблей структур белка для виртуального скрининга были выбраны экспериментальные кристаллографические данные о структуре Mpro SARS-CoV-2 благодаря их доступности и актуальности поиска противокоронавирусных лекарственных средств. Чтобы осуществлять внешний контроль результатов докинга были взяты опубликованные аннотированные библиотеки ингибиторов Mpro SARS-CoV-2, а также сформирована дополнительная библиотека по данным, собранным из научных статей и препринтов.

### 3.1. Библиотека соединений с аннотацией по противокоронавирусной активности

Формирование библиотеки соединений с известной активностью в отношении Mpro SARS-CoV-2 проводили на основе литературы, опубликованной в период с 07.2020 по 03.2021. Литературный поиск проводили в PubMed по запросам: *coronavirus antiviral, (coronavirus inhibitors) NOT coronavirus antiviral, (((((SARS antiviral) OR SARS inhibitor) OR MERS antiviral) OR MERS inhibitor) NOT coronavirus inhibitors) NOT coronavirus antiviral*. Выбирали статьи, содержащие экспериментальные результаты тестирования активности различных соединений против коронавирусов. Соединения, для которых приведена активность против конкретных ферментов (не Mpro), или имеющие известную мишень или механизм действия (не Mpro), отфильтровывали. Проанализировали 286 статей и препринтов, из которых извлекли 2181 результат анализа ингибиторной активности по отношению к протеазе Mpro и/или к репликации различных коронавирусов (Рисунок 25А). Работу с библиотекой вели в программе Instant JChem 20.13.0 (ChemAxon, Hungary). В Таблице 10 приведены поля составленной таким образом базы данных с указанием типов данных в столбцах. Логические поля *Activity virus* и *Activity protein* (флаги активности) заполняли на основании информации об активности из

5 При работе над данным разделом диссертации использованы материалы следующих публикаций автора, в которых, согласно Положению о присуждении ученых степеней в МГУ, отражены основные результаты, положения и выводы исследования:

1. Фомина А.Д., Палюлин В.А., Осолодкин Д.И. Моделирование структуры белка NS1 ортофлавивирусов по гомологии для виртуального скрининга потенциальных лигандов // Биомедицинская химия – 2024 – т. 70 (6) – с. 456–468
2. Fomina A. D., Uvarova V. I., Kozlovskaya L. I., Palyulin V. A., Osolodkin D.I., Ishmukhametov A. A. Ensemble Docking Based Virtual Screening of SARS-CoV-2 Main Protease Inhibitors // Molecular Informatics – 2024 – V.43. – e202300279.
3. Zakharova M.Yu., Kuznetsova A.A., Uvarova V.I., Fomina A.D., Kozlovskaya L.I., Kaliberda E.N., Kurbatskaia I.N., Smirnov I.V., Bulygin A.A., Knorre V.D., Fedorova O.S., Varnek A., Osolodkin D.I., Ishmukhametov A.A., Egorov A.M., Gabibov A.G., Kuznetsov N.A. Pre-Steady-State Kinetics of the SARS-CoV-2 Main Protease as a Powerful Tool for Antiviral Drug Discovery // Frontiers in Pharmacology – 2021 – V. 12. – P. 773198.
4. Фомина А.Д., Осолодкин Д.И. Программный комплекс для систематического ансамблевого докинга // Свидетельство о государственной регистрации программы для ЭВМ № 2024684465. – Номер заявки 2024684027. – Дата поступления 17.10.24. – Дата государственной регистрации в Реестре программ для ЭВМ 17.10.24.
5. Фомина А.Д., Осолодкин Д.И. База данных соединений с экспериментально определенной активностью против главной протеазы β-коронавирусов // Свидетельство о государственной регистрации базы данных № 2024625235. – Номер заявки 2024624786. – Дата поступления 29.10.2024 – Дата государственной регистрации в Реестре баз данных 15.11.2024 Бюл. № 11

оригинальных источников; для отнесения к классу активности было использовано значение  $EC_{50}$  или  $IC_{50}$  10 мкМ. Из собранных соединений 807 (37%) были активны против протеазы, 385 (17,6%) неактивны против протеазы, 717 (32,8%) активны против вируса и 282 (13%) неактивны против вируса (Рисунок 25Б,В). Для 98 соединений была опубликована активность и против протеазы, и против вируса.

Таблица 10. Структура базы данных опубликованных противокоронавирусных соединений [225].

Название поля	Тип данных	Описание поля
CdId	integer	Идентификационный номер молекулы
Structure	structure	Структура молекулы
Mol Weight	float	Молекулярный вес
Molecule name	string	Название молекулы
Virus	string	Аббревиатура вируса
Virus Inhibition	float	Процент ингибирования репродукции вируса при фиксированной концентрации соединения
$EC_{50}$ virus, nM	float	Противовирусная активность в клеточных исследованиях
Activity virus	boolean	Флаг противовирусной активности (True/False)
Protein Inhibition	float	Процент ингибирования Mpro при фиксированной концентрации соединения
$IC_{50}$ protein, nM	float	Mpro $IC_{50}$ в <i>in vitro</i> эксперименте
Ki, nM	float	Константа ингибирования Mpro
Activity protein	boolean	Флаг ингибиторной активности против Mpro (True/False)
Article DOI	string	DOI источника
Covalent	boolean	Флаг наличия ковалентной активной группы (True/False)

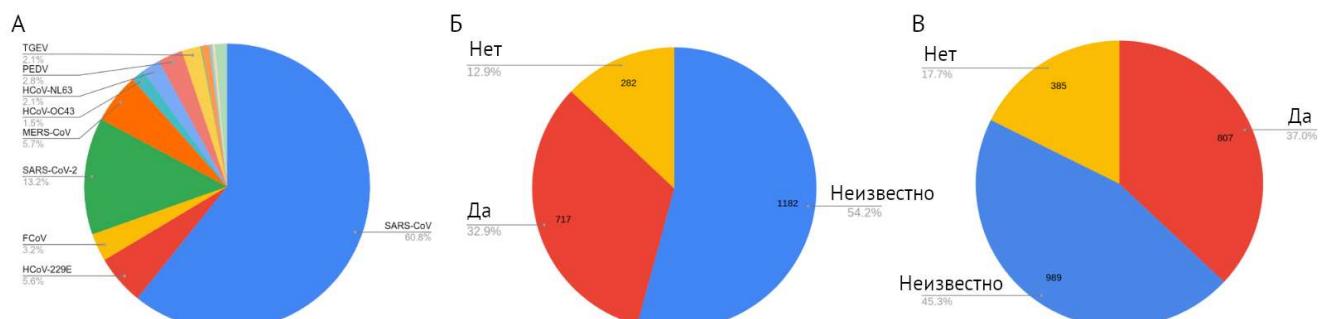


Рисунок 25. Описательная статистика литературных данных. (А) изучавшийся вирус; (Б) значение флага противовирусной активности; (В) значение флага ингибиторной активности.

Собранный библиотека содержит больше активных соединений, чем неактивных (Рисунок 25Б, В, Рисунок 26), в связи с тем, что источник данных – статьи и препринты, в которых чаще публиковали результаты успешных экспериментов. Большинство изученных соединений принадлежит к классу пептидомиметиков и является потенциальными ковалентными ингибиторами. На основании распределения  $IC_{50}$  был выбран более жесткий порог активности, чем обычно используется в литературе: 10 мкМ, в то время как в изученных статьях обычно активным считали соединение с  $IC_{50} < 50$  мкМ.

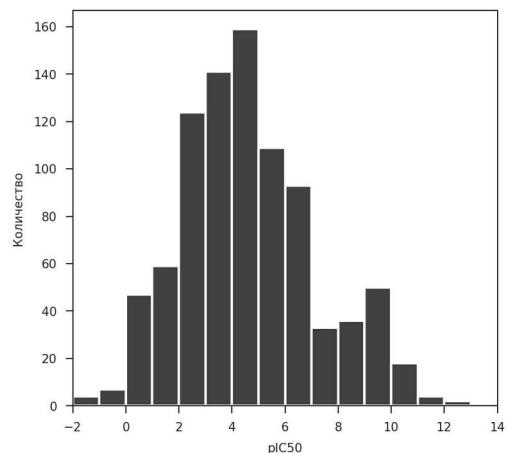


Рисунок 26. Распределение опубликованных значений  $pIC_{50}$  ингибиторов протеазы Mpro различных коронавирусов

Для дальнейшего использования в библиотеке были оставлены только соединения с определенной концентрацией полуингибиции Mpro ( $IC_{50}$ ) для одного из  $\beta$ -коронавирусов (SARS-CoV-2, SARS-CoV, MERS-CoV). В итоговой библиотеке 788 соединений, 195 (24,7%) из которых имеют  $IC_{50} < 10$  мкМ и отнесены к классу активных.

### 3.2. Библиотека разнообразных соединений

Библиотеку DD-ZINC формировали в качестве источника разнообразных, предположительно неактивных соединений для дополнения аннотированных библиотек. Провели отбор по разнообразию из базы данных ZINC15 (Рисунок 27) [226] при помощи программы, написанной на Python 3.7 с использованием библиотек Pandas 1.0.5 [210], Rdkit 2020\_03\_6 и Sklearn 0.23.1 [227]. Из ZINC15 был извлечен набор из ~413 миллионов молекул, подобных соединениям-лидерам (молекулярная масса от 300 до 375 Да, logP от -1 до 3,5), в формате SMILES, разбитый на 21 часть по соотношению молекулярной массы и logP. Для анализа разнообразия молекулы были закодированы в виде моргановских фингерпринтов [228]. В каждой из частей проводили отбор по разнообразию функцией MaxMinPicker на основе

индекса Танимoto, при этом каждую часть делили на 100 последовательных поднаборов, в каждом из которых анализировали разнообразие независимо. Из каждого поднабора выбирали 1% наиболее разнообразных молекул. Далее действовали по двум схемам: выбранные молекулы предварительно перемешивали и объединяли в новые поднаборы, либо объединяли последовательно, без перемешивания, после чего разделяли на 5 наборов, каждый из которых затем делили на 10 поднаборов, в которых снова проводили отбор по разнообразию. Молекулы, выбранные в каждом из пяти наборов (по 8000), объединяли в один общий, из которого выбирали финальную разнообразную выборку из 20000 соединений. Разнообразие контролировали на основе распределения среднего индекса Танимoto между парами молекул в случайных выборках из библиотек, отобранных разными методами. Из наборов молекул, полученных с перемешиванием и без перемешивания, отбирали 1000 случайных наборов по 10000 молекул, для которых рассчитывали средний индекс Танимoto, строили распределения средних индексов и проверяли их нормальность при помощи критерия Шапиро-Уилка [229] и рассчитывали доверительный интервал среднего методом `numpy.percentile`. Различие средних для распределений проверяли по критерию Стьюдента.

Рассчитали 95% доверительный интервал для среднего индекса Танимoto наборов без перемешивания и с перемешиванием, которые составили [0,2314; 0,2346] и [0,1989; 0,2018] соответственно (Рисунок 28). Нормальность распределения средних индексов Танимoto была подтверждена критерием Шапиро-Уилка для зависимых выборок. W-Статистика была равна 0,99 с *p*-значением 0,78, это позволяет принять гипотезу о нормальности распределений средних индексов Танимoto для полученных выборок.

Библиотека, собранная по процедуре с перемешиванием, имеет более низкий средний индекс Танимoto, чем в отсутствие перемешивания. Поэтому для дальнейшей работы была выбрана библиотека с более низким средним индексом Танимoto, как более разнообразная.

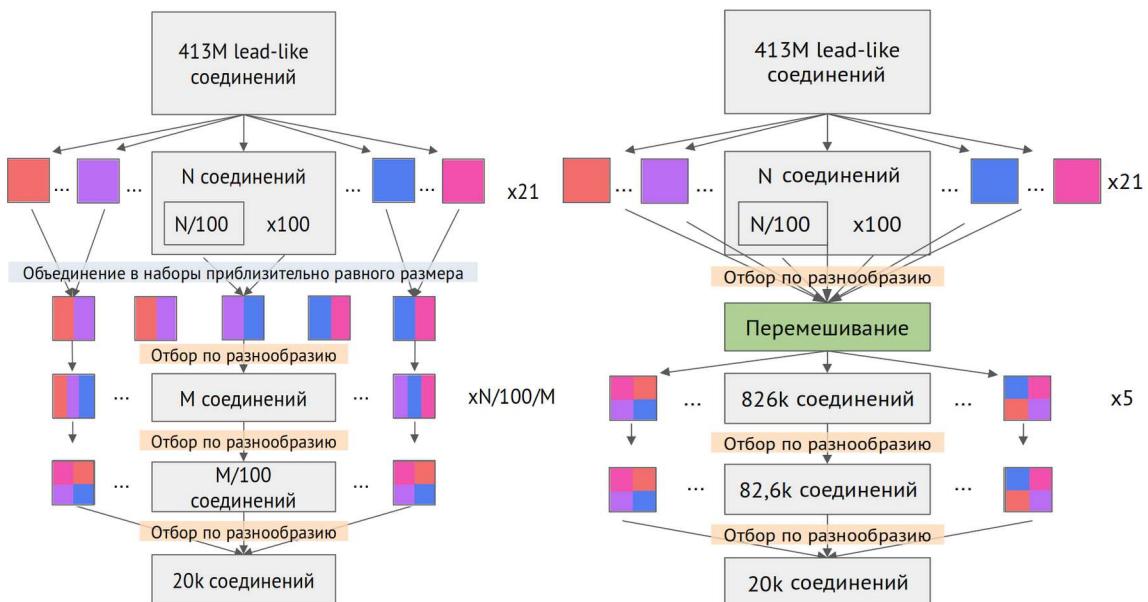


Рисунок 27. Схемы отбора разнообразных библиотек из ZINC – слева без перемешивания, справа с перемешиванием после первичного отбора по разнообразию.

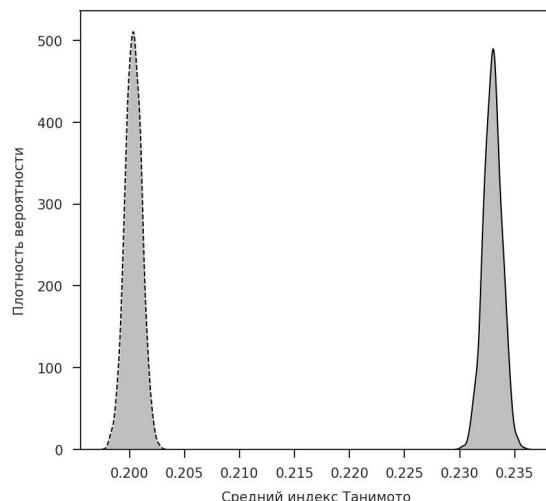


Рисунок 28. Распределения плотности вероятности средних индексов Танимото. Прерывистой линией обозначено распределение для выборки с перемешиванием (слева), непрерывной – для выборки без перемешивания (справа).

### 3.3. Химическое пространство ингибиторов главной протеазы SARS-CoV-2

Выбор библиотек малых молекул для исследования был мотивирован двумя целями: во-первых, необходимо было провести оценку эффективности процедуры ранжирования активных соединений по сравнению с неактивными в рамках процедуры ВС, а во-вторых, выявить перспективные нековалентные ингибиторы Mpro, принадлежащие к новым хемотипам. Таким образом, учитывались критерии активности, химического разнообразия и коммерческой доступности соединений. В работе использовались следующие библиотеки:

1) Аннотированные библиотеки (АА):

- a) AA-Aug20 (53 активных, 763 неактивных, ранее опубликована в [230]): 816 нековалентно-связывающихся соединений, опубликованных в статьях в научных журналах до 30.08.2020. К активным авторы относили соединения с  $IC_{50}$  Mpro  $< 10$  мкМ, а к неактивным – соединения с  $IC_{50} > 20$  мкМ, процентом ингибиования при 20 мкМ  $< 80$  % или процентом ингибиования при 10 мкМ  $< 50$  %..
- b) AA-Feb21 (339 активных, 419 неактивных): 758 соединений, для которых были опубликованы концентрации полуингибиования Mpro SARS-CoV-2 в рецензируемых статьях и препринтах до 09.02.2021 (методику отбора см. ранее в разделе 3.1) [225]. Активными считали соединения с  $IC_{50}$  против Mpro меньше 10 мкМ.
- c) AA-Aug21 (754 активных, 451 неактивных, ранее опубликована в [102]): 758 соединений из статей, опубликованных в научных журналах в период с января 2020 по август 2021, 999 новых соединений из COVID Moonshot initiative [231], доступных на 01.10.2021, и 8 соединений из базы данных ChEMBL. Активными авторы считали соединения с  $IC_{50}$  против Mpro меньше 10 мкМ, если были доступны численные данные.
- d) CHEMBL4495582 (214 активных, 6683 неактивных): Результаты высокопроизводительного скрининга *in vitro* ингибиторной активности лекарственных средств против протеазы Mpro SARS-CoV-2 [232]. Активными считали соединения с процентом ингибиования Mpro SARS-CoV-2  $> 60\%$ . Библиотеку использовали только для построения моделей машинного обучения (глава 5).

2) DD-ZINC (20000 соединений): Библиотека разнообразных соединений из базы данных ZINC15 (процедуру отбора см. ранее в разделе 3.2)

Все структуры малых молекул были стандартизованы при помощи программы Standardizer 20.19.0 (ChemAxon, Hungary). В ходе стандартизации добавляли атомы водорода, удаляли молекулы растворителя и противоионы и оптимизировали энергию структуры. Трехмерные структуры генерировали в DataWarrior 5.5.0 [233]: конформер с наименьшей энергией выбрали с использованием систематического алгоритма, начальные значения торсионных углов извлекали из Crystallography Open Database [234], а минимизацию энергии выполняли с использованием силового поля MMFF94s+ [235]. Файлы были преобразованы из формата SDF в формат MOL2 с помощью openbabel 3.1.0 [236].

Визуализацию и сравнение химического пространства аннотированных библиотек проводили с помощью t-SNE (t-distributed stochastic neighbor embedding, стохастическое

вложение соседей с t-распределением) [237]. Это метод уменьшения размерности, часто применяемый к химическим данным [238], в котором для каждой точки в многомерном пространстве строят центрированное на ней t-распределение евклидовых расстояний до других точек и рассчитывают условные вероятности, отражающие подобие всех остальных точек с ней. Формула вычисления этой условной вероятности параметризована так, чтобы точки в областях с большей плотностью имели меньшую дисперсию. Далее определяют новые положения точек в двумерном пространстве посредством минимизации расстояния Кульбака-Лейблера [239] между распределениями подобия точек в пространстве высокой размерности и в двумерном пространстве.

Карту t-SNE для трех аннотированных библиотек строили на основании 1024-битных фингерпринтов Моргана с радиусом 3 (Рисунок 29). Оказалось, что выборки ингибиторов Mrgo не только мало перекрываются, но и содержат структурно разные классы соединений. Центральная часть облака точек слабо разделена на несколько разреженных кластеров, содержащих пептидомиметики. Область А (Рисунок 29) объединяет молекулы с  $\gamma$ -лактамным фрагментом, в основном из библиотеки AA-Feb21, вследствие чего содержит молекулы, протестированные не только против SARS-CoV-2, но также и против SARS-CoV и MERS-CoV, поскольку этот класс ингибиторов разрабатывали ранее в отношении этих двух вирусов. Кластер Б объединяет аналоги ML-188 — как ковалентные, так и нековалентные, многие из них принадлежат библиотеке AA-Aug21 и были протестированы позднее в ходе пандемии, чем соединения из других аннотированных библиотек. Значительно отделенные от центрального облака точек кластеры В и Г содержат ди- и трипептиды (В) или соединения-фрагменты (Г), имеющие структурные гетеробициклические фрагменты с разными гетероатомами в пятичленном ядре. Некоторые соединения, принадлежащие к одной серии, помещаются не в соответствующие кластеры, а в другие на основе наличия определенных заместителей, например, трифторметила (кластер Д) или N-фениламина (кластер Е). Два плотных кластера Ж и З объединяют «клевероподобные» структуры с пятичленным фрагментом ядра (Ж) и аналоги перампанела с  $\alpha$ -пиридановым ядром (З). Сильно отделенный кластер И — еще один пример кластера, населенного различными соединениями, имеющими специфические заместители — четырехчленные циклы с гетероатомами азота или кислорода. Сильно отделенный кластер К объединяет серии соединений из библиотеки AA-Aug21 с изохинолиновым, тетрагидроизохинолиновым и нафтиридиновым фрагментами, а также включает одно соединение из AA-Aug20.

Чтобы сравнить степень кластеризации трех библиотек, мы рассчитали энтропию Шеннона для проекции результатов t-SNE на квадратную сетку в диапазоне от -87 до 87 с шагом сетки 3. Самая высокая энтропия — 5,61 — наблюдалась для библиотеки AA-Aug21. Энтропия

для AA-Aug20 и AA-Feb-21 составила 5,38 и 5,16 соответственно. Это соотносится с видом K-функции Рипли [240] (Рисунок 30), которая лучше соответствует равномерному распределению для библиотеки AA-Aug21, чем для двух других библиотек. Это указывает на более низкую кластеризацию данных AA-Aug21 по сравнению с библиотеками AA-Aug20 и AA-Feb-21.

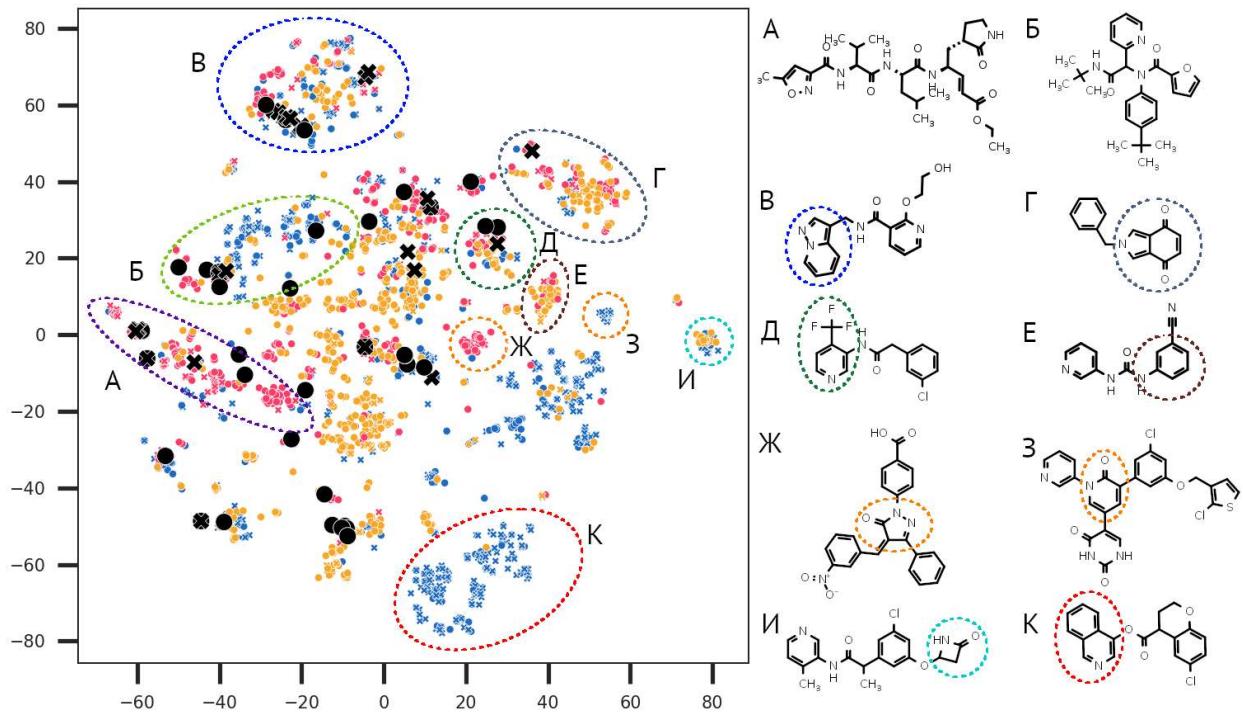


Рисунок 29. Карта t-SNE для экспериментально исследованных ингибиторов Mpro из библиотек AA-Aug20 (желтый), AA-Feb21 (красный) и AA-Aug21 (синий). Маленькие точки соответствуют соединениям, присутствующим в одной библиотеке, большие черные точки представляют соединения из двух библиотек. Активные соединения обозначены крестиками, неактивные кружками.

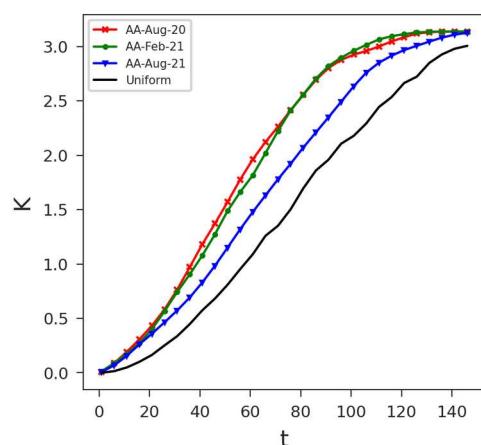


Рисунок 30. Зависимость K-функции Рипли от порогового значения для результатов t-SNE для библиотек AA-Aug21, AA-Feb21 и AA-Aug21 и равномерно распределенных сгенерированных точек.

Разнообразие активных и неактивных соединений для каждой библиотеки оценили, рассчитав для всех молекул распределение попарных индексов Танимото (TI) между фингерпринтами Моргана (Рисунок 31). Молекулы библиотеки AA-Aug20 имеют большее разнообразие, чем молекулы из двух других библиотек. Распределения TI для активных и неактивных молекул библиотеки AA-Aug20 очень похожи. Распределения TI для активных и неактивных молекул библиотек AA-Aug21 и AA-Feb21, как правило, более похожи, чем неактивные (распределение смещено в сторону более высокого TI), при этом активные молекулы AA-Feb21 наименее разнообразны среди трех библиотек. Более высокий процент активных соединений в этих библиотеках при меньшем их разнообразии по сравнению с неактивными, по-видимому, следует из характера источника данных и соответствует тенденции предпочтительно публиковать серии соединений, обладающих активностью. Такие серии обычно появляются в результате целевых программ разработки лекарств, которые не могли быть завершены к середине 2020 года, поэтому AA-Aug20 содержит больше результатов репозиционирования и скрининга природных соединений и фрагментных библиотек.

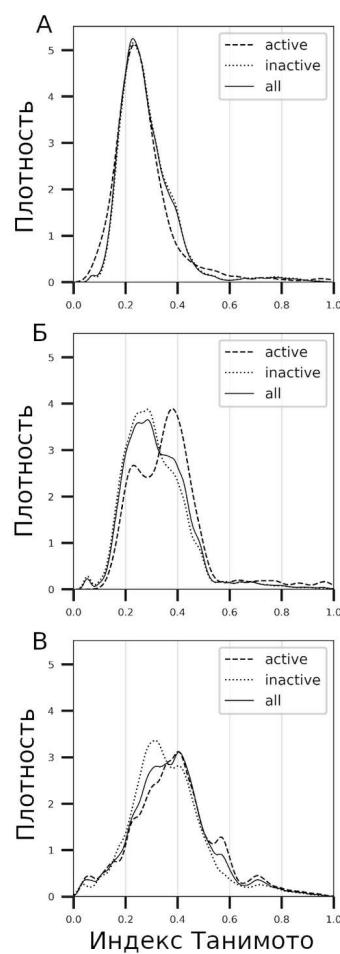


Рисунок 31. Распределение индекса подобия Танимото для молекул в активных (штриховые линии) и неактивных (пунктирные линии) подвыборках и полных (непрерывные линии) библиотеках AA-Aug20 (А), AA-Feb21 (Б) и AA-Aug21 (В).

На базе результатов литературного поиска соединений, экспериментально протестированных в отношении главной протеазы различных коронавирусов, была сформирована “База данных соединений с экспериментально определенной активностью против главной протеазы  $\beta$ -коронавирусов”, на нее получена государственная регистрация в Реестре баз данных № 2024625235 [225]. Эта аннотированная по активности выборка дополняет другие опубликованные базы данных и может быть использована для валидации моделей и методов виртуального скрининга ингибиторов Mpro SARS-CoV-2 и других коронавирусов. Она была применена вместе с выборками, опубликованными другими авторами, в качестве средства внешнего контроля при разработке методики систематического отбора белковых структур для виртуального скрининга потенциальных ингибиторов Mpro SARS-CoV-2 методом ансамблевого докинга.

## Глава 4. Разработка метода систематического ансамблевого докинга<sup>6</sup>

Библиотеки малых молекул, аннотированные по активности в отношении Mpro SARS-CoV-2, использовали для сравнения ансамблей и подбора оптимальной консенсусной оценки виртуального скрининга. Чтобы оценить применимость разработанной процедуры для различных мишеней, по тому же протоколу был выбран ансамбль и проведен ретроспективный виртуальный скрининг библиотеки DUD-е для циклин-зависимой киназы 2. Также было исследовано влияние введения ограничений на конформационный поиск во время докинга на надежность виртуального скрининга.

### 4.1. Виртуальный скрининг методом систематического ансамблевого докинга

Процедура виртуального скрининга была реализована на базе программы DOCK6.9. Подготовку структур белков к докингу проводили при помощи инструмента DockPrep программы UCSF Chimera 1.15: удаляли молекулы растворителя и лиганд, добавляли атомы водорода и рассчитывали частичные заряды на атомах белка по схеме силового поля AMBER ff14SB. Затем генерировали молекулярную поверхность при помощи программы *dms* и обратное представление сайта связывания, заполненного сферами с радиусом от 1.4 до 4 Å при помощи программы *sphgen\_cpp.1.2*. Выбор сфер для докинга в структуры ансамбля осуществляли на основе координат атомов аналога перампанела (16) [132], извлеченных из структуры 7L13, предварительно пространственно выровненной со всеми структурами ансамбля по цепи А. Выбирали сферы в радиусе 9 Å от лиганда при помощи программы *sphere\_select.1.0*. Бокс для докинга генерировали с отступом 5 Å от края сфер (Рисунок 24). Сеточное поле генерировали в программе *grid* с шагом сетки 0.3 Å, с учетом оценки контактов на расстоянии до 4.5 Å, оценки перекрывания атомов (*bump\_filter*) с возможностью сближения атомов до 75% от суммы их вандерваальсовых радиусов, с максимальным расстоянием для расчёта энергии межатомных взаимодействий 9999 Å, с линейной зависимостью

6 При работе над данным разделом диссертации использованы материалы следующих публикаций автора, в которых, согласно Положению о присуждении ученых степеней в МГУ, отражены основные результаты, положения и выводы исследования:

1. Фомина А.Д., Палюлин В.А., Осолодкин Д.И. Моделирование структуры белка NS1 ортофлавивирусов по гомологии для виртуального скрининга потенциальных лигандов // Биомедицинская химия – 2024 – т. 70 (6) – с. 456–468
2. Fomina A. D., Uvarova V. I., Kozlovskaya L. I., Palyulin V. A., Osolodkin D.I., Ishmukhametov A. A. Ensemble Docking Based Virtual Screening of SARS-CoV-2 Main Protease Inhibitors // Molecular Informatics – 2024 – V.43. - e202300279.
3. Zakharova M.Yu., Kuznetsova A.A., Uvarova V.I., Fomina A.D., Kozlovskaya L.I., Kaliberda E.N., Kurbatskaia I.N., Smirnov I.V., Bulygin A.A., Knorre V.D., Fedorova O.S., Varnek A., Osolodkin D.I., Ishmukhametov A.A., Egorov A.M., Gabibov A.G., Kuznetsov N.A. Pre-Steady-State Kinetics of the SARS-CoV-2 Main Protease as a Powerful Tool for Antiviral Drug Discovery // Frontiers in Pharmacology – 2021 – V. 12. – P. 773198.
4. Фомина А.Д., Осолодкин Д.И. Программный комплекс для систематического ансамблевого докинга // Свидетельство о государственной регистрации программы для ЭВМ № 2024684465. – Номер заявки 2024684027. – Дата поступления 17.10.24. – Дата государственной регистрации в Реестре программ для ЭВМ 17.10.24.
5. Фомина А.Д., Осолодкин Д.И. База данных соединений с экспериментально определенной активностью против главной протеазы β-коронавирусов // Свидетельство о государственной регистрации базы данных № 2024625235. – Номер заявки 2024624786. – Дата поступления 29.10.2024 – Дата государственной регистрации в Реестре баз данных 15.11.2024 Бюл. № 11

диэлектрической проницаемости от расстояния с коэффициентом 4 [241].

Был проведен докинг всех соединений библиотек *AA-Aug20*, *AA-Feb21*, *AA-Aug21* и *DD-ZINC* во все структуры ансамблей. Генерировали не более 250 способов связывания, которые кластеризовали по СКО с порогом 2 Å. Из них отбирали до 25 конформеров с наивысшими оценками, по одному для каждого кластера, и для каждой пары лиганд-белок сохраняли конформацию с наилучшей оценкой *grid\_score* (т. е. той, которая имеет наибольшее по модулю отрицательное значение). Примеры входных файлов приведены в дополнительных материалах.

Ранжирование результатов докинга основывалось на консенсусных оценках. Чтобы найти ту, которая позволяет получить наиболее точное разделение активных и неактивных соединений, сравнивали минимальную (*minscore*), максимальную (*maxscore*), среднюю (*meanscore*) и медианную (*median*) оценку по ансамблю, «евклидову» (*euclidean*) оценку, определяемую как отрицательный квадратный корень из суммы квадратов индивидуальных оценок по ансамблю, и «гармоническую» (*harmscore*) оценку, определяемую как отрицательное гармоническое среднее по ансамблю абсолютных значений оценки докинга. «Евклидовы» и «гармонические» оценки брали с обратным знаком, чтобы они соответствовали другим оценкам, поскольку значение *grid\_score* обычно отрицательное. Оценки докинга обрабатывали с помощью пользовательских сценариев для Python 3.9 [209]. Анализировали распределения консенсусных оценок и ROC-кривые.

Была исследована способность ансамблевого докинга отделять истинно активные молекулы от истинно неактивных в аннотированных библиотеках и от случайных соединений в разнообразной библиотеке. Для результатов докинга библиотек *AA-Aug20*, *AA-Feb21*, *AA-Aug21* и *DD-ZINC* рассчитали распределение оценки докинга (Рисунок 32, Приложение А: Рисунки A1-5) для каждой библиотеки и каждого ансамбля. Если между левыми плечами распределений оценок активных и неактивных соединений был разрыв, предполагали, что ансамбль может различать активные и неактивные соединения, давая большую по модулю оценку активным. Только распределения оценок для библиотеки *AA-Feb21* имеют этот разрыв для всех консенсусных функций. Самый большой разрыв наблюдается для распределений минимальных оценок. Другие метрики обеспечивают гораздо более короткие плечи в области наибольших по модулю отрицательных оценок. Для библиотеки *AA-Aug21* почти не наблюдаются различия между распределениями для активных и неактивных соединений, а в случае *AA-Aug20* активные соединения имеют меньшие по модулю оценки, чем неактивные. Тем не менее, распределения оценок докинга всех библиотек характеризуются расположением левых плечей в диапазоне более низких оценок по сравнению с распределением оценок докинга для набора данных *DD-ZINC*. В дальнейшем мы решили использовать минимальную оценку в качестве стандартной

метрики, так как она лучше всего позволяет отделить активные соединения от неактивных на основании сравнения распределений оценок докинга и при этом является наиболее интерпретируемой метрикой, поскольку мы можем напрямую анализировать соответствующие ей способы связывания.

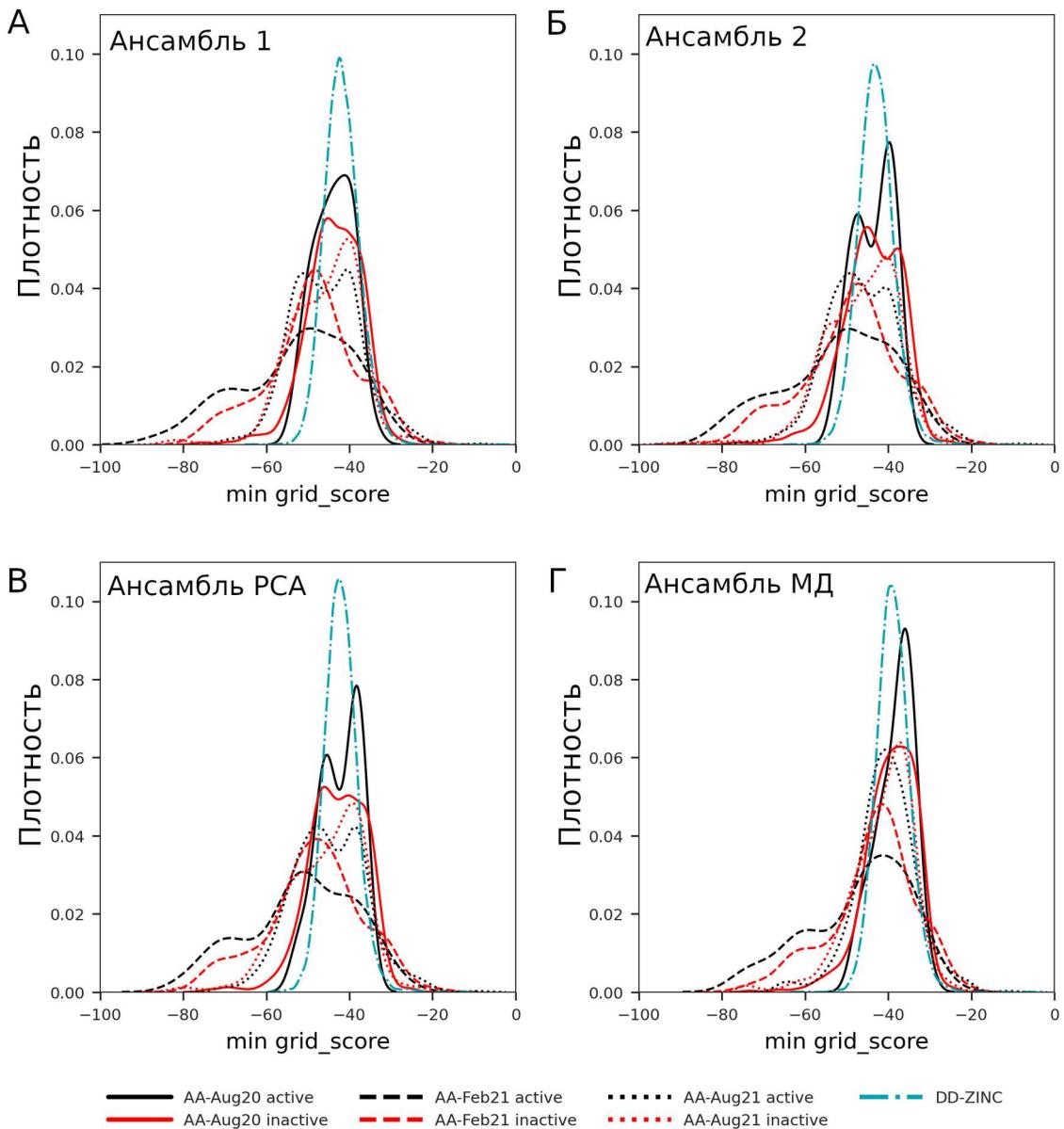


Рисунок 32. Распределения минимальных оценок докинга для активных (active) и неактивных (inactive) соединений из библиотек AA-Aug20, AA-Feb21, AA-Aug21 и DD-ZINC, докированных в Ансамбль 1 (А), Ансамбль 2 (Б), Ансамбль PCA (С) и Ансамбль МД (Д)

Был проведен анализ ROC-кривых результатов докинга библиотек во все рассматриваемые ансамбли. Использовали как исходные библиотеки AA-Aug20, AA-Feb21 и AA-Aug21, так и те же библиотеки, дополненные соединениями из DD-ZINC в качестве неактивных соединений. Наименьшее обогащение и точность классификации наблюдались для библиотеки AA-Aug20, ROC AUC (площадь под ROC-кривой) для результатов докинга которой в ансамбль

PCA и ансамбль МД были ниже 0,5. Докинг библиотеки AA-Feb21 привел к наиболее корректной классификации среди трех наборов данных. Более того, ROC-кривые для этой библиотеки демонстрируют раннее обогащение результатов докинга активными соединениями. ROC-кривые для AA-Aug21 соответствуют почти случайной классификации. Эти результаты подтверждаются ROC-кривыми, опубликованными в других работах по виртуальному скринингу на основе докинга [230], [102], [242]. Несмотря на то, что ансамбли не могут эффективно различать активные и неактивные вещества из серий родственных соединений, все ансамбли хорошо работают в сценарии высокопроизводительного ВС: ROC-кривые, построенные для библиотек, дополненных неактивными соединениями из DD-ZINC, имеют гораздо больший наклон в начале кривой по сравнению с кривыми для исходных наборов данных. Самое высокое раннее обогащение, соответствующее самому высокому наклону в начале кривой, наблюдается для набора данных AA-Feb21.

Судя по сравнению ROC-кривых, качество ранжирования соединений при докинге в разные ансамбли примерно одинаково (Рисунок 33). Самые высокие значения ROC AUC наблюдаются для *Ансамбля 1*, но различия с другими ансамблями незначительны. Однако, с увеличением доли неактивных соединений разница в поведении ансамблей становится более заметной. Результаты докинга в *Ансамбль МД* характеризуются самыми низкими значениями ROC AUC и наименьшими наклонами кривых, что указывает на большее количество ложноположительных результатов при равном с другими ансамблями количестве истинно положительных результатов. Это коррелирует со сравнением распределений оценок докинга: для ансамбля МД наблюдали наименьший среди четырех ансамблей интервал между оценками активных соединений и оценками разнообразных неактивных соединений. Докинг в *Ансамбль 1* в сценарии высокопроизводительного скрининга оказывается наиболее качественным, исходя из сравнения ROC-кривых для четырех ансамблей. Результаты докинга в этот ансамбль имеют наибольшие ROC AUC и наибольшие наклоны кривых на начальных участках, что свидетельствует о значительном обогащении результатов докинга истинно активными соединениями. *Ансамбль 2* и *Ансамбль PCA* демонстрируют схожие результаты. Однако для библиотек AA-Aug20 и AA-Aug21 *Ансамбль 2* имеет большую площадь под ROC-кривой, чем *Ансамбль PCA*.

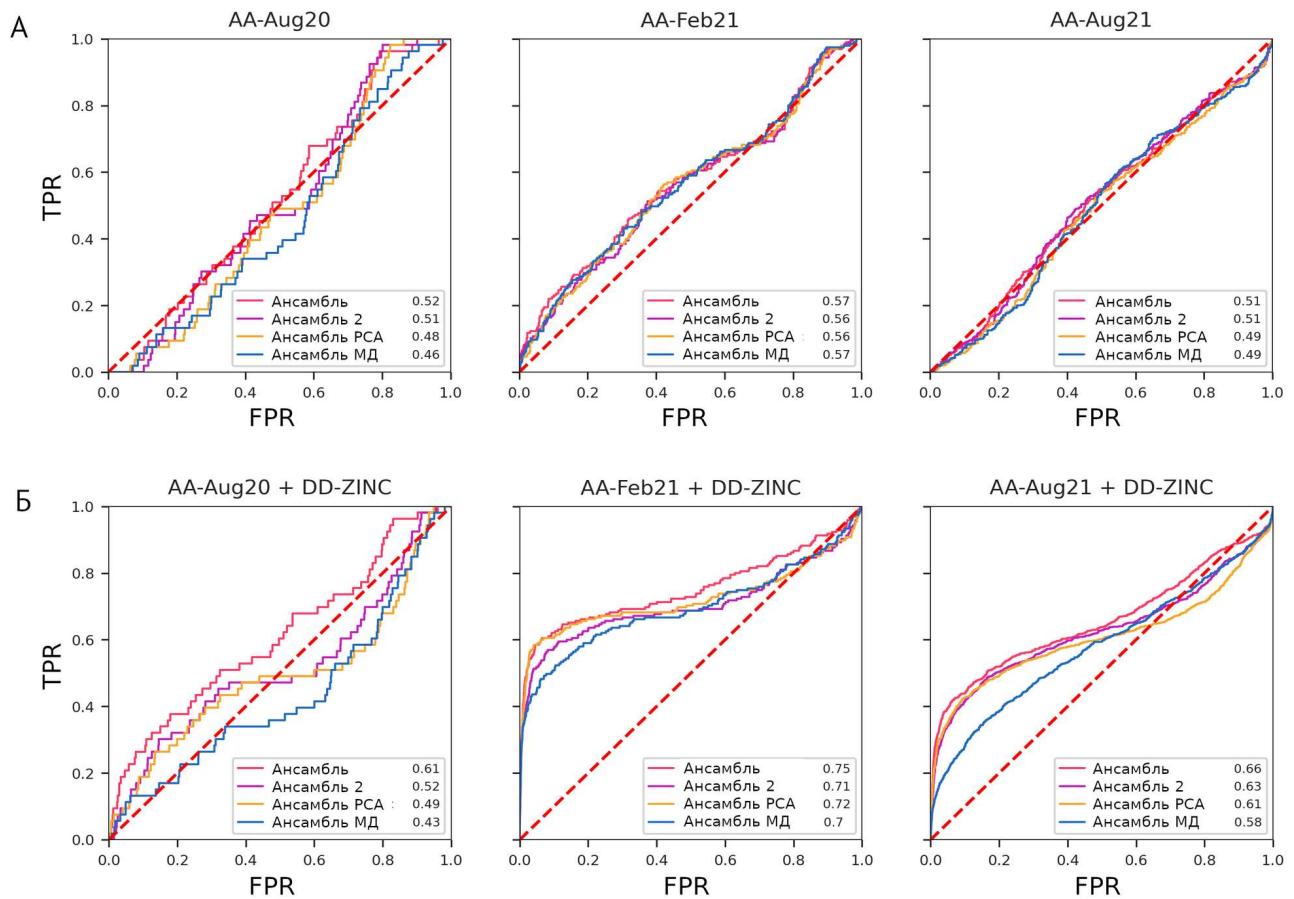


Рисунок 33. ROC-кривые на основе минимальной оценки докинга для библиотек AA-Aug20, AA-Feb21 и AA-Aug21 без дополнения неактивными соединениями (А) и с дополнением (Б). FPR и TPR – true positive rate и false positive rate соответственно.

Эффективность консенсусных метрик также оценили при помощи ROC-анализа. Различные метрики работали одинаково для аннотированных библиотек, однако при дополнении библиотек разнообразными соединениями в качестве неактивных, различия становились более выраженным (Рисунок 34). ROC AUC в случае всех метрик для исходных наборов данных лишь немного превышала 0,5, что свидетельствует о том, что классификация была не лучше случайной (Приложение А: Рисунки A7-A16). После дополнения библиотек разнообразными соединениями ROC-кривые стали явно различаться для разных метрик (Рисунок 33Б, 34Б). “Евклидовы” и минимальные оценки позволяют получить лучшие результаты. Точность классификации результатов докинга, ранжированных при помощи всех остальных консенсусных оценок, была значительно ниже. Использование максимальной оценки по ансамблю для ранжирования результатов докинга дополненных выборок привело к наихудшим результатам со средними значениями ROC AUC < 0.5. Это соотносится с распределениями максимальных оценок докинга (Приложение А: Рисунок A7), где левые плечи графиков распределений для активных и неактивных для всех четырех ансамблей лежат в одном интервале. Это показывает, что минимальная оценка работает приемлемо в сценарии виртуального скрининга и напрямую

интерпретируется, при этом *Ансамбль 1* наиболее эффективен.

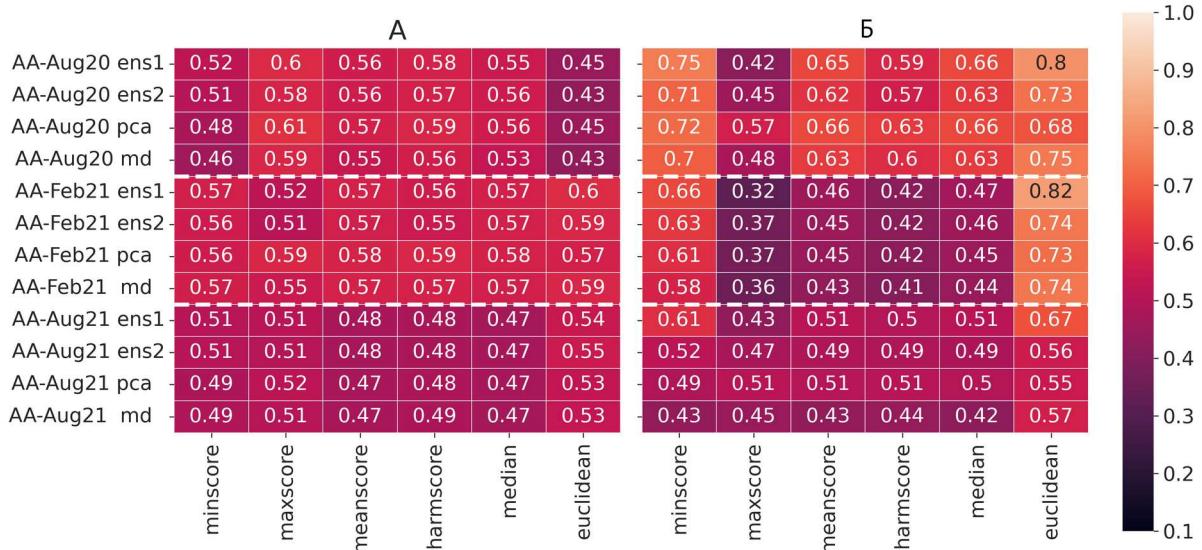


Рисунок 34. ROC AUC для 4 ансамблей с различными консенсусными оценками для библиотек *AA-Aug20*, *AA-Feb21* и *AA-Aug21* без дополнения неактивными соединениями (А) и с дополнением (Б).

Для дальнейшей работы были выбраны *Ансамбль 1* и *Ансамбль 2*. Эти ансамбли позволяют наиболее полно отделить истинно активные соединения в аннотированных библиотеках от неактивных соединений, а также наиболее полно и реалистично отражают конформационную подвижность аминокислотных остатков, составляющих активный центр фермента. Ансамбль, отобранный при помощи анализа главных компонент, обладает менее стабильной ранжирующей способностью и меньшими углами наклона начальных участков ROC-кривых, что свидетельствует о более позднем обогащении хитов активными соединениями; ансамбль на основе траектории молекулярной динамики обладает наиболее низкой точностью ранжирования среди всех ансамблей.

#### 4.2. Применение методики систематического ансамблевого докинга к CDK2

Чтобы подтвердить возможность использования методики систематического отбора ансамбля по СКО конформаций остатков активного центра для различных белковых мишней, она была без изменений применена для структур циклин-зависимой киназы CDK2. Этот фермент хорошо изучен, для него доступно много структур в RSCB PDB и специально разработана аннотированная библиотека DUD-e [47].

Структуры CDK2 искали в PDB по запросу "Full Text = "CDK2" AND (Scientific Name of the Source Organism = "Homo sapiens" AND (Enzyme Classification Name = "Transferases" AND Enzyme Classification Number IS ANY OF (1, 2, 3, 4, 5, 6, 7)))". Были извлечены 254 структуры, 151

структура прошла фильтрацию по наличию мутаций в активном центре и полноте структуры. Структуры были выровнены, очищены от альтернативных конформаций боковых цепей аминокислотных остатков, и для них было рассчитано СКО между аминокислотными остатками, формирующими активный центр: 10-15, 31-33, 64, 80-86, 89, 131-134, 144-148, 162 (Рисунок 35). На тепловой карте выделяются два кластера структур с высоким межклассовым СКО. Причина этого в том, что связывание циклина вызывает конформационные изменения в структуре фермента, из-за которых становится возможным связывание АТФ [243]. Составление ансамбля проводили на основе структур, содержащих ингибитор в активном центре (при подготовке структур все лиганды удаляли). Поскольку CDK2 активна в мономерной форме, при отборе ансамбля потребовалось еще меньше ручных операций по сравнению с Mpro. На основании разнообразия в ансамбль были выбраны 4 структуры: 3UNJ, 6GUB, 8FOW и 8H6T.

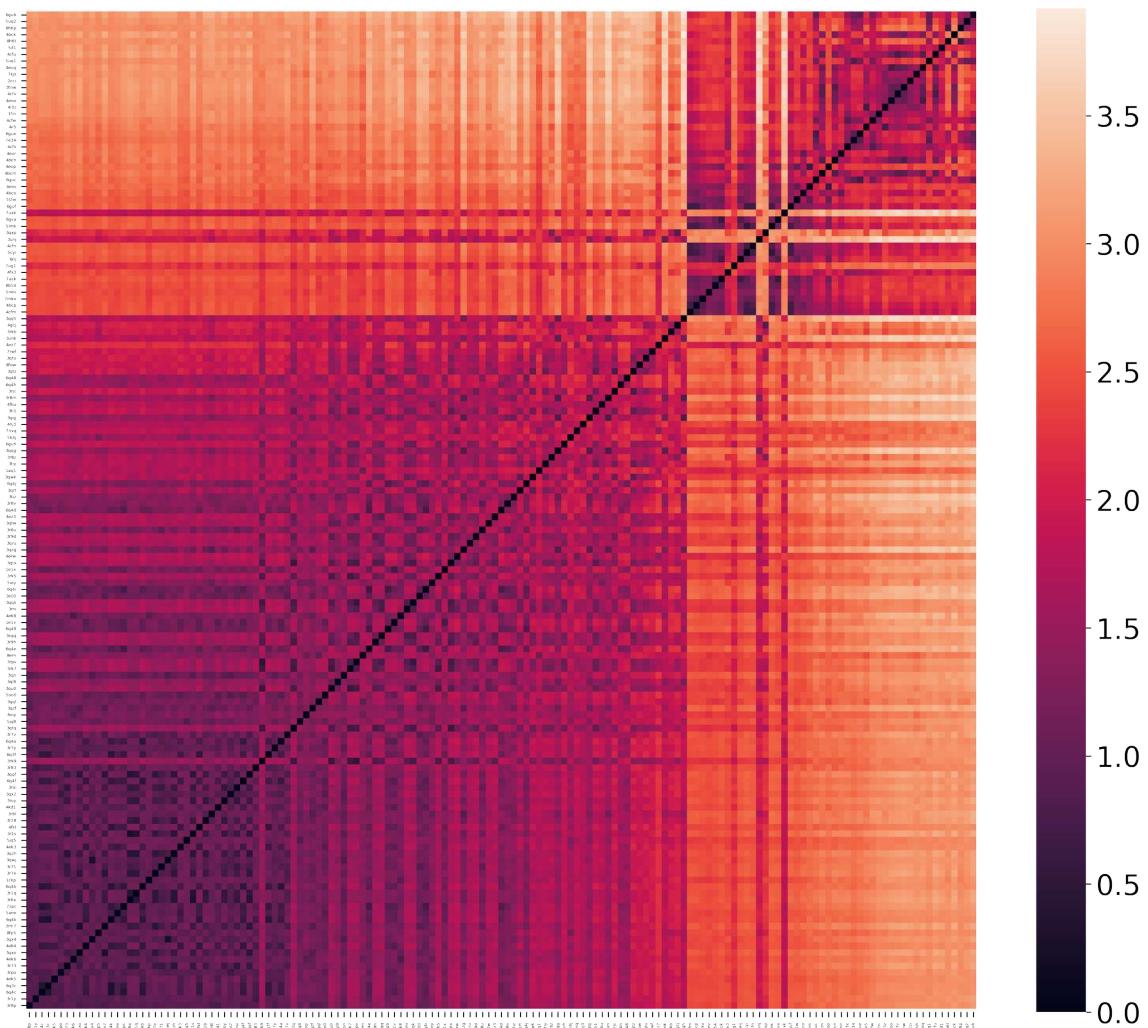


Рисунок 35. Тепловая карта попарных СКО между координатами атомов остатков, формирующих активный центр фермента CDK2.

Был проведен докинг выборки DUD-e для циклин-зависимой киназы CDK2 в структуры

ансамбля и в структуру 1CKP, использованную авторами выборки [47]. Рассчитанное значение ROC AUC 0,65 было одинаковым как для консенсусной оценки по ансамблю, так и для докинга в структуру 1CKP. Индивидуальные ROC AUC для структур ансамбля находились в диапазоне от 0,57 (PDB ID 8H6T) до 0,61 (PDB ID 8FOW) (рисунок 36). Несмотря на то, что значения ROC AUC были довольно низкими из-за неоптимальной оценочной функции, эти результаты показывают, что разработанная процедура позволяет выбрать ансамбль структур, представляющих конформационное пространство белка и обеспечивающих метрики качества виртуального скрининга по крайней мере на уровне одиночной структуры. При этом систематический отбор структур позволяет нивелировать фактор случайности при выборе структуры для виртуального скрининга.

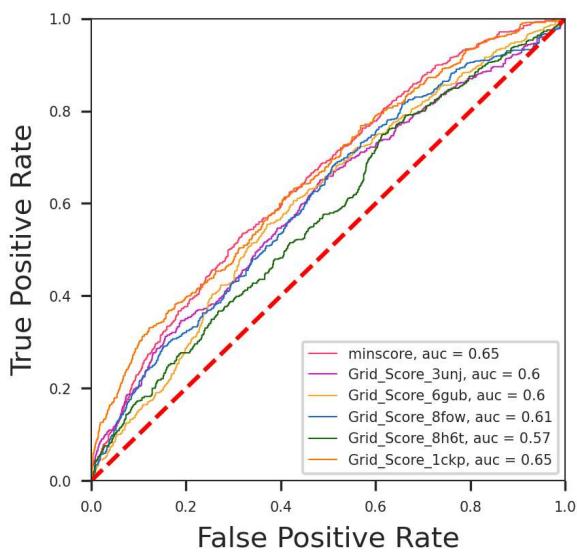


Рисунок 36. ROC-кривые для результатов докинга библиотеки DUD-e в индивидуальные структуры ансамбля CDK2, в структуру 1CKP и с учетом консенсусной оценки minscore.

### 4.3. Докинг со структурными ограничениями

#### 4.3.1. Анализ взаимодействий

С целью подбора конформационных ограничений для докинга был проведен анализ контактов лигандов с белком для структур Mpro SARS-CoV-2, депонированных в PDB по состоянию на март 2021 года. Анализировали 234 структуры, содержащие ковалентные и нековалентные лиганды и не содержащие мутаций, окисленных или незрелых форм белка. В этих структурах были определены пары неводородных атомов лиганда и белка, находящиеся на расстоянии  $\leq 3,5$  Å. Были отдельно выделены атомы, образующие водородные связи, “хорошие” взаимодействия (углерод – углерод) и “плохие” взаимодействия (углерод – гетероатом). Фингерпринты взаимодействия были сформированы в Flare 5.0.0 с помощью пользовательского сценария на языке Python3 (Приложение Б). Ближайшие атомы лиганда и каждого остатка

выделены как контактирующие. Водородные связи задавали с помощью встроенной функции `flare.contacts.h_bonds`. Каждый элемент построенного таким образом вектора соответствовал одному остатку фермента и представлял собой тип контакта лиганда и остатка. Фингерпринты взаимодействия визуализировали с использованием метода тепловых карт, реализованного в Python 3.9 с библиотеками *Pandas 1.4.1*, *NumPy 1.22.2*, *Matplotlib 3.5.1* и *seaborn 0.11.2*. В каждой ячейке карты цветом показан тип взаимодействия, который лиганд образует с соответствующим остатком. Можно видеть, что есть лиганда, образующие связи в активном центре, и аллостерические лиганда. Наибольший интерес представляют остатки, наиболее часто формирующие с лигандами водородные связи.

На основании анализа фингерпринтов взаимодействия (Рисунок 37) в качестве структурных ограничений были выбраны амидный NH основной цепи Glu166 и  $\epsilon$ -NH боковой цепи His163 как доноры водородной связи.

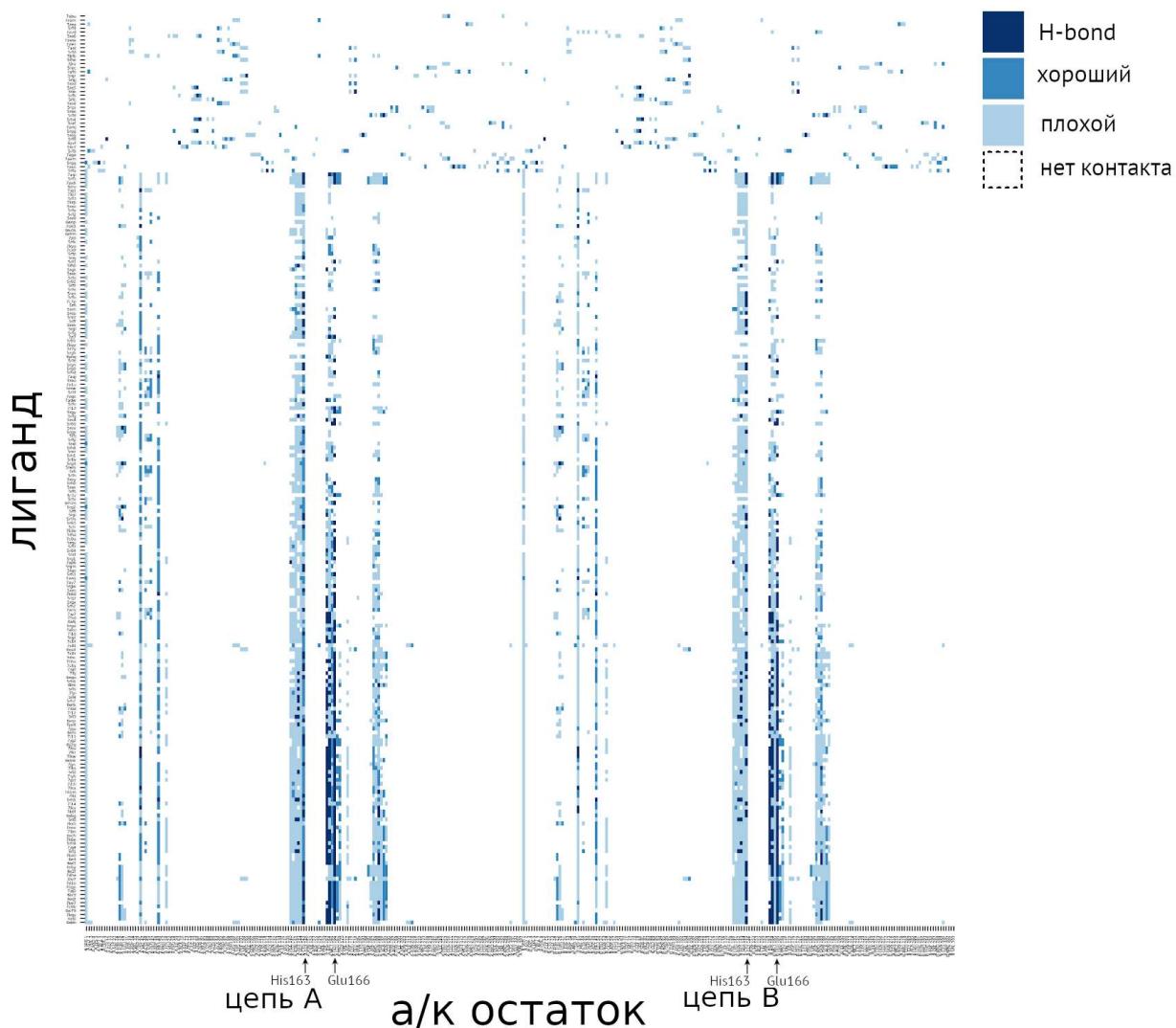


Рисунок 37. Тепловая карта взаимодействий остатков белка в комплексах с лигандами.

#### 4.3.2. Влияние структурных ограничений на результаты докинга

Ограничения на конформационный поиск при докинге вводили при помощи модуля Chemical Matching программного комплекса DOCK6.9. В этой программе подбор способа связывания лиганда осуществляется за счет последовательного размещения его фрагментов в центрах сфер, формирующих обратный образ кармана связывания. Введение ограничений реализовано за счет разметки (“окрашивания”) сфер в областях, где должны формироваться необходимые водородные связи. Таким образом, приоритизируются способы связывания, в которых фрагменты лиганда, содержащие акцептор водородной связи, оказывались в “окрашенных” сферах.

Был проведен докинг библиотеки AA-Feb21 в одни и те же структуры Ансамбля 1 с “окрашенными” (constrained) и “неокрашенными” (simple) сферами. Для сравнения результатов были построены распределения минимальных оценок докинга (Рисунок 38А). В обоих случаях распределение минимальных оценок докинга для активных соединений имеет больший размах, чем для неактивных, и левое плечо в области более низких оценок, однако распределения и ROC-кривые не различаются (Рисунок 38Б) для докинга с ограничениями и без.

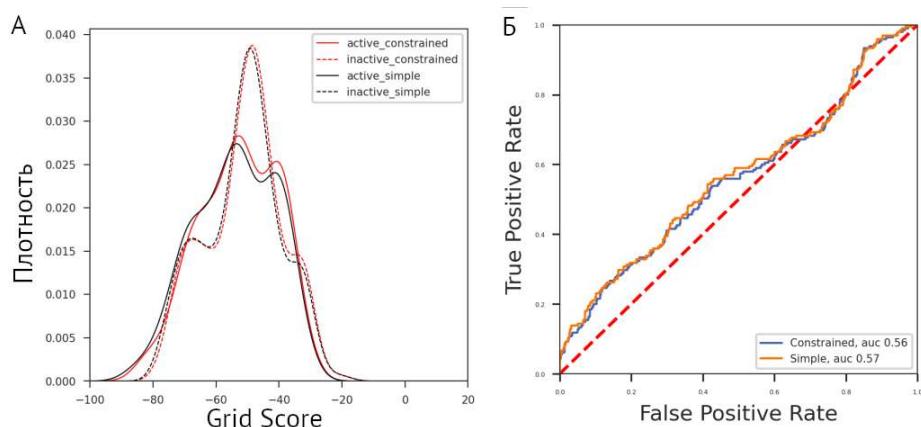


Рисунок 38. (А) – Распределения минимальных оценок докинга библиотеки AA-Feb21 – отдельно для активных (непрерывная линия) и неактивных (пунктирная линия) соединений. Красный – докинг с конформационными ограничениями, черный – докинг без ограничений. (Б) – ROC-кривые результатов докинга библиотеки AA-Feb21, ранжированных по минимальной оценке докинга. Оранжевый – докинг с ограничениями, синий – обычный докинг.

Были рассчитаны СКО между результатами докинга одних и тех же молекул с ограничениями и без ограничений. Учитывали только конформации, соответствующие минимальным оценкам докинга. При значениях  $\text{СКО} \leq 2,5 \text{ \AA}$  можно считать, что молекула принимает в двух системах одинаковые конформации. Распределение значений попарных СКО представлено на Рисунке 39. Наибольшим сходством конформаций, полученных при наличии и в отсутствие ограничений, обладают результаты докинга в структуры 7AKU (49,43%) и 7BFB (45,9%).

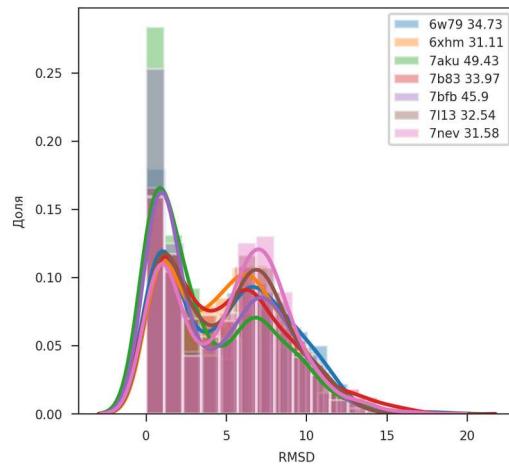


Рисунок 39. Гистограмма без накопления и плотность распределения попарных СКО между результатами докинга для каждой молекулы с ограничениями и без ограничений. Для каждой структуры приведен процент лигандов, для которых СКО ≤ 2,5.

Было проведено сравнение рангов, присваиваемых молекулам на основании оценок докинга в индивидуальные структуры *Ансамбля 1* при наличии и отсутствии структурных ограничений. Тепловая карта коэффициентов корреляции Спирмена для результатов докинга в каждую отдельную структуру представлена на Рисунке 40. Чем ближе значение коэффициента к 1, тем выше корреляция между ранжированными списками. Полученные значения свидетельствуют о высокой корреляции рангов. Наиболее высокие значения (0,96) наблюдаются для корреляции рангов, полученных для одной и той же структуры с ограничениями и без.

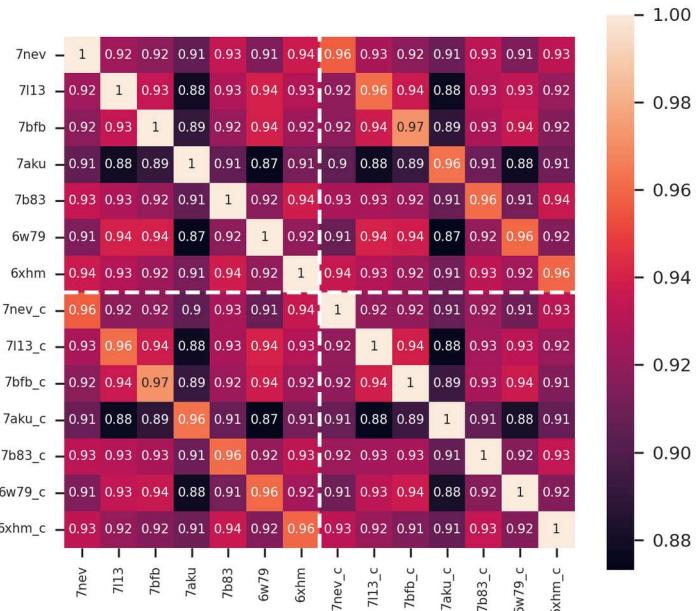


Рисунок 40. Тепловая карта коэффициентов корреляции рангов Спирмена для результатов докинга в индивидуальные структуры с ограничениями (структуры с индексом "с") и без ограничений.

Чтобы выяснить, приводит ли введение ограничений на конформационный поиск к увеличению количества результатов докинга, обладающих желательными взаимодействиями, были построены графики частоты образования водородных связей, а также “хороших” и “плохих” контактов для остатков, формирующих активный центр (Рисунок 41), в зависимости от наличия ограничений. Большинство водородных связей образуется с Glu166 и His163 вне зависимости от наличия или отсутствия ограничений, причем для обоих остатков частоты образования водородных связей приблизительно равны для всех структур. Частоты образования “хороших” и “плохих” контактов для большинства а/к остатков сильнее различаются от структуры к структуре, чем частоты образования водородных связей, что отражает разнообразие конформаций аминокислотных остатков, формирующих активный центр. При этом количество контактов каждого типа примерно равно для результатов докинга с ограничениями и без.

Для дальнейшей работы с Mpro SARS-CoV-2 было решено использовать докинг без конформационных ограничений, так как их использование не приводило к повышению качества виртуального скрининга либо к обогащению результатов докинга конформациями, имеющими больше специфических взаимодействий, но при этом затрудняло автоматизацию процесса, так как метод требует ручного отбора сфер, задающих область, в которых должны присутствовать ограничения.

Результаты работы были имплементированы в программном обеспечении “Программный комплекс для систематического ансамблевого докинга”, зарегистрированном в Реестре программ для ЭВМ за № 2024684465 [209]. Как показано на примере главной протеазы SARS-CoV-2 и CDK2, программа может быть применена для отбора структур и анализа результатов ансамблевого докинга для любого белка, для которого доступно значительное количество структурной информации. Доступно два метода отбора ансамблей – по СКО или по кластеризации результатов анализа главных компонент для координат атомов остатков активного центра, которые можно указать произвольно. В качестве программы для докинга может быть использован программный комплекс DOCK6 или программа Flare, а так же другие программы докинга, которые выдают результаты в похожих форматах. Разработанное программное обеспечение используется для виртуального скрининга потенциальных ингибиторов главной протеазы SARS-CoV-2.

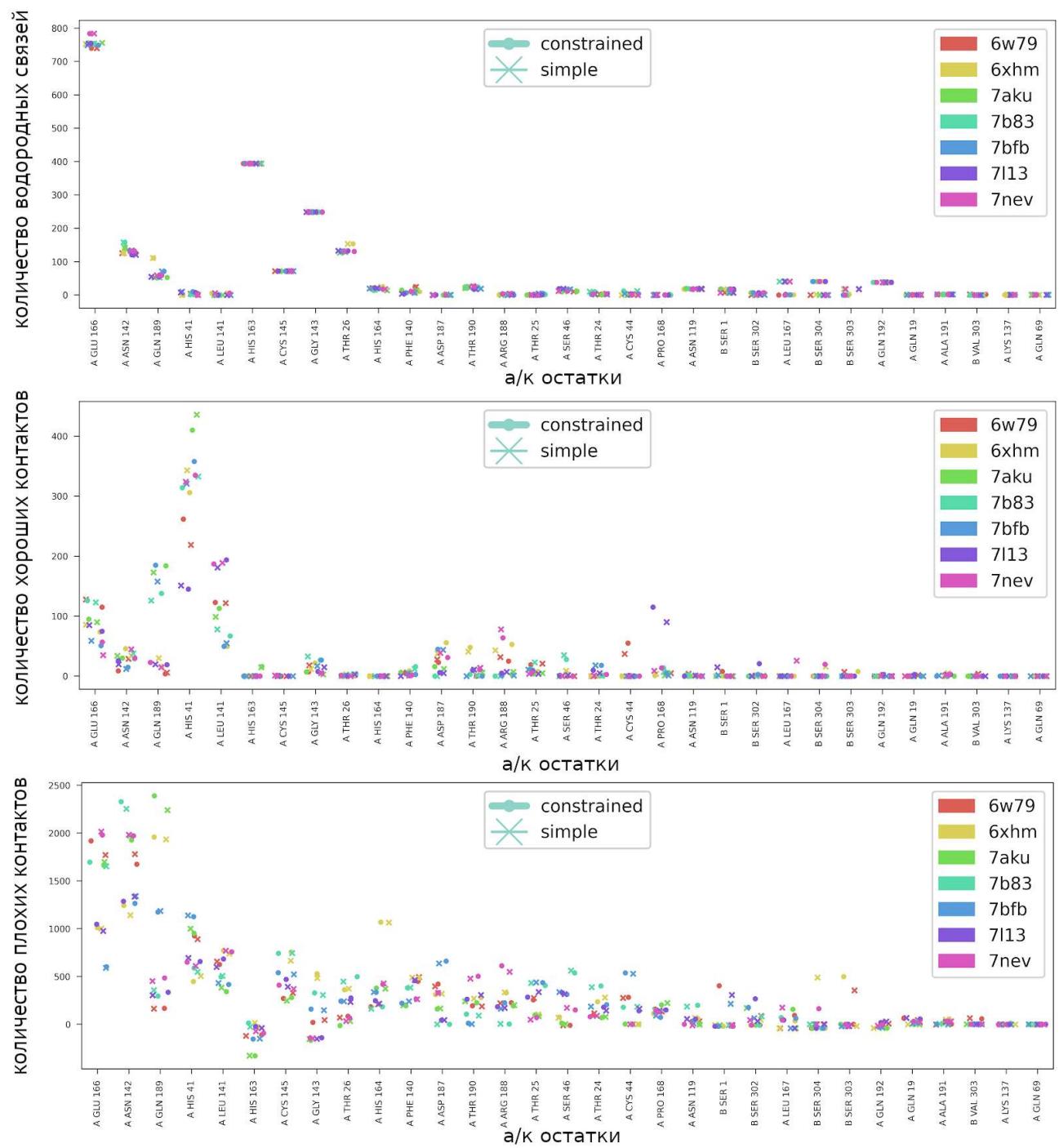


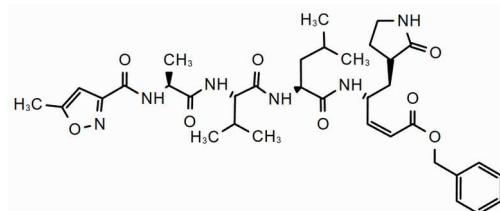
Рисунок 41. Частота встречаемости контактов трех типов с остатками белков.

## Глава 5. Виртуальный скрининг ингибиторов главной протеазы SARS-CoV-2<sup>7</sup>

Поиск потенциальных ингибиторов Mpro SARS-CoV-2 был очень актуальной задачей в ходе пандемии COVID-19. В рамках этой работы была поставлена цель идентификации новых скаффолдов потенциально активных против Mpro коммерчески доступных соединений для экспериментального тестирования. Для этого до разработки методики систематического ансамблевого докинга был проведен предварительный виртуальный скрининг методом молекулярного докинга в одну структуру белка. Так были обнаружены новые скаффолды потенциально активных в отношении Mpro соединений, которые в дальнейшем исследовали при помощи систематического ансамблевого докинга.

### 5.1. Поиск новых хемотипов ингибиторов главной протеазы

Процедура виртуального скрининга была реализована методом молекулярного докинга на базе программного комплекса DOCK 6.9. Для докинга была использована структура 6WQF [244], исследованная методом рентгеноструктурного анализа при комнатной температуре, поэтому предположительно наиболее близкая к структуре белка при физиологических условиях [244]. Подготовку структуры белка к докингу и сам докинг проводили аналогично процедуре, описанной в разделе 3.3.1, но выбор сфер для докинга осуществляли на основе координат атомов ингибитора N3 (20) [202], извлеченных из структуры 6LU7 [202], предварительно пространственно выровненной со структурой 6WQF.



20

Библиотека потенциальных нековалентных ингибиторов протеазы Mpro SARS-CoV-2 была ранее сформирована на основе результатов виртуального лиганд-обоснованного

7 При работе над данным разделом диссертации использованы материалы следующих публикаций автора, в которых, согласно Положению о присуждении ученых степеней в МГУ, отражены основные результаты, положения и выводы исследования:

1. Фомина А.Д., Палюлин В.А., Осолодкин Д.И. Моделирование структуры белка NS1 ортофлавивирусов по гомологии для виртуального скрининга потенциальных лигандов // Биомедицинская химия – 2024 – т. 70 (6) – с. 456–468
2. Fomina A. D., Uvarova V. I., Kozlovskaya L. I., Palyulin V. A., Osolodkin D. I., Ishmukhametov A. A. Ensemble Docking Based Virtual Screening of SARS-CoV-2 Main Protease Inhibitors // Molecular Informatics – 2024 – V.43. – e202300279.
3. Zakharova M.Yu., Kuznetsova A.A., Uvarova V.I., Fomina A.D., Kozlovskaya L.I., Kaliberda E.N., Kurbatskaia I.N., Smirnov I.V., Bulygin A.A., Knorre V.D., Fedorova O.S., Varnek A., Osolodkin D.I., Ishmukhametov A.A., Egorov A.M., Gabibov A.G., Kuznetsov N.A. Pre-Steady-State Kinetics of the SARS-CoV-2 Main Protease as a Powerful Tool for Antiviral Drug Discovery // Frontiers in Pharmacology – 2021 – V. 12. – P. 773198.
4. Фомина А.Д., Осолодкин Д.И. Программный комплекс для систематического ансамблевого докинга // Свидетельство о государственной регистрации программы для ЭВМ № 2024684465. – Номер заявки 2024684027. – Дата поступления 17.10.24. – Дата государственной регистрации в Реестре программ для ЭВМ 17.10.24.
5. Фомина А.Д., Осолодкин Д.И. База данных соединений с экспериментально определенной активностью против главной протеазы β-коронавирусов // Свидетельство о государственной регистрации базы данных № 2024625235. – Номер заявки 2024624786. – Дата поступления 29.10.2024 – Дата государственной регистрации в Реестре баз данных 15.11.2024 Бюл. № 11

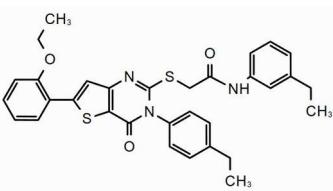
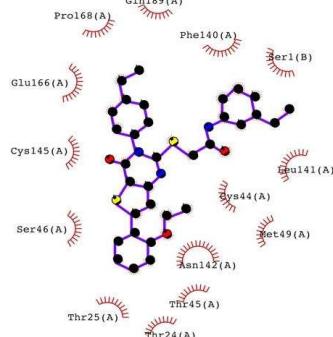
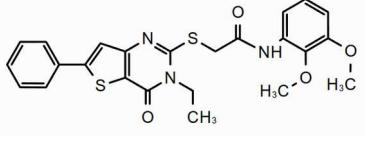
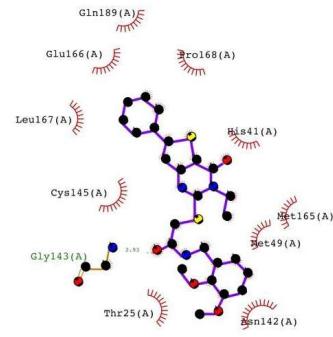
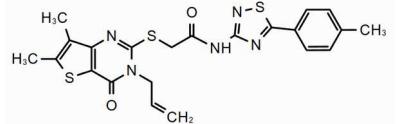
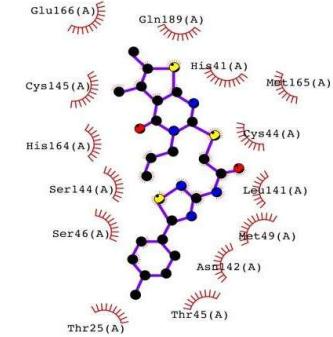
скрининга базы данных ZINC15 методом генеративного топографического картирования [245]. По результатам докинга в структуру 6WQF из 473 соединений были отобраны 39, для которых Grid\_Score был меньше -45. Они были разделены на группы по трем скаффолдам, и из каждой группы были выбраны от 2 до 4 соединений, доступных для непосредственной закупки – одно с лучшим Grid\_Score и еще несколько похожих на него. Результаты докинга отобранных для закупки соединений представлены в Таблице 11. Двумерную визуализацию способа связывания проводили при помощи программы LigPlot+ v.2.2 [246].

Таблица 11. Результаты докинга молекул, закупленных для экспериментального тестирования эффективности ингибиования протеазы Mrgo

Продолжение Таблицы 11.

№	Шифр	ZINC ID	Grid_Score	Структура	Способ связывания по результатам докинга
24	IBS-E0474946	ZINC000045922189	-51,22		
25	IBS-E0475253	ZINC000064699812	-50,82		
<b>Скаффолд 2</b>					
26	IBS-E0530026	ZINC00009320694	-52,31		
27	IBS-E0529844	ZINC00009646241	-51,55		

Продолжение Таблицы 11

№	Шифр	ZINC ID	Grid_Score	Структура	Способ связывания по результатам докинга
Скаффолд 3					
28	IBS-E018342	ZINC000101274344	-51,92		
29	IBS-E0680092	ZINC00033278109	-50,91		
30	IBS-E0148673	ZINC00003655258	-46,65		

Определение ингибиторной активности *in vitro* было выполнено<sup>8</sup> по методике, описанной в [5]. Методика основана на механизме фёрстеровского переноса энергии (FRET) в флуорогенном субстрате между донорной и акцепторной группой. В качестве субстрата был использован пептид с донорно-акцепторной парой Dabcyl/Edans – Dabcyl-KTSAVLQ↓SGFRKM-E(Edans)-NH<sub>2</sub>. Когда донор и акцептор находятся в непосредственной близости друг от друга (10-100 Å), энергия, испускаемая Edans, подавляется Dabcyl, что приводит к низкой флуоресценции или ее отсутствию; когда же расстояние между ними увеличивается при расщеплении субстрата, интенсивность флуоресценции увеличивается. Следовательно, по

8 Эксперименты выполнены в лаборатории противовирусных лекарственных средств ФГАНУ “ФНЦИРИП им. М.П. Чумакова РАН” (Институт полиомиелита) к.х.н В.И. Уваровой.

изменению интенсивности флуоресценции может быть напрямую определена активность фермента. Реакцию инициировали добавлением 16 мкМ FRET-субстрата. Начальную скорость реакции в присутствии и в отсутствие соединения рассчитывали из линейного участка кинетической кривой. Остаточную активность (Рисунок 42) рассчитывали как отношение начальных скоростей в присутствии и в отсутствие ингибитора.

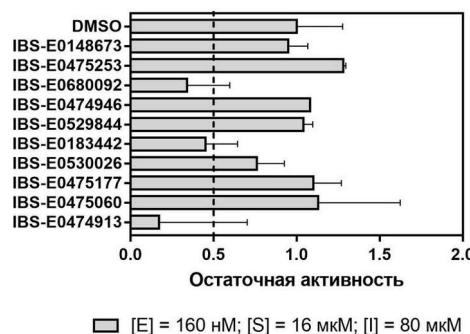


Рисунок 42. Результаты скрининга *in vitro* нековалентных ингибиторов Mpro, предложенных по результатам докинга.

Три соединения (IBS-E0680092, IBS-E0183442 и IBS-E0474913), остаточная активность протеазы в присутствии которых оказалась меньше 0,5, были отобраны для определения  $IC_{50}$  и  $K_i$  (Рисунок 43). Для определения  $IC_{50}$  и  $K_i$  строили зависимость остаточной активности фермента от концентрации ингибитора. Полученные S-образные кривые аппроксимировали с помощью нелинейной регрессии. Константу ингибирования ( $K_i$ ) рассчитывали из уравнения Ченга-Прусоффа [247]

$$K_i = \frac{IC_{50}}{1 + \frac{[S]}{K_m}} \quad (2)$$

где  $IC_{50}$  – концентрация полумаксимального ингибирования,  $[S]$  – концентрация субстрата и  $K_m$  – константа Михаэлиса.

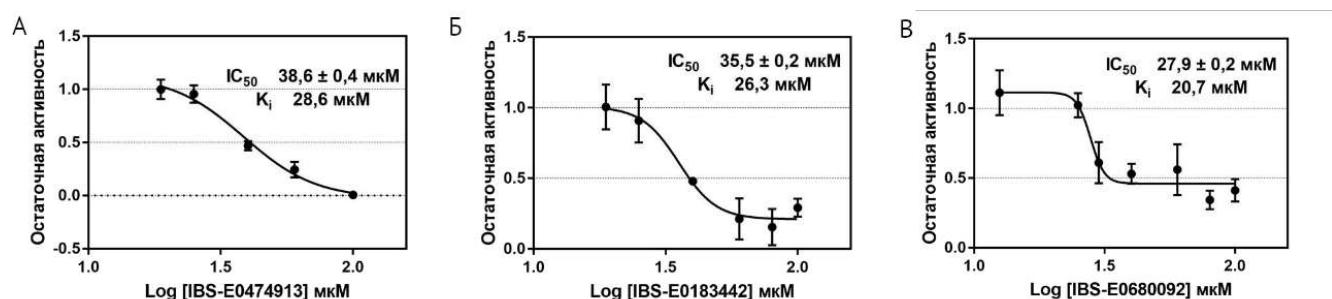
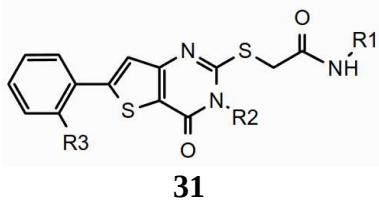


Рисунок 43. Зависимость остаточной активности фермента от концентрации исследованных нековалентных ингибиторов.

По итогам виртуального скрининга фокусированной библиотеки соединений, предложенных методом генеративного топографического картирования, и экспериментальной

валидации десяти отобранных на основе разнообразия хитов, два соединения, IBS-E0680092 и IBS-E0183442, оказались наиболее эффективными ингибиторами Mpro со значениями  $K_i$  около 20 – 30 мкМ [5]. Оба соединения относились к ряду замещённых тиенопиримидинонов, соответствующих общей формуле 31. Соединение IBS-E0474913 тоже проявляло активность с  $K_i$  около 30 мкМ, однако это соединение обладало собственной флуоресценцией и было единственным соединением своего скваффолда, обладавшим активностью.



С целью поиска более активных соединений и установления соотношений «структура — активность» в тиенопиримидиновой серии был проведён анализ коммерческой доступности аналогов. Виртуальный скрининг методом молекулярного докинга был проведён для выявления наиболее перспективных для тестирования соединений из каталога ГК «Алинда» [248], располагающей значительной коллекцией соединений с соответствующим скваффолдом. В каталоге был проведен поиск соединений по подструктуре 31, в результате которого были отобраны 178 соединений. По результатам виртуального скрининга были выбраны соединения, получившие наиболее благоприятные минимальные оценки среди всех структур Ассамбля 1, а также ключевые молекулы для анализа соотношений «структура — активность». Для 72 отобранных таким образом соединений была осуществлена закупка и проведено исследование ингибиторной активности по отношению к Mpro. IBS-E0680092 был родоначальником N3-алкильной серии с незамещённым фенилом ( $R_2 = \text{Alk}$ ,  $R_3 = \text{H}$ ), а IBS-E0183442 — N3-арильной серии с замещённым фенилом ( $R_2 = \text{Ar}$ ,  $R_3 = \text{OEt}$ ).

## 5.2. Расширение серии тиенопиримидинонов

Скрининг *in vitro* тиенопиримидинонов — аналогов IBS-E0680092 и IBS-E0183442 — проводили<sup>9</sup> путём определения остаточной активности фермента (отношения скоростей реакции) в присутствии 100 мкМ соединения. В случае уменьшения активности фермента на 50% и более (Рисунок 44) соединение признавали «хитом» первичного скрининга и проводили для него дальнейшие эксперименты. Для этих экспериментов было отобрано 17 соединений. Полные результаты приведены в Приложении В, в Таблице 12 приведены структуры наиболее активных соединений серии.

9 Эксперименты по скринингу ингибиторной активности и определению  $IC_{50}$  выполнены в Лаборатории противовирусных лекарственных средств ФГАНУ «ФНЦИРИП им. М.П. Чумакова РАН» (Институт полиомиелита) к.х.н В.И. Уваровой. Эксперименты по определению противовирусной активности выполнены в лаборатории полиомиелита и других энтеровирусов с референс-центром ВОЗ по надзору за полиомиелитом ФГАНУ «ФНЦИРИП им. М.П. Чумакова РАН» (Институт полиомиелита) д.б.н. Козловской Л.И.

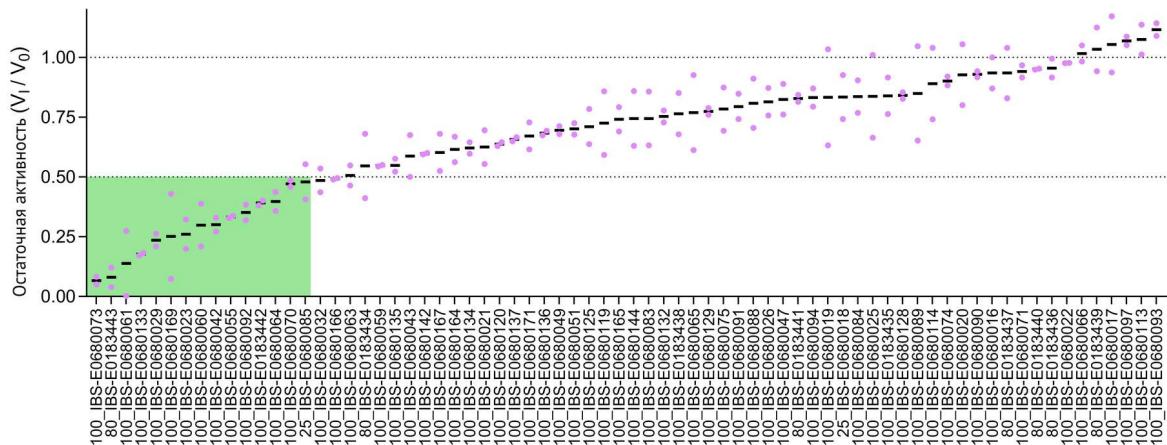


Рисунок 44. Скрининг соединений тиенопиримидинонового ряда. Уменьшение активности Mpro в присутствии соединений на 50% и больше является критерием отбора для дальнейших исследований. Соединения упорядочены по возрастанию средней остаточной активности (чёрные линии) из двух измерений (индивидуальные измерения — фиолетовые круги)

Таблица 12. Структуры активных соединений

№	Шифр	Структура	clogP	Mpro IC <sub>50</sub> (95% доверительный интервал), мкМ	Minscore
31а	IBS-E0680029		4,0857	74,3 (64,2–87,4)	-44,29
31б	IBS-E0680070		4,6917	108 (93–147)	-44,29
31в	IBS-E0680060		3,9213	66,8 (53,6–85,7)	-47,84
31г	IBS-E0680061		3,9213	24,6 (18,9–31,7)	-45,84
31д	IBS-E0680042		5,754	51,0 (42,0–64,5)	-51,97

№	Шифр	Структура	clogP	Mpro IC <sub>50</sub> (95% доверительный интервал), мкМ	Minscore
31е	IBS-E0680064		4,405	92,6 (80,9–115)	-51,72
31ж	IBS-E0680133		4,405	21,2 (16,7–29,8)	-50,15
31з	IBS-E0680085		4,9656	19,7 (13,6–35,6)	-51,28
31и	IBS-E0680055		4,0645	90 (81–107)	-49,89
31к	IBS-E0680169		4,7397	21,9 (16,1–30,3)	-52,91
31л	IBS-E0680023		4,4011	80 (60–130)	-49,06
31м	IBS-E0680092		4,0614	53,9 (40,2–74,2)	-52,21
31н	IBS-E0183443		7,0206	17,2 (12,8–23,3)	-52,52
31о	IBS-E0183442		7,4362	75,8 (65,6–90,2)	-53,83

В серии N-алкилтиенопirimидинонов в положении R1 находятся три возможных заместителя: метил, этил, аллил (-CH<sub>2</sub>CH=CH<sub>2</sub>). Заместители в положении R можно разделить на две серии: фенильные (31 а-к) и бензильные (31 л-о). Ни одно из соединений с заместителем

бензильного типа в положении R не проявляет активности против Mpro SARS-CoV-2 в исследованном диапазоне концентраций.

Результаты докинга абсолютного большинства N-алкилтиенопиримидинонов характеризуются расположением центрального бициклического фрагмента в кармане S3 без специфических взаимодействий, которые бы обеспечивали связывание, и заместителем R1 в кармане S2. Фенильный фрагмент центрального скаффолда при этом ориентирован вовне активного центра или занимает карман S4. Все заместители в положении R1, встречающиеся в серии, не вызывают стерических затруднений при расположении в кармане S2. Карман S2 имеет гидрофобную поверхность во внешней части и гидрофильную поверхность в глубине с потенциальной возможностью образования водородных связей с карбонильными атомами кислорода основной цепи Asp187 и Arg188, которые не реализуются вследствие отсутствия атомов, способных образовывать водородные связи, в рассмотренных заместителях R1 (Рисунок 45).

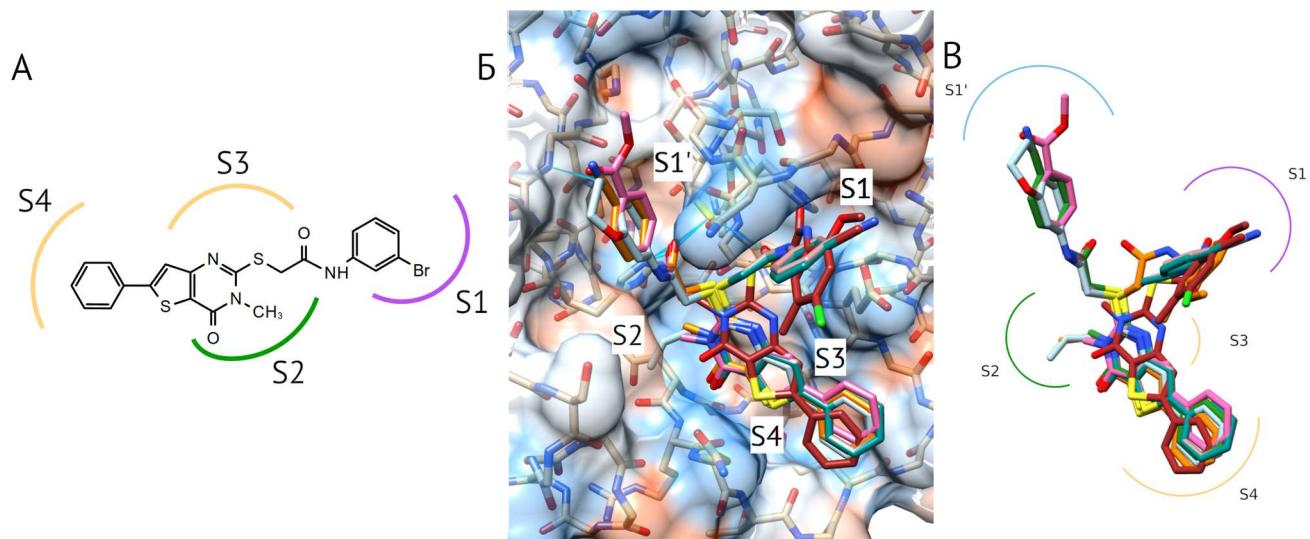


Рисунок 45. Типичная конформация связывания тиенопиримидинонов в активном центре Mpro SARS-CoV-2. (А) – схема расположения IBS-E0680073 (31п) в карманах активного центра. (Б) – активные соединения (31п (сине-зеленый), 31ж (розовый), 31з (бордовый), 31к (голубой), 31г (оранжевый), 31в (бежевый)) в активном центре главной протеазы SARS-CoV-2 (PDB id: 7NEV, окраска поверхности по шкале гидрофобности Кайта-Дулитла [249]. Голубым обозначены наиболее гидрофильные области поверхности, а оранжевым – наиболее гидрофобные). (В) – аналогично Б, но без отображения поверхности.

Другая типичная конформация связывания представлена на Рисунке 46. В ней центральный фрагмент расположен между карманами S1, S1' и S2 и зафиксирован двумя водородными связями с остатком цистеина Cys145. Фенильный заместитель размещен в кармане S2, небольшой алкильный заместитель R1 расположен в кармане S1, а длинный

заместитель R располагается в карманах S3 и S4 и иногда зафиксирован водородными связями. В альтернативном варианте этой конформации молекула развернута на 180°: фенильный заместитель расположен в кармане S1, а заместитель R1 – в кармане S2.

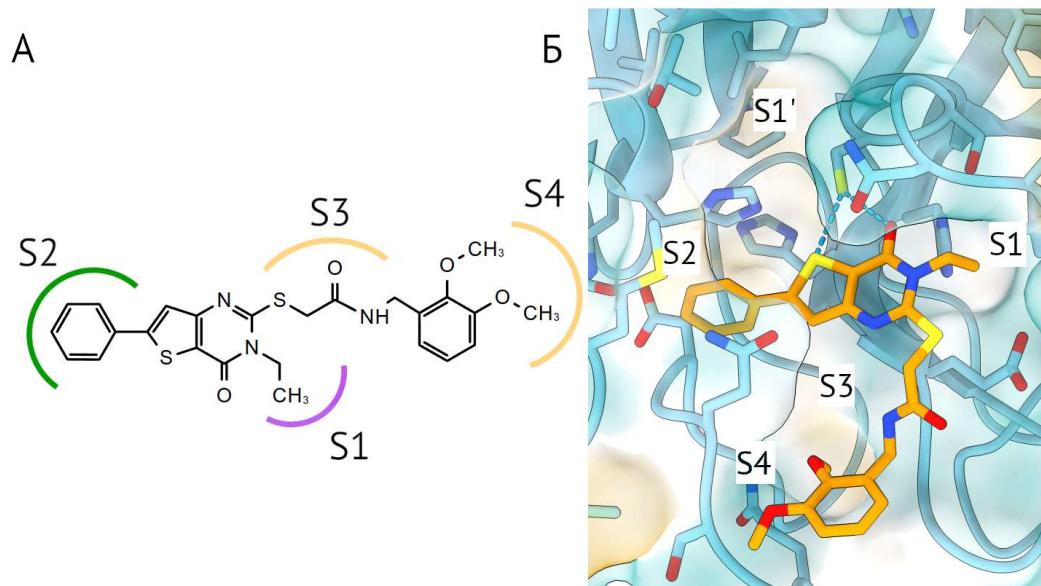


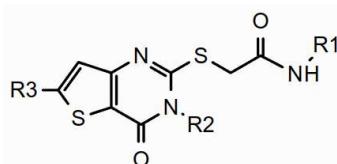
Рисунок 46. Альтернативная конформация связывания тиенопирииминонов в активном центре Mpro SARS-CoV-2. (А) – схема расположения IBS-E0680092 (31м) в карманах активного центра главной протеазы SARS-CoV-2. (Б) – результат докинга (PDB ID: 7AKU). Окраска поверхности аналогично Рисунку 45.

По итогам скрининга коммерчески доступной тиенопирииминоновой серии не было выявлено соединений, обладающих концентрацией полумаксимального ингибирования на уровне ниже 15 мкМ. В наиболее исследованном ряду N3-алкилтиенопирииминонов отсутствует варьирование заместителей по положению R2. В положении R1 присутствуют только 3 вида заместителей: метил, этил и аллил. Таким образом, за изменение ингибиторной активности преимущественно должен отвечать ароматический заместитель, присоединённый по 2-сульфанильной группе. Среди наиболее активных соединений в этом ряду представлены галогенированные производные, метоксипроизводные, нитрильные производные. Активность соединений ряда уменьшается с ростом липофильности. Эта серия соединений была выбрана для анализа потенциальных направлений синтетической модификации на основе результатов молекулярного докинга.

В ряду N3-арилтиенопирииминонов были представлены только гидрофобные соединения, содержащие доноры и акцепторы водородной связи исключительно в центральном фрагменте. Хотя среди них также имелись соединения с ингибиторной активностью на уровне 20 мкМ, недостаток сайтов направленных межмолекулярных взаимодействий и недостаточное покрытие серии затрудняют её развитие и требуют дальнейшего исследования на базе

коммерчески доступных соединений.

С целью изучения возможных направлений варьирования заместителей в тиенопиримидиноновой серии (32) был проведён ансамблевый молекулярный докинг соединений, исследованных экспериментально. Для N3-алкильных производных фенильного типа наблюдаются наиболее однородные результаты докинга. При этом заместитель R1 располагается в кармане S1 или S1', где карбонильный атом кислорода фиксируется водородной связью с Cys145, заместитель R2 – в кармане S2, центральная часть молекулы — в кармане S3, а заместитель R3 – в области кармана S4.



32

На основе результатов докинга были предложены следующие варианты модификации:

1. Варьирование заместителей в ароматическом кольце R1. В рамках серии экспериментально изучено влияние заместителей в ароматическом кольце, выявлены активные соединения. Следует рассмотреть синтез их гетероциклических аналогов с аналогичным характером заместителей.
2. Варьирование подвижного заместителя, включающего R1. В наиболее эффективных ингибиторах Mpro (нирматрелвир, PF-00835231) в кармане S1 располагается характерный  $\gamma$ -лактамный цикл. Следует рассмотреть возможность замены заместителя в целом на  $\gamma$ -лактамный цикл и аналогичные алициклические фрагменты.
3. Варьирование заместителя R2. В наиболее эффективных ингибиторах Mpro (нирматрелвир, PF-00835231) заместителю R2, взаимодействующему с малым гидрофобным карманом S2, соответствуют небольшие гидрофобные заместители. В исследованной серии в соответствующей позиции встречаются только метил, этил и аллил; предлагается исследовать влияние других гидрофобных заместителей.
4. Варьирование заместителя R3. В исследованной серии отсутствуют вариации соответствующего фрагмента. При этом согласно результатам докинга фенильный фрагмент в значительной степени доступен для растворителя и почти не образует специфических взаимодействий. Возможные направления варьирования — введение гетероциклов, замещение атомов водорода, вставка метиленового линкера и их комбинации.

С учётом этого для дальнейшего моделирования были выбраны две начальные структуры, характеризующиеся наиболее высокой активностью в серии ( $IC_{50} \sim 20$  мкМ) и содержащие

акцепторы водородной связи в заместителе R1 (Таблица 12). В исследованной экспериментально серии соединений основной вариабельной позицией был фенильный фрагмент заместителя при экзоциклическом атоме серы. Соединения **31з** и **31ж** имеют разные заместители в этом положении. Можно предположить, что в исследованной серии эти заместители являются оптимальными, и следует рассмотреть возможность модификации других положений центрального фрагмента.

Методом молекулярного докинга проведён систематический анализ допустимости различных заместителей в положении 3 тиенопиримидинонового фрагмента (R1), а также замещения фенильной группы в положении 6 (R3). Исходя из результатов докинга первой серии, в положении R2 исследовалась возможность введения небольших гидрофобных заместителей, а в положении R3 — ароматических заместителей (Рисунок 47). Для каждой «родительской» молекулы сгенерировали 442 аналога с различными комбинациями заместителей R2 и R3. При анализе результатов докинга принимали во внимание консервативность ориентации молекулы потенциального ингибитора в активном центре, изменение оценки докинга по сравнению с исходным соединением (Рисунок 48А, 49А), а также расчётную липофильность (Рисунок 48Б, 49Б). По результатам анализа предложена фокусированная библиотека для синтеза и экспериментальной валидации.

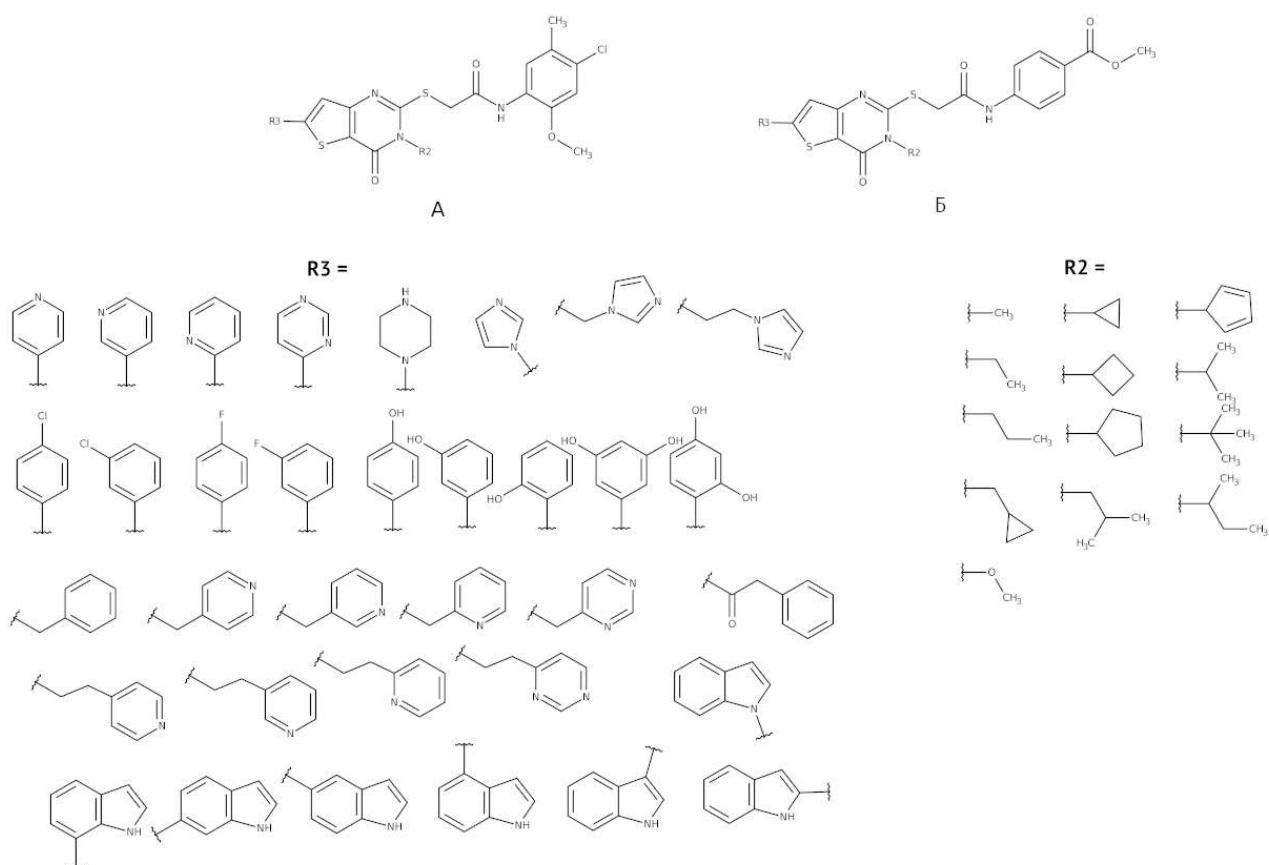


Рисунок 47. Рассмотренные серии аналогов соединений (А) – 31з и (Б) – 31ж.

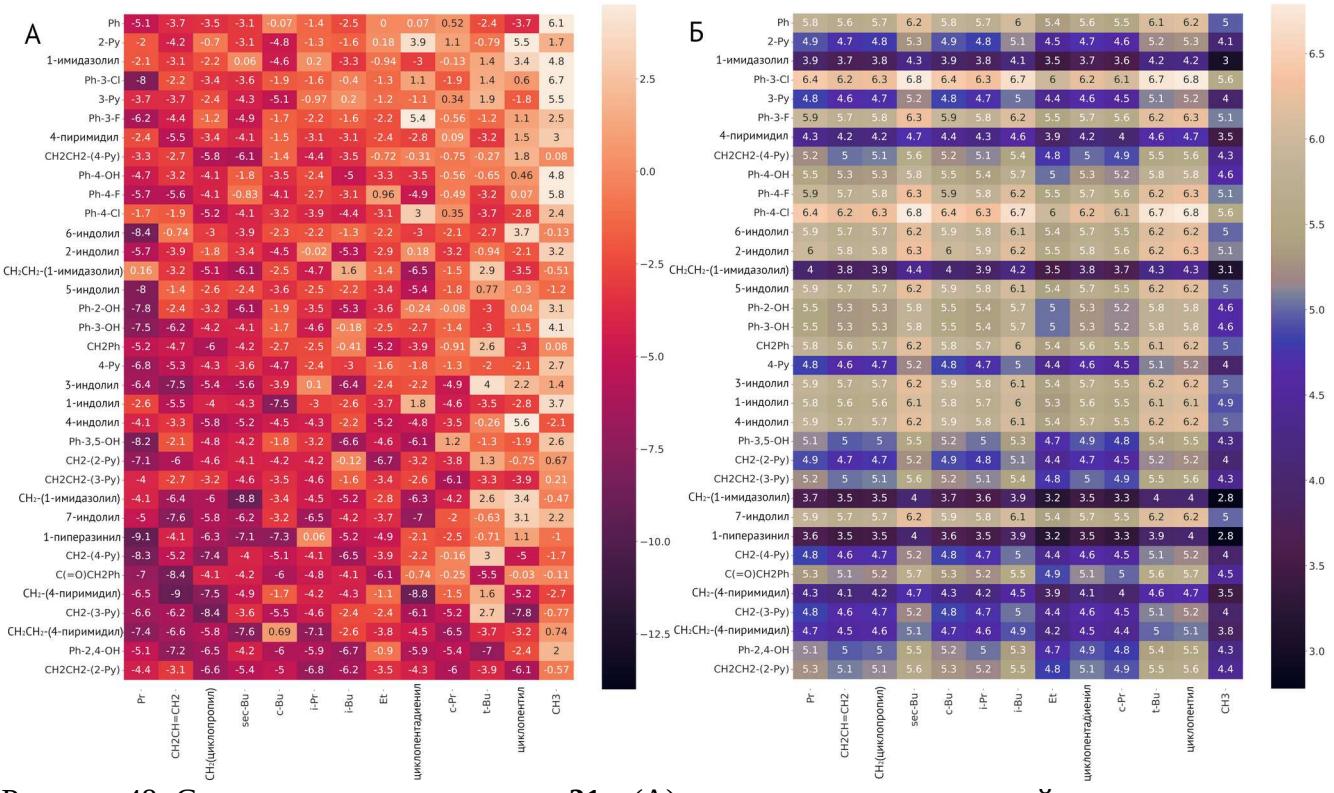


Рисунок 48. Серия аналогов соединения 31з. (А) – изменение минимальной оценки докинга по ансамблю относительно «родительской» молекулы. (Б) – расчетный logP.

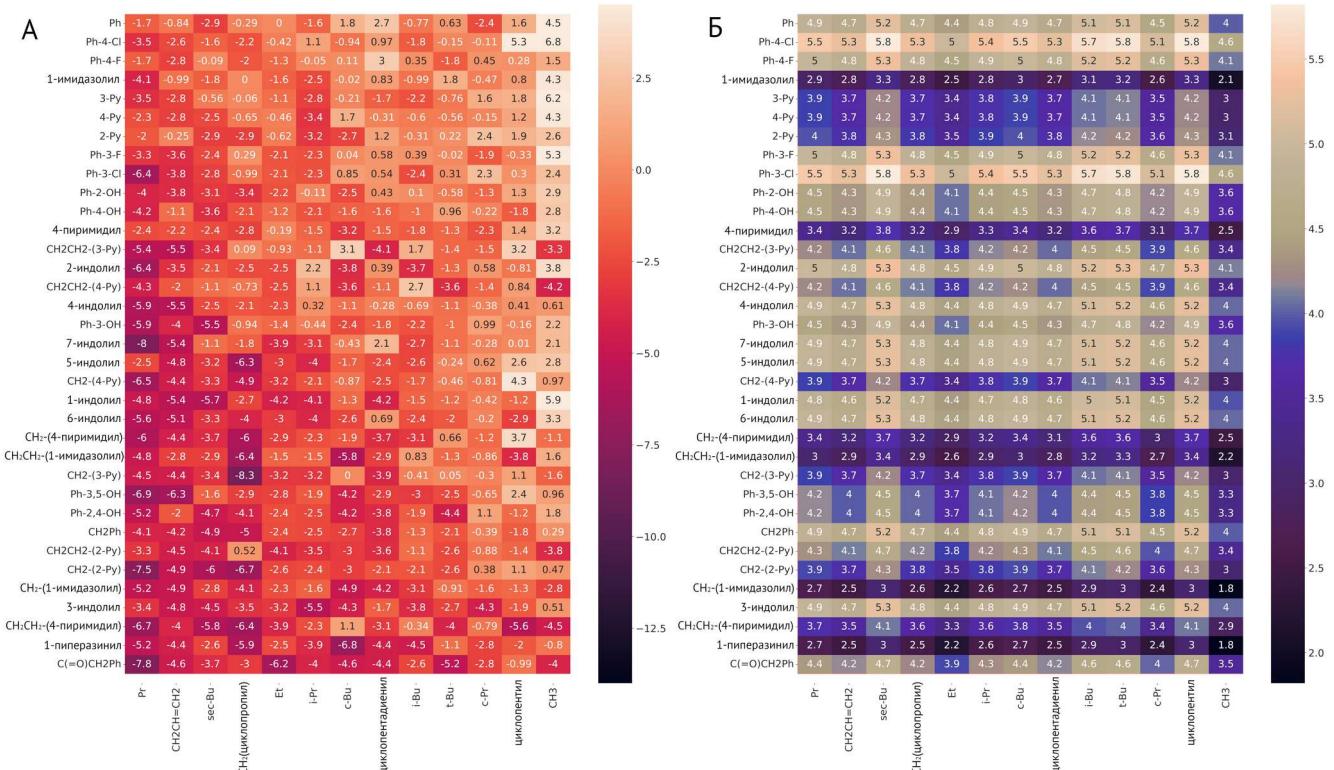


Рисунок 49. Серия аналогов соединения 31ж. (А) – изменение минимальной оценки докинга по ансамблю относительно «родительской» молекулы. (Б) – расчетный logP.

По итогам докинга серий аналогов **31з** и **31ж** можно сделать следующие выводы:

1. Замена метильной или этильной группы в положении R2 **31з** и **31ж** на иные небольшие алифатические заместители допустима и преимущественно энергетически выгодна. Исходя из свойств соответствующей полости, нельзя сделать вывод о предпочтительности конкретного заместителя, и требуется систематическое экспериментальное исследование для характеристики свойств заместителей в положении R1.
2. Замена фенила в положении R3 **31з** и **31ж** на иные заместители допустима и может приводить к изменению энергетически выгодного способа связывания. Поскольку вводимый в это положение заместитель может быть доступен для растворителя, представляется важным систематически исследовать возможность синтеза соответствующих гидрофильных аналогов, прежде всего гетероциклических. Введение одного или нескольких метиленовых звеньев между центральным фрагментом и R3 даёт тенденцию к изменению преимущественного типа ориентации ингибитора и более выгодным оценкам связывания.
3. Заместители в положении R2 **31з** влияют на эффективность связывания менее выраженным и менее однородным образом, чем в случае **31ж**. Отбор соединений для синтеза следует проводить путём выбора оптимальных заместителей в положении R3 и планировать серии с вариацией по R2.

В ходе виртуального скрининга коммерчески доступных низкомолекулярных соединений выявлен новый класс соединений, проявляющих ингибиторную активность против Mpro SARS-CoV-2, – замещенные тиено[3,2-*d*]пиrimидиноны. Хотя наиболее активные соединения серии имеют IC<sub>50</sub> порядка 20 мкМ, изучение их способов связывания, спрогнозированных методом молекулярного докинга, позволяет предположить возможность улучшения показателей активности за счет варьирования заместителей. Соответствующие серии соединений были предложены для последующего экспериментального исследования.

## Глава 6. Прогнозирование класса активности ингибиторов Mpro SARS-CoV-2<sup>10</sup>

Во многих работах было показано, что машинное обучение способно повысить точность виртуального скрининга методом докинга. В качестве входных данных, на которых проводили обучение, использовали как оценочные функции докинга [46], [250], так и данные о спрогнозированных способах связывания молекул с белковой мишенью [251]. Поэтому мы исследовали возможность применения машинного обучения к результатам ансамблевого докинга потенциальных ингибиторов главной протеазы SARS-CoV-2 для прогнозирования их класса активности.

### 6.1. Обучающая выборка и декрипторы

В качестве обучающей выборки были выбраны результаты скрининга *in vitro* ингибиторной активности лекарственных средств в отношении Mpro SARS-CoV-2 CHEMBL4495582 [122], [232], так как на март 2022 года этот набор данных был наибольшим из доступных результатов высокопроизводительного скрининга, выполненных в рамках одного эксперимента в одной лаборатории. Библиотека насчитывала 8702 соединения, для которых была определена эффективность ингибирования Mpro SARS-CoV-2, выраженная в процентах. Для определения активности ингибиторов протеазы Mpro SARS-CoV-2 измеряли флуоресценцию субстрата после 15 минут инкубации фермента с 20 мКМ ингибитора. В качестве положительного контроля использовали пиритион цинка (100% ингибирования), а в качестве отрицательного – ДМСО (0% ингибирования). Распределение процентов ингибирования в выборке представлено на Рисунке 50А. Классы активности назначали по пороговому значению процента ингибирования Mpro SARS-CoV-2 60%. Активными (класс 1) считали соединения с процентом ингибирования Mpro SARS-CoV-2 > 60%, все остальные относили к неактивным (класс 0). Соотношение активных соединений к неактивным составило 0,032.

- 
- 10 При работе над данным разделом диссертации использованы материалы следующих публикаций автора, в которых, согласно Положению о присуждении ученых степеней в МГУ, отражены основные результаты, положения и выводы исследования:
1. Фомина А.Д., Палюлин В.А., Осолодкин Д.И. Моделирование структуры белка NS1 ортофлавивирусов по гомологии для виртуального скрининга потенциальных лигандов // Биомедицинская химия – 2024 – т. 70 (6) – с. 456–468
  2. Fomina A. D., Uvarova V. I., Kozlovskaya L. I., Palyulin V. A., Osolodkin D.I., Ishmukhametov A. A. Ensemble Docking Based Virtual Screening of SARS-CoV-2 Main Protease Inhibitors // Molecular Informatics – 2024 – V.43. - e202300279.
  3. Zakharova M.Yu., Kuznetsova A.A., Uvarova V.I., Fomina A.D., Kozlovskaya L.I., Kaliberda E.N., Kurbatskaia I.N., Smirnov I.V., Bulygin A.A., Knorre V.D., Fedorova O.S., Varnek A., Osolodkin D.I., Ishmukhametov A.A., Egorov A.M., Gabibov A.G., Kuznetsov N.A. Pre-Steady-State Kinetics of the SARS-CoV-2 Main Protease as a Powerful Tool for Antiviral Drug Discovery // Frontiers in Pharmacology – 2021 – V. 12. – P. 773198.
  4. Фомина А.Д., Осолодкин Д.И. Программный комплекс для систематического ансамблевого докинга // Свидетельство о государственной регистрации программы для ЭВМ № 2024684465. – Номер заявки 2024684027. – Дата поступления 17.10.24. – Дата государственной регистрации в Реестре программ для ЭВМ 17.10.24.
  5. Фомина А.Д., Осолодкин Д.И. База данных соединений с экспериментально определенной активностью против главной протеазы β-коронавирусов // Свидетельство о государственной регистрации базы данных № 2024625235. – Номер заявки 2024624786. – Дата поступления 29.10.2024 – Дата государственной регистрации в Реестре баз данных 15.11.2024 Бюл. № 11

Уравнение 3 было использовано авторами для расчета процента ингибиования:

$$X_{\text{new}} = \frac{X_i - \min(X)}{\max(X) - \min(X)}, \quad (3)$$

где  $X_i$  – индивидуальное значение процента ингибиования,  $\min(X)$  – процент ингибиования при отрицательном контроле,  $\max(X)$  – процент ингибиования при положительном контроле. Ингибиование  $>100\%$  связано с погрешностями измерений, однако поскольку к активным соединениям относили все, для которых процент ингибиования был больше 60, эти погрешности не влияют на конечный результат.

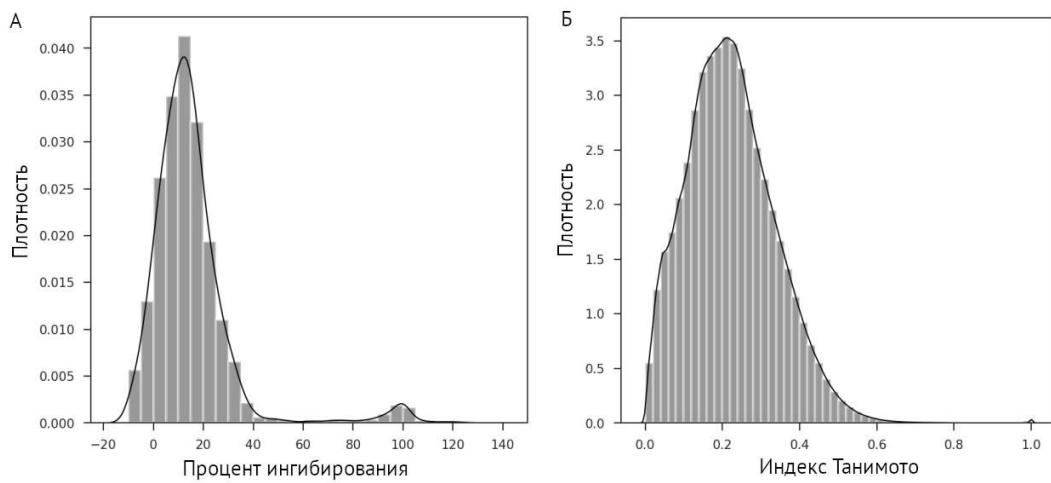


Рисунок 50. Описательная статистика выборки [122]. (А) – Распределение экспериментально определенных процентов ингибиования Mpro SARS-CoV-2; (Б) – Молекулярное разнообразие: распределение попарных индексов Танимото между фингерпринтами Моргана.

Из библиотеки были удалены металлоорганические и неорганические соединения. Все записи, содержащие несколько отдельных молекул (рацематы, координационные соединения с несколькими органическими молекулами, соли органических кислот с органическими противоионами) были найдены в базе данных при помощи сценария на языке Python 3.9. В этих записях были оставлены только самые крупные молекулы, или, в случае рацематов, тот изомер, который был записан в файле первым, поскольку данные о стереоизомерии были необходимы для генерации пространственных координат молекул для докинга. В других расчетах эти данные не учитывали. В итоговой библиотеке осталось 6897 соединений. Среднее значение индекса Танимото между векторными представлениями (1024-битными фингерпринтами Моргана с радиусом 3) молекул в выборке составило  $0,22 \pm 0,11$ , что является показателем высокого разнообразия (Рисунок 50Б). Подготовку библиотеки к докингу осуществляли по процедуре, описанной в разделе 3.

Далее провели докинг соединений в структуры протеазы Mpro SARS-CoV-2 Ансамбля 2

при помощи программы DOCK6.9 по процедуре, описанной в разделе 4. *Ансамбль 2* был выбран потому, что в него входит меньше структур, что понизит размерность данных и вычислительную сложность по сравнению с *Ансамблем 1*.

Из результатов докинга были отобраны оптимальные (наиболее высоко оцененные) ориентации каждой молекулы в каждой структуре белка. Для этих комплексов были рассчитаны фингерпринты взаимодействия методами Flare и ODDT [252]. Первый метод описан ранее в разделе 4.3.1, в его рамках взаимодействие лиганда с каждым из остатков кодировалось одним из четырех значений (Flare-IFP, Рисунок 51А). Второй вид фингерпринтов взаимодействия (ODDT-IFP, Рисунок 51Б) рассчитали при помощи библиотеки Open Drug Discovery Toolkit (ODDT). Были использованы стандартные определения, соответствующие 8 типам взаимодействий лиганда с каждым остатком белка: гидрофобных, π-стекинга, Т-стекинга, водородных связей, в которых аминокислотный остаток выступает донором или акцептором, солевых мостиков, в которых аминокислотный остаток положительно или отрицательно заряжен, а также связей с неорганическими ионами. Таким образом, каждой молекуле соответствовал вектор длины  $8N$ , где  $N$  – число а/к остатков в белке. Столбцы, в которых отсутствовали взаимодействия, были удалены для всех лигандов. Если было невозможно рассчитать фингерпринты для какой-либо молекулы в связи с ограничениями библиотеки RDKit, для этой молекулы во всех полях указывали -1, однако не удаляли. Для векторного описания молекул использовали фингерпринты Моргана [228] радиуса 3 длиной 512 бит и 1024 бит.



Рисунок 51. Схема фингерпринтов взаимодействия: (А) – Flare-IFP и (Б) – ODDT-IFP.

## 6.2. Подбор параметров моделей

Чтобы подобрать параметров обучения было сформировано несколько наборов данных, в которых каждому соединению ставили в соответствие вектор, составленный по принципу, приведенному в Таблице 13, и класс активности. В наборе данных OHF фингерпринты взаимодействия подвергали one-hot-кодированию: каждому элементу вектора ставили в соответствие 4 бита нового вектора, где значению “0” соответствовала 1 в первой позиции при 0 во всех остальных позициях, “-1” – 1 во второй позиции и так далее.

Таблица 13. Наборы входных данных для построения моделей

Название	Дескрипторы	Длины составляющих векторов	Общая длина вектора
OHF	One-hot-закодированные фингерпринты взаимодействия Flare-IFP, индивидуальные оценки докинга, Mw, CLogP	7971, 6, 1, 1	7979
FL	Явно заданные фингерпринты взаимодействия Flare-IFP, индивидуальные оценки докинга, Mw, CLogP	377, 6, 1, 1	385
FLE	Явно заданные фингерпринты взаимодействия Flare-IFP, индивидуальные оценки докинга, отнесенные к числу тяжелых атомов, Mw, CLogP	377, 6, 1, 1	385
M5F	Фингерпринты Моргана (512 б, радиус 3), явно заданные фингерпринты взаимодействия Flare-IFP, индивидуальные оценки докинга, Mw, CLogP	512, 377, 6, 1, 1	897
M5	Фингерпринты Моргана (512 б, радиус 3), индивидуальные оценки докинга, Mw, CLogP	512, 6, 1, 1	520
M5O	Фингерпринты Моргана (512 б, радиус 3), фингерпринты взаимодействия ODDT-IFP, индивидуальные оценки докинга, Mw, CLogP	512, 497, 6, 1, 1	1017
OIF	Фингерпринты взаимодействия ODDT-IFP, индивидуальные оценки докинга, Mw, CLogP	497, 6, 1, 1	505
M1	Фингерпринты Моргана (1024 б, радиус 3), индивидуальные оценки докинга, Mw, CLogP	1024, 6, 1, 1	1032
M1O	Фингерпринты Моргана (1024 б, радиус 3), фингерпринты взаимодействия ODDT-IFP, индивидуальные оценки докинга, Mw, CLogP	1024, 497, 6, 1, 1	1529

Все наборы данных были разделены на обучающую, валидационную и контрольную выборки в соотношении 14:3:3 со стратификацией (Рисунок 52). Соотношение активных соединений к неактивным во всех выборках составило 0,032. Валидационная выборка была использована в экспериментах по построению ансамблевых моделей, результаты которых не представлены.



Рисунок 52. Схема разбиения данных.

Эти данные использовали для построения моделей машинного обучения со стратифицированным 5-кратным скользящим контролем (Рисунок 52). Подбор параметров для всех моделей проводили при помощи функции `GridSearchCV` из библиотеки `sklearn 1.0.2`.

Модели и подбираемые параметры с указанием перебираемых значений:

- 1) Логистическая регрессия:  $C$  от  $10^{-6}$  до 100 с интервалом 5;  $penalty$  – l1 или l2
- 2) Случайный лес:  $class\_weight$  = 'balanced' или 'balanced\_subsample';  $n\_estimators$  = 1000,  $criterion$  – 'gini';  $max\_depth$  = 5,7,10,15;  $min\_samples\_split$  = 3,5,7,10,15;  $min\_samples\_leaf$  = 2, 3, 5, 7
- 3) CatBoost 1.0.5:  $iterations$  = 70,  $learning\_rate$  = 0.5,  $depth$  = 2,3,4,5 или 6;  $l2\_leaf\_reg$  = 2, 5 или 10
- 4) SVM:  $probability$  = True,  $kernel$  = 'sigma',  $class\_weight$  = 'balanced',  $random\_state$  = 42,  $coef0$  = 0,0,1,10
- 5) TabNet 3.1.1:  $n\_d$  = 32,  $n\_a$  = 32,  $n\_steps$  = 4,  $lambda\_sparse$  =  $10^{-4}$

Данные предварительно трансформировали при помощи `MinMaxScaler` из библиотеки `sklearn 1.0.2`, процедуру проводили для каждого разбиения скользящего контроля индивидуально. В качестве целевой метрики использовали F1-меру (4):

$$F1 = \frac{2tp}{2tp + fp + fn}, \quad (4)$$

где  $tp$  – число истинно положительных прогнозов,  $fp$  – число ложноположительных прогнозов,  $fn$  – число ложноотрицательных прогнозов. Для лучших по F1-мере моделей строили ROC-кривую и рассчитывали площадь под ней.

Чтобы исключить влияние разбиения, проводили разбиение на обучающую и контрольную выборки с 9 разными зернами генератора случайных чисел, обучали модели, выбранные с помощью скользящего контроля, а затем находили среднее значение и

стандартное отклонение для ROC AUC, F1-меры, точности и полноты, рассчитанных на каждом разбиении (Таблица Г1, Приложение Г). Стандартное отклонение на контрольной выборке было значительно выше, чем на обучающей (Рисунок 53), что может быть связано с высоким разнообразием выборки и значительным дисбалансом классов вследствие использования репозиционированных лекарств в исходном скрининге. Кроме того, все модели характеризуются высоким переобучением.

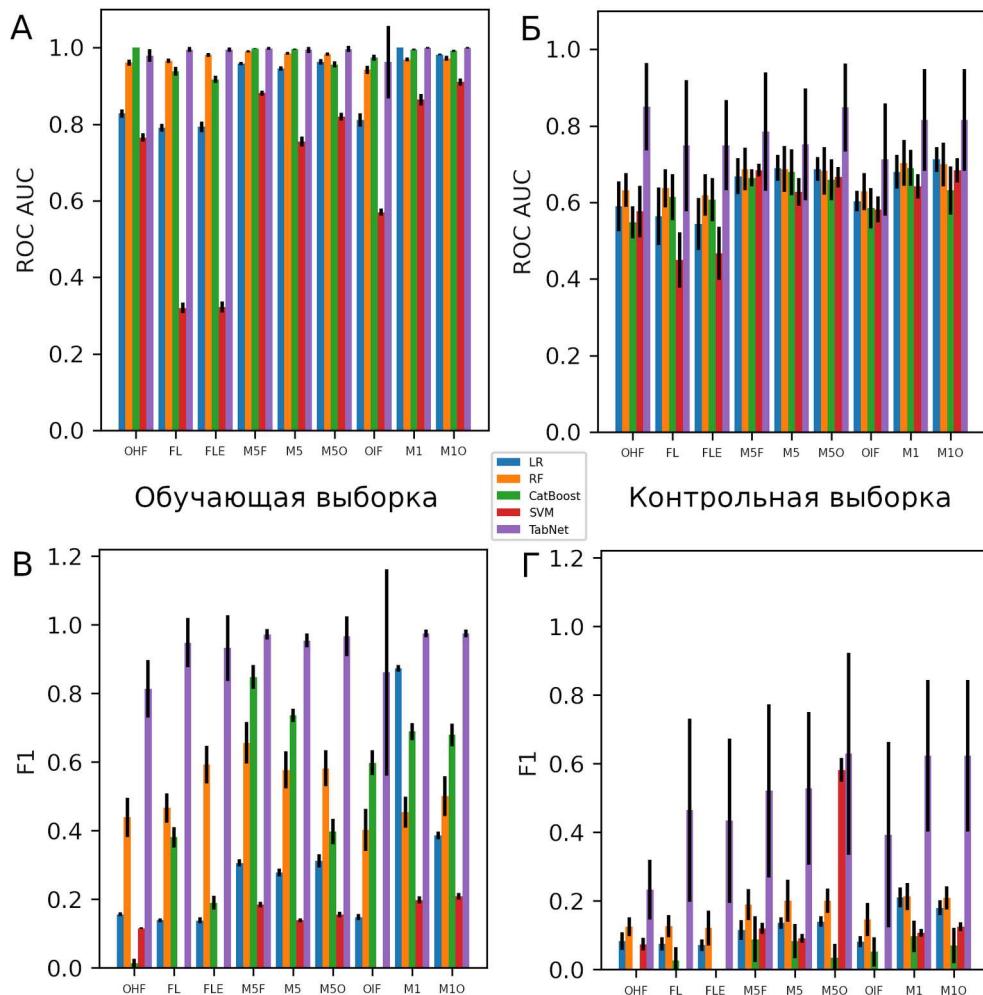


Рисунок 53. Средние значения ROC AUC (А, Б) и F1-меры (В, Г) на обучающей (А, В) и контрольной (Б, Г) выборках для лучших моделей логистической регрессии, случайного леса, градиентного бустинга, SVM и TabNet, построенных с F1-мерой в качестве целевой метрики

Добавление фингерпринтов взаимодействия Flare и ODDT не приводит к значительному повышению качества классификации по сравнению с моделями, построенными на наборах данных, не содержащих фингерпринтов взаимодействия. Результаты моделей, построенных на наборах данных, включающих фингерпринты взаимодействия в оригинальном виде и в виде категориальных дескрипторов, подвергнутых one-hot-кодированию, заметно не отличаются, хотя расчетное время во втором случае значительно увеличивается. Таким образом, использование one-hot-кодирования для решения этой задачи нецелесообразно. Для всех

моделей TabNet (нейросети, архитектура которой оптимизирована для обработки табличных данных) систематически наблюдали самое большое стандартное отклонение на контрольной выборке, что не позволяет делать из прогнозов модели достоверные выводы. Они были исключены из дальнейшего рассмотрения.

На рисунке 54 показаны средние значения точности (precision) и полноты (recall) на классе активных соединений. Полнота на классе активных соединений использовалась как метрика для определения лучших моделей, так как в виртуальном скрининге важно извлечь как можно больше активных молекул и поместить их как можно выше в рейтинге. Самые высокие показатели полноты наблюдали для моделей SVM и логистической регрессии. Самые низкие показатели были у CatBoost, также значения точности для этой модели имели очень большой разброс.

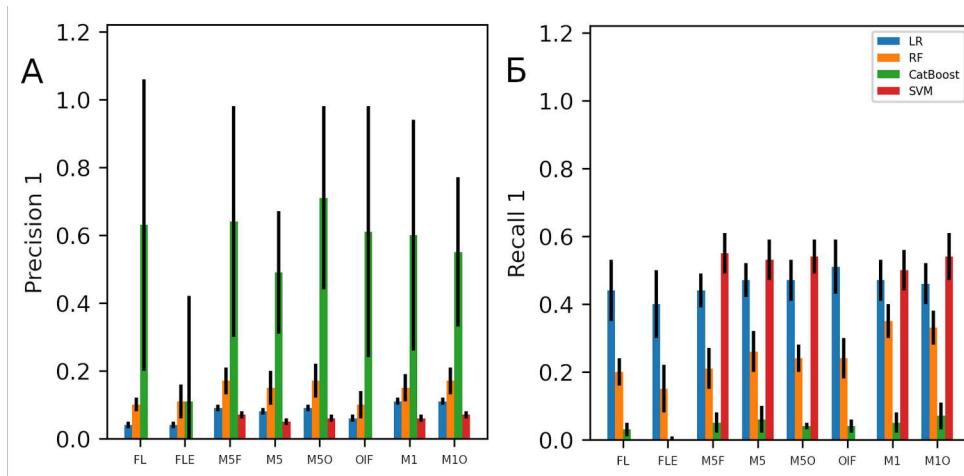


Рисунок 54. Средние значения точности (А) и полноты (Б) прогнозов моделей для класса активных соединений на контрольной выборке лучших при скользящем контроле моделей, построенных с F1-мерой в качестве целевой метрики

Исходя из анализа ROC-кривых (Рисунок 55), лучшие результаты получены для моделей логистической регрессии, случайного леса и CatBoost. Однако CatBoost дает очень низкую полноту на классе активных молекул, что не подходит для задачи виртуального скрининга. Кроме того, модели CatBoost были наиболее склонны к переобучению в широком диапазоне параметров, поэтому метод был исключен из дальнейшего рассмотрения. При этом все модели МО характеризуются более высокой точностью и ранним обогащением, чем классификация на основании минимальной оценки докинга по ансамблю, которая разделяет соединения на активные и неактивные хуже, чем модель, назначающая классы случайнным образом.

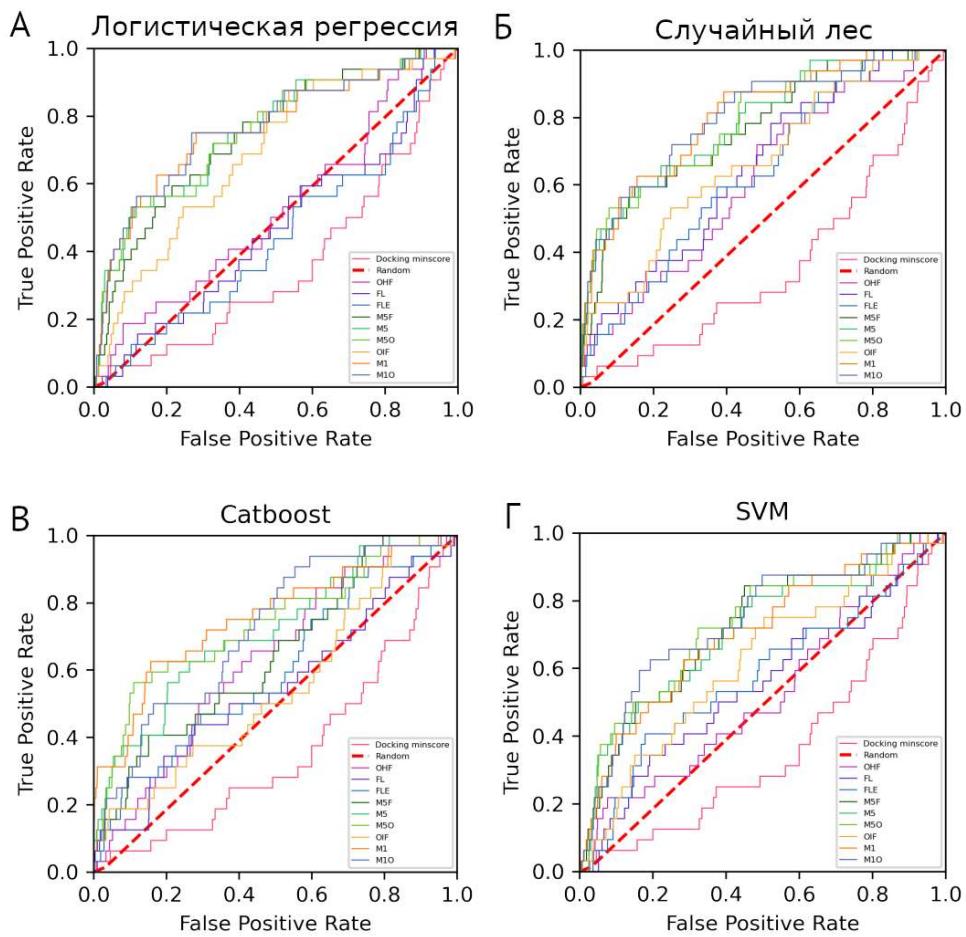


Рисунок 55. ROC-кривые для моделей логистической регрессии, случайного леса, градиентного бустинга и SVM, построенных на всех изученных наборах данных (Таблица 13). Ранжирование произведено на основании вероятности принадлежности к классу активных. Для сравнения представлена ROC-кривая для классификации активных соединений на основании минимальной оценки докинга по ансамблю (розовый).

Чтобы оптимизировать протокол виртуального скрининга, для наборов данных, на которых были получены наилучшие результаты на предыдущем этапе, повторно провели подбор параметров моделей логистической регрессии, случайного леса и SVM в более широких интервалах с оптимизацией полноты. При этом проблему переобучения решить полностью так и не удалось (Рисунок 56), хотя относительно моделей, для которых подбирали параметры с оптимизацией F1-меры полнота увеличилась (Рисунок 57, Рисунок 54). Самые высокие значения F1-меры наблюдали для моделей случайного леса. При этом значения ROC AUC на контрольной выборке практически не отличаются для моделей, построенных по различным алгоритмам. Также удалось добиться уменьшения разброса значений F1-меры и ROC AUC на разных разбиениях. Хотя самые высокие значения полноты получаются на моделях логистической регрессии и SVM, они также характеризуются очень низкими значениями точности (Рисунок 57) и поздним обогащением. Это делает их непригодными для виртуального

скрининга – на больших выборках такие модели будут давать слишком много ложноположительных результатов.

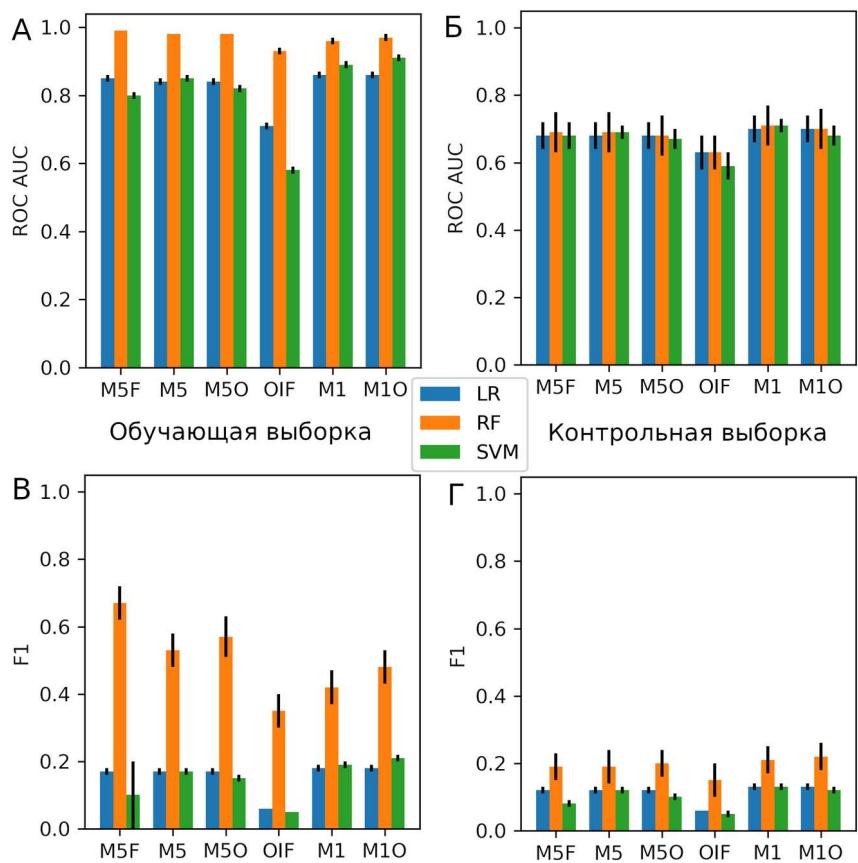


Рисунок 56. Средние значения ROC AUC (А, Б) и F1-меры (В, Г) на обучающей (А, В) и контрольной (Б, Г) выборках для моделей логистической регрессии, случайного леса и метода опорных векторов, построенных с полнотой в качестве целевой метрики.

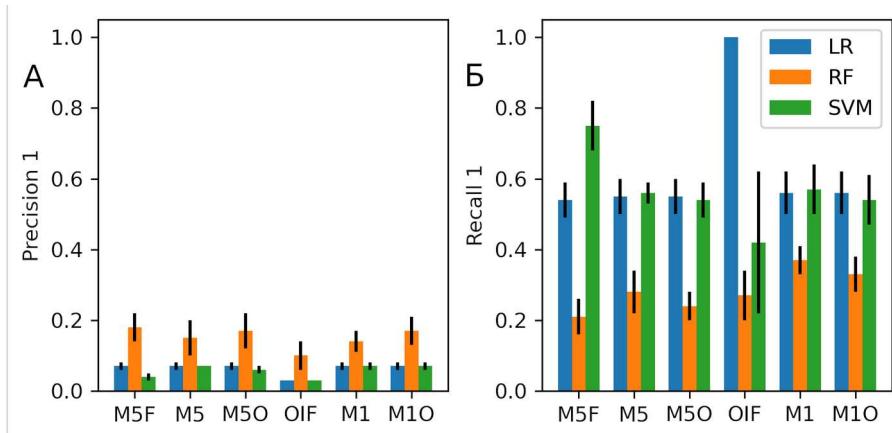


Рисунок 57. Значения средней точности (А) и полноты (Б) для класса активных соединений в контрольной выборке моделей логистической регрессии, случайного леса и метода опорных векторов, построенных с полнотой в качестве целевой метрики.

Одним из важных параметров виртуального скрининга является раннее обогащение

ранжированного списка потенциальных хитов скрининга активными соединениями. Анализ раннего обогащения проводили путем построения кривых зависимости точности и полноты на первых  $k$  элементах (precision@ $k$  и recall@ $k$ ) от  $k$  (Рисунок 58). Лучшие показатели раннего обогащения (высокие значения precision@ $k$  и recall@ $k$  при малых  $k$ ) демонстрирует модель случайного леса, построенная на наборе данных M5. Однако точность и полнота для моделей случайного леса, обученных на наборе данных M1, выше и имеет меньший разброс. Оба эти набора данных не содержат информации о способах связывания.

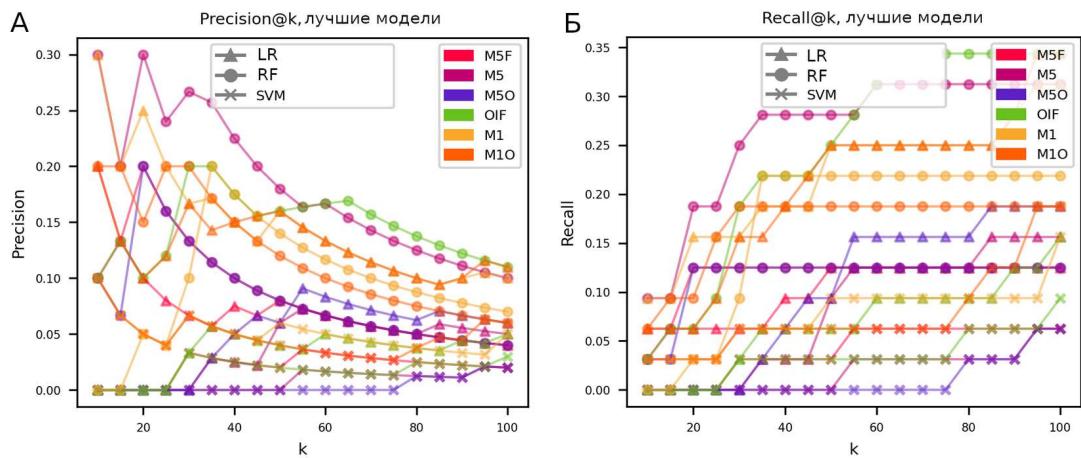


Рисунок 58. Результаты прогнозирования активности лучшими моделями, выбранными методом скользящего контроля, ранжированными по уменьшению спрогнозированной вероятности принадлежности к классу активных соединений. (А) – Precision@ $k$  и (Б) – Recall@ $k$ .

Анализ ROC-кривых (Рисунок 59) моделей, отобранных при скользящем контроле с полнотой в качестве метрики, подтверждает наблюдавшуюся и ранее тенденцию, что добавление фингерпринтов Моргана приводит к построению моделей, обладающих более высокой прогностической способностью.

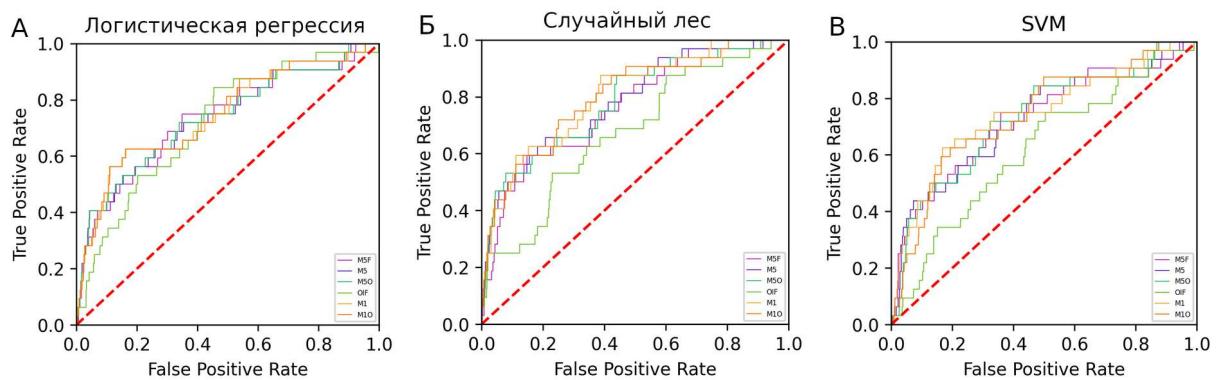


Рисунок 59. ROC-кривые для лучших моделей: (А) – логистической регрессии, (Б) – случайного леса, (В) – SVM, выбранных методом скользящего контроля с полнотой в качестве метрики.

### 6.3. Анализ вкладов дескрипторов в прогноз моделей машинного обучения

Для моделей случайного леса, обученных на наборах данных M5F, M5, M5O, OIF, M1, M1O, был проведен анализ вклада дескрипторов в прогноз при помощи пакета SHAP 0.41.0 [253] (Рисунок 60). Для всех моделей, основанных на 512-битных фингерпринтах Моргана, наибольший вклад в прогноз вносит 66-й бит, соответствующий у большинства молекул гидроксильной группе при ароматическом кольце. Для моделей, содержащих 1024-битные фингерпринты Моргана, 578-й бит, вносящий наибольший вклад в прогноз, также соответствует такой гидроксильной группе. Это часто встречающийся в лекарственноподобных молекулах фрагмент, и его высокий вклад обусловлен природой обучающей выборки.

Второй по величине вклада (а для модели, построенной на наборе данных OIF, – первый) дескриптор – cLogP, рассчитанный при помощи программы DataWarrior [233]. Это один из важных молекулярных дескрипторов, описывающий липофильность молекулы, и его вклад в прогнозирование ее активности закономерен. Единственный элемент фингерпринтов взаимодействия, вошедший в топ-20 важных дескрипторов, — это взаимодействие с Arg188 для модели случайного леса, построенной на наборе данных M5F, содержавшей 512-битные фингерпринты Моргана с радиусом 3 и фингерпринты взаимодействия Flare-IFP. Этот остаток находится в кармане S2, взаимодействия реализуются с карбонильным кислородом или амидным NH основной цепи (Рисунок 61). По всей видимости, важность данного взаимодействия обусловлена небольшими размерами и труднодоступностью кармана S2 для растворителя.

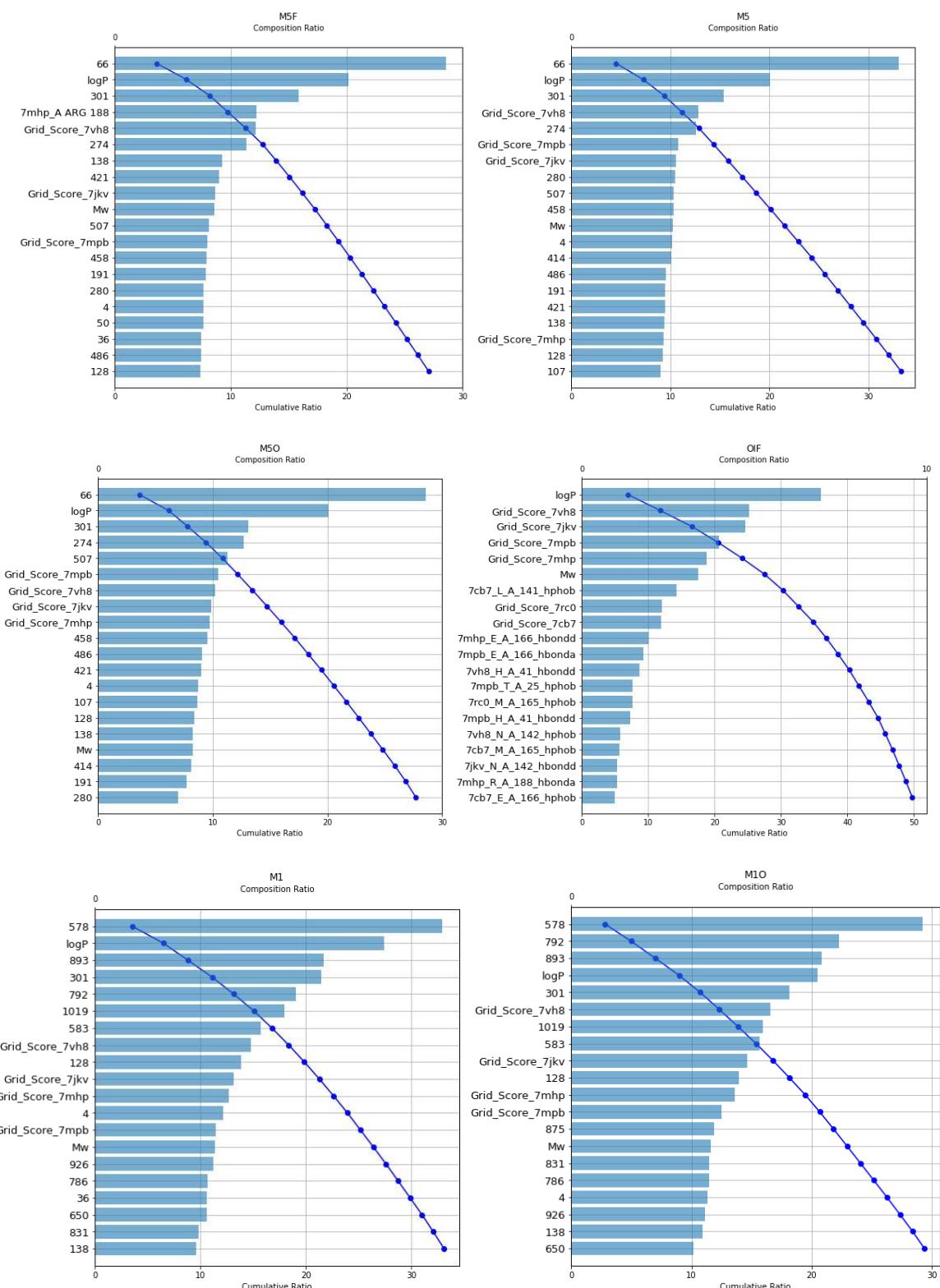


Рисунок 60. Вклады дескрипторов в прогнозы лучших моделей случайного леса, выбранных методом скользящего контроля. Цифрами обозначены номера битов в фингерпринтах Моргана.

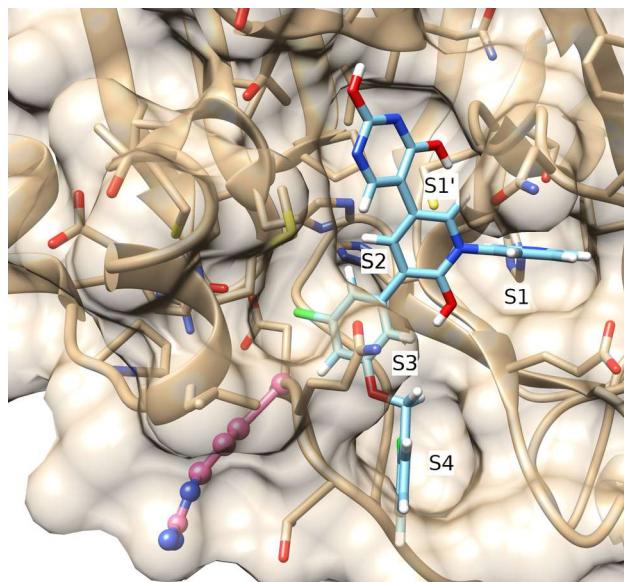


Рисунок 61. Структура Mpro SARS-CoV-2 (PDB id 7CB7) с лигандом **16** из структуры 7L13 (голубой) [132]. Розовым цветом выделен остаток Arg188. Индексами S1-S4 отмечены карманы активного центра.

#### 6.4. Влияние релевантности выборки на качество прогнозирования

Качество моделей, построенных на данных *CHEMBL4495582*, оказалось относительно невысоким. Для сравнения мы повторили все описанные ранее процедуры с молекулами из библиотеки *AA-Feb21*, которая была сформирована и аннотирована вручную по материалам статей и препринтов. Как и ранее, к классу активных относили соединения с  $IC_{50} \leq 10$  мкМ, остальные соединения относили к классу неактивных. Подготовку данных проводили аналогично: для результатов докинга рассчитывали фингерпринты взаимодействия методами Flare и ODDT, для лигандов – фингерпринты Моргана, и формировали обучающие выборки аналогично Таблице 13. Набор данных FLE, содержащий фингерпринты взаимодействия и оценочные функции докинга, нормированные на число тяжелых атомов, не формировали, поскольку результаты построения моделей в предыдущей стадии на наборе FLE были хуже, чем на наборе данных FL, где использовали сами оценочные функции докинга. Дополнительно формировали набор данных GS, состоящий только из оценочных функций докинга в индивидуальные структуры, молекулярной массы лигандов и clogP. Деление выборок на обучающую, контрольную и валидационную проводили аналогично ранее описанной процедуре со стратификацией. Параметры моделей подбирали методом скользящего контроля аналогично процедуре, описанной в разделе 6.2. Методом TabNet модели не строили, так как ранее они отличались самым низким качеством.

Исходя из анализа средних значений ROC AUC и F1-меры (Таблица Г2, Приложение Г), модели *AA-Feb21* характеризуются меньшим переобучением, чем модели, обученные на данных

*CHEMBL4495582* (Рисунок 62). При этом самое низкое качество классификации и наибольшее стандартное отклонение наблюдается для моделей SVM, в связи с чем эти модели были исключены из рассмотрения. Модели CatBoost наиболее переобучены, как и в случае наборов данных, сформированных на основе результатов докинга *CHEMBL4495582*. Точность и полнота на классе активных для моделей *AA-Feb21* (Рисунок 63) также превышают таковые для моделей, обученных на данных *CHEMBL4495582*.

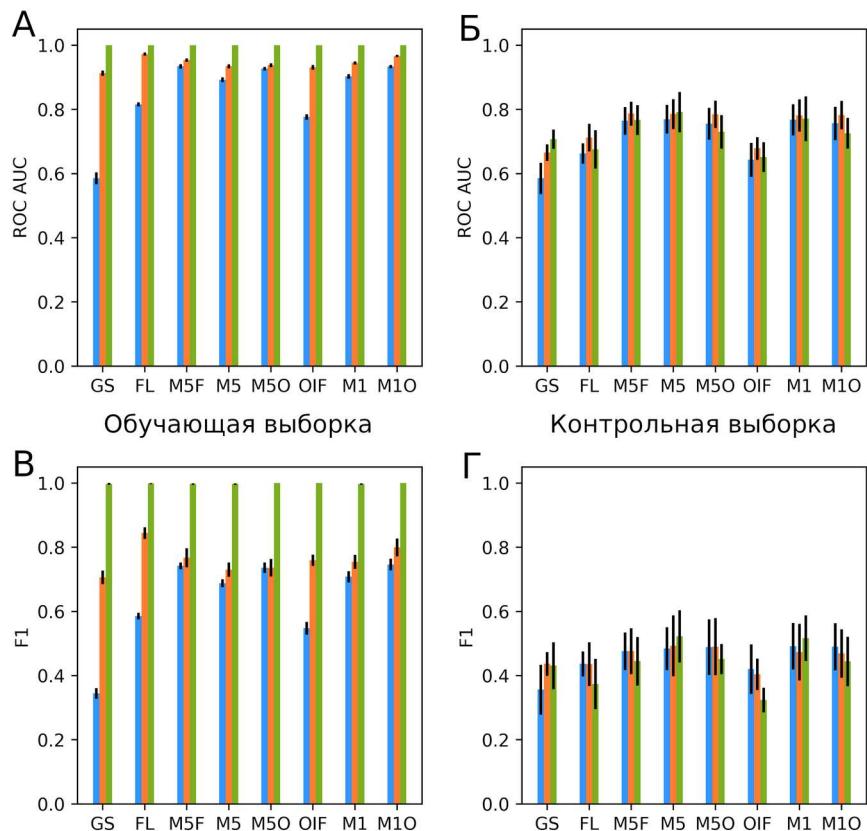


Рисунок 62. Средние значения ROC AUC и F1-меры в обучающей и контрольной выборке для моделей логистической регрессии (голубой), случайного леса (оранжевый) и CatBoost (зеленый), построенных по данным из библиотеки *AA-Feb21*.

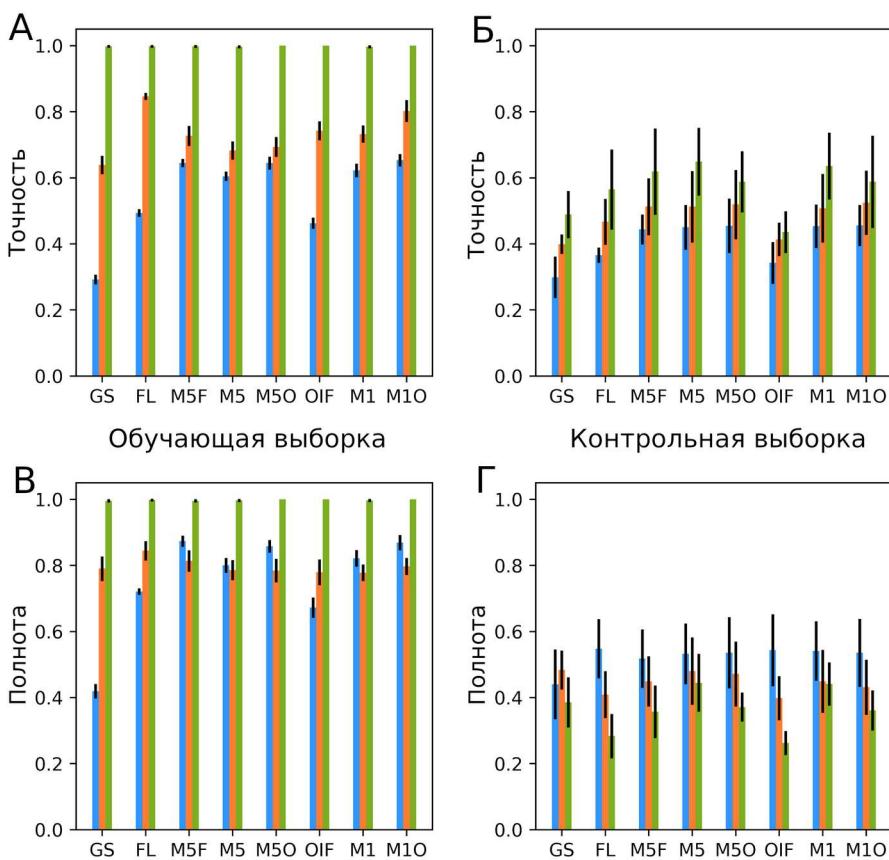


Рисунок 63. Средние значения точности (А, Б) и полноты (В, Г) для класса активных соединений в обучающей и контрольной выборке для моделей логистической регрессии (голубой), случайного леса (оранжевый) и CatBoost (зеленый), построенных на выборках из библиотеки AA-Feb21.

Значения F1-меры, точности и полноты зависят от соотношения активных и неактивных соединений в выборках. Для того, чтобы сравнивать метрики для моделей, построенных на CHEMBL4495582 и на AA-Feb21 с учетом дисбаланса классов, контрольные выборки в последних были дополнены 2000 соединений, отобранными по разнообразию из библиотеки DD-ZINC, которые относили к классу неактивных. Отношение количества активных соединений к неактивным в таких наборах данных составило 0,014. Это дополнение данных практически не влияет на значения полноты, но значительно уменьшает точность прогнозов на контрольной выборке. Таким образом, модели, построенные на AA-Feb21, прогнозируют много ложноположительных результатов, однако достаточно хорошо идентифицируют истинно положительные результаты. При этом значения точности и полноты для класса активных соединений выше для моделей AA-Feb21 (Рисунок 64, 65), чем для моделей CHEMBL4495582 (Рисунок 54), особенно методом случайного леса. В случае моделей CatBoost, построенных на данных AA-Feb21, разброс значений метрик значительно меньше, чем для аналогичных моделей, построенных на данных CHEMBL4495582. Это показывает положительное влияние

релевантности данных на прогностическую способность моделей.

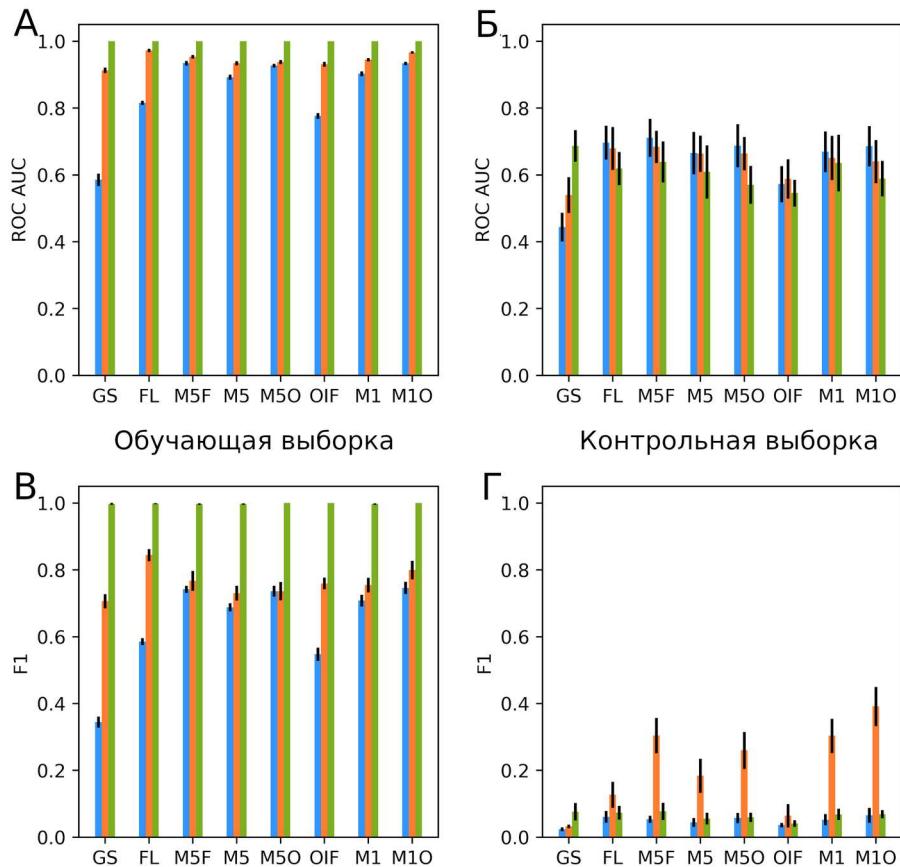


Рисунок 64. Средние значения ROC AUC (А, Б) и F1-меры (В, Г) в обучающей и контрольной выборке для моделей логистической регрессии (голубой), случайного леса (оранжевый) и CatBoost (зеленый), построенных на выборках из библиотеки *AA-Feb21*. Контрольные выборки были дополнены разнообразными соединениями из библиотеки *DD-ZINC*.

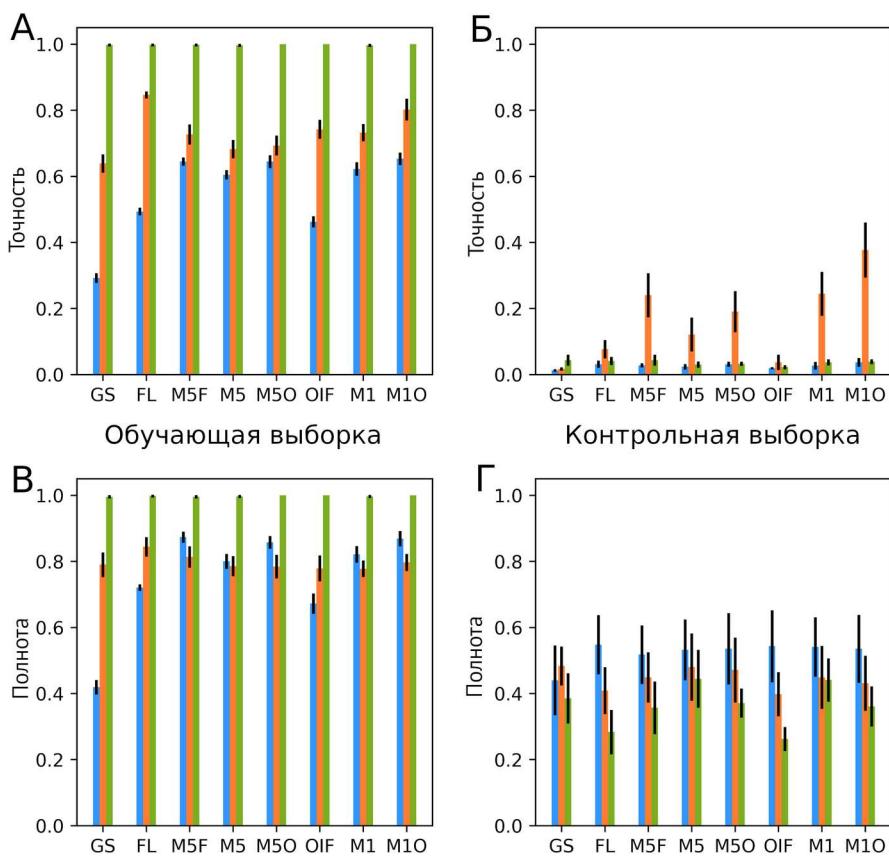


Рисунок 65. Средние значения точности (А, Б) и полноты (В, Г) для класса активных соединений в обучающей и контрольной выборке для моделей логистической регрессии (голубой), случайного леса (оранжевый) и CatBoost (зеленый), построенных на выборках из библиотеки AA-Feb21. Контрольные выборки дополнены разнообразными соединениями из библиотеки DD-ZINC.

В случае моделей AA-Feb21 добавление фингерпринтов взаимодействия позволило получить более высокие значения целевых метрик по сравнению с моделями, построенными только на основе Моргановских фингерпринтов, оценочных функций докинга и молекулярных дескрипторов. Для моделей CHEMBL4495582 такой закономерности не наблюдалось, что можно связать с различиями состава библиотек. Соединения в библиотеке AA-Feb21 в основном принадлежат к сериям, разрабатывавшимся мишень-направленными методами против главной протеазы бетакоронавирусов, и обладают специфичностью по отношению к этому ферменту. Включение в модели информации о взаимодействиях молекул в активном центре фермента и значительное количество примеров специфических взаимодействий способствует тому, что в прогнозах таких моделей меньше ложноположительных результатов, чем в прогнозах моделей, содержащих информацию только о структуре малых молекул, а также в моделях на основе неспецифичных соединений.

В задаче виртуального скрининга важную роль играет раннее обогащение

ранжированного списка соединений активными молекулами. Для анализа обогащения были построены ROC-кривые для прогнозов на контрольных выборках (Рисунок 66А) и контрольных выборках, дополненных разнообразными соединениями (Рисунок 66Б). Для первых характерны большие площади под кривыми, что говорит о более высоком качестве классификации, а для вторых – быстрое возрастание кривой, что свидетельствует о высоком раннем обогащении. Так же наблюдается более высокое качество классификации на основании консенсусной оценки докинга для библиотеки AA-Feb21. Это подчеркивает более высокую релевантность данных в библиотеке AA-Feb21 по сравнению с CHEMBL4495582. В моделях случайного леса и CatBoost наборы данных, содержащие фингерпринты взаимодействия, имеют большие площади под ROC-кривыми, чем аналогичные модели LR. Самым высоким качеством классификации на недополненных выборках обладают модели, построенные на наборах данных M5, M1, M5F, M5O. Набор данных M1O имеет высокое качество на логистической регрессии, однако более низкое на моделях случайного леса и CatBoost, что может быть связано с высокой размерностью данных. Количество полезной информации в более длинных векторах оказывается таким же, если не меньшим, чем в более коротких.

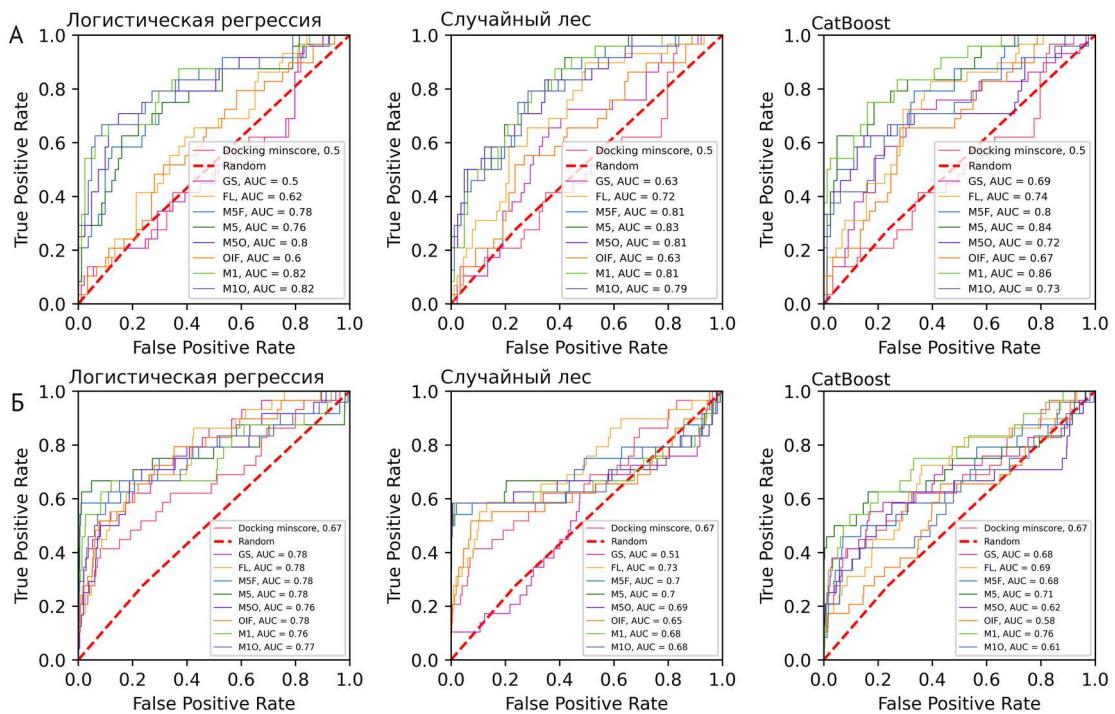


Рисунок 66. ROC-кривые для моделей машинного обучения, построенных на данных AA-Feb21. (А) – для контрольной выборки без дополнений, (Б) – для контрольной выборки, дополненной соединениями из DD-ZINC, отобранными по разнообразию.

ROC-кривые для всех моделей логистической регрессии на AA-Feb21, кроме FL и OIF, ведут себя одинаково на контрольных выборках, дополненных разнообразными соединениями. Модели логистической регрессии, построенные на наборах данных, включающих только

фингерпринты взаимодействия (FL, OIF), делают прогнозы, близкие к случайным. При дополнении разнообразными соединениями лучшие результаты по площади под ROC-кривой и раннему обогащению наблюдаются для набора данных M5, M1 и M5F. Использование фингерпринтов взаимодействия Flare, фиксирующих только наличие водородных связей, хороших и плохих контактов, а также отсутствие контактов, позволяет добиться более высокого качества классификации, чем использование более сложных фингерпринтов ODDT. Возможно, это связано с более высокой размерностью и низкой информативностью фингерпринтов ODDT по сравнению с фингерпринтами Flare. В первых отсутствует отображение “плохих” контактов, препятствующих связыванию, и почти все позиции векторов заняты 0.

Для моделей случайного леса, прогнозы которых, как и ранее, характеризовались наилучшими соотношениями точности и полноты на классе активных соединений, был проведен анализ вкладов дескрипторов (Рисунок 67). Как и для моделей, построенных на данных CHEMBL4495582, значительный вклад в прогноз вносит расчетная липофильность молекулы. Для всех наборов данных, включающих фингерпринты Моргана длиной 512 и 1024 бит, среди самых значимых дескрипторов находятся биты, соответствующие гидроксильной группе в ароматическом кольце, ароматическим пяти- и шестичленным циклам,  $sp^3$ -гибридизованным атомам углерода, одним из заместителей при которых является водород, и фрагментам N-замещенного бензотриазола. В наборах данных, содержащих фингерпринты взаимодействия, важный вклад в прогноз вносят взаимодействия с Arg188 и Ala189 в кармане S2.

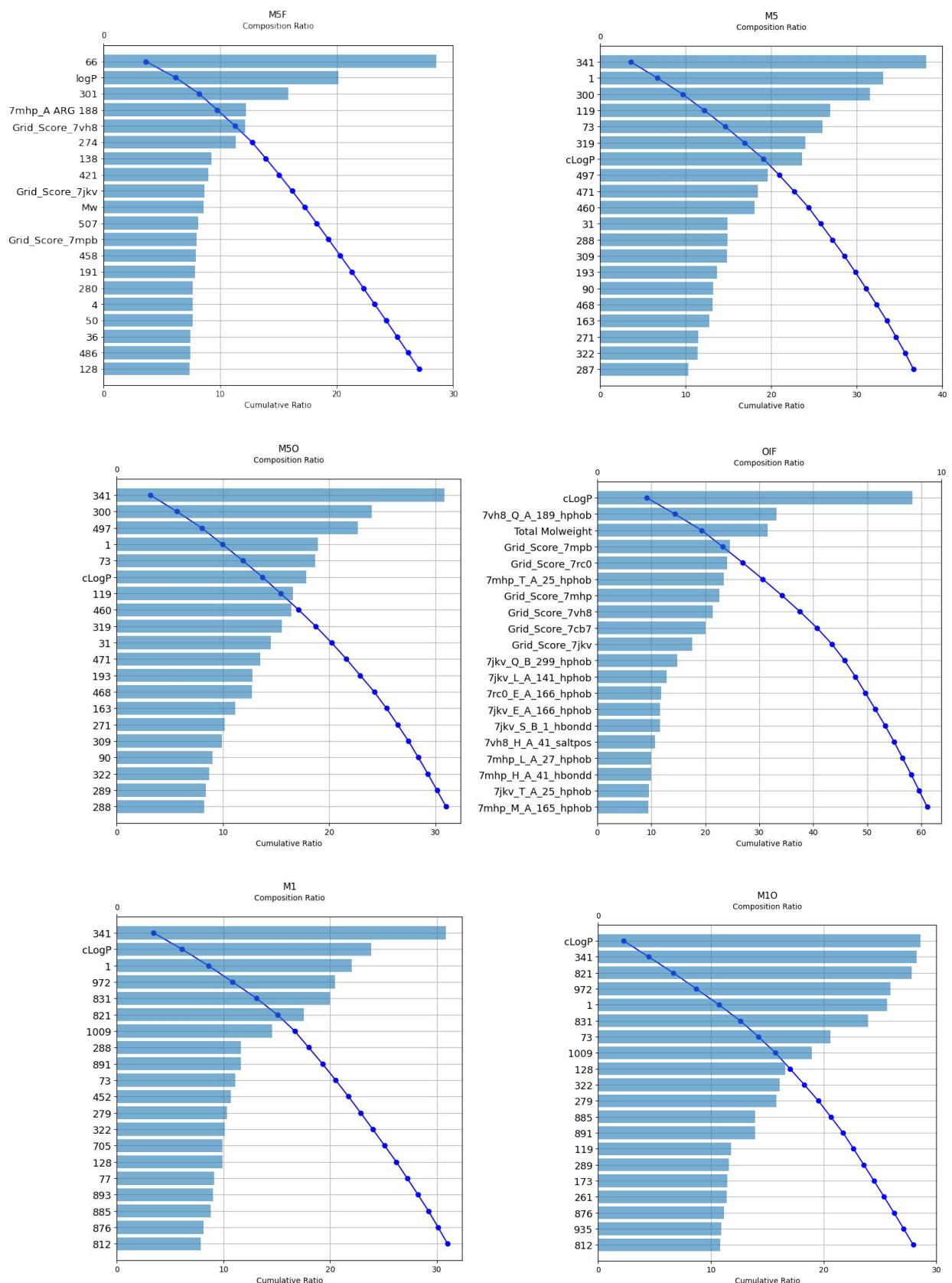


Рисунок 67. Вклады дескрипторов в прогнозы лучших моделей случайного леса, обученных на выборках из библиотеки AA-Feb21. Цифрами обозначены номера битов в фингерпринтах Моргана.

Библиотека *AA-Feb21* меньше по размеру и менее разнообразна, чем результаты высокопроизводительного скрининга *CHEMBL4495582*, кроме того, она собрана из разных источников, из-за чего данные, полученные в разных условиях, могут иметь разные распределения [254], нормировать которые не представляется возможным. Однако в этой библиотеке используется более надежная количественная характеристика активности соединений –  $IC_{50}$ , а не проценты ингибиции при фиксированной концентрации, как в высокопроизводительном скрининге. Соединения библиотеки *AA-Feb21* обладают более высокой специфичностью к главной протеазе коронавирусов и имеют значительно меньший дисбаланс классов, чем результаты высокопроизводительного скрининга. Выборка разнообразных лекарств *CHEMBL4495582* обладает высоким разнообразием, которое должно вести к более высокой обобщающей способности модели. Однако этого не происходит. Задача классификации изначально была выбрана исходя из распределения показателей активности (Рисунок 50А), на котором два класса очень четко разделены, а в задаче регрессии недостаток примеров с целевым значением в некотором диапазоне приводит к понижению качества прогноза. Однако проценты ингибиции *Mrgo* при фиксированной концентрации соединения оказались недостаточно надежной характеристикой активности даже в задаче классификации.

Хотя обе выборки соответствуют рекомендациям по отбору и подготовке данных для машинного обучения [255], качество итоговых моделей недостаточно для их применения на практике. Модели, для которых проводили подбор параметров в широком диапазоне значений, характеризуются высоким переобучением и не обладают достаточной обобщающей способностью. Несмотря на то, что точность классификационных моделей превышает точность виртуального скрининга методом ансамблевого докинга, и, более того, докинга в одну структуру, целесообразность применения построенных моделей для виртуального скрининга затруднена ввиду значительного количества дополнительных вычислительных процедур, необходимых для их использования, при незначительном повышении качества. Использование фингерпринтов взаимодействия улучшает точность классификации, если обучающая выборка содержит достаточно примеров специфических взаимодействий. Дальнейшая разработка моделей машинного обучения, использующих информацию о связывании молекулы с целевым белком, перспективна. Преимущество такого подхода в возможности приоритизации соединений новых хемотипов, отличающихся от тех соединений, на которых обучалась модель. Это расширяет прогностическую способность модели по отношению к лиганд-ориентированному виртуальному скринингу. Однако использование фингерпринтов взаимодействия в машинном обучении требует разработки аннотированных библиотек большего размера, содержащих информацию о разнообразных соединениях с более надежными показателями активности (например,  $IC_{50}$ ).

## Глава 7. Виртуальный скрининг лигандов белка NS1 ортофлавивирусов<sup>11</sup>

Разработка противоортовирусных лекарственных средств – актуальная задача медицинской химии. Однако эта тема не так хорошо разработана, как разработка противокоронавирусных соединений, ставшая очень востребованной в период пандемии COVID-19, и для белков ортофлавивирусов доступно меньше структурной информации. На примере белка NS1 ортофлавивирусов мы рассмотрели возможность использования моделирования по гомологии в качестве источника структурных данных для ансамблевого докинга.

### 7.1. Сравнительный анализ структур белка NS1 ортофлавивирусов

Структуры белка NS1 ортофлавивирусов (OFV-NS1) извлекали из БД PDB путём анализа результатов поиска по таксономии организмов-источников (Source organism). Для извлечения некорректно аннотированных структур, которые не идентифицируются поиском среди белков вирусов рода *Orthoflavivirus*, была использована программа BLASTP [256], в качестве запроса для поиска по базу данных PDB была использована последовательность белка NS1 серотипа 1 вириуса денге.

В Таблице 14 представлены общие сведения о структурах белков NS1 различных ортофлавивирусов, депонированных в PDB. Только 11 структур имеют аминокислотную последовательность всех структурных доменов (в дальнейшем будем называть их “полными структурами”). Остальные представляют собой только фрагмент белка с β-лестницей и спагетти-петлей (см. раздел 1.3, указаны в таблице 14 как С-домен). На момент проведения изначального анализа структуры 7WUS, 7WUT, 7WUU, 7WUV и 7WUR еще не были опубликованы, однако процедуры были повторены впоследствии с включением этих структур, и значительных отличий от приведенных далее результатов не было обнаружено.

---

11 При работе над данным разделом диссертации использованы материалы следующих публикаций автора, в которых, согласно Положению о присуждении ученых степеней в МГУ, отражены основные результаты, положения и выводы исследования:

1. Фомина А.Д., Палюлин В.А., Осолодкин Д.И. Моделирование структуры белка NS1 ортофлавивирусов по гомологии для виртуального скрининга потенциальных лигандов // Биомедицинская химия – 2024 – т. 70 (6) – с. 456–468
2. Fomina A. D., Uvarova V. I., Kozlovskaya L. I., Palyulin V. A., Osolodkin D.I., Ishmukhametov A. A. Ensemble Docking Based Virtual Screening of SARS-CoV-2 Main Protease Inhibitors // Molecular Informatics – 2024 – V.43. – e202300279.
3. Zakharova M.Yu., Kuznetsova A.A., Uvarova V.I., Fomina A.D., Kozlovskaya L.I., Kaliberda E.N., Kurbatskaia I.N., Smirnov I.V., Bulygin A.A., Knorre V.D., Fedorova O.S., Varnek A., Osolodkin D.I., Ishmukhametov A.A., Egorov A.M., Gabibov A.G., Kuznetsov N.A. Pre-Steady-State Kinetics of the SARS-CoV-2 Main Protease as a Powerful Tool for Antiviral Drug Discovery // Frontiers in Pharmacology – 2021 – V. 12. – P. 773198.
4. Фомина А.Д., Осолодкин Д.И. Программный комплекс для систематического ансамблевого докинга // Свидетельство о государственной регистрации программы для ЭВМ № 2024684465. – Номер заявки 2024684027. – Дата поступления 17.10.24. – Дата государственной регистрации в Реестре программ для ЭВМ 17.10.24.
5. Фомина А.Д., Осолодкин Д.И. База данных соединений с экспериментально определенной активностью против главной протеазы β-коронавирусов // Свидетельство о государственной регистрации базы данных № 2024625235. – Номер заявки 2024624786. – Дата поступления 29.10.2024 – Дата государственной регистрации в Реестре баз данных 15.11.2024 Бюл. № 11

Таблица 14. Основная информация о структурах NS1, доступных в БД PDB.

Код PDB	Вирус	Год	Метод*	Полнота структуры (указанные разрешенные участки)**	Олигомерное состояние	Разрешение, Å	Источник
4O6B	ДЕНВ2	2014	P	A: 0-7, 11-107, 129-158, 166-349 B: 1-106, 131-161, 164-349	Димер	3,00	[168]
4OIG	ДЕНВ1	2014	P	С-домен: 178-352	Димер	2,69	[257]
5IY3	ЗИКВ	2016	P	С-домен: 176-352	Димер	2,20	[184]
5K6K	ЗИКВ	2016	P	A:-5 – 112, 121-352 B: -1 – 352	Димер	1,89	[258]
5GS6	ЗИКВ	2016	P	A: 1-352 B: -4 – 25, 34-352	Димер	2,85	[259]
5X8Y	ЗИКВ	2017	P	С-домен: A,B: 176-351	Димер	2,82	[260]
5YXA	ВЖЛ	2018	P	С-домен, мономер: 176-352	Димер	2,10	[261]
4O6D	ВЛЗН	2014	P	A:-6 – 107, 129-352 B: -4 – 107, 130-352	Димер	2,59	[168]
4O6C	ВЛЗН	2014	P	A,B: 0-109, 125-352	Димер	2,75	[168]
4OII	ВЛЗН	2014	P	С-домен+антитело A,B: 176-352	Комплекс димера с антителами	3,0	[257]
4OIE	ВЛЗН	2014	P	С-домен, мономер 176-352	Димер	1,85	[257]
4TPL	ВЛЗН	2014	P	A: -10-107, 129-352 B: -20 – 108,130-352	Димер	2,90	[262]
5O19	ВЯЭ	2018	P	С-домен, мономер 177-352	Димер	2,10	[263]
5O36	ВЯЭ	2018	P	С-домен, мономер 177-354	Димер	2,60	[263]
7WUS	ДЕНВ2	2022	K	Полная структура	Димер	3,40	[182]
7WUT	ДЕНВ2	2022	K	Полная структура	Тетramer	3,5	[182]
7WUU	ДЕНВ2	2022	K	Полная структура	Тетramer	8,0	[182]
7WUV	ДЕНВ2	2022	K	Полная структура	Тетramer	8,3	[182]
7WUR	ДЕНВ2	2022	K	Полная структура	Комплекс димера с антителами	3,5	[182]

\* P – Рентгеноструктурный анализ, K – Криоэлектронная микроскопия

\*\* A, B – субъединицы димерного белка

Контроль стереохимического качества полных структур OFV-NS1 осуществляли на основе карт Рамачандрана, реализованных в ProCheck [264] (Таблица 15). Структуры обладают

достаточно высоким качеством: лишь отдельные остатки расположены в неблагоприятных областях карт, что позволяет сделать вывод об общей корректности данных.

Таблица 15. Результаты стереохимической валидации структур NS1. Указаны количества и проценты остатков, относящихся к соответствующей области карт Рамачандрана.

Структура	4O6B	4O6D	4TPL	4O6C	5K6K	5GS6
предпочтительные области карты	475 (85,7%)	510 (86,6%)	525 (86,9%)	536 (91,5%)	524 (89,0%)	492 (81,6%)
разрешенные области карты	76 (13,7%)	76 (12,9%)	76 (12,6%)	46 (7,8%)	64 (10,5%)	104 (17,2%)
допустимые области карты	1 (0,2%)	1 (0,2%)	3 (0,5%)	2 (0,3%)	3 (0,5%)	6 (1,0%)
неблагоприятные области карты	2 (0,4%)	2 (0,3%)	0 (0,0%)	2 (0,3%)	0 (0,0%)	1 (0,2%)

Выравнивание аминокислотных последовательностей проводили с помощью сервера MUSCLE [205]. Использовали последовательности NS1 BKЭ (штамм Абсеттаров, UniProtKB – Q01299) и ДЕНВ-1 (штамм Nauru/West Pac/1974, UniProtKB – P17763), а также последовательности, извлечённые из pdb-файлов. Последовательности были извлечены из pdb-файлов для корректного описания состояния белка, наличия инженерных конструкций и мутаций. Выравнивание аминокислотных остатков с окраской по идентичности приведено на Рисунке Д1 (Приложение Д). Матрица гомологичности последовательностей, рассчитанная на основе выравнивания, приведена на Рисунке 68. Белок NS1 BKЭ обладает идентичностью с белками NS1 других ортофлавивирусов на уровне 39-50%, что позволяет сделать вывод о высокой гомологии. Белок NS1 ВЖЛ (PDB ID 5YXA) обладает примерно тем же уровнем подобия с остальными последовательностями. С-Домены белка NS1 ВЯЭ (5O36, 5O19) имеют повышенную идентичность с NS1 ВЛЗН (4TPL, 4OIE, 4OII, 4O6C, 4O6D). Эти вирусы принадлежат к одному серокомплексу [265], поэтому закономерна более высокая идентичность последовательностей в области белка, в которой располагаются антигенные детерминанты. Во всех случаях идентичность последовательностей значительно выше внутри вида, чем между видами. Использование структур биомицелии из разных вирусов может способствовать выявлению потенциальных противовирусных препаратов широкого спектра действия, поэтому разнообразие аминокислотных последовательностей учитывали при выборе структур ансамбля.

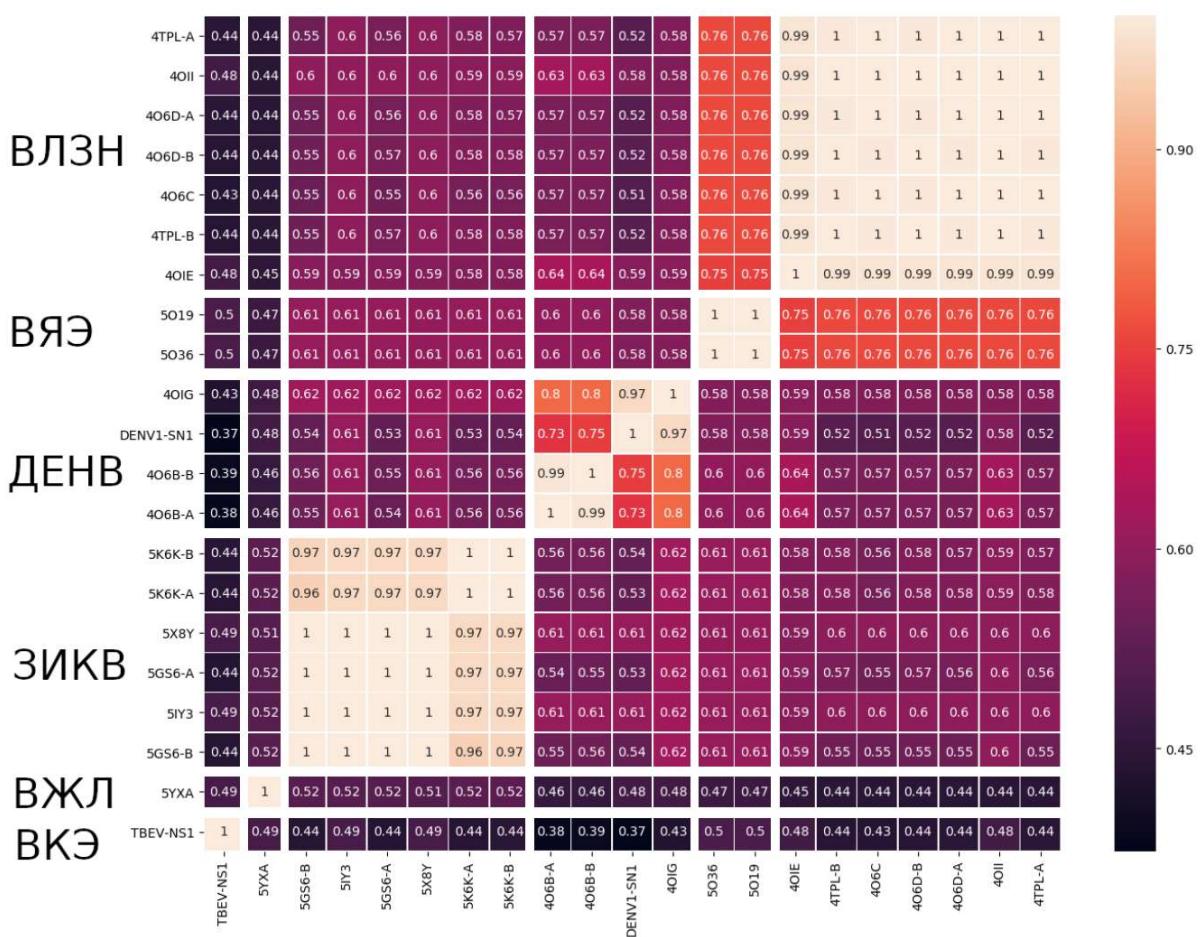


Рисунок 68. Матрица гомологичности выровненных участков последовательностей OFV-NS1. Приведены доли идентичных остатков.

Пространственное выравнивание по мономеру проводили в программе UCSF Chimera [201] при помощи инструмента Match->Align со следующими параметрами: порог расстояния между остатками 5,0 Å, выравнивание остатков при наличии хотя бы одного на той же позиции выравнивания, итерации наложения и выравнивания до полной конвергенции. В качестве неподвижной структуры использовали цепь А структуры 4O6B (ДЕНВ-2).

Выравнивание димеров полных структур было сделано методом Matchmaker со следующими параметрами: наложение по оптимально выравнивающейся цепи, присвоение элементов вторичной структуры, учёт вторичной структуры в оценке с весом 30%.

После выравнивания структур рассчитывали значения среднеквадратичного отклонения (СКО) координат С $\alpha$ -атомов для пар структур. Визуализацию попарных СКО осуществляли методом тепловых карт, реализованным на языке Python 3.9 с использованием библиотек Pandas 1.4.1 [210], NumPy 1.21.6 [211], Matplotlib 3.5.1 [212] и seaborn 0.11.2 [213].

Пространственное выравнивание мономеров проводили для всех структур NS1 из PDB и отдельно – только для полных структур с целью выявления конформационных различий структур на уровне мономера, а также определения степени симметричности мономеров в

одной структуре. Для количественного анализа результатов выравнивания были построены тепловые карты попарных СКО (Рисунок 69).

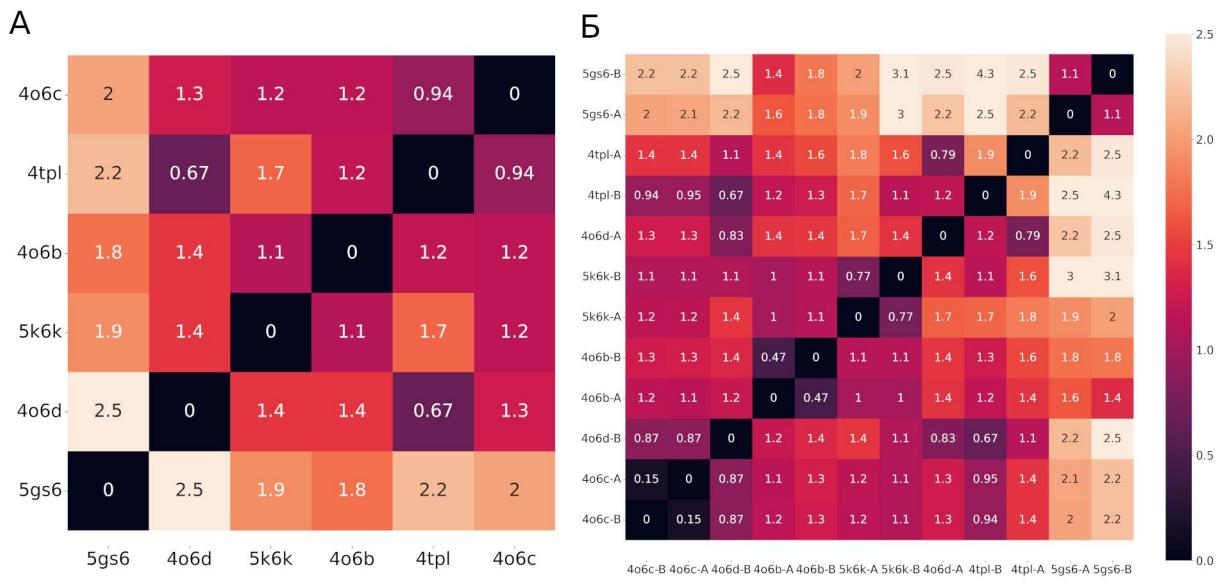


Рисунок 69. Тепловая карта СКО координат C<sub>α</sub>-атомов пространственного выравнивания димеров (А) и мономеров (Б) полных структур [158]. Цветовая шкала соответствует значениям СКО от 0 до 2,5 Å.

При пространственном выравнивании мономеров полных структур NS1, извлеченных из PDB, не было выявлено существенных различий: для подавляющего большинства структур СКО не превышает 1.2 Å, что говорит об очень высокой степени сходства структур. На рисунке 69Б визуализированы попарные СКО для мономеров полных структур (4O6B, 5K6K, 5GS6, 4O6D, 4O6C, 4TPL). Структура 5GS6 наиболее сильно отличается от остальных – значения СКО значительно превышают 1,2 Å.

Простое выравнивание димеров проводили для определения общего СКО и отбора структур для ансамблевого докинга. Все структуры очень консервативны, что подтверждается низкими значениями попарного СКО: для большинства пар  $\leq 1,4$  Å. Такие значения СКО объясняются разрешением структур и общей подвижностью полипептидных цепей [266].

При выравнивании димеров наиболее заметные конформационные различия структур 5GS6 и 5K6K обусловлены отклонениями в области β-бочки (5DS6-B, остатки 6-14) и α-спирального крыла (5DS6-B, А остатки 62-82), где находятся подвижные петли (Рисунок 70). Для всех структур не наблюдали значительных конформационных отличий во взаимном расположении мономеров в димере.

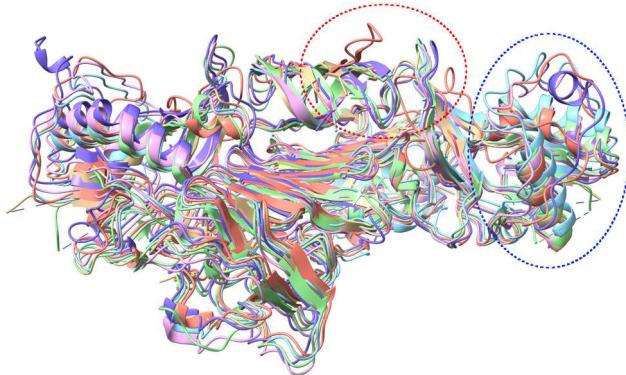


Рисунок 70. Пространственное выравнивание димеров полных кристаллических структур белка NS1 ортофлавивирусов [158]. Выделены области белка, обуславливающие наибольшие вклады в увеличение попарного СКО: красный –  $\beta$ -бочка, остатки 6-14; синий –  $\alpha$ -спиральное крыло, остатки 62-82.

На основе стереохимической корректности и полноты разрешения структур, пространственных выравниваний и выравниваний аминокислотных последовательностей, для дальнейшей работы были выбраны наиболее разнообразные и качественные структуры: 4O6C (ВЛЗН), 4O6D (ВЛЗН), 5K6K (ЗИКВ) и 5GS6 (ЗИКВ).

## 7.2. Поиск потенциальных сайтов связывания малых молекул

Поскольку экспериментальные данные о способах связывания малых молекул с OFV-NS1 отсутствуют, мы применили расчётные методы поиска карманов, способных связывать потенциальные лиганды. Поиск проводили для структур 4O6C, 4O6D, 5K6K и 5GS6 — отдельно для димера и мономера каждой структуры. Для структур 4O6D и 5GS6 поиск карманов проводили до и после удаления фрагментов дополнительных генно-инженерных конструкций, однако результаты совпали, поэтому далее наличие дополнительных фрагментов не учитывается.

Поиск карманов проводили для всех структур белка, из которых были удалены молекулы растворителя, при помощи двух программ, в основе которых лежат различные методы идентификации сайтов связывания: FTSite (FTMap) [267] и ProteinsPlus (DoGSiteScorer) [268]. Обе программы использовали через веб-интерфейс. Найденные карманы представляют собой наборы точек или сфер, заполняющие потенциальные сайты связывания.

### 7.2.1. Поиск методом FTSite

Метод, реализованный в программе FTSite, основан на энергетическом методе выявления карманов. В его рамках 16 малых молекул используются в качестве зондов и для каждой из них рассчитываются оптимальные по энергии позиции при помощи эмпирических функций

свободной энергии. Кластеры каждого типа зондов ранжируются на основе средней свободной энергии. Первичные кластеры разных зондов объединяются, и объединенные кластеры ранжируются на основе общего количества взаимодействий между белком и всеми зондами кластера.

В результате работы программы FTSite выводятся только три наиболее высоко оцененных кармана. Для каждого из карманов при помощи программы UCSF Chimera была рассчитана общая площадь поверхности и площадь поверхности, доступной для растворителя (Таблица 16). Расчеты были проведены отдельно для димеров и мономеров, однако результаты оказались похожи. Для всех структур, как димерных, так и мономерных, обнаруживаются карманы в области между  $\beta$ -бочкой, доменом-коннектором,  $\alpha$ -спиральным крылом и  $\beta$ -лестницей, а также карманы на внешней поверхности  $\alpha$ -спирального крыла (Рисунок 71).

Таблица 16. Сайты FTSite. 1 – Площадь поверхности,  $\text{\AA}^2$  ; 2 – Площадь поверхности, доступной для растворителя,  $\text{\AA}^2$

Димеры								
Карман	4O6C		4O6D		5K6K		5GS6	
	1	2	1	2	1	2	1	2
FT1	254,523	496,587	378,484	684,588	317,113	559,968	241,590	472,430
FT2	278,533	521,433	66,372	202,736	286,198	543,746	215,907	435,082
FT3	207,685	428,336	309,400	565,472	124,496	293,716	124,955	296,750
Мономеры								
Карман	4O6C		4O6D		5K6K		5GS6	
	1	2	1	2	1	2	1	2
FT1	229,094	453,904	317,113	559,968	304,575	550,396	324,625	563,421
FT2	232,496	454,591	286,198	543,746	191,382	392,054	324,261	588,642
FT3	272,384	511,209	124,496	293,716	189,763	391,433	147,632	333,571

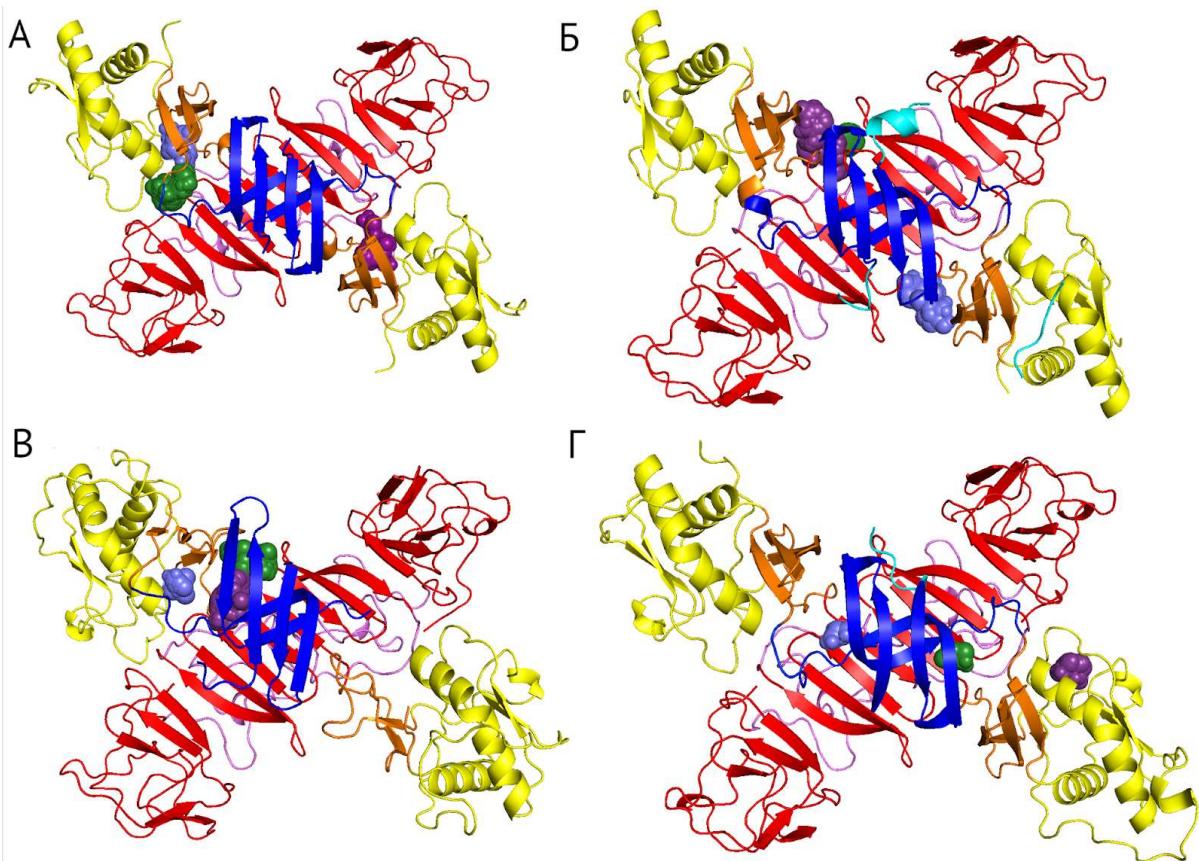


Рисунок 71. Карманы FTSite для димерных структур, окрашенные по убыванию оценки: фиолетовым, темно-зеленым и светло-синим [158]. А – 4O6C (ВЛЗН), Б – 4O6D (ВЛЗН), В – 5GS6 (ЗИКВ), Г – 5K6K (ЗИКВ). Окраска доменов: синий –  $\beta$ -бочка, оранжевый – домен-коннектор, желтый –  $\alpha$ -спиральное крыло, красный –  $\beta$ -лестница), розовый – “спагетти-петля”.

### 7.2.2. Поиск методом DoGSiteScorer

Алгоритм программы ProteinsPlus (DoGSiteScorer) относится к геометрическим методам и использует анализ молекулярной формы для идентификации карманов. В ходе работы алгоритма белок помещается в сеточное поле, после чего точки сетки помечаются как занятые, если они лежат на расстоянии ван-дер-ваальсового радиуса от какого-либо атома белка, и как свободные в противном случае. DoGSiteScorer на основании этой разметки идентифицирует полости, которые достаточно велики, чтобы вместить хотя бы один тяжелый атом. В качестве результата работы программы DoGSiteScorer выводится 20 карманов, ранжированных по объему. Из них отбирали от 3 до 5 наиболее высоко оцененных карманов (они получали индекс a), а также симметричные им (указаны с индексом b). В случае, если два кармана с меньшим рейтингом находятся в области, симметричной карману с большим индексом, им присваивали индексы b1 и b2).

Карманы, обнаруженные при помощи программы DoGSiteScorer, большего размера, чем обнаруженные при помощи FTSite (Таблица 17). Однако местоположения карманов схожи –

между  $\beta$ -бочкой, доменом-коннектором,  $\alpha$ -спиральным крылом и  $\beta$ -лестницей, и на внешней поверхности  $\alpha$ -спирального крыла (Рисунок 72). Результаты поиска для димеров и мономеров гораздо более похожи между собой, чем таковые у FTSite.

Для мономерных структур 4O6C (ВЛЗН) и 4O6D (ВЛЗН) карманы, найденные DoGSiteScorer в области  $\beta$ -бочки, явно имеют дефекты расчета, поскольку в обоих случаях поверхность кармана пересекается со структурой белка. В остальных случаях подобных ошибок не наблюдалось. Однако это дает основания предположить, что объем потенциальных сайтов связывания может быть завышен.

Таблица 17. Характеристика карманов DoGSiteScorer для мономеров. SimpleScore и drugScore – оценки программы, соответствующие общей характеристике кармана и его пригодности для связывания низкомолекулярных органических веществ.

	Объем, $\text{\AA}^3$	Площадь поверхности, $\text{\AA}^2$	Глубина, $\text{\AA}$	Гидрофобность	simpleScore	drugScore
Мономеры						
4O6C	1	999,17	1044,15	26,18	0,21	0,57
	2	345,98	673,26	10	0,4	0,19
	3	294,08	620,58	12,46	0,47	0,17
	4	265,02	534,23	15,23	0,52	0,15
4O6D	1	441,41	303,44	19,88	0,14	0,17
	2	300,61	522,24	12,14	0,48	0,17
	3	275,26	570,65	14,32	0,43	0,13
	4	270,02	298,72	12,76	0,3	0,07
5K6K	1	1222,46	1542,01	27,04	0,34	0,62
	2	268,29	378,66	10,8	0,42	0,12
	3	244,93	640,01	10,2	0,44	0,1
	4	143,36	353,74	7,82	0,26	0
5GS6	1	569,73	519,83	20,29	0,23	0,31
	2	500,86	890,33	12,27	0,36	0,3
	3	270,08	427,18	13,08	0,38	0,11
	4	228,99	555,95	10,75	0,41	0,07
	5	221,7	288,07	12,2	0,34	0,04
Димеры						
4O6C	1.a	1034,5	965,84	23,79	0,1	0,53
	1.b1	811,61	891,58	31,44	0,15	0,44
	2.a	306,42	568,33	10,97	0,44	0,16

		Объем, $\text{\AA}^3$	Площадь поверхности, $\text{\AA}^2$	Глубина, $\text{\AA}$	Гидрофобность	simpleScore	drugScore
4O6D	2.b	288,97	574,57	9,76	0,39	0,13	0,43
	1.b2	259,92	255,98	17,32	0,09	0	0,68
	3.a	247,27	512,7	15,13	0,53	0,14	0,61
	3.b	212,36	500	14,69	0,55	0,11	0,57
5K6K	1.b	683,14	652,83	24,76	0,11	0,34	0,86
	1.a	420,66	316,57	18,96	0,12	0,15	0,78
	2.a	286,69	222,41	16,38	0,05	0	0,66
	2.b	261,47	203,41	16,63	0,04	0	0,66
	3	213,44	286,62	12,36	0,19	0	0,42
	4	206,21	453,81	13,73	0,17	0	0,52
5GS6	1.b	902,45	1008,15	25,22	0,26	0,53	0,83
	2.b	889,25	1135,66	33,77	0,31	0,55	0,84
	1.a	778,21	852,62	23,79	0,21	0,44	0,85
	2.a	752,7	843,88	31,09	0,27	0,45	0,88
	3.a	260,16	386,31	11,14	0,44	0,12	0,47
	3.b	258,64	414,12	9,4	0,46	0,13	0,41

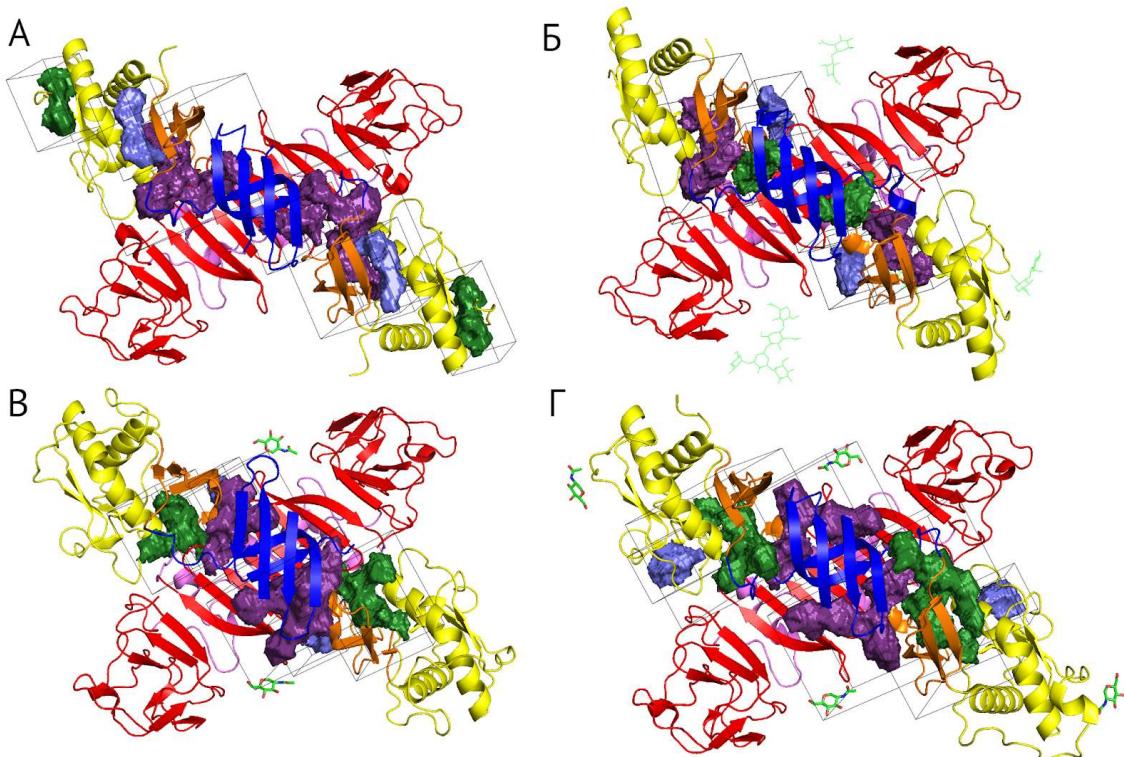


Рисунок 72. Карманы DoGSiteScorer для димеров OFV-NS1 [158]. А – 4O6C (ВЛЗН), Б – 4O6D (ВЛЗН), В – 5GS6 (ЗИКВ), Г – 5K6K (ЗИКВ). Окраска доменов и карманов аналогична рисунку 71.

### 7.2.3. Выбор потенциального сайта связывания

Обнаруженные программой FTSite карманы в структуре димера 4O6C находились в той же области, что и карманы 1.a и 1.b, обнаруженные DoGSiteScorer, хотя карманы DoGSiteScorer больше по размеру. FT1 лежит в нижней части кармана 1.a – между коннектором, спагетти-петлей и  $\alpha$ -спиральным крылом цепи А. Карманы FT2 и FT3 расположены внутри кармана 1.b: FT2 между  $\beta$ -бочкой,  $\beta$ -лестницей и  $\alpha$ -спиральным крылом, а FT3 – симметрично FT1. Карманы 1.a и 1.b1 имеют оценку Drug Score  $> 0,8$ , что указывает на их перспективность в качестве сайтов связывания низкомолекулярных лигандов.

Все карманы FTSite для мономерной структуры 4O6C лежат внутри кармана 1 DoGSiteScorer, Drug Score которого составляет 0,81. По сравнению с димером, DoGSiteScorer дополнительно находит карман в области  $\beta$ -бочки, а в случае FTSite – карман 4O6C-A FT1, лежащий между  $\beta$ -бочкой,  $\beta$ -лестницей и коннектором. Карман 1 DoGSiteScorer мономера также больше в этой области, чем карман димера.

Примечательными отличиями структуры 4O6D от структуры 4O6C, которая принадлежит белку того же вируса, являются более низкая оценка программой DoGSiteScorer областей между  $\beta$ -бочкой и  $\beta$ -лестницей и разделение карманов, для 4O6C определявшихся как 1.a и 1.b, на четыре кармана в случае 4O6D.

В случае мономера 4O6D карман 1 DoGSiteScorer, единственный с Drug Score  $\geq 0,8$ , находится там же, где карман 1 для структуры 4O6C, однако FTSite не определяет карманов в этой области. FT1 совпадает с карманом 4, Drug Score которого 0,54. По сравнению с димером, программой FTSite выделен дополнительный карман в области над  $\beta$ -лестницей, а ранг кармана, находящегося рядом с FT1, понижен со 2 места до 3.

Для мономера 5GS6 результаты работы программы DoGSiteScorer свидетельствуют о более высокой доступности для растворителя пространства между  $\alpha$ -спиральным крылом, соединительным доменом и спагетти-петлей. Находящийся там карман обладает самым высоким рейтингом. FTSite ставит аналогичный карман на второе место. Карман FT1 получил, напротив, низкую оценку DoGSiteScorer, его Drug Score  $\sim 0,5$ . Карман 2 DoGSiteScorer расположен на боковой стороне  $\beta$ -лестницы и хотя и выявляется этой программой для всех структур, обычно имеет довольно низкую оценку. Метод FTSite не позволяет обнаружить карман в этом месте ни для одной из структур. Аналогичная ситуация наблюдается с карманом 3. Карман FT3 находится приблизительно в той же боковой области  $\alpha$ -спирального крыла, однако с карманом 3 не пересекается.

В случае димера 5GS6 карман FT1, имеющий наиболее высокую оценку FTSite, находится в области кармана 3.b DoGSiteScorer – на боковой части  $\alpha$ -спирального крыла. Drug Score кармана 3.b составляет 0,41, что является недостаточно высоким показателем для рассмотрения его в качестве сайта связывания потенциальных лигандов. Однако другие два кармана FTSite лежат в области между  $\beta$ -бочкой и  $\beta$ -лестницей, где расположены карманы, получившие от DoGSiteScorer наивысшую оценку и Drug Score  $> 0,8$ . Следует отметить, что для этой структуры FTSite выявил карманы очень маленького размера.

Для мономера 5K6K обе программы ранжируют карманы одинаково. Карман FT1, расположенный между  $\alpha$ -спиральным крылом, коннектором и  $\beta$ -лестницей, лежит внутри области, обозначенной DoGSiteScorer как карман 1. Только он имеет Drug Score  $> 0,8$ . Карман FT2 совпадает с карманом 2 и лежит на боковой стороне  $\alpha$ -спирального крыла. Также DoGSiteScorer выявляет карман 3 в верхней части  $\alpha$ -спирального крыла, под его гибкой петлей, однако его Drug Score меньше 0,5. Карман 4, имеющий наиболее близкое расположение к карману FT3, имеет низкий Drug Score, и облако точек, формирующее карман, перекрывается с атомами белка, что, вероятно, является ошибкой расчета.

Обоими методами был обнаружен карман в области интерфейса димеризации между  $\beta$ -бочкой, доменом-коннектором,  $\alpha$ -спиральным крылом и  $\beta$ -лестницей (Рисунок 73А), образованный остатками 1-15, 186-193, 202-207 одной цепи и 1, 16-26, 156-162, 179-185, 193-200 другой цепи (Рисунок 73Б). Боковые цепи остатков  $\beta$ -бочки направлены вовне кармана,

образуя гидрофобную поверхность, которая способствует взаимодействию белка с мембранами (Рисунок 73В), при этом внутрь кармана ориентированы в основном гидрофильные остатки и атомы основной цепи, способные выступать в качестве доноров или акцепторов водородных связей. В более ранних работах по моделированию молекулярной динамики [193] и докингу [194] в этой области была показана возможность связывания лигандов для мономера белка NS1. Этот карман – наибольший по размеру, обнаруживается во всех кристаллических структурах белка NS1 и находится в области сайта димеризации, который важен для функционирования белка, поэтому он был выбран в качестве потенциального сайта связывания лигандов для виртуального скрининга и последующего анализа.

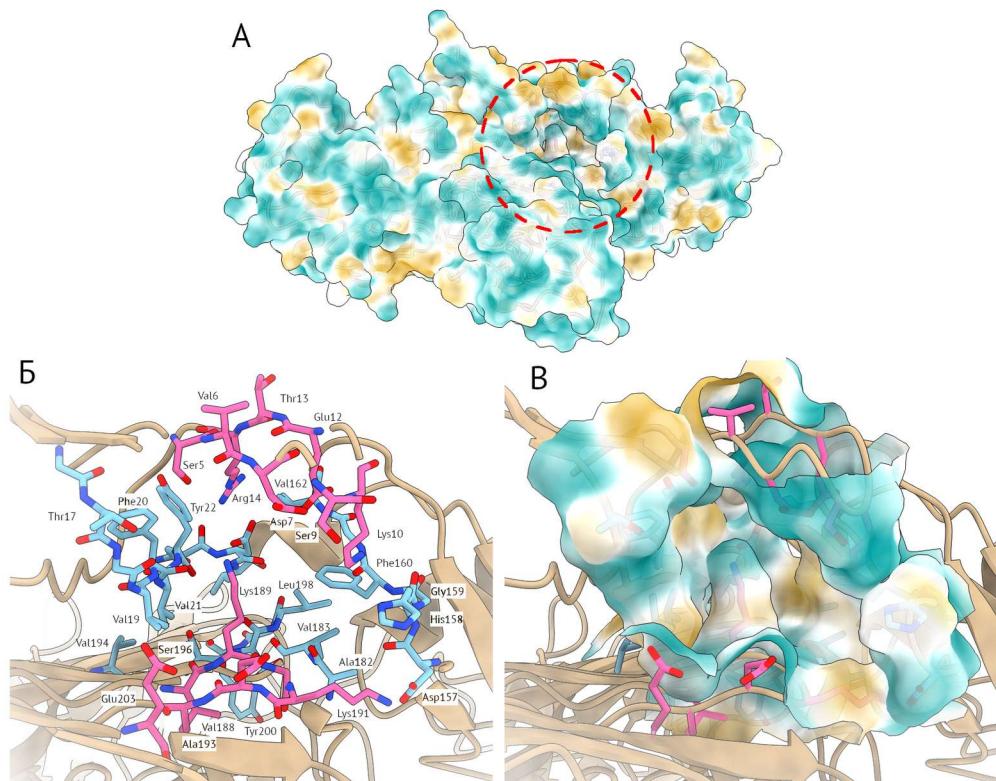


Рисунок 73. Потенциальный сайт связывания малых молекул [158]. (А) – Расположение кармана, (Б) – остатки, формирующие карман, окрашенные в соответствии с цепью, где цепь А розовая, а цепь В – светло-голубая, (В) – поверхность кармана, окрашенная по шкале гидрофобности Кайта-Дулитла, где желтый цвет отображает высокую, а голубой – низкую гидрофобность.

### 7.3. Моделирование ансамбля

Несмотря на малое количество экспериментальной структурной информации, систематический отбор ансамбля может быть применен для моделей OFV-NS1, построенных по гомологии. Использование для виртуального скрининга такого ансамбля, составленного из структур белка разных вирусов, могло бы способствовать приоритизации потенциальных лигандов широкого спектра действия.

### 7.3.1. Построение моделей структур белка NS1 ортофлавивирусов по гомологии

Выравнивание аминокислотных последовательностей белка NS1 эпидемиологически значимых ортофлавивирусов [156], [269] (Приложение Д, Т2), полученных из базы данных GenBank, проводили с помощью веб-интерфейса MUSCLE [205], [206]. Дополнительно использовали последовательность NS1 BKЭ штамма Абсеттаров (GenBank KU885457.1) (Приложение Д, рисунок Д2). Матрица гомологичности последовательностей (Рисунок 74) построена на основании оценки BLOSUM62. При такой визуализации вирусы, переносимые клещами (TBFV) и комарами (MBFV), разделены на две группы с низким перекрёстным подобием. Внутриклассовое подобие белков NS1 TBFV значительно выше, чем внутриклассовое подобие белков NS1 MBFV [158]. Как следствие, доступные в PDB кристаллические структуры, принадлежащие только белкам MBFV, могут недостаточно полно отражать разнообразие структур белка NS1 различных ортофлавивирусов, ограничивая достоверность моделей структуры NS1 TBFV, построенных по доступным шаблонам (Таблица 14). Существование двух различных групп последовательностей белка NS1 соответствует филогенетическому делению рода *Orthoflavivirus* (Рисунок 12). Из-за этого потенциальные лиганды белка NS1 могут оказаться селективны относительно вирусов с разными переносчиками.

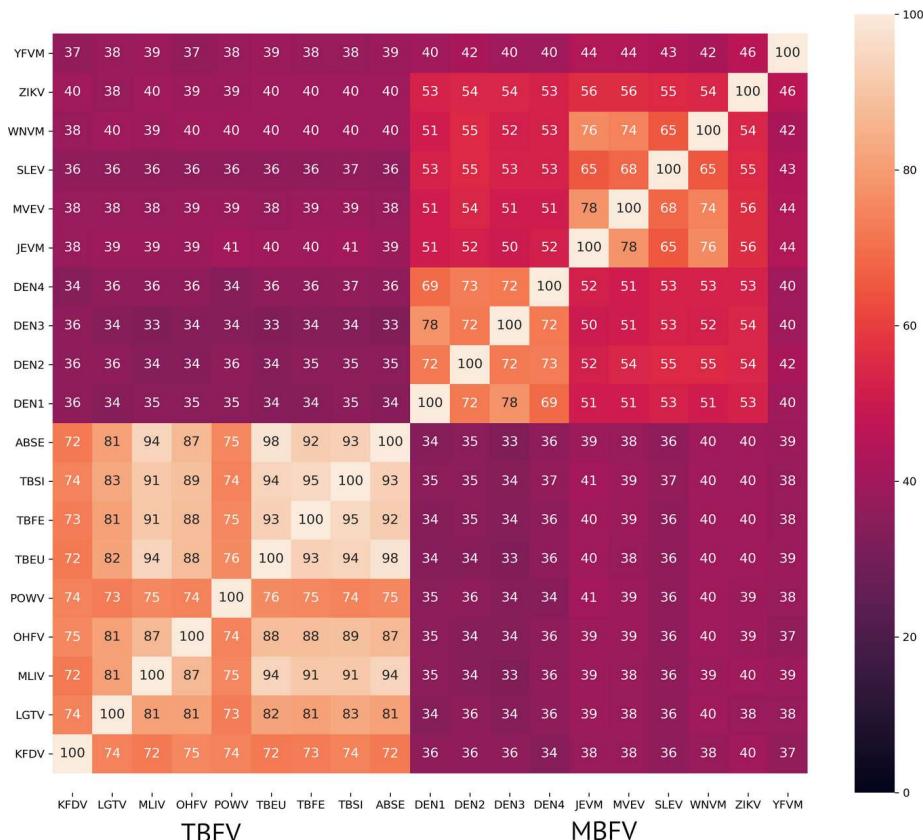


Рисунок 74. Матрица гомологичности аминокислотных последовательностей белка NS1 ортофлавивирусов [158]. Расшифровка сокращений приведена в Приложении Д.

Моделирование по гомологии проводили при помощи программы Modeller 10.5 [270]. На основе каждого индивидуального шаблона генерировали 100 моделей, также генерировали 300 моделей по всем шаблонам одновременно с функциями DOPE и GA341 в качестве оценки [158]. Моделирование проводили с неявно заданными атомами водорода (класс `automodel`). Использовали “медленный” алгоритм оптимизации с максимальным количеством итераций, равным 300. Оптимизацию повторяли 2 раза.

Для дальнейшей работы необходимо было выбрать одну оптимальную модель для каждого шаблона и каждого вируса. Стандартный критерий отбора основан на значении оценочной функции DOPE [271]. Эта функция приоритизирует структуры, наиболее близкие к нативному состоянию белка, на основании статистической информации о плотности вероятности расстояний между атомами. Однако оптимальное значение DOPE не гарантирует отсутствия значительных конформационных отклонений от шаблона, особенно в подвижных областях. Для всех сгенерированных моделей значения DOPE находились в приемлемом диапазоне, поэтому модель для дальнейшего исследования выбирали не по значению DOPE, а по значению СКО С $\alpha$ -атомов моделей, пространственно выровненных относительно структуры 7WUS, которая не содержит неразрешенных участков [158].

### 7.3.2. Выбор моделей по среднеквадратичному отклонению

В рамках описанной в главе 2 процедуры формирования ансамбля [95] для всех моделей были рассчитаны попарные СКО С $\alpha$ -атомов остатков, формирующих карман в области интерфейса димеризации (V1-G15, A186-A193, E202-D207 цепи А и V1, T16-W26, D156-F162, D179-T185, V193-W200 цепи В). При этом в широком диапазоне пороговых значений сходства и разнообразия структур в ансамбль отбирались всего две модели, что недостаточно для презентации конформационного пространства. Причина неуспешного отбора ансамбля по значениям СКО между координатами атомов а/к остатков, формирующих потенциальный сайт связывания может заключаться в слишком высоком разнообразии структур.

Модели, построенные по гомологии со структурой 4TPL, имеют особенно высокие средние значения СКО (Рисунок 75, выделены оранжевым). Значительное отличие от остальных моделей негативно сказывается на отборе ансамбля по конформационному разнообразию.

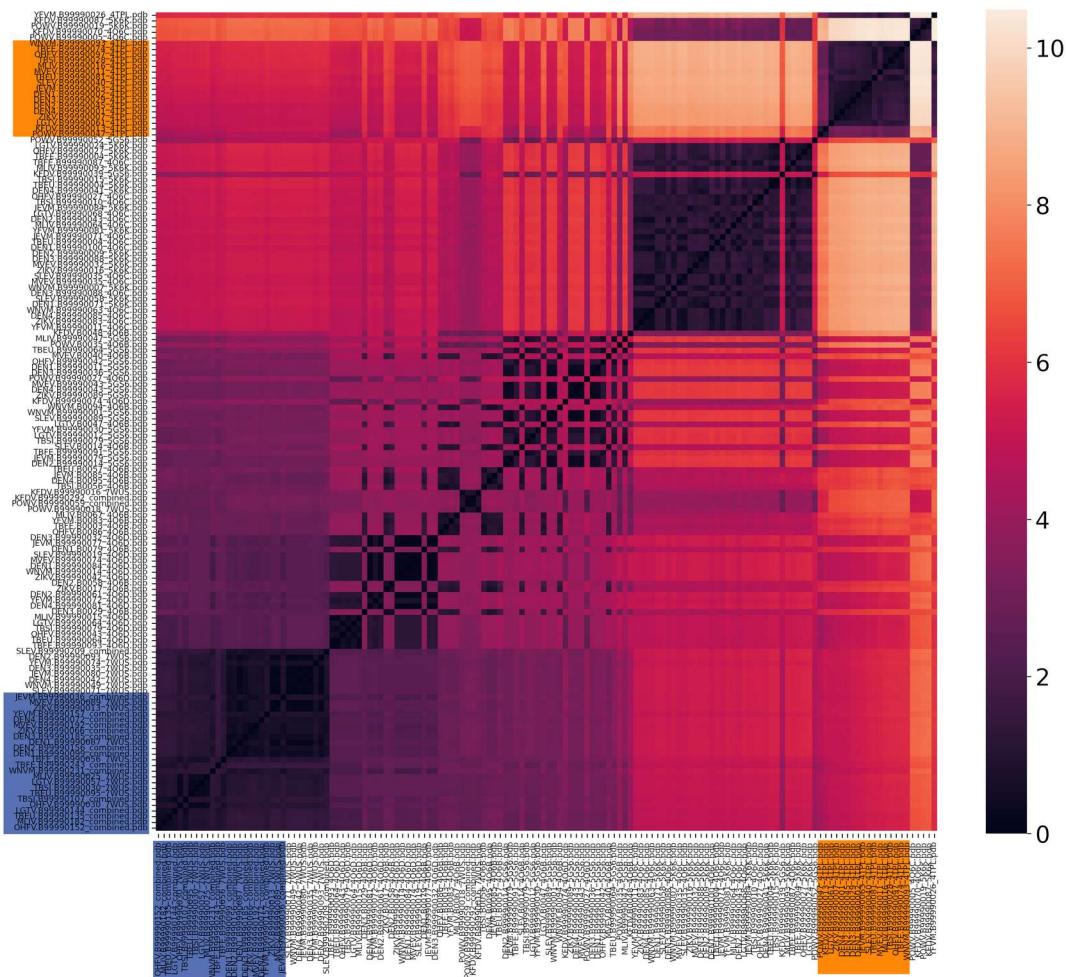


Рисунок 75. Термальная карта попарных СКО координат атомов а/к остатков, формирующих потенциальный сайт связывания в моделях структур белка NS1 эпидемиологически значимых ортофлавивирусов [158]. Оранжевым выделены модели, построенные по шаблону 4TPL, голубым – модели, построенные по “усредненному” шаблону.

При раздельном рассмотрении моделей белка NS1 MBFV (Рисунок 76) и TBFV (Рисунок 77), заметно, что вторые более конформационно-разнообразны. Наибольшие конформационные отличия у моделей, построенных по одинаковым шаблонам, наблюдаются для вирусов Повассан (POVV) и киассанурской лесной болезни (KFDV). Белки NS1 этих вирусов имеют наименьшее сходство аминокислотных последовательностей как с белками NS1 других OFV, переносимых клещами, так и с белками NS1 MBFV. Это, а также наличие множественных неконсервативных замен в области, выбранной в качестве потенциального сайта связывания (Рисунок 78), ставит под сомнение достоверность моделирования этих белков по шаблонным структурам белков NS1 MBFV [158].

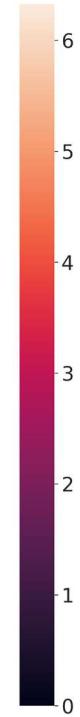
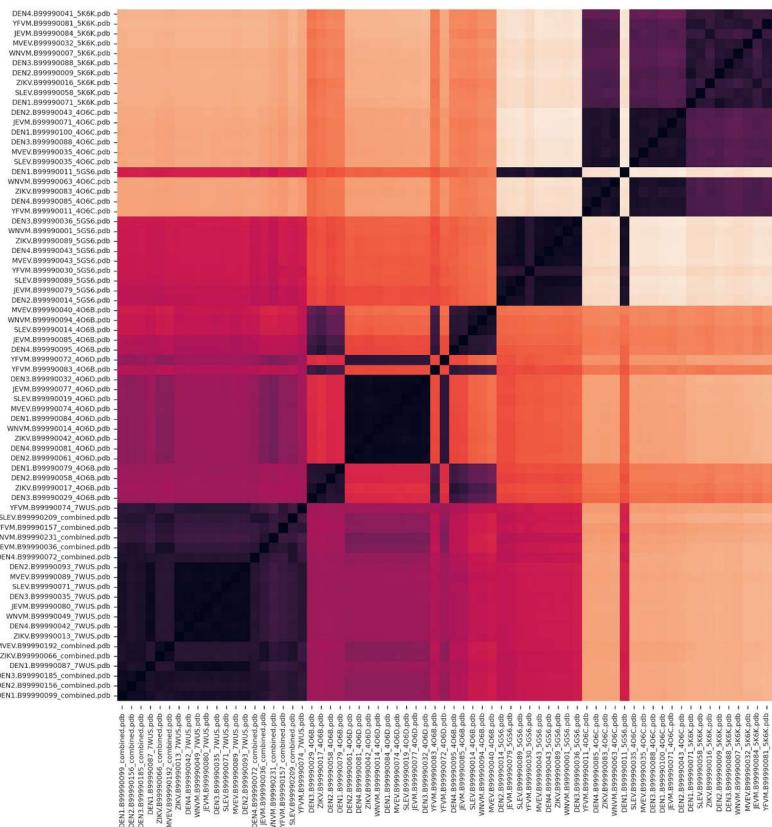


Рисунок 76. Тепловая карта попарных СКО между Са-атомами, формирующими целевой карман, для моделей белков NS1 MBFV [158]

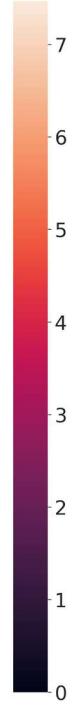
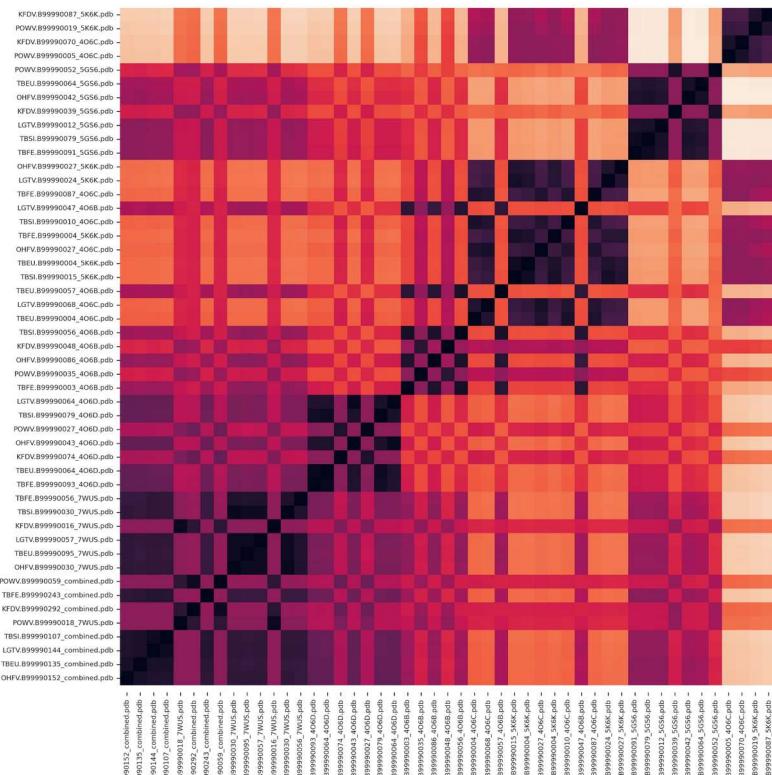


Рисунок 77. Тепловая карта попарных СКО между Са-атомами, формирующими целевой карман, для моделей белков NS1 TBFV [158]

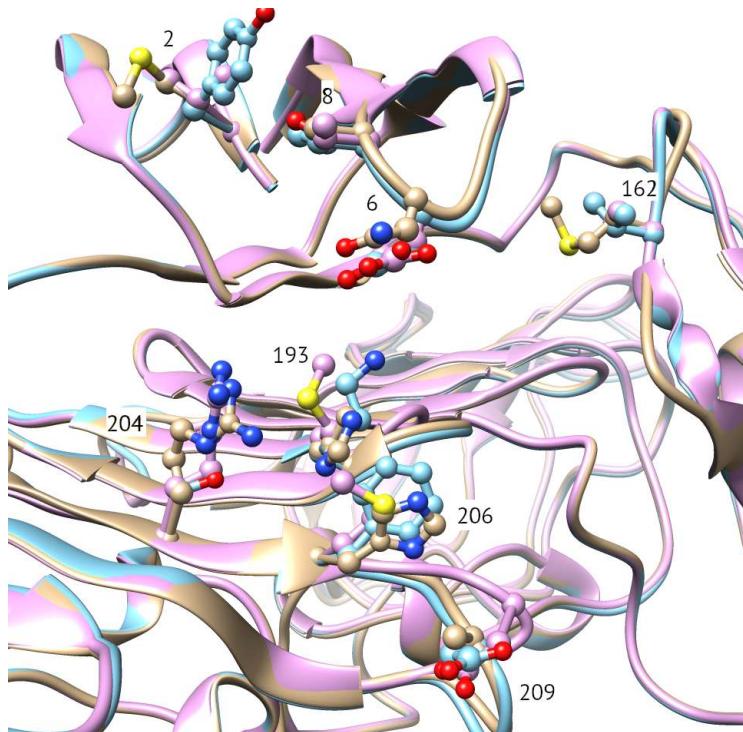


Рисунок 78. Модели белков NS1 POWV (бежевый), KFDV (голубой) и TBEV Eu (розовый): показаны аминокислотные остатки, содержащие неконсервативные замены у NS1 POWV и KFDV, их номера приведены в соответствии с последовательностью белка NS1 POWV [158].

С целью более подробного исследования конформационного пространства моделей белка NS1 был проведён анализ главных компонент (PCA) координат С $\alpha$ -атомов а/к остатков, формирующих целевой карман (Рисунок 79). Модели по шаблонам 4O6C и 5K6K выделяются в отдельно лежащий кластер. Визуальный анализ этих моделей, однако, не позволяет выявить значительных отличий от других структур. Модели, построенные по шаблону 4TPL, выделяются в отдельный кластер, сильнее удалённый от основного множества моделей, чем кластер, содержащий модели по шаблонам 4O6C и 5K6K. При этом сама шаблонная структура 4TPL расположена рядом с остальными кристаллическими структурами (отмечены крестиками, Рисунок 79А). Структура 4TPL содержит наибольший по длине неразрешенный участок (остатки 108-130  $\alpha$ -спирального крыла) – следовательно, соответствующие фрагменты структуры моделируются стохастически, что и вызывает значительные конформационные отличия. Из-за этого модели по шаблону 4TPL не применяли в дальнейшей работе. “Усредненные” модели, построенные по объединенному шаблону из семи структур, лежат в проекции на главные компоненты между моделями, построенными по индивидуальным шаблонам, и обладают наиболее высоким сходством со всеми остальными (ср. Рисунок 75, выделены голубым). При отборе ансамбля по разнообразию эти модели имеют низкий приоритет и не попадают в ансамбли [158].

Наблюдавшиеся ранее по тепловым картам попарных СКО конформационные отличия

моделей белков NS1 POWV и KFDV от моделей белков других OFV подтверждаются при помощи PCA – эти модели лежат в проекции на главные компоненты (Рисунок 79Б) отдельно от основных кластеров моделей, построенных по тем же шаблонам, при этом повторяя их характер расположения [158]. Кроме того, последовательности белка NS1 этих вирусов характеризуются низкой гомологией с последовательностями белков NS1 других TBFV. Для отбора ансамблей белковых структур имеет смысл исключать из рассмотрения такие модели, поскольку они не репрезентативны относительно всего набора структур.

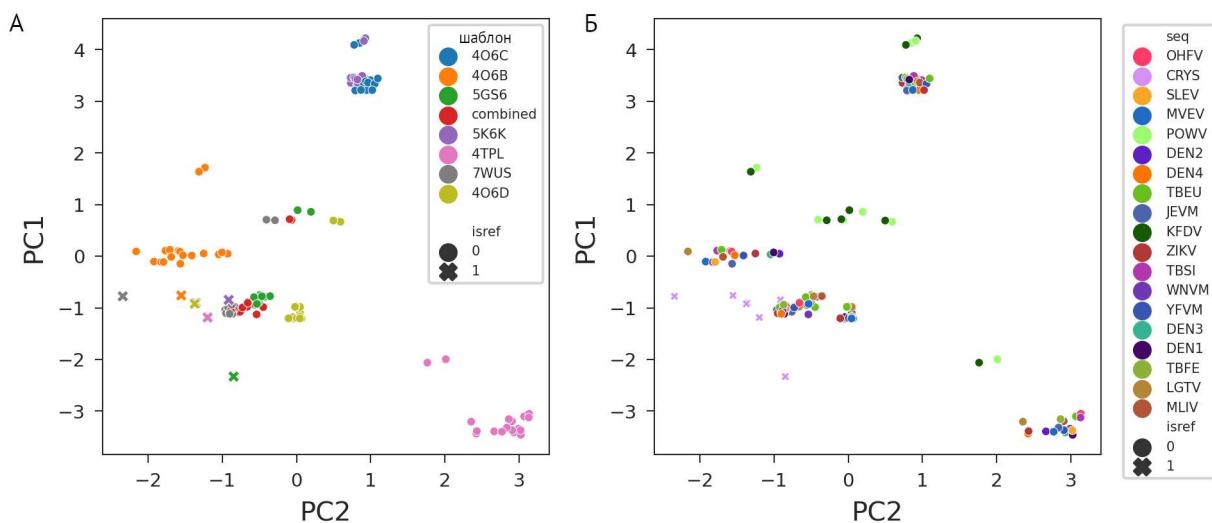


Рисунок 79. PCA координат  $\text{C}\alpha$ -атомов остатков, формирующих целевой карман OFV-NS1 моделей и шаблонных кристаллических структур [158]. (А) – модели окрашены по шаблону, (Б) – модели окрашены по вирусу. Категория “isref” показывает, является ли структура моделью или шаблоном: модели показаны кружками, шаблоны – крестиками.

Удаление моделей, построенных по шаблону 4TPL, и моделей белков NS1 POWV и KFDV из набора данных не способствовало улучшению результатов автоматического составления ансамбля по СКО остатков сайта связывания. Поэтому в дальнейшем для работы с моделями OFV-NS1 был рассмотрен метод составления ансамблей фиксированного размера, максимизирующих корреляцию рангов результатов докинга разнообразных соединений.

### 7.3.3. Составление ансамбля моделей по результатам докинга

В качестве альтернативы подходу к формированию ансамбля на основе белковой структуры был применен подход, основанный на результатах молекулярного докинга. Он подразумевает неявное описание конформационного пространства белков через корреляцию рангов результатов докинга библиотеки разнообразных соединений: модели, демонстрирующие высокую степень корреляции, считаются условно идентичными, тогда как низкая корреляция указывает на то, что модели принадлежат принципиально различным ансамблям. Модели с

промежуточными значениями коэффициента корреляции могут быть использованы для консенсусного ранжирования. Для оценки ранжирующей способности различных моделей был проведен докинг библиотеки 5000 соединений, отобранных по разнообразию из выборки *DD-ZINC*, в карман в области сайта димеризации всех сгенерированных моделей и кристаллических структур. Для выполнения этой задачи использовалась программа Flare.

Подготовку белковых структур и молекулярный докинг проводили при помощи *pyflare* [272] – библиотеки на языке Python 3, позволяющей писать и исполнять сценарии для программы Flare 5.0.0, Cresset [273]. Подготовку белков осуществляли при помощи сценария *proteinprep.py* со стандартными настройками. Докинг проводили для каждой структуры белка индивидуально при помощи сценария *docking.py*. Генерацию сеточного поля для докинга проводили на расстоянии 6 Å от молекулы ZINC000734046780 (**48**), размещенной вручную при помощи UCSF Chimera в кармане между β-бочкой, β-лестницей и α-спиральным крылом (Рисунок 80). Докинг проводили с параметром качества “normal”, – то есть по стандартному алгоритму докинга, реализованному в программе Lead Finder [274]. Программа использует генетический алгоритм генерации результатов докинга. Лиганд представляется в виде “хромосомы”, каждый “ген” в которой – возможные степени свободы лиганда. В ходе оптимизации положения лиганда гены “мутируют” так, чтобы расчетная энергия связывания лиганда в активном центре белка была наименьшей. Генерировали 5 способов связывания для каждого соединения. Обработку результатов ансамблевого докинга проводили при помощи сценария на языке Python 3.9. Оценки считывали из sdf-файлов и для конформеров, наиболее высоко оцененных по Rank score (оценка, оптимизированная для корректного энергетического ранжирования результатов докинга конкретной молекулы), проводили ранжирование на основании уменьшения модуля оценки VS score (оценка, разработанная для надежного ранжирования активных молекул по сравнению с неактивными в ходе виртуального скрининга) [274].

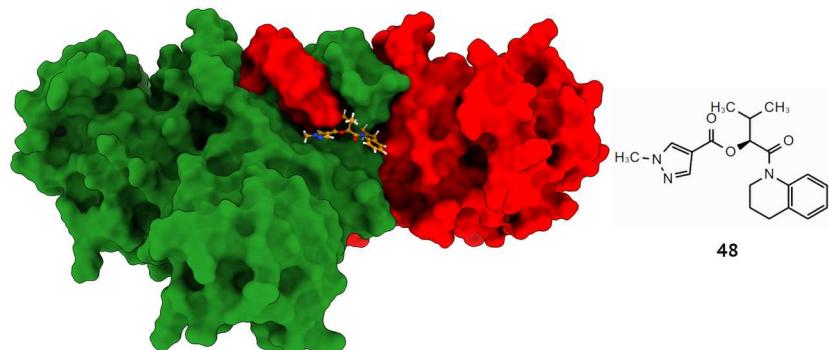


Рисунок 80. Расположение сайта для докинга [158]. Положение референсного лиганда ZINC000734046780 (**48**) в кармане на поверхности белка NS1 (PDB ID: 4O6D).

Была изучена корреляция между попарным СКО либо гомологией аминокислотных последовательностей и коэффициентом корреляции рангов. При рассмотрении всех пар структур не наблюдается четкой зависимости между этими параметрами (Рисунок 81). Однако стоит отметить, что попарные коэффициенты корреляции рангов не превышают значения 0,739. Это свидетельствует о том, что существующие модели имеют достаточно различные конформации и могут быть использованы для консенсусного ранжирования. В качестве примера ансамбля можно выделить три модели с наибольшей суммой коэффициентов корреляции рангов, равной 2,188: модель NS1 ДЕНВ4 (шаблон 4O6C), модель NS1 ВЯЭ (шаблон 5K6K), модель NS1 вируса шотландского энцефаломиелита овец (шаблон 5GS6) (Рисунок 82А). Этот ансамбль включает вирусы, которые переносятся как клещами, так и комарами, и может быть полезен для поиска лигандов с широким спектром действия. Аналогичным образом можно формировать ансамбли из четырех и более структур.

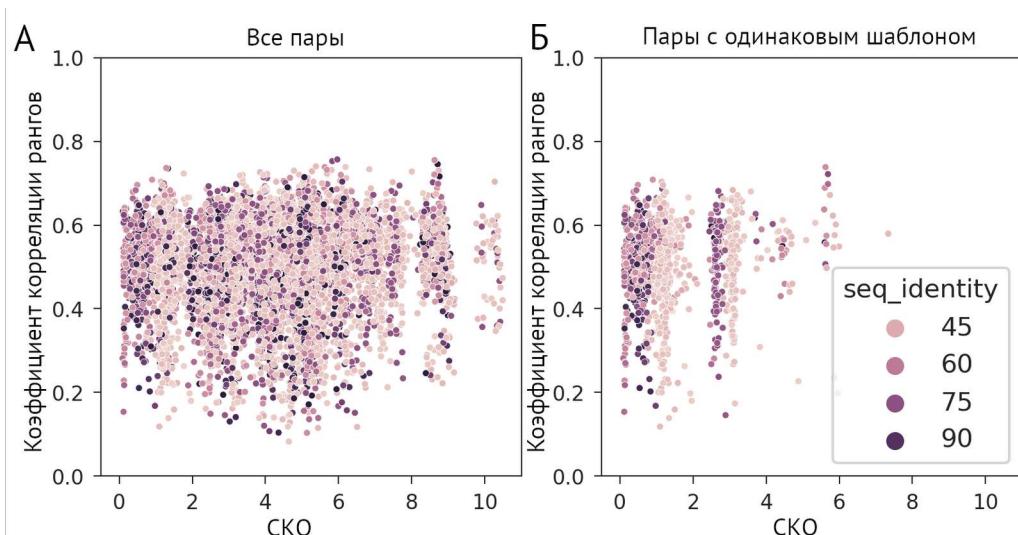


Рисунок 81. Зависимость попарного коэффициента корреляции рангов Спирмана результатов докинга в структуры моделей белков NS1 ортофлавивирусов от попарного СКО между соответствующими структурами, окраска по проценту сходства между соответствующими последовательностями (seq\_identity) [158]. (А) — для всех пар структур, (Б) — для пар моделей разных белков с одинаковым шаблоном.

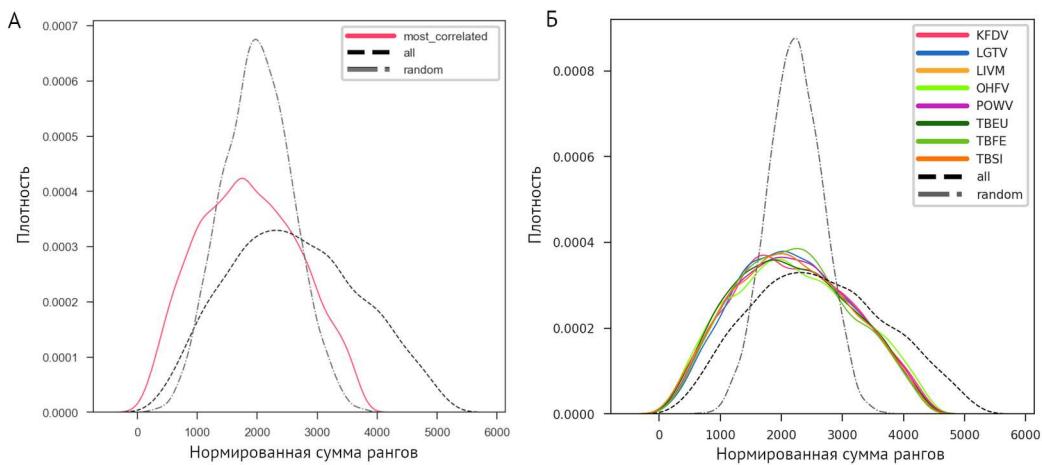


Рисунок 82. Распределение суммы рангов для результатов докинга. (А) — нормированная сумма по ансамблю, составленному из модели NS1 ДЕНВ4 по шаблону 4О6С, модели NS1 ВЯЭ по шаблону 5К6К, модели NS1 вируса шотландского энцефаломиелита овец по шаблону 5GS6 [158]. (Б) — нормированная сумма по моделям, принадлежащим белку NS1 одного вируса (для вирусов, распространяемых клещами).

При анализе пар моделей, построенных на основе одинаковых шаблонов, можно выделить три основные группы, которые различаются по СКО. Эти группы располагаются в диапазонах: от 0 до 2, от 2 до 4 и от 4 до 6. В каждой из этих групп структуры, которые демонстрируют более высокое сходство в аминокислотных последовательностях, как и ожидалось, характеризуются более низкими значениями СКО. Тем не менее, стоит подчеркнуть, что структурное сходство оказывает минимальное влияние на характер ранжирования соединений при докинге в эти модели (Рисунок 81Б).

При рассмотрении распределений оценок докинга (Рисунок 83) не выявлено корреляции между оценками, типами шаблонных структур и аминокислотными последовательностями. Это согласуется с другими результатами, подтверждающими высокий вклад стохастической генерации координат при моделировании по гомологии. В таких случаях целесообразно использовать сумму рангов по ансамблю в качестве консенсусной оценки. Анализ суммы рангов проводили для ансамблей, основанных на общем шаблоне или аминокислотной последовательности, чтобы выявить закономерности, связанные с различием последовательностей белка NS1. Распределение нормированной суммы рангов по ансамблям моделей, построенных по одному шаблону либо для одного вируса (Рисунок 82Б), имеет одновершинную асимметричную форму. В сравнении с распределениями случайно сгенерированных рангов, распределения для всех исследованных ансамблей имели плечи в области низких сумм рангов, что свидетельствует о их способности приоритизировать предположительно активные молекулы. У ансамблей из 7 моделей NS1 одного вируса по

различным шаблонам плечо в области малых сумм рангов смещено влево по сравнению с распределением суммы рангов по всем моделям, что указывает на консенсусное приоритизирование одних и тех же молекул в результатах докинга.

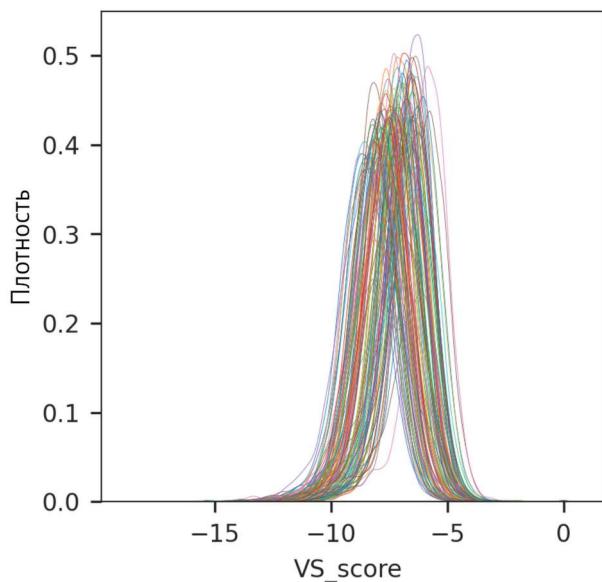


Рисунок 83. Распределения VS score для результатов докинга в индивидуальные модели белка NS1 эпидемиологически значимых ортофлавивирусов.

Распределения нормированной суммы рангов по ансамблям моделей NS1 разных вирусов по каждому шаблону в отдельности и ансамблям моделей NS1 индивидуальных вирусов по всем отдельным шаблонам сравнили при помощи критерия Колмогорова-Смирнова [275], в котором в качестве нулевой гипотезы принимается принадлежность двух выборок одному распределению (Таблица 18). Результаты сравнения позволяют отклонить гипотезу об идентичности распределений нормированной суммы рангов для каждого переносчика в отдельности и для всех вирусов в целом. При этом для разделенных по переносчику ансамблей число молекул с нормированной суммой рангов  $\leq 500$  (потенциальные хиты виртуального скрининга) больше, чем для ансамблей, включающих белки NS1 всех исследованных вирусов (Таблица 19). Разделение MBFV и TBFV при построении ансамблей следует считать эффективной стратегией при отборе низкомолекулярных соединений для экспериментального исследования. Таким образом на основании максимизации суммы коэффициентов корреляции рангов в тройках были сформированы два ансамбля (Таблица 20).

Таблица 18. Статистика критерия Колмогорова-Смирнова для нормированной суммы рангов по ансамблям моделей белка NS1, объединенных по шаблону.

Шаблон	MBFV VS OFV		TBFV VS OFV		MBFV VS TBFV	
	Статистика	р-значение	Статистика	р-значение	Статистика	р-значение
4O6B	0,4239	0,0	0,6406	0,0	0,2784	$4,29*10^{-170}$
4O6C	0,4215	0,0	0,6671	0,0	0,3051	$8,35*10^{-205}$
4O6D	0,4207	0,0	0,6418	0,0	0,2766	$7,25*10^{-168}$
5GS6	0,5747	0,0	0,6744	0,0	0,1363	$1,12*10^{-40}$
5K6K	0,4623	0,0	0,5688	0,0	0,1316	$5,54*10^{-38}$
7WUS	0,4587	0,0	0,6113	0,0	0,2036	$1,26*10^{-90}$

Таблица 19. Число молекул, для которых нормированная сумма рангов меньше или равна 500, для ансамблей моделей белка OFV-NS1, построенных по одному шаблону.

Шаблон	OFV	MBFV	TBFV
4O6B	81	95	170
4O6C	67	117	100
4O6D	79	91	129
5K6K	40	54	63
5GS6	121	193	145
7WUS	124	173	167

Таблица 20. Ансамбли моделей белка NS1 TBFV и MBFV.

MB3		TB3	
Вирус	Шаблон	Вирус	Шаблон
Вирус японского энцефалита	4O6C	Вирус клещевого энцефалита дальневосточного подтипа	7WUS
Вирус японского энцефалита	5K6K	Вирус клещевого энцефалита сибирского подтипа	Объединенный шаблон
Вирус энцефалита Сент-Луис	4O6C	Вирус энцефаломиелита овец	5GS6

## 7.4. Виртуальный скрининг лигандов белка NS1 ортофлавивирусов

Виртуальный скрининг проводили по процедуре, описанной в разделе 6.3.3. Хитами виртуального скрининга считали первые 500 молекул в списке, ранжированном по сумме рангов по ансамблю. Поскольку для белка NS1 OFV нет известных лигандов, анализировали химическое пространство разнообразных соединений из базы данных ZINC15.

Для 5000 соединений, выбранных по разнообразию из библиотеки *DD-ZINC*, при помощи библиотеки RDKit 2020\_03\_6 был рассчитан набор молекулярных дескрипторов (Рисунок 84). Ансамбль *TB3* приоритизирует соединения с большей молекулярной массой, чем ансамбль *MB3*, что может быть связано с большим объемом карманов у ансамбля структур белка NS1 *TBFV*. Распределения расчетной липофильности (*clogP*) и растворимости (*clogS*) для хитов виртуального скрининга смещены относительно распределения полной библиотеки в области большей липофильности и меньшей растворимости соответственно. Чтобы оценить форму молекулы использовался индекс молекулярной формы, который определяется как отношение числа атомов в самой длинной цепи к общему числу атомов в молекуле [233]. Для идеальных линейных молекул индекс равен 1; чем больше циклов и ответвлений в молекуле, тем ниже индекс. У хитов виртуального скрининга распределение индекса молекулярной формы смещено в сторону больших значений по сравнению с полной библиотекой, что указывает на приоритизацию более длинных линейных молекул, соответствующих продолговатой форме потенциального сайта связывания.

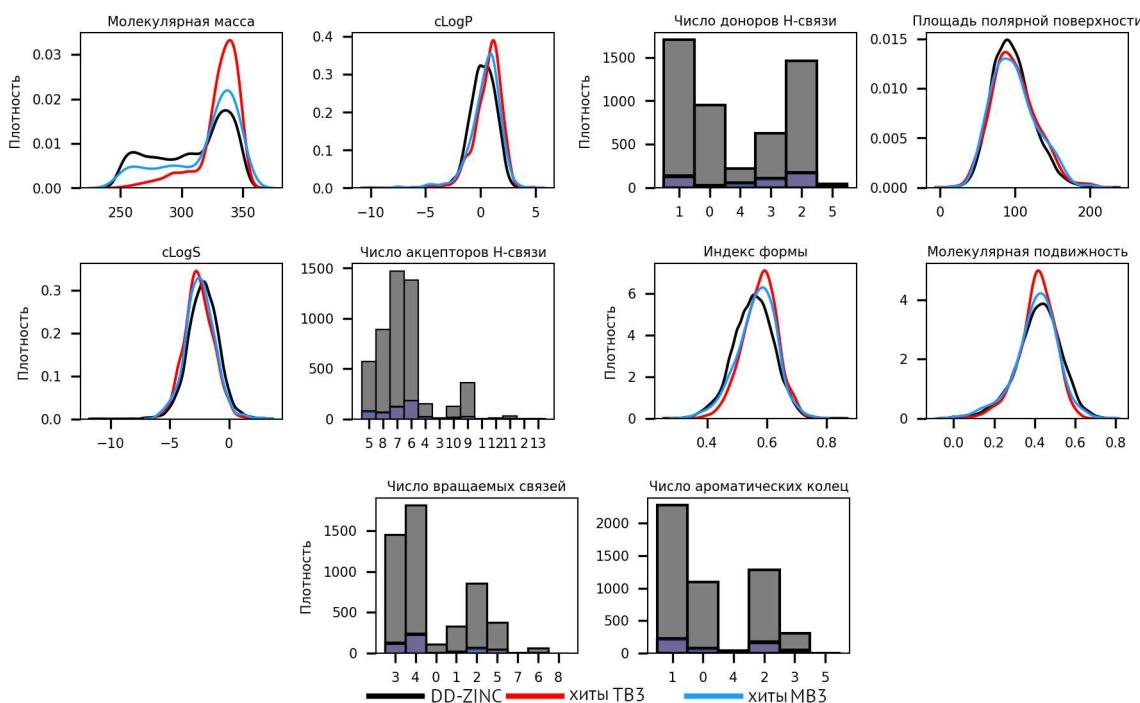


Рисунок 84. Распределение плотности дескрипторов для полной библиотеки *DD-ZINC5000* (черный) и хитов виртуального скрининга этой библиотеки в ансамбле *MB3* (красный) и *TB3* (голубой).

На основании частоты контактов лигандов и остатков, формирующих потенциальный сайт связывания в результатах виртуального скрининга была сформулирована модель фармакофора. Фингерпринты взаимодействия были рассчитаны при помощи библиотеки ODDT (см. раздел 5.1). Далее для каждого остатка и типа взаимодействия рассчитывали частоту контактов с хитами виртуального скрининга во всех структурах ансамбля и отфильтровывали взаимодействия, встречающиеся реже 200 раз. Остальные, наиболее частые взаимодействия составили модель фармакофора (Рисунок 85, 86). Можно заметить, что все остатки, формирующие наибольшее количество контактов в ансамбле *MB3*, находятся на периферии кармана. Это связано с его меньшим объемом и более закрытой конформацией петли  $\beta$ -бочки (a\k 13-15). При этом закрытое положение петли поддерживается водородными связями между остатками  $\beta$ -бочки цепи В (Asp2, Asp8, Phe21) с остатком Lys190A. В моделях ансамбля *TB3* карман пребывает в более открытой конформации, фиксируемой только водородной связью между остатками Lys190A и Asp2B. Этим также можно объяснить смещение распределения молекулярных масс хитов виртуального скрининга в ансамбле *MB3* в область меньших значений и их менее линейную форму.

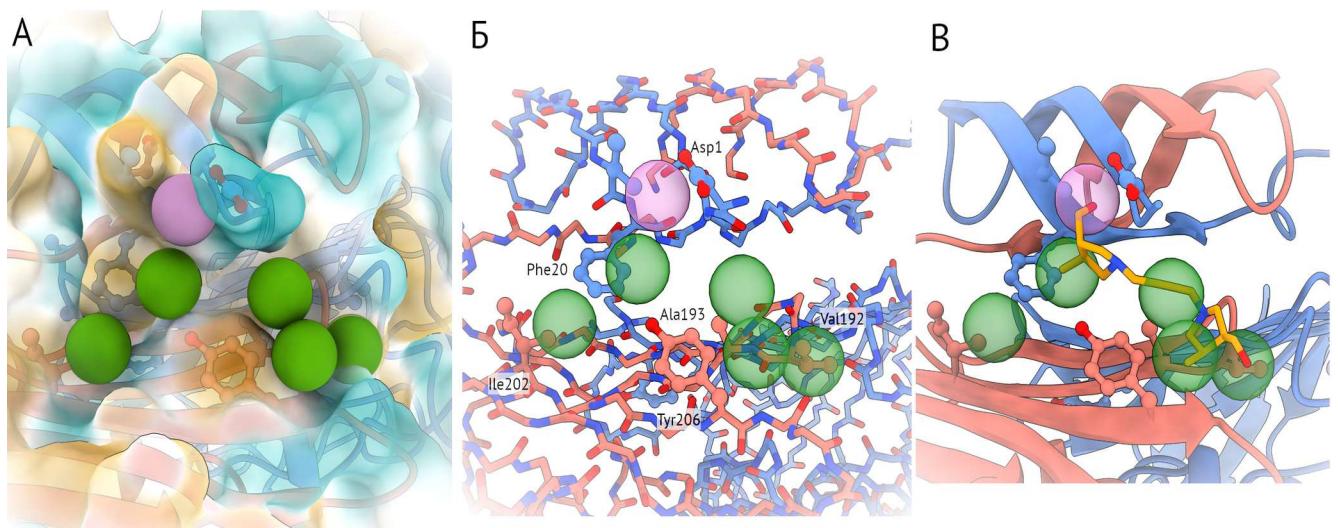


Рисунок 85. Остатки, формирующие наибольшее число контактов с результатами виртуального скрининга методом ансамблевого докинга в ансамбль *MB3* (модель белка NS1 ВЯЭ JEV\_4O6C). Зеленые сферы – гидрофобные взаимодействия, оранжевые – акцепторы водородной связи, фиолетовые – доноры водородной связи. (А) – проекция с отображением поверхности, окрашенной по гидрофобности, (Б) – проекция с указанием ключевых остатков и атомов основной цепи, (В) – проекция с лигандом в активном центре.

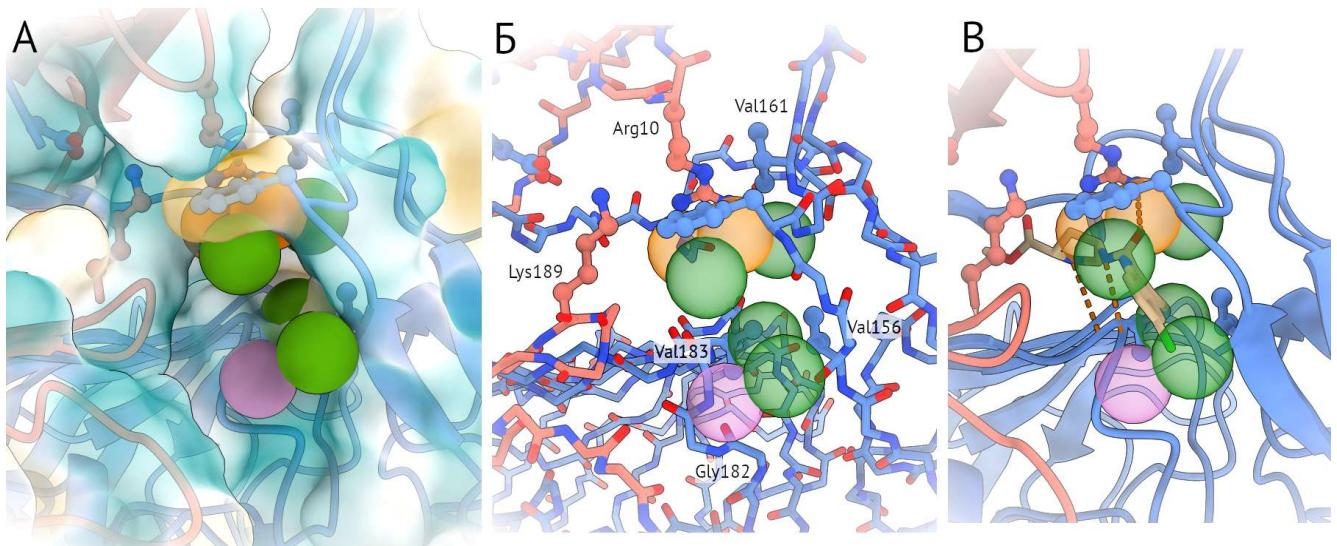


Рисунок 86. Остатки, формирующие наибольшее число контактов с результатами виртуального скрининга методом ансамблевого докинга в ансамбль TB3 (модель белка NS1 ВКЭ TBFE\_7WUS). Окраска сфер и панели аналогичны рисунку 85.

Хиты виртуального скрининга лигандов белка NS1 ортофлавивирусов, переносимых комарами и ортофлавивирусами, переносимых клещами, приведены в Приложении Е. Однако для дальнейшей разработки таких соединений требуется создание методов определения эффективности связывания малых молекул с белком NS1, которые можно будет использовать для экспериментального скрининга. Дальнейшая разработка требует создания систем для анализа связывания низкомолекулярных органических соединений с белком NS1 OFV.

## **Заключение**

Молекулярный докинг играет важную роль в разработке противовирусных лекарственных средств. Доступность все большего количества структурных и биологических данных подталкивает развитие методологии структурно-обоснованного виртуального скрининга и требует разработки и применения стандартизованных и автоматизированных подходов к ансамблевому докингу.

В настоящей работе разработан метод систематического отбора ансамблей для молекулярного докинга, основанный на разнообразии конформаций аминокислотных остатков активного центра белка. Его автоматизированная имплементация, реализованная в программном обеспечении “Программный комплекс для систематического ансамблевого докинга”, универсально применима для любых белков, для которых доступно большое количество однородной структурной информации, полученной из экспериментальных данных либо моделирования. На примере моделей структур эпидемиологически значимых ортофлавивирусов продемонстрированы ограничения метода: на отбор по конформационному разнообразию сильно влияет стохастическая генерация координат. В этом случае закономерно использовать корреляцию результатов докинга разнообразных соединений в индивидуальные структуры для составления ансамбля, поскольку для относительно небольшого количества структур это занимает приемлемое количество вычислительных ресурсов.

На основании результатов виртуального скрининга методом систематического ансамблевого докинга потенциальных ингибиторов Mpro SARS-CoV-2 был обнаружен новый класс соединений, проявляющих активность в отношении этого фермента, не описанный ранее в литературе. Кроме того, предложены фокусированные библиотеки потенциальных лигандов ортофлавивирусов, переносимых комарами и клещами, дальнейшее исследование которых требует разработки экспериментальных методик.

По результатам работы сделаны следующие выводы:

1. Разработана и автоматизирована процедура отбора белковых структур по среднеквадратичному отклонению координат атомов аминокислотных остатков активного центра для систематического ансамблевого докинга.
2. На основании литературной информации об ингибировании активности главной протеазы и репродукции бетакоронавирусов составлена количественно аннотированная база данных для валидации методик виртуального скрининга.
3. Разработана процедура генерации и обработки результатов виртуального скрининга методом систематического ансамблевого докинга. На примере аннотированных библиотек низкомолекулярных органических соединений показана корректность виртуального

скрининга при помощи разработанной методики.

4. Учет результатов систематического ансамблевого докинга в виде фингерпринтов взаимодействия при построении моделей машинного обучения в задаче классификации потенциальных ингибиторов главной протеазы SARS-CoV-2 приводит к повышению прогностической способности моделей случайного леса.
5. На основании виртуального скрининга методом систематического ансамблевого докинга выявлена серия новых ингибиторов главной протеазы SARS-CoV-2 — замещенные тиено[3,2-*d*]пиrimидиноны. Предложены пути расширения серии для увеличения специфичности связывания лигандов в активном центре белка.
6. Методом моделирования по гомологии сгенерирован набор конформаций белка NS1 ортофлавивирусов для составления ансамблей. Отбор структур проводили на основании коэффициентов корреляции рангов результатов докинга. Показана значимость разделения ансамблей структур по переносчику вируса. Сформирована фокусированная библиотека потенциальных лигандов белка NS1 для экспериментального исследования.

## **Список сокращений**

АДФ – аденозиндифосфат

αк – аминокислота

ВЖЛ – вирус желтой лихорадки, *Orthoflavivirus flavi*

ВКЭ – вирус клещевого энцефалита, *Orthoflavivirus encephalitis*

ВЛЗН – вирус лихорадки Западного Нила, *Orthoflavivirus nilense*

ВЯЭ – вирус японского энцефалита, *Orthoflavivirus japonicum*

ДЕНВ – вирус денге, *Orthoflavivirus denguei*

ЗИКВ – вирус Зика, *Orthoflavivirus zikaense*

мРНК – матричная рибонуклеиновая кислота

РНК – рибонуклеиновая кислота

СКО – среднеквадратичное отклонение

AUC – area under curve, площадь под кривой

CDK2 – Циклин-зависимая киназа 2

IC<sub>50</sub> – концентрация полуингибиции

K<sub>i</sub> – константа ингибиции

LR – logistic regression, логистическая регрессия

MBFV – mosquito-borne orthoflaviviruses, ортофлавивирусы, переносимые комарами

MERS-CoV – Middle East respiratory syndrome coronavirus, коронавирус ближневосточного острого респираторного синдрома

Mpro – main protease, главная протеаза

RF – random forest, метод случайного леса

ROC – receiver operating characteristic

SARS-CoV-2 – Severe acute respiratory syndrome-related coronavirus 2, коронавирус тяжёлого острого респираторного синдрома 2

SMILES – Simplified Molecular Input Line Entry System, система упрощенной записи структуры молекул в строке ввода

SVM – support vector machine, метод опорных векторов

TBFV – tick-borne orthoflaviviruses, ортофлавивирусы, переносимые клещами

## Список литературы

1. Homology modeling in drug discovery: Overview, current applications, and future perspectives / M.T. Muhammed, E. Aki-Yalcin // Chemical Biology & Drug Design. – 2019. – Vol. 93. – Homology modeling in drug discovery. – № 1. – P. 12-20.
2. Highly accurate protein structure prediction with AlphaFold / J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S.A.A. Kohl, A.J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A.W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis // Nature. – 2021. – Vol. 596. – № 7873. – P. 583-589.
3. Ensemble docking into multiple crystallographically derived protein structures: an evaluation based on the statistical analysis of enrichments / I.R. Craig, J.W. Essex, K. Spiegel // Journal of Chemical Information and Modeling. – 2010. – Vol. 50. – Ensemble docking into multiple crystallographically derived protein structures. – № 4. – P. 511-524.
4. Enhancement of Zika virus pathogenesis by preexisting antiflavivirus immunity / S.V. Bardina, P. Bunduc, S. Tripathi, J. Duehr, J.J. Frere, J.A. Brown, R. Nachbagauer, G.A. Foster, D. Krysztof, D. Tortorella, S.L. Stramer, A. García-Sastre, F. Krammer, J.K. Lim // Science (New York, N.Y.). – 2017. – Vol. 356. – № 6334. – P. 175-180.
5. Pre-Steady-State Kinetics of the SARS-CoV-2 Main Protease as a Powerful Tool for Antiviral Drug Discovery / M.Yu. Zakharova, A.A. Kuznetsova, V.I. Uvarova, A.D. Fomina, L.I. Kozlovskaya, E.N. Kaliberda, I.N. Kurbatskaia, I.V. Smirnov, A.A. Bulygin, V.D. Knorre, O.S. Fedorova, A. Varnek, D.I. Osolodkin, A.A. Ishmukhametov, A.M. Egorov, A.G. Gabibov, N.A. Kuznetsov // Frontiers in Pharmacology. – 2021. – Vol. 12. – P. 773198.
6. Software for molecular docking: a review / N.S. Pagadala, K. Syed, J. Tuszyński // Biophysical Reviews. – 2017. – Vol. 9. – Software for molecular docking. – № 2. – P. 91-102.
7. Ultralarge Virtual Screening Identifies SARS-CoV-2 Main Protease Inhibitors with Broad-Spectrum Activity against Coronaviruses / A. Luttens, H. Gullberg, E. Abdurakhmanov, D.D. Vo, D. Akaberi, V.O. Talibov, N. Nekhotiaeva, L. Vangeel, S. De Jonghe, D. Jochmans, J. Krambrich, A. Tas, B. Lundgren, Y. Gravenfors, A.J. Craig, Y. Atilaw, A. Sandström, L.W.K. Moodie, Å. Lundkvist, M.J. van Hemert, J. Neyts, J. Lennerstrand, J. Kihlberg, K. Sandberg, U.H. Danielson, J. Carlsson // Journal of the American Chemical Society. – 2022. – Vol. 144. – № 7. – P. 2905-2920.
8. Critical evaluation of search algorithms for automated molecular docking and database screening / T.J.A. Ewing, I.D. Kuntz // Journal of Computational Chemistry. – 1997. – Vol. 18. – № 9. – P. 1175-1189.
9. Improved FlexX docking using FlexS-determined base fragment placement / S.S.J. Cross // Journal of Chemical Information and Modeling. – 2005. – Vol. 45. – № 4. – P. 993-1001.
10. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility / G.M. Morris, R. Huey, W. Lindstrom, M.F. Sanner, R.K. Belew, D.S. Goodsell, A.J. Olson // Journal of Computational Chemistry. – 2009. – Vol. 30. – AutoDock4 and AutoDockTools4. – № 16. – P. 2785-2791.
11. Improved protein-ligand docking using GOLD / M.L. Verdonk, J.C. Cole, M.J. Hartshorn, C.W. Murray, R.D. Taylor // Proteins. – 2003. – Vol. 52. – № 4. – P. 609-623.
12. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of

docking accuracy / R.A. Friesner, J.L. Banks, R.B. Murphy, T.A. Halgren, J.J. Klicic, D.T. Mainz, M.P. Repasky, E.H. Knoll, M. Shelley, J.K. Perry, D.E. Shaw, P. Francis, P.S. Shenkin // Journal of Medicinal Chemistry. – 2004. – Vol. 47. – Glide. – № 7. – P. 1739-1749.

13. Docking and scoring with ICM: the benchmarking results and strategies for improvement / M.A.C. Neves, M. Totrov, R. Abagyan // Journal of Computer-Aided Molecular Design. – 2012. – Vol. 26. – Docking and scoring with ICM. – № 6. – P. 675-686.

14. Молекулярное моделирование / Х.Д.Хёльтье, В. Зиппль, Д. Роньян, Г. Фолькерс. – БИНОМ. Лаборатория знаний, 2010. – 318 р.

15. Docking of Macrocycles: Comparing Rigid and Flexible Docking in Glide / H. Alogheli, G. Olanders, W. Schaal, P. Brandt, A. Karlén // Journal of Chemical Information and Modeling. – 2017. – Vol. 57. – Docking of Macrocycles. – № 2. – P. 190-202.

16. Reliable and Accurate Solution to the Induced Fit Docking Problem for Protein-Ligand Binding / E.B. Miller, R.B. Murphy, D. Sindhikara, K.W. Borrelli, M.J. Grisewood, F. Ranalli, S.L. Dixon, S. Jerome, N.A. Boyles, T. Day, P. Ghanakota, S. Mondal, S.B. Rafi, D.M. Troast, R. Abel, R.A. Friesner // Journal of Chemical Theory and Computation. – 2021. – Vol. 17. – № 4. – P. 2630-2639.

17. 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.

18. FlexE: efficient molecular docking considering protein structure variations / H. Claussen, C. Buning, M. Rarey, T. Lengauer // Journal of Molecular Biology. – 2001. – Vol. 308. – FlexE. – № 2. – P. 377-395.

19. Method for Including the Dynamic Fluctuations of a Protein in Computer-Aided Drug Design / H.A. Carlson, K.M. Masukawa, J.A. McCammon // The Journal of Physical Chemistry A. – 1999. – Vol. 103. – № 49. – P. 10213-10219.

20. Predictive power of molecular dynamics receptor structures in virtual screening / S.E. Nichols, R. Baron, A. Ivetac, J.A. McCammon // Journal of Chemical Information and Modeling. – 2011. – Vol. 51. – № 6. – P. 1439-1446.

21. Large conformational changes in proteins: signaling and other functions / B.J. Grant, A.A. Gorfe, J.A. McCammon // Current Opinion in Structural Biology. – 2010. – Vol. 20. – Large conformational changes in proteins. – № 2. – P. 142-147.

22. Lessons in molecular recognition: the effects of ligand and protein flexibility on molecular docking accuracy / J.A. Erickson, M. Jalaie, D.H. Robertson, R.A. Lewis, M. Vieth // Journal of Medicinal Chemistry. – 2004. – Vol. 47. – Lessons in molecular recognition. – № 1. – P. 45-55.

23. Ligand-induced conformational changes: improved predictions of ligand binding conformations and affinities / T.M. Frimurer, G.H. Peters, L.F. Iversen, H.S. Andersen, N.P.H. Møller, O.H. Olsen // Biophysical Journal. – 2003. – Vol. 84. – Ligand-induced conformational changes. – № 4. – P. 2273-2281.

24. Recipes for the selection of experimental protein conformations for virtual screening / M. Rueda, G. Bottegoni, R. Abagyan // Journal of Chemical Information and Modeling. – 2010. – Vol. 50. – № 1. – P. 186-193.

25. Unveiling the full potential of flexible receptor docking using multiple crystallographic structures / X. Barril, S.D. Morley // Journal of Medicinal Chemistry. – 2005. – Vol. 48. – № 13. – P. 4432-4443.

26. Optimization Rules for SARS-CoV-2 Mpro Antivirals: Ensemble Docking and Exploration of the Coronavirus Protease Active Site / S.V. Stoddard, S.D. Stoddard, B.K. Oelkers, K. Fitts, K. Whalum,

- K. Whalum, A.D. Hemphill, J. Manikonda, L.M. Martinez, E.G. Riley, C.M. Roof, N. Sarwar, D.M. Thomas, E. Ulmer, F.E. Wallace, P. Pandey, S. Roy // *Viruses*. – 2020. – Vol. 12. – № 9. – P. 942.
27. Supercomputer-Based Ensemble Docking Drug Discovery Pipeline with Application to Covid-19 / A. Acharya, R. Agarwal, M.B. Baker, J. Baudry, D. Bhowmik, S. Boehm, K.G. Byler, S.Y. Chen, L. Coates, C.J. Cooper, O. Demerdash, I. Daidone, J.D. Eblen, S. Ellingson, S. Forli, J. Glaser, J.C. Gumbart, J. Gunnels, O. Hernandez, S. Irle, D.W. Kneller, A. Kovalevsky, J. Larkin, T.J. Lawrence, S. LeGrand, S.-H. Liu, J.C. Mitchell, G. Park, J.M. Parks, A. Pavlova, L. Petridis, D. Poole, L. Pouchard, A. Ramanathan, D.M. Rogers, D. Santos-Martins, A. Scheinberg, A. Sedova, Y. Shen, J.C. Smith, M.D. Smith, C. Soto, A. Tsaris, M. Thavappiragasam, A.F. Tillack, J.V. Vermaas, V.Q. Vuong, J. Yin, S. Yoo, M. Zahran, L. Zanetti-Polzi // *Journal of Chemical Information and Modeling*. – 2020. – Vol. 60. – № 12. – P. 5832-5852.
28. Accelerating AutoDock4 with GPUs and Gradient-Based Local Search / D. Santos-Martins, L. Solis-Vasquez, A.F. Tillack, M.F. Sanner, A. Koch, S. Forli // *Journal of Chemical Theory and Computation*. – 2021. – Vol. 17. – № 2. – P. 1060-1073.
29. Improving AutoDock Vina Using Random Forest: The Growing Accuracy of Binding Affinity Prediction by the Effective Exploitation of Larger Data Sets / H. Li, K.-S. Leung, M.-H. Wong, P.J. Ballester // *Molecular Informatics*. – 2015. – Vol. 34. – Improving AutoDock Vina Using Random Forest. – № 2-3. – P. 115-126.
30. Exploring experimental sources of multiple protein conformations in structure-based drug design / K.L. Damm, H.A. Carlson // *Journal of the American Chemical Society*. – 2007. – Vol. 129. – № 26. – P. 8225-8235.
31. Structure-Based Discovery of Novel Nonpeptide Inhibitors Targeting SARS-CoV-2 Mpro / J. Yang, X. Lin, N. Xing, Z. Zhang, H. Zhang, H. Wu, W. Xue // *Journal of Chemical Information and Modeling*. – 2021. – Vol. 61. – № 8. – P. 3917-3926.
32. Computational identification of a transiently open L1/S3 pocket for reactivation of mutant p53 / C.D. Wassman, R. Baronio, Ö. Demir, B.D. Wallentine, C.-K. Chen, L.V. Hall, F. Salehi, D.-W. Lin, B.P. Chung, G.W. Hatfield, A. Richard Chamberlin, H. Luecke, R.H. Lathrop, P. Kaiser, R.E. Amaro // *Nature Communications*. – 2013. – Vol. 4. – P. 1407.
33. Enhancing Virtual Screening Performance of Protein Kinases with Molecular Dynamics Simulations / T.L. Offutt, R.V. Swift, R.E. Amaro // *Journal of Chemical Information and Modeling*. – 2016. – Vol. 56. – № 10. – P. 1923-1935.
34. An insight into the inhibitory mechanism of phytochemicals and FDA-approved drugs on the ACE2-Spike complex of SARS-CoV-2 using computational methods / V. Jani, S. Koulgi, V.N.M. Uppuladinne, U. Sonavane, R. Joshi // *Chemicke Zvesti*. – 2021. – Vol. 75. – № 9. – P. 4625-4648.
35. On Quality Thresholds for the Clustering of Molecular Structures / X. Daura, O. Conchillo-Solé // *Journal of Chemical Information and Modeling*. – 2022. – Vol. 62. – № 22. – P. 5738-5745.
36. Principal component analysis: a method for determining the essential dynamics of proteins / C.C. David, D.J. Jacobs // *Methods in Molecular Biology* (Clifton, N.J.). – 2014. – Vol. 1084. – P. 193-226.
37. Analysis of a complex of statistical variables into principal components. / H. Hotelling // *Journal of Educational Psychology*. – 1933. – Vol. 24. – № 6. – P. 417-441.
38. Initializing K-means Batch Clustering: A Critical Evaluation of Several Techniques / D. Steinley, M.J. Brusco // *Journal of Classification*. – 2007. – Vol. 24. – № 1. – P. 99-121.
39. DINC-COVID: A webserver for ensemble docking with flexible SARS-CoV-2 proteins / S. Hall-

- Swan, D. Devaurs, M.M. Rigo, D.A. Antunes, L.E. Kavraki, G. Zanatta // Computers in Biology and Medicine. – 2021. – Vol. 139. – DINC-COVID. – P. 104943.
40. Ligand-guided receptor optimization / V. Katritch, M. Rueda, R. Abagyan // Methods in Molecular Biology (Clifton, N.J.). – 2012. – Vol. 857. – P. 189-205.
41. Utilizing experimental data for reducing ensemble size in flexible-protein docking / M. Xu, M.A. Lill // Journal of Chemical Information and Modeling. – 2012. – Vol. 52. – № 1. – P. 187-198.
42. Identification of a minimal subset of receptor conformations for improved multiple conformation docking and two-step scoring / S. Yoon, W.J. Welsh // Journal of Chemical Information and Computer Sciences. – 2004. – Vol. 44. – № 1. – P. 88-96.
43. Knowledge-Based Methods To Train and Optimize Virtual Screening Ensembles / R.V. Swift, S.A. Jusoh, T.L. Offutt, E.S. Li, R.E. Amaro // Journal of Chemical Information and Modeling. – 2016. – Vol. 56. – № 5. – P. 830-842.
44. Comparison of Data Fusion Methods as Consensus Scores for Ensemble Docking / D. Bajusz, A. Rácz, K. Héberger // Molecules. – 2019. – Vol. 24. – № 15. – P. 2690.
45. Evaluating Virtual Screening Methods: Good and Bad Metrics for the “Early Recognition” Problem / J.-F. Truchon, C.I. Bayly // Journal of Chemical Information and Modeling. – 2007. – Vol. 47. – № 2. – P. 488-508.
46. Using machine learning to improve ensemble docking for drug discovery / T. Chandak, J.P. Mayginnes, H. Mayes, C.F. Wong // Proteins. – 2020. – Vol. 88. – № 10. – P. 1263-1270.
47. Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking / M.M. Mysinger, M. Carchia, J.J. Irwin, B.K. Shoichet // Journal of Medicinal Chemistry. – 2012. – Vol. 55. – № 14. – P. 6582-6594.
48. Consensus scoring for protein–ligand interactions / M. Feher // Drug Discovery Today. – 2006. – Vol. 11. – № 9. – P. 421-428.
49. Automated discovery of noncovalent inhibitors of SARS-CoV-2 main protease by consensus Deep Docking of 40 billion small molecules / F. Gentile, M. Fernandez, F. Ban, A.-T. Ton, H. Mslati, C.F. Perez, E. Leblanc, J.C. Yaacoub, J. Gleave, A. Stern, B. Wong, F. Jean, N. Strynadka, A. Cherkasov // Chemical Science. – 2021. – Vol. 12. – № 48. – P. 15960-15974.
50. Molecular docking to ensembles of protein structures / R.M. Knegtel, I.D. Kuntz, C.M. Oshiro // Journal of Molecular Biology. – 1997. – Vol. 266. – № 2. – P. 424-440.
51. Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock / F. Osterberg, G.M. Morris, M.F. Sanner, A.J. Olson, D.S. Goodsell // Proteins. – 2002. – Vol. 46. – Automated docking to multiple target structures. – № 1. – P. 34-40.
52. Virology: Coronaviruses | Nature. – 1968. – Vol. 220. – P. 650.
53. Origin and evolution of pathogenic coronaviruses | Nature Reviews Microbiology / Cui, J., Li, F. , Shi, ZL. – 2019. – Vol. 17. – P. 181-192.
54. Discovery of seven novel Mammalian and avian coronaviruses in the genus deltacoronavirus supports bat coronaviruses as the gene source of alphacoronavirus and betacoronavirus and avian coronaviruses as the gene source of gammacoronavirus and deltacoronavirus / P.C.Y. Woo, S.K.P. Lau, C.S.F. Lam, C.C.Y. Lau, A.K.L. Tsang, J.H.N. Lau, R. Bai, J.L.L. Teng, C.C.C. Tsang, M. Wang, B.-J. Zheng, K.-H. Chan, K.-Y. Yuen // Journal of Virology. – 2012. – Vol. 86. – № 7. – P. 3995-4008.

55. Visualizing in deceased COVID-19 patients how SARS-CoV-2 attacks the respiratory and olfactory mucosae but spares the olfactory bulb / M. Khan, S.-J. Yoo, M. Clijsters, W. Backaert, A. Vanstapel, K. Speleman, C. Lietaer, S. Choi, T.D. Hether, L. Marcelis, A. Nam, L. Pan, J.W. Reeves, P. Van Bulck, H. Zhou, M. Bourgeois, Y. Debaveye, P. De Munter, J. Gunst, M. Jorissen, K. Lagrou, N. Lorent, A. Neyrinck, M. Peetermans, D.R. Thal, C. Vandebrielle, J. Wauters, P. Mombaerts, L. Van Gerven // *Cell.* – 2021. – Vol. 184. – № 24. – P. 5932-5949.e15.
56. Why Does the Omicron Variant Largely Spare Olfactory Function? Implications for the Pathogenesis of Anosmia in Coronavirus Disease 2019 / R. Butowt, K. Bilińska, C. von Bartheld // *The Journal of Infectious Diseases.* – 2022. – Vol. 226. – № 8. – P. 1304-1308.
57. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China / C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu, Z. Cheng, T. Yu, J. Xia, Y. Wei, W. Wu, X. Xie, W. Yin, H. Li, M. Liu, Y. Xiao, H. Gao, L. Guo, J. Xie, G. Wang, R. Jiang, Z. Gao, Q. Jin, J. Wang, B. Cao // *Lancet.* – 2020. – Vol. 395. – № 10223. – P. 497-506.
58. SARS-CoV-2 pathogenesis / M.M. Lamers, B.L. Haagmans // *Nature Reviews. Microbiology.* – 2022. – Vol. 20. – № 5. – P. 270-284.
59. Severity of COVID-19 cases in the months of predominance of the Alpha and Delta variants / D. Florensa, J. Mateo, R. Spaimoc, C. Miret, S. Godoy, F. Solsona, P. Godoy // *Scientific Reports.* – 2022. – Vol. 12. – № 1. – P. 15456.
60. The isolated effect of age on the risk of COVID-19 severe outcomes: a systematic review with meta-analysis / K. Romero Starke, D. Reissig, G. Petereit-Haack, S. Schmauder, A. Nienhaus, A. Seidler // *BMJ global health.* – 2021. – Vol. 6. – № 12. – P. e006434.
61. Coronavirus biology and replication: implications for SARS-CoV-2 / P. V'kovski, A. Kratzel, S. Steiner, H. Stalder, V. Thiel // *Nature Reviews. Microbiology.* – 2021. – Vol. 19. – № 3. – P. 155-170.
62. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor / M. Hoffmann, H. Kleine-Weber, S. Schroeder, N. Krüger, T. Herrler, S. Erichsen, T.S. Schiergens, G. Herrler, N.-H. Wu, A. Nitsche, M.A. Müller, C. Drosten, S. Pöhlmann // *Cell.* – 2020. – Vol. 181. – № 2. – P. 271-280.e8.
63. Structural biology of SARS-CoV-2 and implications for therapeutic development / H. Yang, Z. Rao // *Nature Reviews. Microbiology.* – 2021. – Vol. 19. – № 11. – P. 685-700.
64. Coronavirus genome structure and replication / D.A. Brian, R.S. Baric // *Current Topics in Microbiology and Immunology.* – 2005. – Vol. 287. – P. 1-30.
65. The molecular biology of coronaviruses / P.S. Masters // *Advances in Virus Research.* – 2006. – Vol. 66. – P. 193-292.
66. The mechanism of RNA capping by SARS-CoV-2 / G.J. Park, A. Osinski, G. Hernandez, J.L. Eitson, A. Majumdar, M. Tonelli, K. Henzler-Wildman, K. Pawłowski, Z. Chen, Y. Li, J.W. Schoggins, V.S. Tagliabuaci // *Nature.* – 2022. – Vol. 609. – № 7928. – P. 793-800.
67. The coding capacity of SARS-CoV-2 / Y. Finkel, O. Mizrahi, A. Nachshon, S. Weingarten-Gabbay, D. Morgenstern, Y. Yahalom-Ronen, H. Tamir, H. Achdout, D. Stein, O. Israeli, A. Beth-Din, S. Melamed, S. Weiss, T. Israely, N. Paran, M. Schwartz, N. Stern-Ginossar // *Nature.* – 2021. – Vol. 589. – № 7840. – P. 125-130.
68. Nsp3 of coronaviruses: Structures and functions of a large multi-domain protein / J. Lei, Y. Kusov, R. Hilgenfeld // *Antiviral Research.* – 2018. – Vol. 149. – P. 58-74.
69. SARS-CoV-2 Nsp1 binds the ribosomal mRNA channel to inhibit translation / K. Schubert, E.D.

Karousis, A. Jomaa, A. Scaiola, B. Echeverria, L.-A. Gurzeler, M. Leibundgut, V. Thiel, O. Mühlemann, N. Ban // *Nature Structural & Molecular Biology*. – 2020. – Vol. 27. – № 10. – P. 959-966.

70. Severe acute respiratory syndrome coronavirus nonstructural proteins 3, 4, and 6 induce double-membrane vesicles / M.M. Angelini, M. Akhlaghpour, B.W. Neuman, M.J. Buchmeier // *mBio*. – 2013. – Vol. 4. – № 4. – P. e00524-13.
71. Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors / R.N. Kirchdoerfer, A.B. Ward // *Nature Communications*. – 2019. – Vol. 10. – № 1. – P. 2342.
72. Coronavirus replication-transcription complex: Vital and selective NMPylation of a conserved site in nsp9 by the NiRAN-RdRp subunit / H. Slanina, R. Madhugiri, G. Bylapudi, K. Schultheiß, N. Karl, A. Gulyaeva, A.E. Gorbalyena, U. Linne, J. Ziebuhr // *Proceedings of the National Academy of Sciences of the United States of America*. – 2021. – Vol. 118. – № 6. – P. e2022310118.
73. Structural basis of ribosomal frameshifting during translation of the SARS-CoV-2 RNA genome / P.R. Bhatt, A. Scaiola, G. Loughran, M. Leibundgut, A. Kratzel, R. Meurs, R. Dreos, K.M. O'Connor, A. McMillan, J.W. Bode, V. Thiel, D. Gatfield, J.F. Atkins, N. Ban // *Science*. – 2021. – Vol. 372. – № 6548. – P. 1306-1313.
74. Ma, Y. Structural basis and functional analysis of the SARS coronavirus nsp14-nsp10 complex / Y. Ma, L. Wu, N. Shaw, Y. Gao, J. Wang, Y. Sun, Z. Lou, L. Yan, R. Zhang, Z. Rao // *Proceedings of the National Academy of Sciences of the United States of America*. – 2015. – Vol. 112. – № 30. – P. 9436-9441.
75. Crystal structure and functional analysis of the SARS-coronavirus RNA cap 2'-O-methyltransferase nsp10/nsp16 complex / E. Decroly, C. Debarnot, F. Ferron, M. Bouvet, B. Coutard, I. Imbert, L. Gluais, N. Papageorgiou, A. Sharff, G. Bricogne, M. Ortiz-Lombardia, J. Lescar, B. Canard // *PLoS pathogens*. – 2011. – Vol. 7. – № 5. – P. e1002059.
76. Multiple enzymatic activities associated with severe acute respiratory syndrome coronavirus helicase / K.A. Ivanov, V. Thiel, J.C. Dobbe, Y. van der Meer, E.J. Snijder, J. Ziebuhr // *Journal of Virology*. – 2004. – Vol. 78. – № 11. – P. 5619-5632.
77. Structural basis for backtracking by the SARS-CoV-2 replication-transcription complex / B. Malone, J. Chen, Q. Wang, E. Llewellyn, Y.J. Choi, P.D.B. Olinares, X. Cao, C. Hernandez, E.T. Eng, B.T. Chait, D.E. Shaw, R. Landick, S.A. Darst, E.A. Campbell // *Proceedings of the National Academy of Sciences of the United States of America*. – 2021. – Vol. 118. – № 19. – P. e2102516118.
78. Discovery of an RNA virus 3'->5' exoribonuclease that is critically involved in coronavirus RNA synthesis / E. Minskaia, T. Hertzig, A.E. Gorbalyena, V. Campanacci, C. Cambillau, B. Canard, J. Ziebuhr // *Proceedings of the National Academy of Sciences of the United States of America*. – 2006. – Vol. 103. – № 13. – P. 5108-5113.
79. Functional screen reveals SARS coronavirus nonstructural protein nsp14 as a novel cap N7 methyltransferase / Y. Chen, H. Cai, J. Pan, N. Xiang, P. Tien, T. Ahola, D. Guo // *Proceedings of the National Academy of Sciences of the United States of America*. – 2009. – Vol. 106. – № 9. – P. 3484-3489.
80. Major genetic marker of nidoviruses encodes a replicative endoribonuclease / K.A. Ivanov, T. Hertzig, M. Rozanov, S. Bayer, V. Thiel, A.E. Gorbalyena, J. Ziebuhr // *Proceedings of the National Academy of Sciences of the United States of America*. – 2004. – Vol. 101. – № 34. – P. 12694-12699.
81. Coronavirus endoribonuclease targets viral polyuridine sequences to evade activating host sensors /

- M. Hackbart, X. Deng, S.C. Baker // *Proceedings of the National Academy of Sciences of the United States of America*. – 2020. – Vol. 117. – № 14. – P. 8094-8103.
82. Coronavirus nonstructural protein 16 is a cap-0 binding enzyme possessing (nucleoside-2’O)-methyltransferase activity / E. Decroly, I. Imbert, B. Coutard, M. Bouvet, B. Selisko, K. Alvarez, A.E. Gorbatenya, E.J. Snijder, B. Canard // *Journal of Virology*. – 2008. – Vol. 82. – № 16. – P. 8071-8084.
83. Coronaviruses use discontinuous extension for synthesis of subgenome-length negative strands / S.G. Sawicki, D.L. Sawicki // *Advances in Experimental Medicine and Biology*. – 1995. – Vol. 380. – P. 499-506.
84. Cryo-EM Structure of an Extended SARS-CoV-2 Replication and Transcription Complex Reveals an Intermediate State in Cap Synthesis / L. Yan, J. Ge, L. Zheng, Y. Zhang, Y. Gao, T. Wang, Y. Huang, Y. Yang, S. Gao, M. Li, Z. Liu, H. Wang, Y. Li, Y. Chen, L.W. Guddat, Q. Wang, Z. Rao, Z. Lou // *Cell*. – 2021. – Vol. 184. – № 1. – P. 184-193.e10.
85. In Vitro Reconstitution of SARS-Coronavirus mRNA Cap Methylation / M. Bouvet, C. Debarnot, I. Imbert, B. Selisko, E.J. Snijder, B. Canard, E. Decroly // *PLoS Pathogens*. – 2010. – Vol. 6. – № 4. – P. e1000863.
86. SARS-CoV-2 structure and replication characterized by in situ cryo-electron tomography / S. Klein, M. Cortese, S.L. Winter, M. Wachsmuth-Melm, C.J. Neufeldt, B. Cerikan, M.L. Stanifer, S. Boulant, R. Bartenschlager, P. Chlanda // *Nature Communications*. – 2020. – Vol. 11. – № 1. – P. 5885.
87. The intracellular sites of early replication and budding of SARS-coronavirus / S. Stertz, M. Reichelt, M. Spiegel, T. Kuri, L. Martínez-Sobrido, A. García-Sastre, F. Weber, G. Kochs // *Virology*. – 2007. – Vol. 361. – № 2. – P. 304-315.
88.  $\beta$ -Coronaviruses Use Lysosomes for Egress Instead of the Biosynthetic Secretory Pathway / S. Ghosh, T.A. Dellibovi-Ragheb, A. Kerviel, E. Pak, Q. Qiu, M. Fisher, P.M. Takvorian, C. Bleck, V.W. Hsu, A.R. Fehr, S. Perlman, S.R. Achar, M.R. Straus, G.R. Whittaker, C.A.M. de Haan, J. Kehrl, G. Altan-Bonnet, N. Altan-Bonnet // *Cell*. – 2020. – Vol. 183. – № 6. – P. 1520-1535.e14.
89. Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra alpha-helical domain / K. Anand, G.J. Palm, J.R. Mesters, S.G. Siddell, J. Ziebuhr, R. Hilgenfeld // *The EMBO journal*. – 2002. – Vol. 21. – № 13. – P. 3213-3224.
90. The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor / H. Yang, M. Yang, Y. Ding, Y. Liu, Z. Lou, Z. Zhou, L. Sun, L. Mo, S. Ye, H. Pang, G.F. Gao, K. Anand, M. Bartlam, R. Hilgenfeld, Z. Rao // *Proceedings of the National Academy of Sciences of the United States of America*. – 2003. – Vol. 100. – № 23. – P. 13190-13195.
91. Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs / K. Anand, J. Ziebuhr, P. Wadhwani, J.R. Mesters, R. Hilgenfeld // *Science (New York, N.Y.)*. – 2003. – Vol. 300. – № 5626. – P. 1763-1767.
92. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved  $\alpha$ -ketoamide inhibitors / L. Zhang, D. Lin, X. Sun, U. Curth, C. Drosten, L. Sauerhering, S. Becker, K. Rox, R. Hilgenfeld // *Science (New York, N.Y.)*. – 2020. – Vol. 368. – № 6489. – P. 409-412.
93. The catalysis of the SARS 3C-like protease is under extensive regulation by its extra domain / J. Shi, J. Song // *The FEBS journal*. – 2006. – Vol. 273. – № 5. – P. 1035-1045.
94. Structural basis for the in vitro efficacy of nirmatrelvir against SARS-CoV-2 variants / S.E. Greasley, S. Noell, O. Plotnikova, R. Ferre, W. Liu, B. Bolanos, K. Fennell, J. Nicki, T. Craig, Y. Zhu,

A.E. Stewart, C.M. Steppan // The Journal of Biological Chemistry. – 2022. – Vol. 298. – № 6. – P. 101972.

95. Ensemble docking based virtual screening of SARS-CoV-2 main protease inhibitors / A.D. Fomina, V.I. Uvarova, L.I. Kozlovskaya, V.A. Palyulin, D.I. Osolodkin, A.A. Ishmukhametov // Molecular Informatics. – 2024. – Vol. 43. – № 8. – P. e202300279.

96. Conserved coronavirus proteins as targets of broad-spectrum antivirals / C.C. Melo-Filho, T. Bobrowski, H.-J. Martin, Z. Sessions, K.I. Popov, N.J. Moorman, R.S. Baric, E.N. Muratov, A. Tropsha // Antiviral Research. – 2022. – Vol. 204. – P. 105360.

97. Potential Inhibitors of SARS-CoV-2 Main Protease (Mpro) Identified from the Library of FDA-Approved Drugs Using Molecular Docking Studies / D.K. Verma, S. Kapoor, S. Das, K.G. Thakur // Biomedicines. – 2022. – Vol. 11. – № 1. – P. 85.

98. The P132H mutation in the main protease of Omicron SARS-CoV-2 decreases thermal stability without compromising catalysis or small-molecule drug inhibition / M.D. Sacco, Y. Hu, M.V. Gongora, F. Meilleur, M.T. Kemp, X. Zhang, J. Wang, Y. Chen // Cell Research. – 2022. – Vol. 32. – № 5. – P. 498-500.

99. Structural basis for replicase polyprotein cleavage and substrate specificity of main protease from SARS-CoV-2 / Y. Zhao, Y. Zhu, X. Liu, Z. Jin, Y. Duan, Q. Zhang, C. Wu, L. Feng, X. Du, J. Zhao, M. Shao, B. Zhang, X. Yang, L. Wu, X. Ji, L.W. Guddat, K. Yang, Z. Rao, H. Yang // Proceedings of the National Academy of Sciences of the United States of America. – 2022. – Vol. 119. – № 16. – P. e2117142119.

100. Comprehensive Insights into the Catalytic Mechanism of Middle East Respiratory Syndrome 3C-Like Protease and Severe Acute Respiratory Syndrome 3C-Like Protease / H. Wang, S. He, W. Deng, Y. Zhang, G. Li, J. Sun, W. Zhao, Y. Guo, Z. Yin, D. Li, L. Shang // ACS catalysis. – 2020. – Vol. 10. – P. 5871-5890.

101. Revealing the molecular mechanisms of proteolysis of SARS-CoV-2 Mpro by QM/MM computational methods / K. Świderek, V. Moliner // Chemical Science. – 2020. – Vol. 11. – № 39. – P. 10626-10630.

102. A Review of the Current Landscape of SARS-CoV-2 Main Protease Inhibitors: Have We Hit the Bullseye Yet? / G. Macip, P. Garcia-Segura, J.Mestres-Truyol, B.Saldivar-Espinoza, S.Garcia-Vallvé // International Journal of Molecular Sciences. – 2021. – Vol. 23. – № 1. – P. 259.

103. Covalent inhibitors: a rational approach to drug discovery / F. Sutanto, M. Konstantinidou, A. Dömling // RSC medicinal chemistry. – 2020. – Vol. 11. – № 8. – P. 876-884.

104. Targeting SARS-CoV-2 Main Protease for Treatment of COVID-19: Covalent Inhibitors Structure-Activity Relationship Insights and Evolution Perspectives / G. La Monica, A. Bono, A. Lauria, A. Martorana // Journal of Medicinal Chemistry. – 2022. – Vol. 65. – № 19. – P. 12500-12534.

105. Design and synthesis of dipeptidyl glutaminyl fluoromethyl ketones as potent severe acute respiratory syndrome coronavirus (SARS-CoV) inhibitors / H.-Z. Zhang, H. Zhang, W. Kemnitzer, B. Tseng, J. Cinatl, M. Michaelis, H.W. Doerr, S.X. Cai // Journal of Medicinal Chemistry. – 2006. – Vol. 49. – № 3. – P. 1198-1201.

106. Conservation of substrate specificities among coronavirus main proteases / A. Hegyi, J. Ziebuhr // The Journal of General Virology. – 2002. – Vol. 83. – № Pt 3. – P. 595-599.

107. Feline coronavirus drug inhibits the main protease of SARS-CoV-2 and blocks virus replication / W. Vuong, M.B. Khan, C. Fischer, E. Arutyunova, T. Lamer, J. Shields, H.A. Saffran, R.T. McKay,

- M.J. van Belkum, M.A. Joyce, H.S. Young, D.L. Tyrrell, J.C. Vederas, M.J. Lemieux // *Nature Communications*. – 2020. – Vol. 11. – № 1. – P. 4282.
108. Boceprevir, GC-376, and calpain inhibitors II, XII inhibit SARS-CoV-2 viral replication by targeting the viral main protease / Chunlong Ma.
109. Preclinical characterization of an intravenous coronavirus 3CL protease inhibitor for the potential treatment of COVID19 / B. Boras, R.M. Jones, B.J. Anson, D. Arenson, L. Aschenbrenner, M.A. Bakowski, N. Beutler, J. Binder, E. Chen, H. Eng, H. Hammond, J. Hammond, R.E. Haupt, R. Hoffman, E.P. Kadar, R. Kania, E. Kimoto, M.G. Kirkpatrick, L. Lanyon, E.K. Lendy, J.R. Lillis, J. Logue, S.A. Luthra, C. Ma, S.W. Mason, M.E. McGrath, S. Noell, R.S. Obach, M.N. O' Brien, R. O'Connor, K. Ogilvie, D. Owen, M. Pettersson, M.R. Reese, T.F. Rogers, R. Rosales, M.I. Rossulek, J.G. Sathish, N. Shirai, C. Steppan, M. Ticehurst, L.W. Updyke, S. Weston, Y. Zhu, K.M. White, A. García-Sastre, J. Wang, A.K. Chatterjee, A.D. Mesecar, M.B. Frieman, A.S. Anderson, C. Allerton // *Nature Communications*. – 2021. – Vol. 12. – P. 6055.
110. Hepatitis C virus drugs that inhibit SARS-CoV-2 papain-like protease synergize with remdesivir to suppress viral replication in cell culture / K. Bafna, K. White, B. Harish, R. Rosales, T.A. Ramelot, T.B. Acton, E. Moreno, T. Kehrer, L. Miorin, C.A. Royer, A. García-Sastre, R.M. Krug, G.T. Montelione // *Cell Reports*. – 2021. – Vol. 35. – № 7. – P. 109133.
111. Owen. An oral SARS-CoV-2 Mpro inhibitor clinical candidate for the treatment of COVID-19 / Dafydd R. Owen // *Science*. – 2021. – Vol. 374. – № 6575. – P. 1586-1593.
112. Structural Basis of the Main Proteases of Coronavirus Bound to Drug Candidate PF-07321332 / J. Li, C. Lin, X. Zhou, F. Zhong, P. Zeng, Y. Yang, Y. Zhang, B. Yu, X. Fan, P.J. McCormick, R. Fu, Y. Fu, H. Jiang, J. Zhang // *Journal of Virology*. – 2022. – Vol. 96. – № 8. – P. e0201321.
113. Discovery of Ketone-Based Covalent Inhibitors of Coronavirus 3CL Proteases for the Potential Therapeutic Treatment of COVID-19 / R.L. Hoffman, R.S. Kania, M.A. Brothers, J.F. Davies, R.A. Ferre, K.S. Gajiwala, M. He, R.J. Hogan, K. Kozminski, L.Y. Li, J.W. Lockner, J. Lou, M.T. Marra, L.J. Mitchell, B.W. Murray, J.A. Nieman, S. Noell, S.P. Planken, T. Rowe, K. Ryan, G.J. Smith, J.E. Solowiej, C.M. Steppan, B. Taggart // *Journal of Medicinal Chemistry*. – 2020. – Vol. 63. – № 21. – P. 12725-12747.
114. The history, mechanism, and perspectives of nirmatrelvir (PF-07321332): an orally bioavailable main protease inhibitor used in combination with ritonavir to reduce COVID-19-related hospitalizations / R.P. Joyce, V.W. Hu, J. Wang // *Medicinal Chemistry Research*. – 2022. – Vol. 31. – № 10. – P. 1637-1646.
115. Oral Nirmatrelvir for High-Risk, Nonhospitalized Adults with Covid-19 / J. Hammond, H. Leister-Tebbe, A. Gardner, P. Abreu, W. Bao, W. Wisemandle, M. Baniecki, V.M. Hendrick, B. Damle, A. Simón-Campos, R. Pypstra, J.M. Rusnak, EPIC-HR Investigators // *The New England Journal of Medicine*. – 2022. – Vol. 386. – № 15. – P. 1397-1408.
116. Identification of Vitamin K3 and its analogues as covalent inhibitors of SARS-CoV-2 3CLpro / R. Wang, Q. Hu, H. Wang, G. Zhu, M. Wang, Q. Zhang, Y. Zhao, C. Li, Y. Zhang, G. Ge, H. Chen, L. Chen // *International Journal of Biological Macromolecules*. – 2021. – Vol. 183. – P. 182-192.
117. Identification of pyrogallol as a warhead in design of covalent inhibitors for the SARS-CoV-2 3CL protease / H. Su, S. Yao, W. Zhao, Y. Zhang, J. Liu, Q. Shao, Q. Wang, M. Li, H. Xie, W. Shang, C. Ke, L. Feng, X. Jiang, J. Shen, G. Xiao, H. Jiang, L. Zhang, Y. Ye, Y. Xu // *Nature Communications*. – 2021. – Vol. 12. – № 1. – P. 3623.

118. GRL-0920, an Indole Chloropyridinyl Ester, Completely Blocks SARS-CoV-2 Infection / S.-I. Hattori, N. Higashi-Kuwata, J. Raghavaiah, D. Das, H. Bulut, D.A. Davis, Y. Takamatsu, K. Matsuda, N. Takamune, N. Kishimoto, T. Okamura, S. Misumi, R. Yarchoan, K. Maeda, A.K. Ghosh, H. Mitsuya // *mBio*. – 2020. – Vol. 11. – № 4. – P. e01833-20.
119. Chloropyridinyl Esters of Nonsteroidal Anti-Inflammatory Agents and Related Derivatives as Potent SARS-CoV-2 3CL Protease Inhibitors / A.K. Ghosh, D. Shahabi, M. Yadav, S. Kovela, B.J. Anson, E.K. Lendy, C. Bonham, D. Sirohi, C.A. Brito-Sierra, S. Hattori, R. Kuhn, H. Mitsuya, A.D. Mesecar // *Molecules*. – 2021. – Vol. 26. – № 19. – P. 5782.
120. Inhibition mechanism of SARS-CoV-2 main protease by ebselen and its derivatives / K. Amporndanai, X. Meng, W. Shang, Z. Jin, M. Rogers, Y. Zhao, Z. Rao, Z.-J. Liu, H. Yang, L. Zhang, P.M. O'Neill, S. Samar Hasnain // *Nature Communications*. – 2021. – Vol. 12. – № 1. – P. 3061.
121. Myricetin Inhibits SARS-CoV-2 Viral Replication by Targeting Mpro and Ameliorates Pulmonary Inflammation / T. Xiao, M. Cui, C. Zheng, M. Wang, R. Sun, D. Gao, J. Bao, S. Ren, B. Yang, J. Lin, X. Li, D. Li, C. Yang, H. Zhou // *Frontiers in Pharmacology*. – 2021. – Vol. 12. – P. 669642.
122. Identification of Inhibitors of SARS-CoV-2 3CL-Pro Enzymatic Activity Using a Small Molecule in Vitro Repurposing Screen / M. Kuzikov, E. Costanzi, J. Reinshagen, F. Esposito, L. Vangeel, M. Wolf, B. Ellinger, C. Claussen, G. Geisslinger, A. Corona, D. Iaconis, C. Talarico, C. Manelfi, R. Cannalire, G. Rossetti, J. Gossen, S. Albani, F. Musiani, K. Herzog, Y. Ye, B. Giabbai, N. Demitri, D. Jochmans, S.D. Jonghe, J. Rymenants, V. Summa, E. Tramontano, A.R. Beccari, P. Leyssen, P. Storici, J. Neyts, P. Gribbon, A. Zaliani // *ACS pharmacology & translational science*. – 2021. – Vol. 4. – № 3. – P. 1096-1110.
123. Indole Chloropyridinyl Ester-Derived SARS-CoV-2 3CLpro Inhibitors: Enzyme Inhibition, Antiviral Efficacy, Structure-Activity Relationship, and X-ray Structural Studies / A.K. Ghosh, J. Raghavaiah, D. Shahabi, M. Yadav, B.J. Anson, E.K. Lendy, S.-I. Hattori, N. Higashi-Kuwata, H. Mitsuya, A.D. Mesecar // *Journal of Medicinal Chemistry*. – 2021. – Vol. 64. – № 19. – P. 14702-14714.
124. Ebselen, a glutathione peroxidase mimetic seleno-organic compound, as a multifunctional antioxidant. Implication for inflammation-associated carcinogenesis / Y. Nakamura, Q. Feng, T. Kumagai, K. Torikai, H. Ohigashi, T. Osawa, N. Noguchi, E. Niki, K. Uchida // *The Journal of Biological Chemistry*. – 2002. – Vol. 277. – № 4. – P. 2687-2694.
125. Ebsulfur and Ebselen as highly potent scaffolds for the development of potential SARS-CoV-2 antivirals / L.-Y. Sun, C. Chen, J. Su, J.-Q. Li, Z. Jiang, H. Gao, J.-Z. Chigan, H.-H. Ding, L. Zhai, K.-W. Yang // *Bioorganic Chemistry*. – 2021. – Vol. 112. – P. 104889.
126. Discovery of highly potent SARS-CoV-2 Mpro inhibitors based on benzoisothiazolone scaffold / W. Chen, B. Feng, S. Han, P. Wang, W. Chen, Y. Zang, J. Li, Y. Hu // *Bioorganic & Medicinal Chemistry Letters*. – 2022. – Vol. 58. – P. 128526.
127. The resurgence of covalent drugs / J. Singh, R.C. Petter, T.A. Baillie, A. Whitty // *Nature Reviews. Drug Discovery*. – 2011. – Vol. 10. – № 4. – P. 307-317.
128. Chapter Three - Chemistry perspectives of reversible covalent drugs / J. Gao, V. Nobile // *Annual Reports in Medicinal Chemistry : The Design of Covalent-Based Inhibitors* / R.A. Ward, N.P. Grimster eds. . – Academic Press, 2021. – Vol. 56. – P. 75-94.
129. A Phase 3, Multicenter, Randomized, Double-Blind, 24-Week Study of the Clinical and Antiviral

Effect of S-217622 Compared With Placebo in Non-Hospitalized Participants With COVID-19 / Shionogi. – clinicaltrials.gov, 2023. – Mode of access: <https://clinicaltrials.gov/study/> NCT05305547 (date of access: 08.12.2023). – [Electronic resource].

130. Discovery of S-217622, a Noncovalent Oral SARS-CoV-2 3CL Protease Inhibitor Clinical Candidate for Treating COVID-19 / Y. Unoh, S. Uehara, K. Nakahara, H. Nobori, Y. Yamatsu, S. Yamamoto, Y. Maruyama, Y. Taoda, K. Kasamatsu, T. Suto, K. Kouki, A. Nakahashi, S. Kawashima, T. Sanaki, S. Toba, K. Uemura, T. Mizutare, S. Ando, M. Sasaki, Y. Orba, H. Sawa, A. Sato, T. Sato, T. Kato, Y. Tachibana // Journal of Medicinal Chemistry. – 2022. – Vol. 65. – № 9. – P. 6499-6512.
131. S-217622, a SARS-CoV-2 main protease inhibitor, decreases viral load and ameliorates COVID-19 severity in hamsters / M. Sasaki, K. Tabata, M. Kishimoto, Y. Itakura, H. Kobayashi, T. Ariizumi, K. Uemura, S. Toba, S. Kusakabe, Y. Maruyama, S. Iida, N. Nakajima, T. Suzuki, S. Yoshida, H. Nobori, T. Sanaki, T. Kato, T. Shishido, W.W. Hall, Y. Orba, A. Sato, H. Sawa // Science Translational Medicine. – 2023. – Vol. 15. – № 679. – P. eabq4064.
132. Potent Noncovalent Inhibitors of the Main Protease of SARS-CoV-2 from Molecular Sculpting of the Drug Perampanel Guided by Free Energy Perturbation Calculations / C.-H. Zhang, E.A. Stone, M. Deshmukh, J.A. Ippolito, M.M. Ghahremanpour, J. Tirado-Rives, K.A. Spasov, S. Zhang, Y. Takeo, S.N. Kudalkar, Z. Liang, F. Isaacs, B. Lindenbach, S.J. Miller, K.S. Anderson, W.L. Jorgensen // ACS central science. – 2021. – Vol. 7. – № 3. – P. 467-475.
133. Crystal Structure of SARS-CoV-2 Main Protease in Complex with the Non-Covalent Inhibitor ML188 / G.J. Lockbaum, A.C. Reyes, J.M. Lee, R. Tilwawala, E.A. Nalivaika, A. Ali, N. Kurt Yilmaz, P.R. Thompson, C.A. Schiffer // Viruses. – 2021. – Vol. 13. – № 2. – P. 174.
134. Identification of 14 Known Drugs as Inhibitors of the Main Protease of SARS-CoV-2 / M.M. Ghahremanpour, J. Tirado-Rives, M. Deshmukh, J.A. Ippolito, C.-H. Zhang, I. Cabeza de Vaca, M.-E. Liosi, K.S. Anderson, W.L. Jorgensen // ACS medicinal chemistry letters. – 2020. – Vol. 11. – № 12. – P. 2526-2533.
135. Discovery, synthesis, and structure-based optimization of a series of N-(tert-butyl)-2-(N-arylamido)-2-(pyridin-3-yl) acetamides (ML188) as potent noncovalent small molecule inhibitors of the severe acute respiratory syndrome coronavirus (SARS-CoV) 3CL protease / J. Jacobs, V. Grum-Tokars, Y. Zhou, M. Turlington, S.A. Saldanha, P. Chase, A. Eggler, E.S. Dawson, Y.M. Baez-Santos, S. Tomar, A.M. Mielech, S.C. Baker, C.W. Lindsley, P. Hodder, A. Mesecar, S.R. Stauffer // Journal of Medicinal Chemistry. – 2013. – Vol. 56. – № 2. – P. 534-546.
136. An automatic pipeline for the design of irreversible derivatives identifies a potent SARS-CoV-2 Mpro inhibitor / D. Zaidman, P. Gehrtz, M. Filep, D. Fearon, R. Gabizon, A. Douangamath, J. Prilusky, S. Duberstein, G. Cohen, C.D. Owen, E. Resnick, C. Strain-Damerell, P. Lukacik, Covid-Moonshot Consortium, H. Barr, M.A. Walsh, F. von Delft, N. London // Cell Chemical Biology. – 2021. – Vol. 28. – № 12. – P. 1795-1806.e5.
137. The continued threat of emerging flaviviruses / T.C. Pierson, M.S. Diamond // Nature Microbiology. – 2020. – Vol. 5. – № 6. – P. 796-812.
138. Лихорадка Западного Нила в Российской Федерации в 2022 г., прогноз заболеваемости на 2023 г. / Е.В. Путинцева, С.К. Удовиченко, Д.Н. Никитин, Н.В. Бородай, А.А. Батурина, А.Ю. Мачнева, А.С. Антонов, Н.А. Зарубин, А.В. Топорков // Проблемы особо опасных инфекций. – 2023. – Vol. 0. – № 1. – P. 75-84.
139. Современное состояние проблемы клещевого энцефалита в России и мире / Н.М.

- Колясникова, А.А. Ишмухаметов, В.Г. Акимкин // Эпидемиология и Вакцинопрофилактика. – 2023. – Vol. 22. – № 1. – P. 104-123.
140. Pathogenic flaviviruses / E.A. Gould, T. Solomon // Lancet. – 2008. – Vol. 371. – № 9611. – P. 500-509.
141. Pathogenesis and virulence of flavivirus infections / S.W. van Leur, T. Heunis, D. Munnur, S. Sanyal // Virulence. – 2021. – Vol. 12. – № 1. – P. 2814-2838.
142. Flaviviruses are neurotropic, but how do they invade the CNS? / J.W. Neal // The Journal of Infection. – 2014. – Vol. 69. – № 3. – P. 203-215.
143. Flavivirus Receptors: Diversity, Identity, and Cell Entry / M. Laureti, D. Narayanan, J. Rodriguez-Andres, J.K. Fazakerley, L. Kedzierski // Frontiers in Immunology. – 2018. – Vol. 9. – Flavivirus Receptors. – P. 2180.
144. Zika virus NS1 affects the junctional integrity of human brain microvascular endothelial cells / M. Rastogi, S.K. Singh // Biochimie. – 2020. – Vol. 176. – P. 52-61.
145. Flavivirus NS1 Triggers Tissue-Specific Vascular Endothelial Dysfunction Reflecting Disease Tropism / H. Puerta-Guardo, D.R. Glasner, D.A. Espinosa, S.B. Biering, M. Patana, K. Ratnasiri, C. Wang, P.R. Beatty, E. Harris // Cell Reports. – 2019. – Vol. 26. – № 6. – P. 1598-1613.e8.
146. Dengue Virus NS1 Disrupts the Endothelial Glycocalyx, Leading to Hyperpermeability / H. Puerta-Guardo, D.R. Glasner, E. Harris // PLoS pathogens. – 2016. – Vol. 12. – № 7. – P. e1005738.
147. Antagonism of the complement component C4 by flavivirus nonstructural protein NS1 / P. Avirutnan, A. Fuchs, R.E. Hauhart, P. Somnuke, S. Youn, M.S. Diamond, J.P. Atkinson // The Journal of Experimental Medicine. – 2010. – Vol. 207. – № 4. – P. 793-806.
148. Chen, S. Innate Immune Evasion Mediated by Flaviviridae Non-Structural Proteins / S. Chen, Z. Wu, M. Wang, A. Cheng // Viruses. – 2017. – Vol. 9. – № 10. – P. 291.
149. Chapter One - Discovery of small molecule antivirals targeting tick-borne encephalitis virus / A.S. Goryashchenko, V.I. Uvarova, D.I. Osolodkin, A.A. Ishmukhametov // Annual Reports in Medicinal Chemistry : Medicinal Chemistry of Tick-Borne Encephalitis. – Academic Press, 2022. – Vol. 58. – P. 1-54.
150. Tick-borne encephalitis in Europe and Russia: Review of pathogenesis, clinical features, therapy, and vaccines / D. Ružek, T. Avšič Županc, J. Borde, A. Chrdle, L. Eyer, G. Karganova, I. Kholodilov, N. Knap, L. Kozlovskaya, A. Matveev, A.D. Miller, D.I. Osolodkin, A.K. Överby, N. Tikunova, S. Tkachev, J. Zajkowska // Antiviral Research. – 2019. – Vol. 164. – P. 23-51.
151. Dengue: Status of current and under-development vaccines / M. Redoni, S. Yacoub, L. Rivino, D.R. Giacobbe, R. Luzzati, S. Di Bella // Reviews in Medical Virology. – 2020. – Vol. 30. – № 4. – P. e2101.
152. Current status and future prospects of yellow fever vaccines / A.S. Beck, A.D.T. Barrett // Expert Review of Vaccines. – 2015. – Vol. 14. – № 11. – P. 1479-1492.
153. Japanese encephalitis: the virus and vaccines / S.-I. Yun, Y.-M. Lee // Human Vaccines & Immunotherapeutics. – 2014. – Vol. 10. – № 2. – P. 263-279.
154. Structural perspectives of antibody-dependent enhancement of infection of dengue virus / S.R. Morrone, S.-M. Lok // Current Opinion in Virology. – 2019. – Vol. 36. – P. 1-8.
155. Lethal antibody enhancement of dengue disease in mice is prevented by Fc modification / S.J. Balsitis, K.L. Williams, R. Lachica, D. Flores, J.L. Kyle, E. Mehlhop, S. Johnson, M.S. Diamond, P.R.

- Beatty, E. Harris // PLoS pathogens. – 2010. – Vol. 6. – № 2. – P. e1000790.
156. Genus: Orthoflavivirus | ICTV. – Mode of access: <https://ictv.global/report/chapter/flaviviridae/flaviviridae/orthoflavivirus> (date of access: 01.10.2024). – [Electronic resource].
157. Flavivirus infection—A review of immunopathogenesis, immunological response, and immunodiagnosis / H.Y. Chong, C.Y. Leow, A.B. Abdul Majeed, C.H. Leow // Virus Research. – 2019. – Vol. 274. – P. 197770.
158. Моделирование структуры белка NS1 ортофлавивирусов по гомологии для виртуального скрининга потенциальных лигандов / Фомина А.Д., Палюлин В.А., Осолодкин Д.И. // Биомедицинская химия. – 2024. – Vol. 70. – № 6. – P. 456-468.
159. Flavivirus entry receptors: an update / M. Perera-Lecoin, L. Meertens, X. Carnec, A. Amara // Viruses. – 2013. – Vol. 6. – Flavivirus entry receptors. – № 1. – P. 69-88.
160. Molecular mechanisms of flavivirus membrane fusion / K. Stiasny, R. Fritz, K. Pangerl, F.X. Heinz // Amino Acids. – 2011. – Vol. 41. – № 5. – P. 1159-1163.
161. Posttranslational signal peptidase cleavage at the flavivirus C-prM junction in vitro / C.E. Stocks, M. Lobigs // Journal of Virology. – 1995. – Vol. 69. – № 12. – P. 8123-8126.
162. A novel interaction between dengue virus nonstructural protein 1 and the NS4A-2K-4B precursor is required for viral RNA replication but not for formation of the membranous replication organelle / A. Płaszczyca, P. Scaturro, C.J. Neufeldt, M. Cortese, B. Cerikan, S. Ferla, A. Brancale, A. Pichlmair, R. Bartenschlager // PLOS Pathogens. – 2019. – Vol. 15. – № 5. – P. e1007736.
163. A structural perspective of the flavivirus life cycle / S. Mukhopadhyay, R.J. Kuhn, M.G. Rossmann // Nature Reviews. Microbiology. – 2005. – Vol. 3. – № 1. – P. 13-22.
164. Tick-Borne Encephalitis Virus: A Structural View / L.I.A. Pulkkinen, S.J. Butcher, M. Anastasina // Viruses. – 2018. – Vol. 10. – Tick-Borne Encephalitis Virus. – № 7. – P. 350.
165. Transport and budding at two distinct sites of visible nucleocapsids of West Nile (Sarafend) virus / M.L. Ng, S.H. Tan, J.J. Chu // Journal of Medical Virology. – 2001. – Vol. 65. – № 4. – P. 758-764.
166. Structural and Functional Properties of the Capsid Protein of Dengue and Related Flavivirus / A.F. Faustino, A.S. Martins, N. Karguth, V. Artilheiro, F.J. Enguita, J.C. Ricardo, N.C. Santos, I.C. Martins // International Journal of Molecular Sciences. – 2019. – Vol. 20. – № 16. – P. 3870.
167. PrM- and cell-binding domains of the dengue virus E protein / S. Wang, R. He, R. Anderson // Journal of Virology. – 1999. – Vol. 73. – № 3. – P. 2547-2551.
168. Flavivirus NS1 structures reveal surfaces for associations with membranes and the immune system / D.L. Akey, W.C. Brown, S. Dutta, J. Konwerski, J. Jose, T.J. Jurkiw, J. DelProposto, C.M. Ogata, G. Skiniotis, R.J. Kuhn, J.L. Smith // Science (New York, N.Y.). – 2014. – Vol. 343. – № 6173. – P. 881-885.
169. Flavivirus NS1 protein in infected host sera enhances viral acquisition by mosquitoes / J. Liu, Y. Liu, K. Nie, S. Du, J. Qiu, X. Pang, P. Wang, G. Cheng // Nature Microbiology. – 2016. – Vol. 1. – № 9. – P. 16087.
170. Role of nonstructural protein NS2A in flavivirus assembly / J.Y. Leung, G.P. Pijlman, N. Kondratieva, J. Hyde, J.M. Mackenzie, A.A. Khromykh // Journal of Virology. – 2008. – Vol. 82. – № 10. – P. 4731-4741.
171. Architects of assembly: roles of Flaviviridae non-structural proteins in virion morphogenesis /

C.L. Murray, C.T. Jones, C.M. Rice // *Nature Reviews. Microbiology.* – 2008. – Vol. 6. – Architects of assembly. – № 9. – P. 699-708.

172. Structure and functionality in flavivirus NS-proteins: perspectives for drug design / M. Bollati, K. Alvarez, R. Assenberg, C. Baronti, B. Canard, S. Cook, B. Coutard, E. Decroly, X. de Lamballerie, E.A. Gould, G. Grard, J.M. Grimes, R. Hilgenfeld, A.M. Jansson, H. Malet, E.J. Mancini, E. Mastrangelo, A. Mattevi, M. Milani, G. Moureau, J. Neyts, R.J. Owens, J. Ren, B. Selisko, S. Speroni, H. Steuber, D.I. Stuart, T. Unge, M. Bolognesi // *Antiviral Research.* – 2010. – Vol. 87. – № 2. – P. 125-148.

173. Processing of the yellow fever virus nonstructural polyprotein: a catalytically active NS3 proteinase domain and NS2B are required for cleavages at dibasic sites / T.J. Chambers, A. Grakoui, C.M. Rice // *Journal of Virology.* – 1991. – Vol. 65. – № 11. – P. 6042-6050.

174. Japanese encephalitis virus nonstructural protein NS3 has RNA binding and ATPase activities / T. Takegami, D. Sakamuro, T. Furukawa // *Virus Genes.* – 1995. – Vol. 9. – № 2. – P. 105-112.

175. Characterization of dengue virus NS4A and NS4B protein interaction / J. Zou, X. Xie, Q.-Y. Wang, H. Dong, M.Y. Lee, C. Kang, Z. Yuan, P.-Y. Shi // *Journal of Virology.* – 2015. – Vol. 89. – № 7. – P. 3455-3470.

176. Inhibition of alpha/beta interferon signaling by the NS4B protein of flaviviruses / J.L. Muñoz-Jordán, M. Laurent-Rolle, J. Ashour, L. Martínez-Sobrido, M. Ashok, W.I. Lipkin, A. García-Sastre // *Journal of Virology.* – 2005. – Vol. 79. – № 13. – P. 8004-8013.

177. Disulfide Reduction Allosterically Destabilizes the  $\beta$ -Ladder Subdomain Assembly within the NS1 Dimer of ZIKV / P. Roy, S. Roy, N. Sengupta // *Biophysical Journal.* – 2020. – Vol. 119. – № 8. – P. 1525-1537.

178. Secreted dengue virus nonstructural protein NS1 is an atypical barrel-shaped high-density lipoprotein / I. Gutsche, F. Coulibaly, J.E. Voss, J. Salmon, J. d'Alayer, M. Ermonval, E. Larquet, P. Charneau, T. Krey, F. Mégret, E. Guittet, F.A. Rey, M. Flamand // *Proceedings of the National Academy of Sciences of the United States of America.* – 2011. – Vol. 108. – № 19. – P. 8003-8008.

179. N130, N175 and N207 are N-linked glycosylation sites of duck Tembusu virus NS1 that are important for viral multiplication, viremia and virulence in ducklings / S. Zhang, X. Wang, Y. He, T. Hu, J. Guo, M. Wang, R. Jia, D. Zhu, M. Liu, X. Zhao, Q. Yang, Y. Wu, S. Zhang, J. Huang, S. Mao, X. Ou, Q. Gao, D. Sun, Y. Liu, L. Zhang, S. Chen, A. Cheng // *Veterinary Microbiology.* – 2021. – Vol. 261. – P. 109215.

180. N-linked glycosylation of dengue virus NS1 protein modulates secretion, cell-surface expression, hexamer stability, and interactions with human complement / P. Somnuke, R.E. Hauhart, J.P. Atkinson, M.S. Diamond, P. Avirutnan // *Virology.* – 2011. – Vol. 413. – № 2. – P. 253-264.

181. Glycosylation mutants of dengue virus NS1 protein / M.J. Pryor, P.J. Wright // *The Journal of General Virology.* – 1994. – Vol. 75 ( Pt 5). – P. 1183-1187.

182. CryoEM structures of the multimeric secreted NS1, a major factor for dengue hemorrhagic fever / B. Shu, J.S.G. Ooi, A.W.K. Tan, T.-S. Ng, W. Dejnirattisai, J. Mongkolsapaya, G. Fibriansah, J. Shi, V.A. Kostyuchenko, G.R. Screamton, S.-M. Lok // *Nature Communications.* – 2022. – Vol. 13. – № 1. – P. 6756.

183. The step-by-step assembly mechanism of secreted flavivirus NS1 tetramer and hexamer captured at atomic resolution / Q. Pan, H. Jiao, W. Zhang, Q. Chen, G. Zhang, J. Yu, W. Zhao, H. Hu // *Science Advances.* – 2024. – Vol. 10. – № 18. – P. eadm8275.

184. Zika virus NS1 structure reveals diversity of electrostatic surfaces among flaviviruses / H. Song, J. Qi, J. Haywood, Y. Shi, G.F. Gao // *Nature Structural & Molecular Biology*. – 2016. – Vol. 23. – № 5. – P. 456-458.
185. Identifying Candidate Targets of Immune Responses in Zika Virus Based on Homology to Epitopes in Other Flavivirus Species / X. Xu, K. Vaughan, D. Weiskopf, A. Grifoni, M.S. Diamond, A. Sette, B. Peters // *PLoS currents*. – 2016. – Vol. 8. – P. ecurrents.outbreaks.9aa2e1fb61b0f632f58a098773008c4b.
186. Structure-guided insights on the role of NS1 in flavivirus infection / D.L. Akey, W.C. Brown, J. Jose, R.J. Kuhn, J.L. Smith // *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*. – 2015. – Vol. 37. – № 5. – P. 489-494.
187. Clinical and Laboratory Diagnosis of Dengue Virus Infection / D.A. Muller, A.C.I. Depelsenaire, P.R. Young // *The Journal of Infectious Diseases*. – 2017. – Vol. 215. – № suppl\_2. – P. S89-S95.
188. Virus-specific cytolytic antibodies to nonstructural protein 1 of Japanese encephalitis virus effect reduction of virus output from infected cells / V.D. Krishna, M. Rangappa, V. Satchidanandam // *Journal of Virology*. – 2009. – Vol. 83. – № 10. – P. 4766-4777.
189. The structural immunology of antibody protection against West Nile virus / M.S. Diamond, T.C. Pierson, D.H. Fremont // *Immunological Reviews*. – 2008. – Vol. 225. – P. 212-225.
190. The flavivirus NS1 protein: molecular and structural biology, immunology, role in pathogenesis and application as a diagnostic biomarker / D.A. Muller, P.R. Young // *Antiviral Research*. – 2013. – Vol. 98. – The flavivirus NS1 protein. – № 2. – P. 192-208.
191. Viral Entry and NS1 as Potential Antiviral Drug Targets / A.M. de Silva, F.A. Rey, P.R. Young, R. Hilgenfeld, S.G. Vasudevan // *Advances in Experimental Medicine and Biology*. – 2018. – Vol. 1062. – P. 107-113.
192. Inhibition of Japanese encephalitis virus NS1 protein expression in cell by small interfering RNAs / X. Liu, S. Cao, R. Zhou, G. Xu, S. Xiao, Y. Yang, M. Sun, Y. Li, H. Chen // *Virus Genes*. – 2006. – Vol. 33. – № 1. – P. 69-75.
193. Peptides targeting dengue viral nonstructural protein 1 inhibit dengue virus production / P. Songprakhon, T. Thaingtamtanha, T. Limjindaporn, C. Puttikhunt, C. Srisawat, P. Luangaram, T. Dechtaewewat, C. Uthaipibull, S. Thongsima, P.-T. Yenchitsomanus, P. Malasit, S. Noisakran // *Scientific Reports*. – 2020. – Vol. 10. – № 1. – P. 12933.
194. Screening Pipeline for Flavivirus Based Inhibitors for Zika Virus NS1 / S. Raza, G. Abbas, S.S. Azam // *IEEE/ACM Transactions on Computational Biology and Bioinformatics*. – 2020. – Vol. 17. – № 5. – P. 1751-1761.
195. Dual targeting of dengue virus virions and NS1 protein with the heparan sulfate mimic PG545 / N. Modhiran, N.S. Gandhi, N. Wimmer, S. Cheung, K. Stacey, P.R. Young, V. Ferro, D. Watterson // *Antiviral Research*. – 2019. – Vol. 168. – P. 121-127.
196. Computational screening of medicinal plant phytochemicals to discover potent pan-serotype inhibitors against dengue virus / M. Tahir Ul Qamar, A. Maryam, I. Muneer, F. Xing, U.A. Ashfaq, F.A. Khan, F. Anwar, M.H. Geesi, R.R. Khalid, S.A. Rauf, A.R. Siddiqi // *Scientific Reports*. – 2019. – Vol. 9. – № 1. – P. 1433.
197. Multiple pathways for SARS-CoV-2 resistance to nirmatrelvir / S. Iketani, H. Mohri, B. Culbertson, S.J. Hong, Y. Duan, M.I. Luck, M.K. Annavajhala, Y. Guo, Z. Sheng, A.-C. Uhlemann, S.P. Goff, Y. Sabo, H. Yang, A. Chavez, D.D. Ho // *Nature*. – 2023. – Vol. 613. – № 7944. – P. 558-564.

198. Surely you are joking, Mr Docking! / F. Gentile, T.I. Oprea, A. Tropsha, A. Cherkasov // Chemical Society Reviews. – 2023. – Vol. 52. – № 3. – P. 872-878.
199. The Protein Data Bank / H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weisig, I.N. Shindyalov, P.E. Bourne // Nucleic Acids Research. – 2000. – Vol. 28. – № 1. – P. 235-242.
200. PyChimera: use UCSF Chimera modules in any Python 2.7 project / J. Rodríguez-Guerra Pedregal, J.-D. Maréchal // Bioinformatics. – 2018. – Vol. 34. – PyChimera. – № 10. – P. 1784-1785.
201. Pettersen. UCSF Chimera—A visualization system for exploratory research and analysis / Eric F. Pettersen, Thomas D. Goddard, Gregory S. Couch, Daniel M. Greenblatt, Elaine C. Meng, Thomas E. Ferrin // Journal of Computational Chemistry. – 2004. – Vol. 25. – № 13. – P. 1605-1612.
202. Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors / Z. Jin, X. Du, Y. Xu, Y. Deng, M. Liu, Y. Zhao, B. Zhang, X. Li, L. Zhang, C. Peng, Y. Duan, J. Yu, L. Wang, K. Yang, F. Liu, R. Jiang, X. Yang, T. You, X. Liu, X. Yang, F. Bai, H. Liu, X. Liu, L.W. Guddat, W. Xu, G. Xiao, C. Qin, Z. Shi, H. Jiang, Z. Rao, H. Yang // Nature. – 2020. – Vol. 582. – № 7811. – P. 289-293.
203. VMD: Visual molecular dynamics / W. Humphrey, A. Dalke, K. Schulten // Journal of Molecular Graphics. – 1996. – Vol. 14. – VMD. – № 1. – P. 33-38.
204. pdb-tools: a swiss army knife for molecular structures / J.P.G.L.M. Rodrigues, J.M.C. Teixeira, M. Trellet, A.M.J.J. Bonvin // F1000Research. – 2018. – Vol. 7. – pdb-tools. – P. 1961.
205. MUSCLE: multiple sequence alignment with high accuracy and high throughput / R.C. Edgar // Nucleic Acids Research. – 2004. – Vol. 32. – MUSCLE. – № 5. – P. 1792-1797.
206. MUSCLE < Multiple Sequence Alignment < EMBL-EBI. – Mode of access: <https://www.ebi.ac.uk/Tools/msa/muscle/> (date of access: 17.08.2023). – [Electronic resource].
207. A discussion of the solution for the best rotation to relate two sets of vectors / W. Kabsch // Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography. – 1978. – Vol. 34. – № 5. – P. 827-828.
208. Crystallographic and electrophilic fragment screening of the SARS-CoV-2 main protease / A. Douangamath, D. Fearon, P. Gehrtz, T. Krojer, P. Lukacik, C.D. Owen, E. Resnick, C. Strain-Damerell, A. Aimon, P. Ábrányi-Balogh, J. Brandão-Neto, A. Carbery, G. Davison, A. Dias, T.D. Downes, L. Dunnett, M. Fairhead, J.D. Firth, S.P. Jones, A. Keeley, G.M. Keserü, H.F. Klein, M.P. Martin, M.E.M. Noble, P. O'Brien, A. Powell, R.N. Reddi, R. Skyner, M. Snee, M.J. Waring, C. Wild, N. London, F. von Delft, M.A. Walsh // Nature Communications. – 2020. – Vol. 11. – № 1. – P. 5047.
209. Программный комплекс для систематического ансамблевого докинга / Фомина А.Д., Осолодкин Д.И. – Свидетельство о государственной регистрации программы для ЭВМ № 2024684465. — Номер заявки 2024684027. — Дата поступления 17.10.24. — Дата государственной регистрации в Реестре программ для ЭВМ 17.10.24., .
210. Data Structures for Statistical Computing in Python / W. McKinney // Proceedings of the 9th Python in Science Conference. – 2010. – P. 56-61.
211. Array programming with NumPy / C.R. Harris, K.J. Millman, S.J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N.J. Smith, R. Kern, M. Picus, S. Hoyer, M.H. van Kerkwijk, M. Brett, A. Haldane, J.F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, T.E. Oliphant // Nature. – 2020. – Vol. 585. – № 7825. – P. 357-362.
212. Matplotlib: A 2D Graphics Environment / J.D. Hunter // Computing in Science & Engineering. – 2007. – Vol. 9. – № 3. – P. 90-95.

213. seaborn: statistical data visualization / M.L. Waskom // Journal of Open Source Software. – 2021. – Vol. 6. – № 60. – P. 3021.
214. X-ray screening identifies active site and allosteric inhibitors of SARS-CoV-2 main protease / S. Günther, P.Y.A. Reinke, Y. Fernández-García [et al.] // Science. – 2021.
215. The temperature-dependent conformational ensemble of SARS-CoV-2 main protease (Mpro) / A. Ebrahim, B.T. Riley, D. Kumaran, B. Andi, M.R. Fuchs, S. McSweeney, D.A. Keedy // IUCrJ. – 2022. – Vol. 9. – № Pt 5. – P. 682-694.
216. Vitamin C Binds to SARS Coronavirus-2 Main Protease Essential for Viral Replication / T.N. Malla, S. Pandey, L. Aldama, D. Feliz, M. Noda, I. Poudyal, G.N. Phillips, E.A. Stojković, M. Schmidt. – bioRxiv, 2021.
217. Structural basis of SARS-CoV-2 main protease inhibition by a broad-spectrum anti-coronaviral drug / Y.-C. Wang, W.-H. Yang, C.-S. Yang, M.-H. Hou, C.-L. Tsai, Y.-Z. Chou, M.-C. Hung, Y. Chen // American Journal of Cancer Research. – 2020. – Vol. 10. – № 8. – P. 2535-2545.
218. Crystal structure of SARS-CoV-2 main protease in complex with protease inhibitor PF-07321332 / Y. Zhao, C. Fang, Q. Zhang, R. Zhang, X. Zhao, Y. Duan, H. Wang, Y. Zhu, L. Feng, J. Zhao, M. Shao, X. Yang, L. Zhang, C. Peng, K. Yang, D. Ma, Z. Rao, H. Yang // Protein & Cell. – 2022. – Vol. 13. – № 9. – P. 689-693.
219. A small molecule compound with an indole moiety inhibits the main protease of SARS-CoV-2 and blocks virus replication / S.-I. Hattori, N. Higashi-Kuwata, H. Hayashi, S.R. Allu, J. Raghavaiah, H. Bulut, D. Das, B.J. Anson, E.K. Lendy, Y. Takamatsu, N. Takamune, N. Kishimoto, K. Murayama, K. Hasegawa, M. Li, D.A. Davis, E.N. Kodama, R. Yarchoan, A. Wlodawer, S. Misumi, A.D. Mesecar, A.K. Ghosh, H. Mitsuya // Nature Communications. – 2021. – Vol. 12. – № 1. – P. 668.
220. SARS-CoV-2 Mpro inhibitors with antiviral activity in a transgenic mouse model / J. Qiao, Y.-S. Li, R. Zeng, F.-L. Liu, R.-H. Luo, C. Huang, Y.-F. Wang, J. Zhang, B. Quan, C. Shen, X. Mao, X. Liu, W. Sun, W. Yang, X. Ni, K. Wang, L. Xu, Z.-L. Duan, Q.-C. Zou, H.-L. Zhang, W. Qu, Y.-H.-P. Long, M.-H. Li, R.-C. Yang, X. Liu, J. You, Y. Zhou, R. Yao, W.-P. Li, J.-M. Liu, P. Chen, Y. Liu, G.-F. Lin, X. Yang, J. Zou, L. Li, Y. Hu, G.-W. Lu, W.-M. Li, Y.-Q. Wei, Y.-T. Zheng, J. Lei, S. Yang // Science. – 2021. – Vol. 371. – № 6536. – P. 1374-1378.
221. k-means++: the advantages of careful seeding / D. Arthur, S. Vassilvitskii // Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms : SODA '07. – USA : Society for Industrial and Applied Mathematics, 2007. – k-means++. – P. 1027-1035.
222. A Quick Route to Multiple Highly Potent SARS-CoV-2 Main Protease Inhibitors / K.S. Yang, X.R. Ma, Y. Ma, Y.R. Alugubelli, D.A. Scott, E.C. Vatansever, A.K. Drelich, B. Sankaran, Z.Z. Geng, L.R. Blankenship, H.E. Ward, Y.J. Sheng, J.C. Hsu, K.C. Kratch, B. Zhao, H.S. Hayatshahi, J. Liu, P. Li, C.A. Fierke, C.-T.K. Tseng, S. Xu, W.R. Liu // ChemMedChem. – 2021. – Vol. 16. – № 6. – P. 942-948.
223. COVID-19 related trajectory data of 10 microseconds all atom molecular dynamics simulation of SARS-CoV-2 dimeric main protease / T.S. Komatsu, Y.M. Koyama, N. Okimoto, G. Morimoto, Y. Ohno, M. Taiji. – [Electronic resource]. – 2020. – Vol. 1. – Mode of access: <https://data.mendeley.com/datasets/vpps4vhryg/1> (date of access: 18.08.2023).
224. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit / S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M.R. Shirts, J.C. Smith, P.M. Kasson, D. van der Spoel, B. Hess, E. Lindahl // Bioinformatics. – 2013. – Vol. 29. – № 7. – P. 845-

225. База данных соединений с экспериментально определенной активностью против главной протеазы  $\beta$ -коронавирусов / Фомина А.Д., Осолодкин Д.И. – Свидетельство о государственной регистрации базы данных № 2024625235. — Номер заявки 2024624786. — Дата поступления 29.10.2024 — Дата государственной регистрации в Реестре баз данных 15.11.2024 Бюл. № 11, .
226. ZINC 15--Ligand Discovery for Everyone / T. Sterling, J.J. Irwin // Journal of Chemical Information and Modeling. – 2015. – Vol. 55. – № 11. – P. 2324-2337.
227. Scikit-learn: Machine Learning in Python / F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau // Journal of Machine Learning Research. – 2011. – Vol. 12. – № 85. – P. 2825–2830.
228. The Generation of a Unique Machine Description for Chemical Structures-A Technique Developed at Chemical Abstracts Service. / H.L. Morgan // Journal of Chemical Documentation. – 1965. – P. 107.
229. An analysis of variance test for normality (complete samples) / S.S. Shapiro, M.B. Wilk // Biometrika. – 1965. – Vol. 52. – № 3-4. – P. 591-611.
230. Strengths and Weaknesses of Docking Simulations in the SARS-CoV-2 Era: the Main Protease (Mpro) Case Study / M.A. Llanos, M.E. Gantner, S. Rodriguez, L.N. Alberca, C.L. Bellera, A. Talevi, L. Gavernet // Journal of Chemical Information and Modeling. – 2021. – Vol. 61. – № 8. – P. 3758-3770.
231. COVID Moonshot: open science discovery of SARS-CoV-2 main protease inhibitors by combining crowdsourcing, high-throughput experiments, computational simulations, and machine learning / H. Achdout, A. Aimon, E. Bar-David, G.M. Morris // bioRxiv. – 2020.
232. Assay Report Card. – Mode of access: [https://www.ebi.ac.uk/chembl/assay\\_report\\_card/CHEMBL4495582/](https://www.ebi.ac.uk/chembl/assay_report_card/CHEMBL4495582/) (date of access: 18.08.2023). – [Electronic resource].
233. DataWarrior: an open-source program for chemistry aware data visualization and analysis / T. Sander, J. Freyss, M. von Korff, C. Rufener // Journal of Chemical Information and Modeling. – 2015. – Vol. 55. – № 2. – P. 460-473.
234. Crystallography Open Database (COD): an open-access collection of crystal structures and platform for world-wide collaboration / S. Gražulis, A. Daškevič, A. Merkys, D. Chateigner, L. Lutterotti, M. Quirós, N.R. Serebryanaya, P. Moeck, R.T. Downs, A. Le Bail // Nucleic Acids Research. – 2012. – Vol. 40. – № Database issue. – P. D420-427.
235. Accuracy evaluation and addition of improved dihedral parameters for the MMFF94s / J. Wahl, J. Freyss, M. von Korff, T. Sander // Journal of Cheminformatics. – 2019. – Vol. 11. – № 1. – P. 53.
236. Open Babel: An open chemical toolbox / N.M. O’Boyle, M. Banck, C.A. James, C. Morley, T. Vandermeersch, G.R. Hutchison // Journal of Cheminformatics. – 2011. – Vol. 3. – № 1. – P. 33.
237. Visualizing Data using t-SNE / L. van der Maaten, G. Hinton // Journal of Machine Learning Research. – 2008. – Vol. 9. – № 86. – P. 2579-2605.
238. Progress in visual representations of chemical space / D.I. Osolodkin, E.V. Radchenko, A.A. Orlov, A.E. Voronkov, V.A. Palyulin, N.S. Zefirov // Expert Opinion on Drug Discovery. – 2015. – Vol. 10. – № 9. – P. 959-973.
239. On Information and Sufficiency / S. Kullback, R.A. Leibler // The Annals of Mathematical Statistics. – 1951. – Vol. 22. – № 1. – P. 79-86.

240. Ripley's K Function / Philip M. Dixon // Encyclopedia of Environmetrics. – 2002. – Vol. 3. – P. 1796-1803.
241. Effect of immature tick-borne encephalitis virus particles on antiviral activity of 5-aminoisoxazole-3-carboxylic acid adamantylmethyl esters / K.K. Tuchynskaya, A.D. Fomina, N.A. Nikitin, V.V. Illarionova, V.P. Volok, L.I. Kozlovskaya, A.A. Rogova, D.A. Vasilenko, E.B. Averina, D.I. Osolodkin, G.G. Karganova // Journal of General Virology. – 2021. – Vol. 102. – № 9. – P. 001658.
242. A critical overview of computational approaches employed for COVID-19 drug discovery / E.N. Muratov, R. Amaro, C.H. Andrade, N. Brown, S. Ekins, D. Fourches, O. Isayev, D. Kozakov, J.L. Medina-Franco, K.M. Merz, T.I. Oprea, V. Poroikov, G. Schneider, M.H. Todd, A. Varnek, D.A. Winkler, A.V. Zakharov, A. Cherkasov, A. Tropsha // Chemical Society Reviews. – 2021. – Vol. 50. – № 16. – P. 9121-9151.
243. Bound to activate: conformational consequences of cyclin binding to CDK2 / E.R. Andzelm, J. Lew, S. Taylor // Structure. – 1995. – Vol. 3. – № 11. – P. 1135-1141.
244. Structural plasticity of SARS-CoV-2 3CL Mpro active site cavity revealed by room temperature X-ray crystallography / D.W. Kneller, G. Phillips, H.M. O'Neill, R. Jedrzejczak, L. Stols, P. Langan, A. Joachimiak, L. Coates, A. Kovalevsky // Nature Communications. – 2020. – Vol. 11. – № 1. – P. 3202.
245. A Chemographic Audit of anti-Coronavirus Structure-activity Information from Public Databases (ChEMBL) / D. Horvath, A. Orlov, D.I. Osolodkin, A.A. Ishmukhametov, G. Marcou, A. Varnek // Molecular Informatics. – 2020. – Vol. 39. – № 12. – P. e2000080.
246. LigPlot+: multiple ligand-protein interaction diagrams for drug discovery / R.A. Laskowski, M.B. Swindells // Journal of Chemical Information and Modeling. – 2011. – Vol. 51. – LigPlot+. – № 10. – P. 2778-2786.
247. Relationship between the inhibition constant ( $K_I$ ) and the concentration of inhibitor which causes 50 per cent inhibition ( $I_{50}$ ) of an enzymatic reaction / C. Yung-Chi, W.H. Prusoff // Biochemical Pharmacology. – 1973. – Vol. 22. – № 23. – P. 3099-3108.
248. Alinda. – Mode of access: [https://www.alinda.ru/SCREENING\\_en.html](https://www.alinda.ru/SCREENING_en.html) (date of access: 18.08.2023). – [Electronic resource].
249. A simple method for displaying the hydropathic character of a protein / Jack Kyte, Russell F. Doolittle // Journal of Molecular Biology. – Vol. 157. – P. 105-132.
250. Exploring Protein Flexibility: Incorporating Structural Ensembles From Crystal Structures and Simulation into Virtual Screening Protocols / D.J. Osguthorpe, W. Sherman, A.T. Hagler // The Journal of Physical Chemistry B. – 2012. – Vol. 116. – № 23. – P. 6952-6959.
251. Combining Docking Pose Rank and Structure with Deep Learning Improves Protein-Ligand Binding Mode Prediction over a Baseline Docking Approach / J.A. Morrone, J.K. Weber, T. Huynh, H. Luo, W.D. Cornell // Journal of Chemical Information and Modeling. – 2020. – Vol. 60. – № 9. – P. 4170-4179.
252. Open Drug Discovery Toolkit (ODDT): a new open-source player in the drug discovery field / M. Wójcikowski, P. Zielenkiewicz, P. Siedlecki // Journal of Cheminformatics. – 2015. – Vol. 7. – P. 26.
253. Explaining prediction models and individual predictions with feature contributions / E. Štrumbelj, I. Kononenko // Knowledge and Information Systems. – 2014. – Vol. 41. – № 3. – P. 647-665.
254. Batch Effect Confounding Leads to Strong Bias in Performance Estimates Obtained by Cross-

Validation / C. Soneson, S. Gerster, M. Delorenzi // PLOS ONE. – 2014. – Vol. 9. – № 6. – P. e100335.

255. Best practices in machine learning for chemistry / N. Artrith, K.T. Butler, F.-X. Coudert, S. Han, O. Isayev, A. Jain, A. Walsh // Nature Chemistry. – 2021. – Vol. 13. – № 6. – P. 505-508.

256. Magic-BLAST, an accurate RNA-seq aligner for long and short reads / G.M. Boratyn, J. Thierry-Mieg, D. Thierry-Mieg, B. Busby, T.L. Madden // BMC bioinformatics. – 2019. – Vol. 20. – № 1. – P. 405.

257. Structural basis of Flavivirus NS1 assembly and antibody recognition / M.A. Edeling, M.S. Diamond, D.H. Fremont // Proceedings of the National Academy of Sciences of the United States of America. – 2014. – Vol. 111. – № 11. – P. 4285-4290.

258. Extended surface for membrane association in Zika virus NS1 structure / W.C. Brown, D.L. Akey, J.R. Konwerski, J.T. Tarrasch, G. Skiniotis, R.J. Kuhn, J.L. Smith // Nature Structural & Molecular Biology. – 2016. – Vol. 23. – № 9. – P. 865-867.

259. Contribution of intertwined loop to membrane association revealed by Zika virus full-length NS1 structure / X. Xu, H. Song, J. Qi, Y. Liu, H. Wang, C. Su, Y. Shi, G.F. Gao // The EMBO journal. – 2016. – Vol. 35. – № 20. – P. 2170-2178.

260. Mutation Identified in Neonatal Microcephaly Destabilizes Zika Virus NS1 Assembly in Vitro / D. Wang, C. Chen, S. Liu, H. Zhou, K. Yang, Q. Zhao, X. Ji, C. Chen, W. Xie, Z. Wang, L.-Z. Mi, H. Yang // Scientific Reports. – 2017. – Vol. 7. – P. 42580.

261. Crystal structure of the C-terminal fragment of NS1 protein from yellow fever virus / H. Wang, M. Han, J. Qi, R. Hilgenfeld, T. Luo, Y. Shi, G.F. Gao, H. Song // Science China. Life Sciences. – 2017. – Vol. 60. – № 12. – P. 1403-1406.

262. Use of massively multiple merged data for low-resolution S-SAD phasing and refinement of flavivirus NS1 / D.L. Akey, W.C. Brown, J.R. Konwerski, C.M. Ogata, J.L. Smith // Acta Crystallographica Section D: Biological Crystallography. – 2014. – Vol. 70. – № 10. – P. 2719-2729.

263. Structural Study of the C-Terminal Domain of Nonstructural Protein 1 from Japanese Encephalitis Virus / T. Poonsiri, G.S.A. Wright, M.S. Diamond, L. Turtle, T. Solomon, S.V. Antonyuk // Journal of Virology. – 2018. – Vol. 92. – № 7. – P. e01868-17.

264. PROCHECK: a program to check the stereochemical quality of protein structures / R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton // Journal of Applied Crystallography. – 1993. – Vol. 26. – PROCHECK. – № 2. – P. 283-291.

265. Cross-Reactive Immunity Among Flaviviruses / A.P.S. Rathore, A.L. St. John // Frontiers in Immunology. – 2020. – Vol. 11. – P. 334.

266. Methods of protein structure comparison / I. Kufareva, R. Abagyan // Methods in Molecular Biology (Clifton, N.J.). – 2012. – Vol. 857. – P. 231-257.

267. FTSite: high accuracy detection of ligand binding sites on unbound protein structures / C.-H. Ngan, D.R. Hall, B. Zerbe, L.E. Grove, D. Kozakov, S. Vajda // Bioinformatics (Oxford, England). – 2012. – Vol. 28. – № 2. – P. 286-287.

268. Analyzing the topology of active sites: on the prediction of pockets and subpockets / A. Volkamer, A. Griewel, T. Grombacher, M. Rarey // Journal of Chemical Information and Modeling. – 2010. – Vol. 50. – № 11. – P. 2041-2052.

269. Recent changes to virus taxonomy ratified by the International Committee on Taxonomy of Viruses (2022) / P.J. Walker, S.G. Siddell, E.J. Lefkowitz, A.R. Mushegian, E.M. Adriaenssens, P.

- Alfenas-Zerbini, D.M. Dempsey, B.E. Dutilh, M.L. García, R. Curtis Hendrickson, S. Junglen, M. Krupovic, J.H. Kuhn, A.J. Lambert, M. Łobocka, H.M. Oksanen, R.J. Orton, D.L. Robertson, L. Rubino, S. Sabanadzovic, P. Simmonds, D.B. Smith, N. Suzuki, K. Van Doorslaer, A.-M. Vandamme, A. Varsani, F.M. Zerbini // Archives of Virology. – 2022. – Vol. 167. – № 11. – P. 2429-2440.
270. Comparative Protein Structure Modeling Using MODELLER / B. Webb, A. Sali. – 2016. – Vol. 54. – P. 5.6.1-5.6.37.
271. Statistical potential for assessment and prediction of protein structures / M. Shen, A. Sali. – 2006. – Vol. 15. – № 11. – P. 2507-2524.
272. Cresset / flare-python-pyflare · GitLab. – Mode of access: <https://gitlab.com/cresset/flare-python-pyflare> (date of access: 01.10.2024). – [Electronic resource].
273. Ligand and Structure-Based Drug Design Software | Flare. – Mode of access: <https://www.cresset-group.com/software/flare/> (date of access: 01.10.2024). – [Electronic resource].
274. Lead finder: an approach to improve accuracy of protein-ligand docking, binding energy estimation, and virtual screening / O.V. Stroganov, F.N. Novikov, V.S. Stroylov, V. Kulkov, G.G. Chilov // Journal of Chemical Information and Modeling. – 2008. – Vol. 48. – № 12. – P. 2371-2385.
275. Handbook of Methods of Applied Statistics. / Karson, M. // Journal of the American Statistical Association. – 1967. – Vol. 63. – № 323. – P. 1047-1049.

## Приложение А

### Характеристика результатов докинга в ансамбли структур главной протеазы SARS-CoV-2

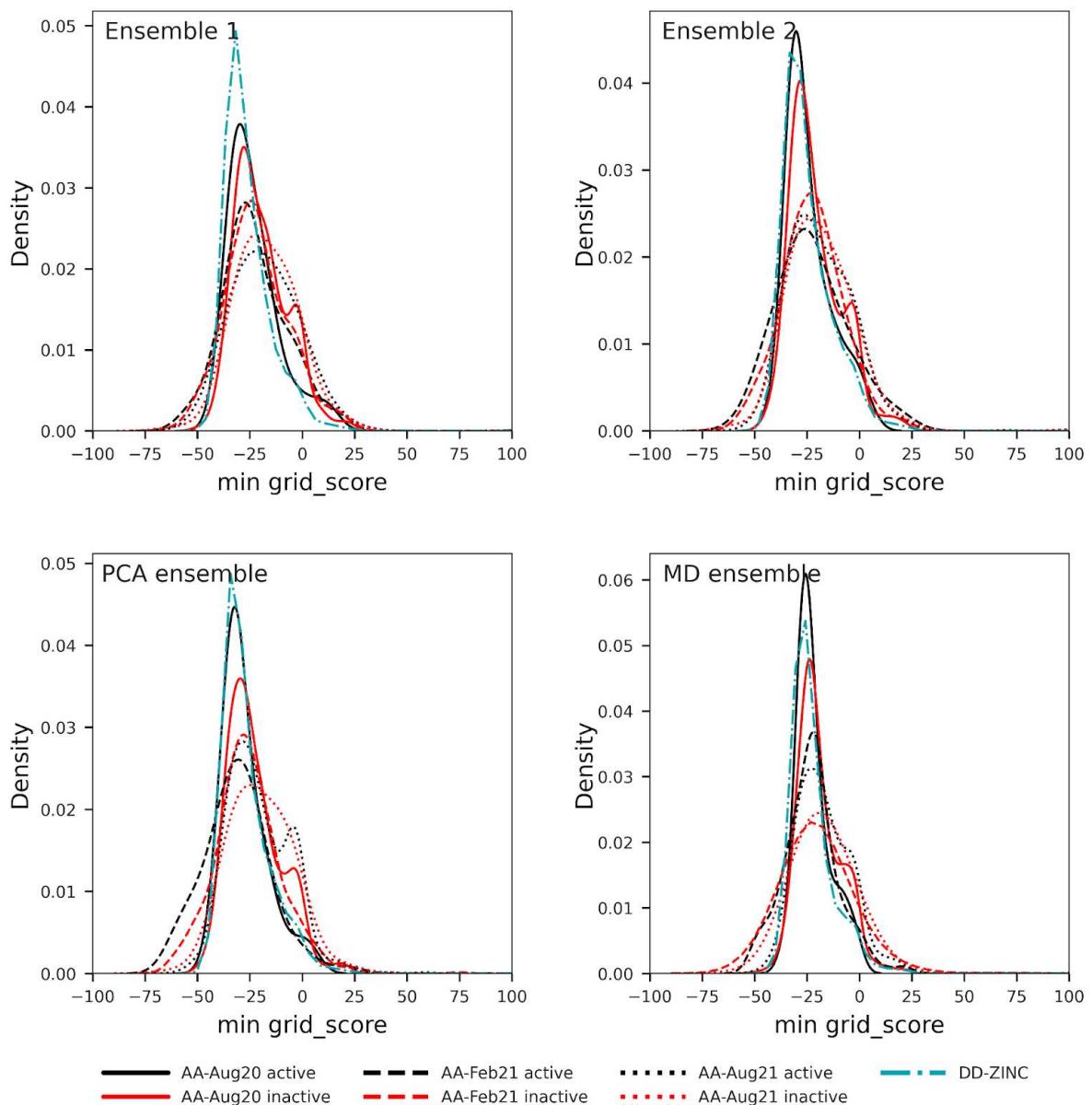


Рисунок А1. Распределение максимальных оценок докинга активных и неактивных соединений библиотек *AA-Aug20*, *AA-Feb21*, *AA-Aug21* и *DD-ZINC* в ансамбли 1, 2, PCA и МД. Здесь и далее: Density – плотность распределения, active – активные соединения, inactive – неактивные, Ensemble 1 – Ансамбль 1, Ensemble 2 – Ансамбль 2, PCA Ensemble – Ансамбль PCA, MD Ensemble – Ансамбль МД.

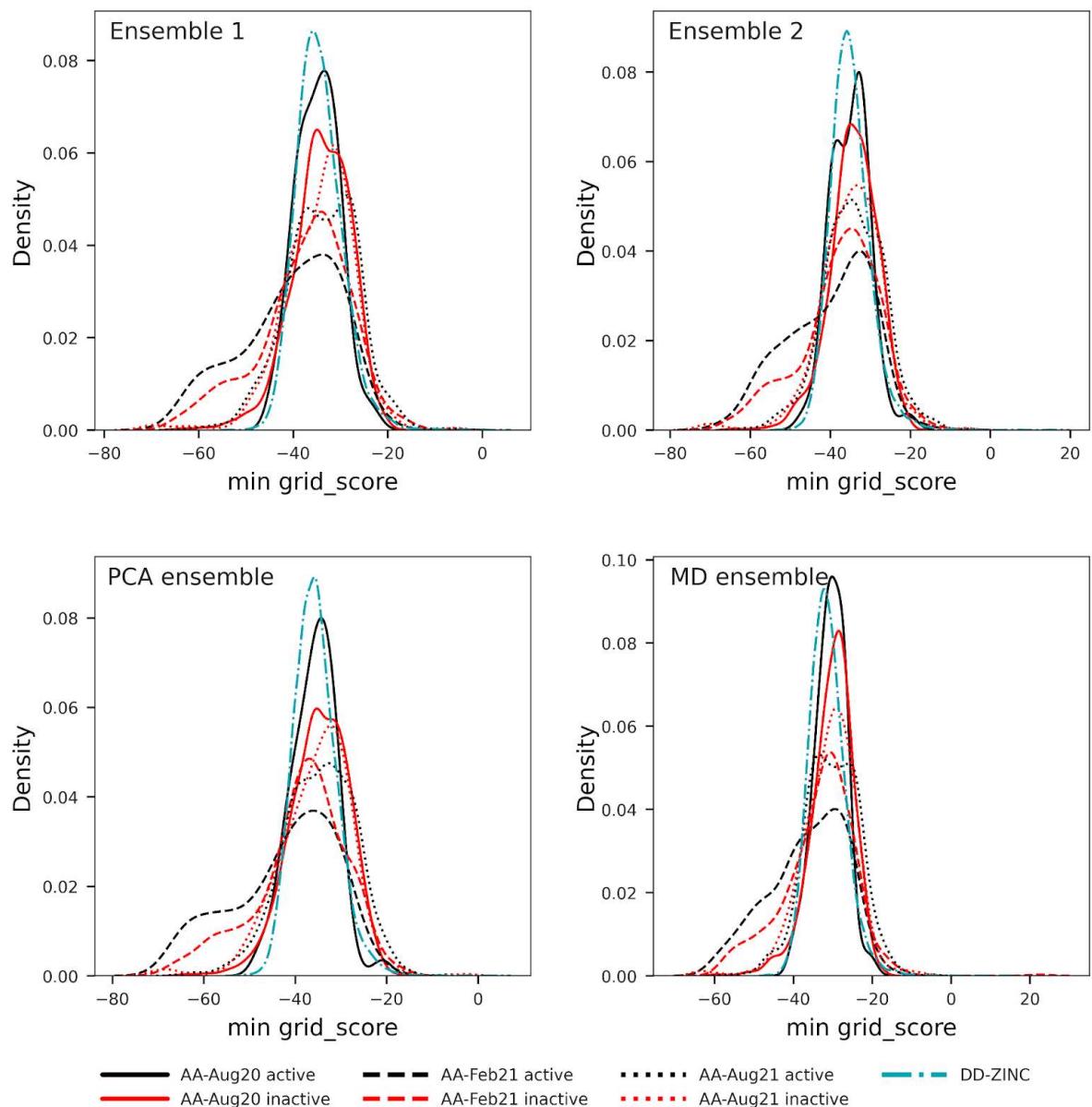


Рисунок А2. Распределение средних оценок докинга активных и неактивных соединений библиотек AA-Aug20, AA-Feb21, AA-Aug21 и DD-ZINC в ансамбли 1, 2, PCA и МД

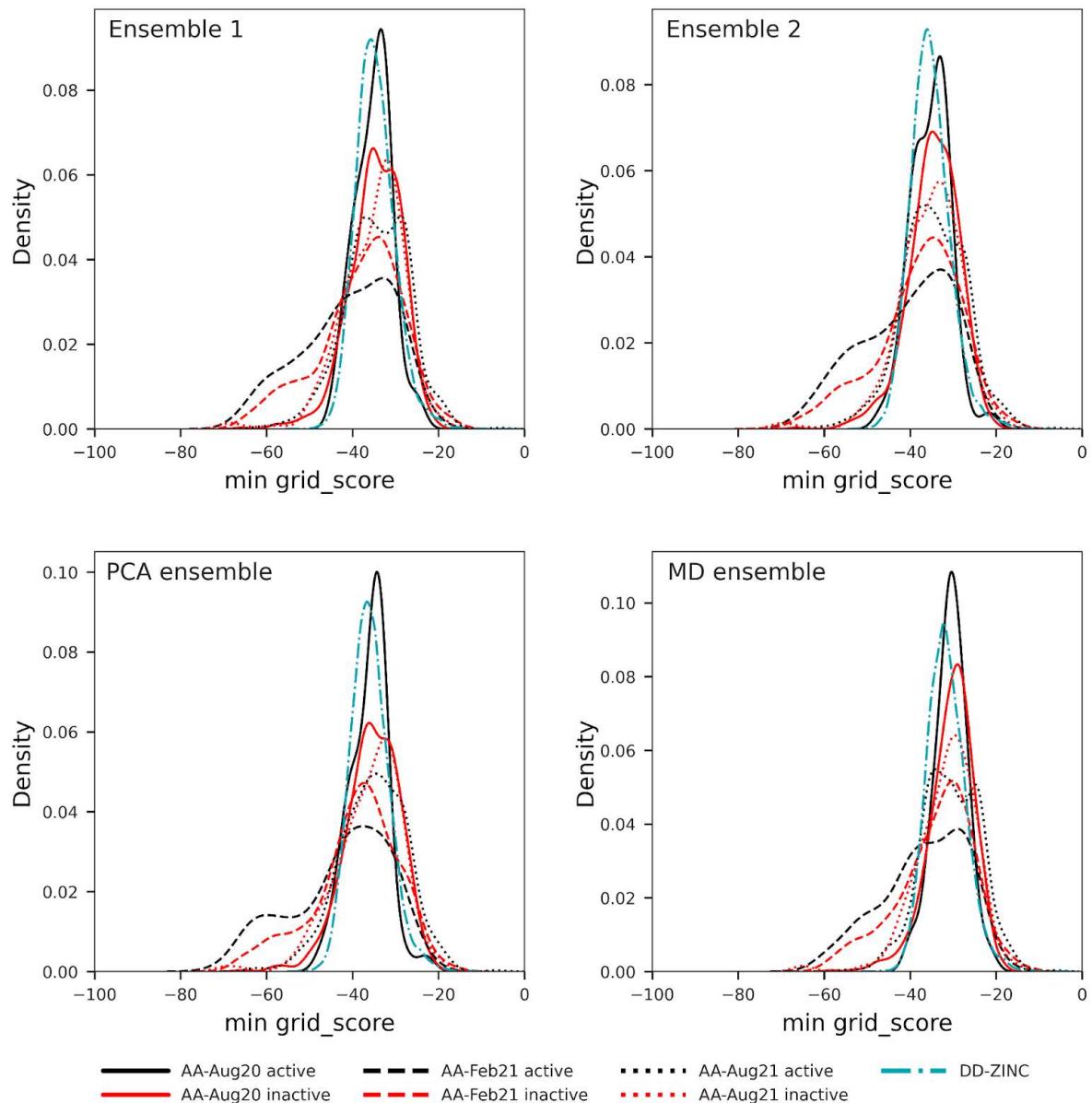


Рисунок А3. Распределение медианных оценок докинга активных и неактивных соединений библиотек AA-Aug20, AA-Feb21, AA-Aug21 и DD-ZINC в ансамблях 1, 2, PCA и МД

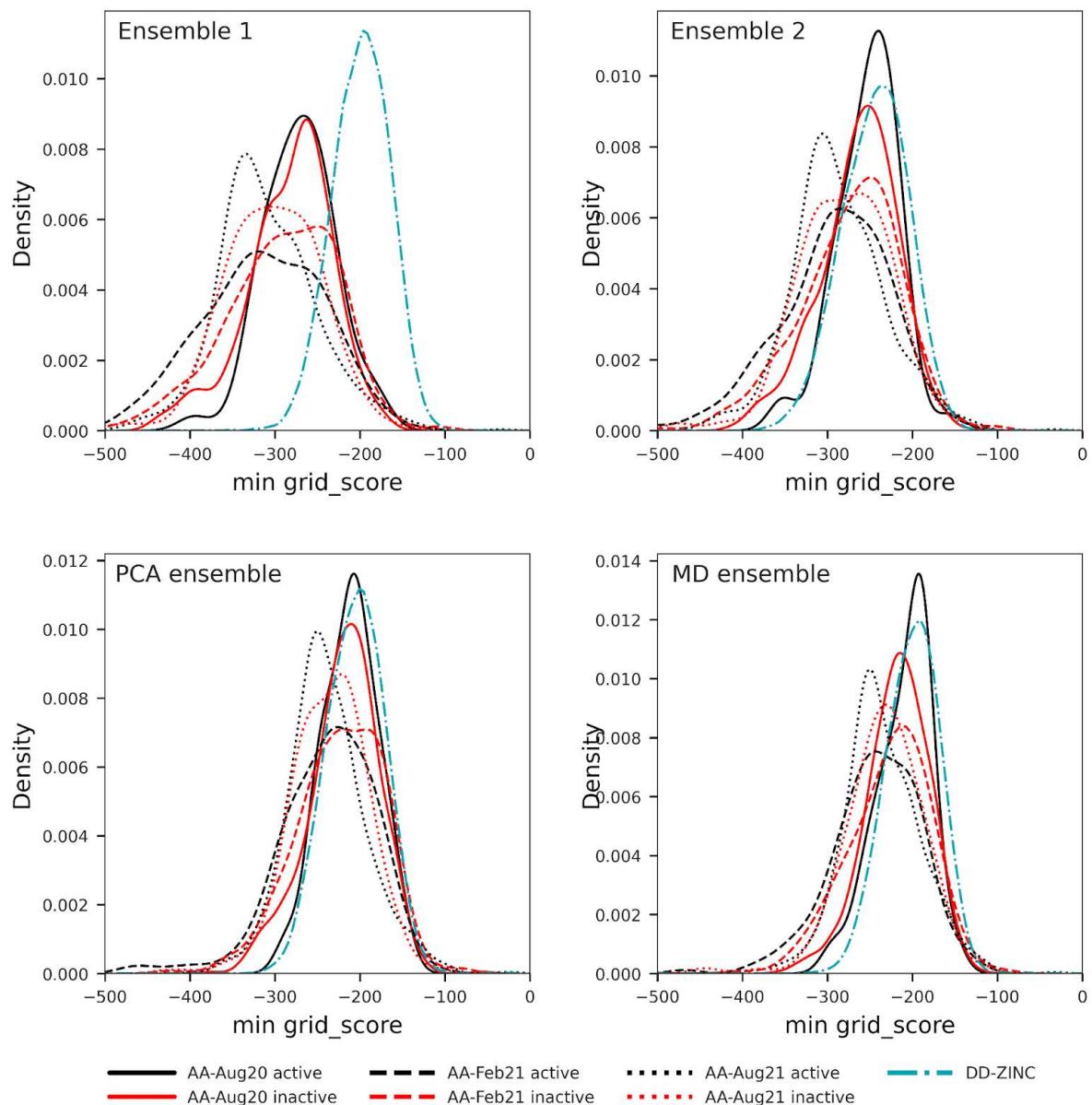


Рисунок А4. Распределение “евклидовых” оценок докинга активных и неактивных соединений библиотек AA-Aug20, AA-Feb21, AA-Aug21 и DD-ZINC в ансамблях 1, 2, PCA и МД

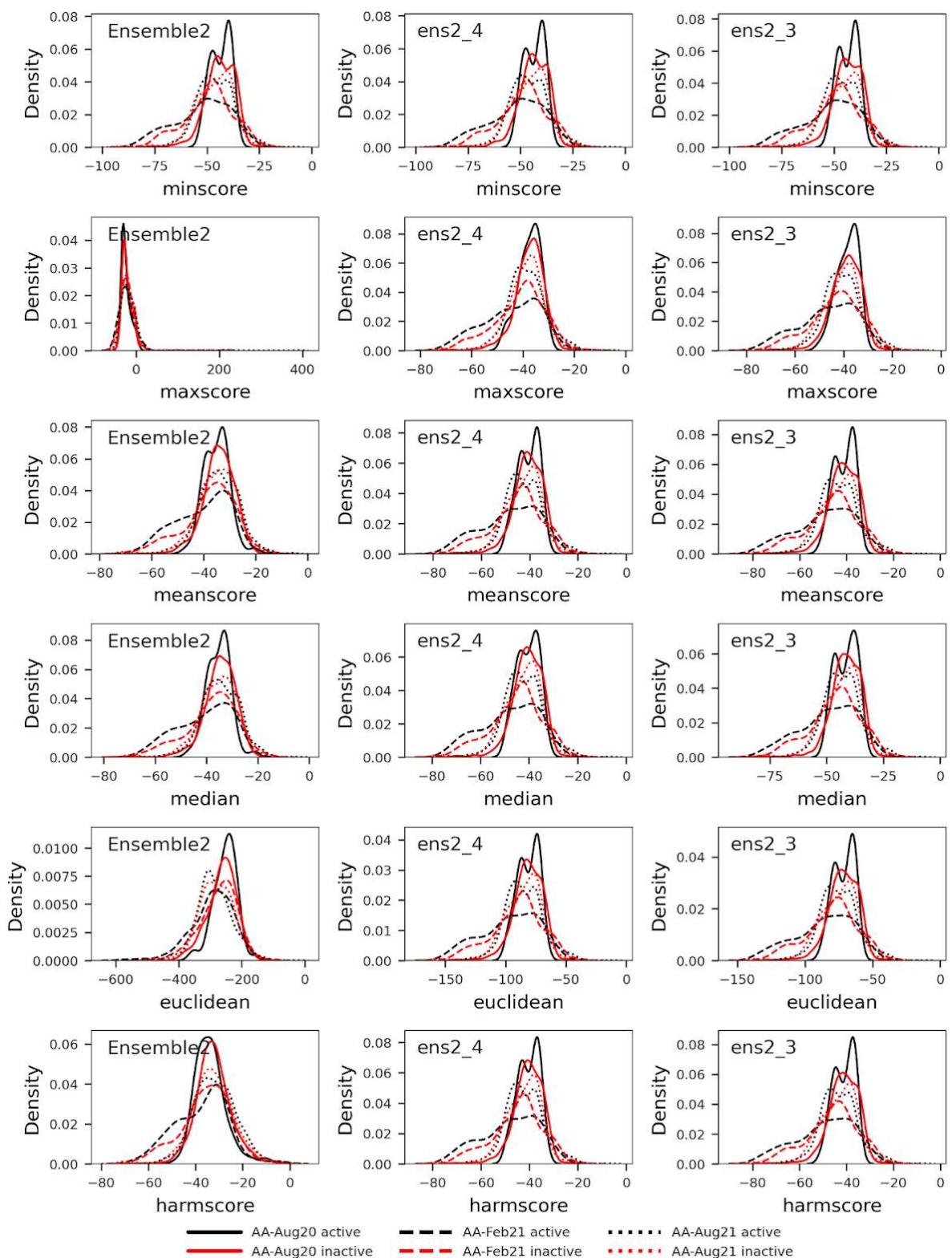


Рисунок А5. Распределение различных оценок докинга активных и неактивных соединений библиотек AA-Aug20, AA-Feb21, AA-Aug21 и DD-ZINC в ансамбли размеров 3, 4 и 6.

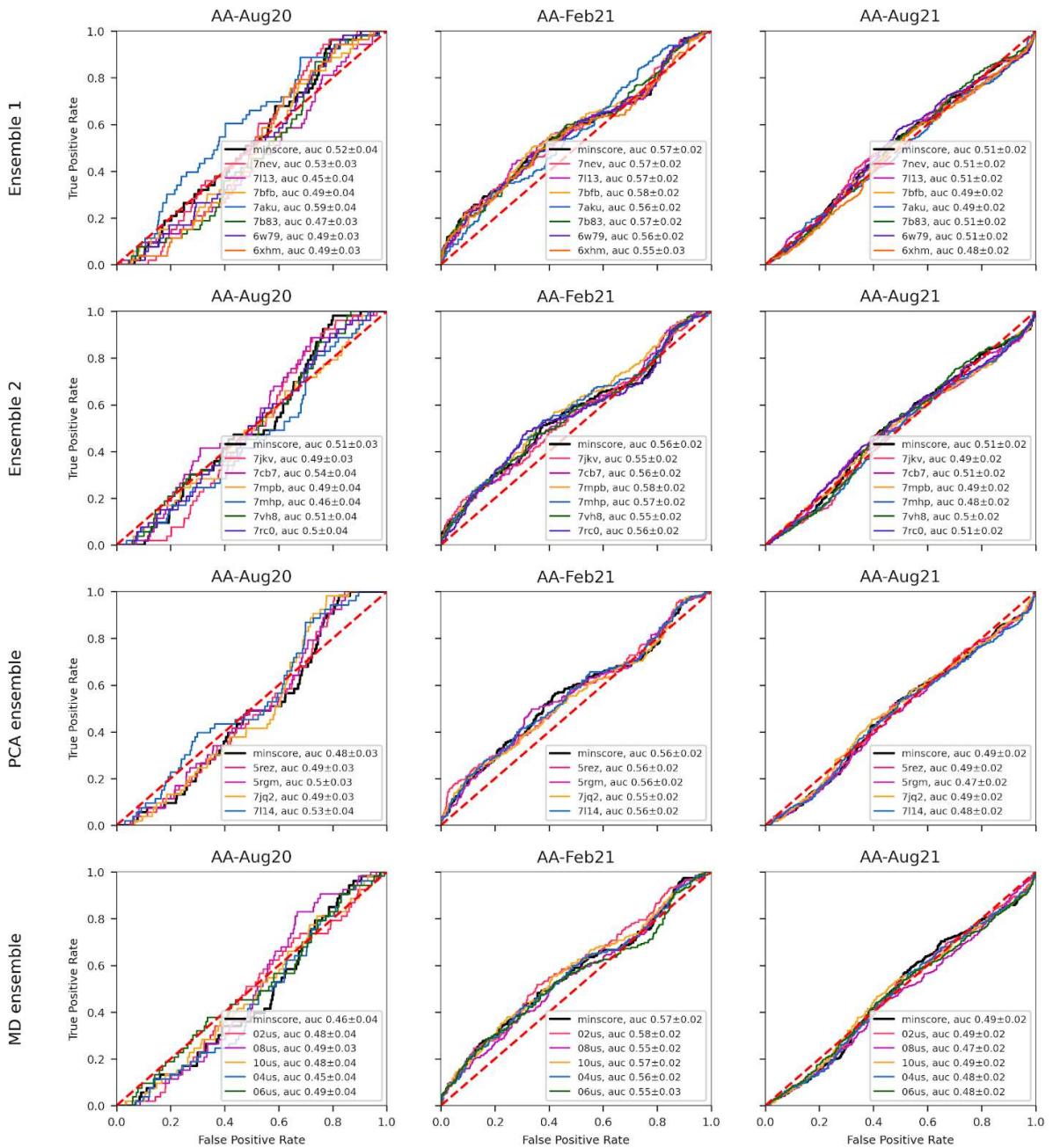


Рисунок А6. ROC-кривые для результатов докинга AA-Feb21, AA-Aug20 и AA-Aug21, основанные на ранжировании по оценкам докинга в отдельные структуры четырех ансамблей и минимальной оценке по ансамблю.

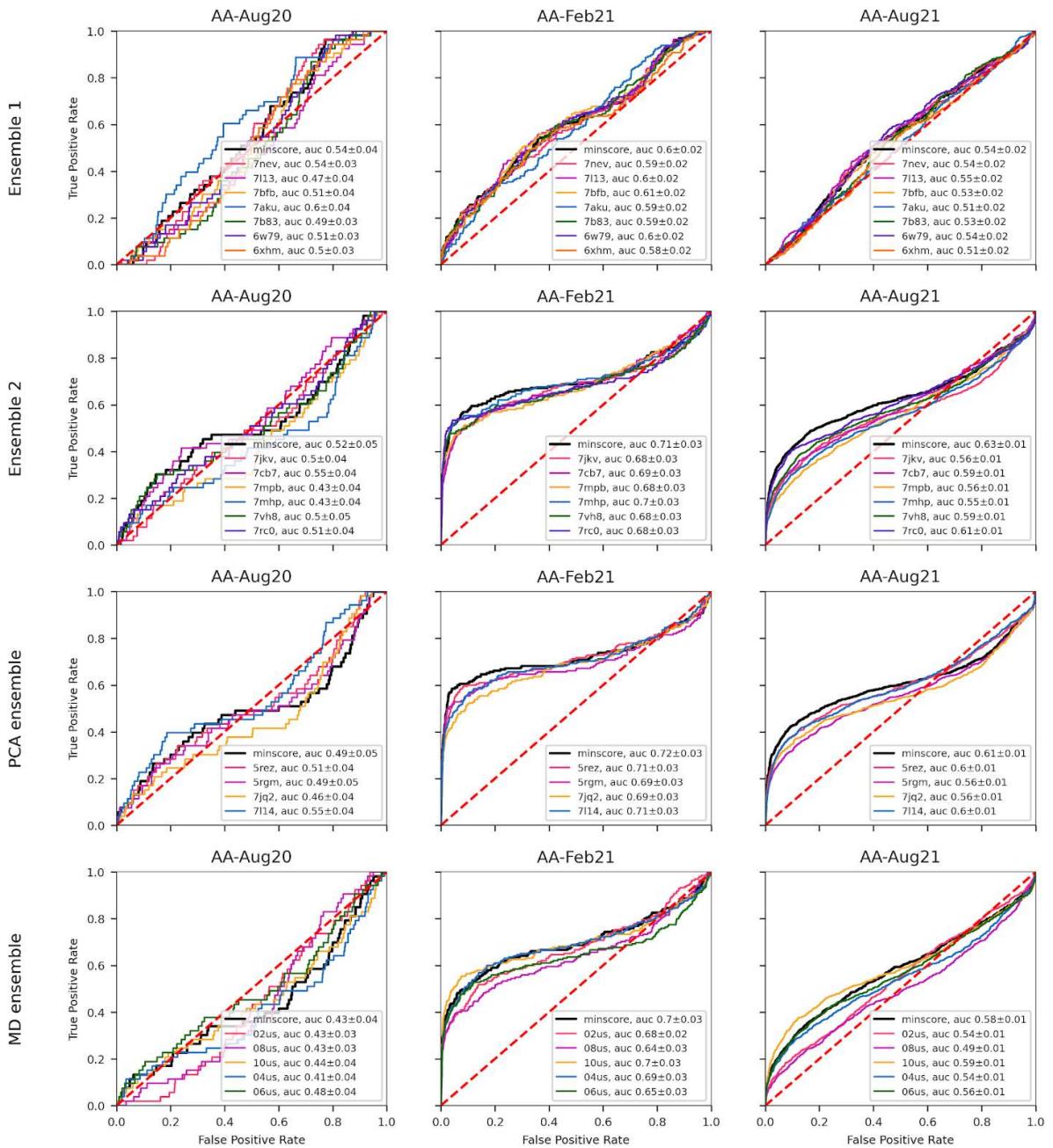


Рисунок A7. ROC-кривые для результатов докинга AA-Feb21, AA-Aug20 и AA-Aug21, обогащенных разнообразными соединениями из DD-ZINC, основанные на ранжировании по оценкам докинга в отдельные структуры четырех ансамблей и минимальной оценке по ансамблю.

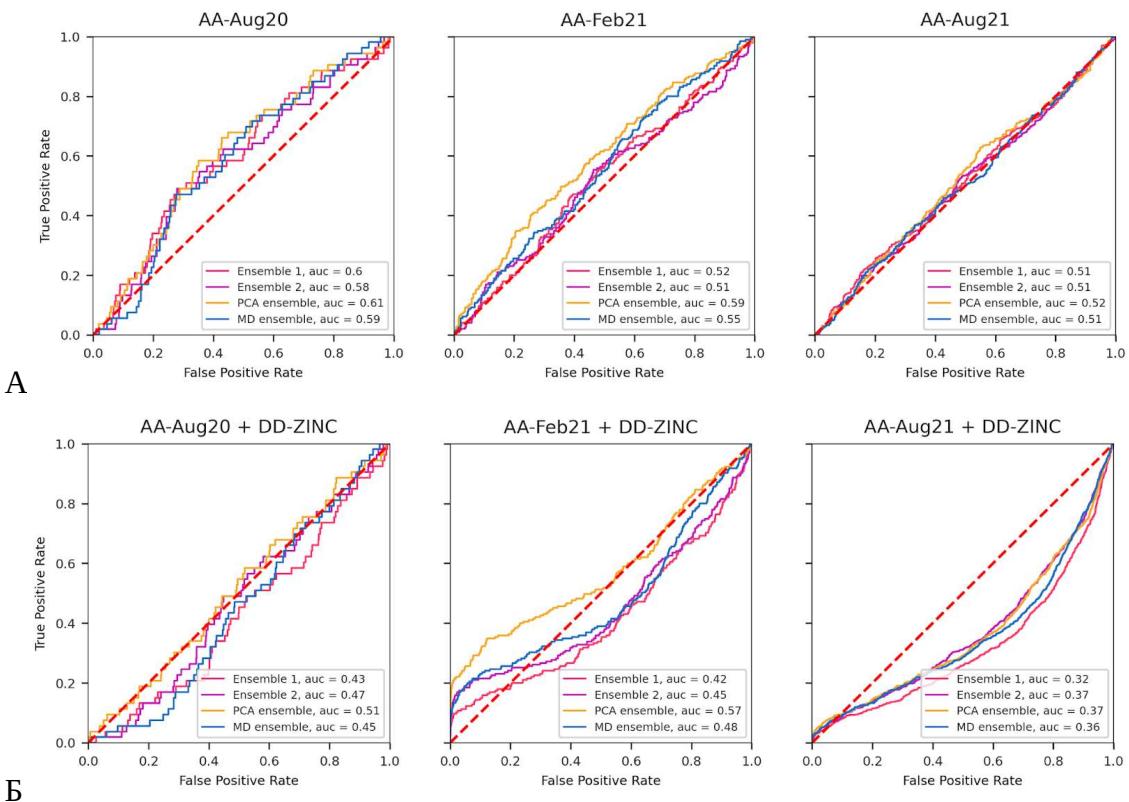


Рисунок A8. ROC-кривые для результатов докинга AA-Feb21, AA-Aug20 и AA-Aug21 (А), и тех же библиотек, обогащенных разнообразными соединениями из DD-ZINC, основанные на ранжировании по максимальным оценкам докинга (Б).

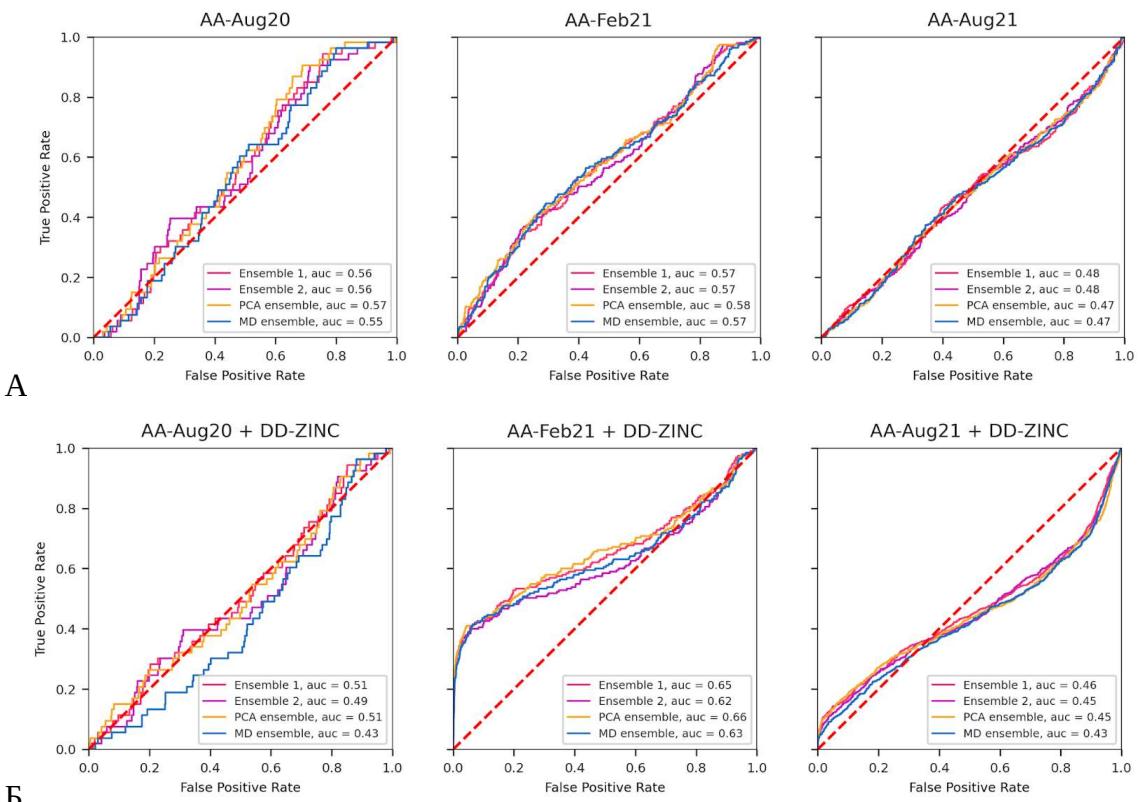


Рисунок A9. ROC-кривые для результатов докинга AA-Feb21, AA-Aug20 и AA-Aug21 (А), и тех же библиотек, обогащенных разнообразными соединениями из DD-ZINC, основанные на ранжировании по средним оценкам докинга (Б).

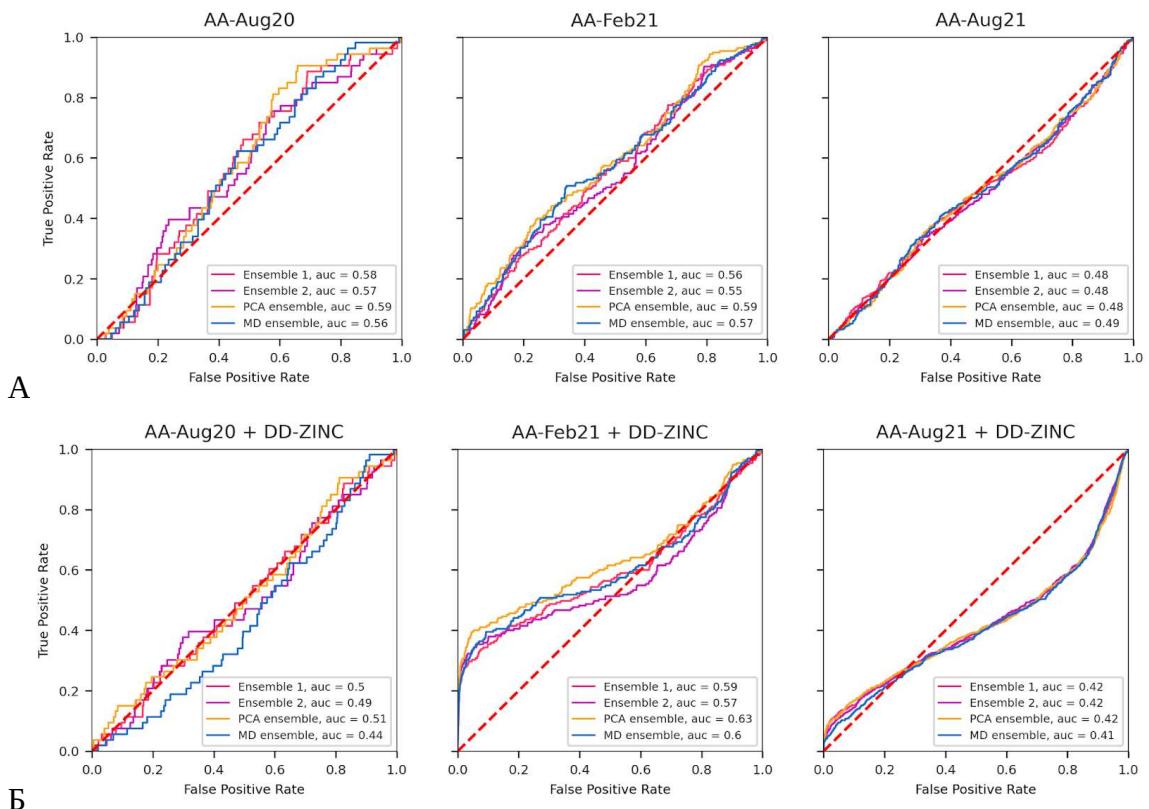


Рисунок A10. ROC-кривые для результатов докинга AA-Feb21, AA-Aug20 и AA-Aug21 (А), и тех же библиотек, обогащенных разнообразными соединениями из DD-ZINC, основанные на ранжировании по “гармоническим” оценкам докинга (Б).

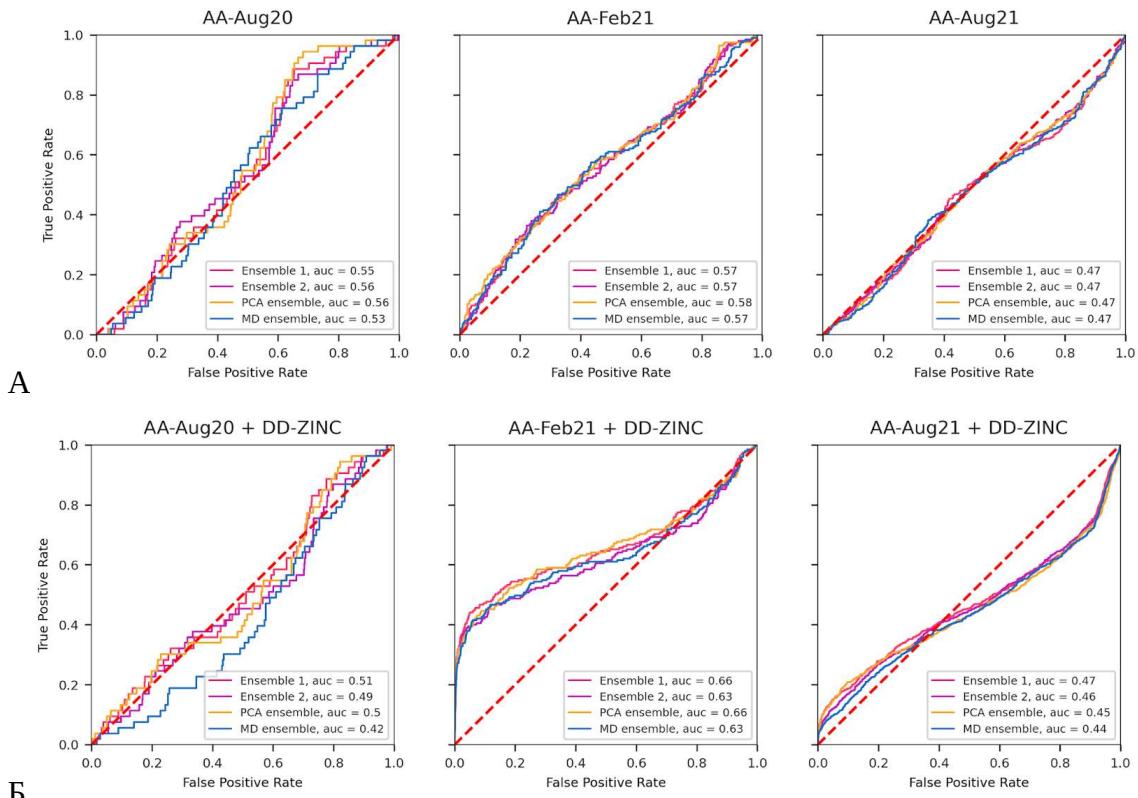


Рисунок A11. ROC-кривые для результатов докинга AA-Feb21, AA-Aug20 и AA-Aug21 (А), и тех же библиотек, обогащенных разнообразными соединениями из DD-ZINC, основанные на ранжировании по медианным оценкам докинга (Б).

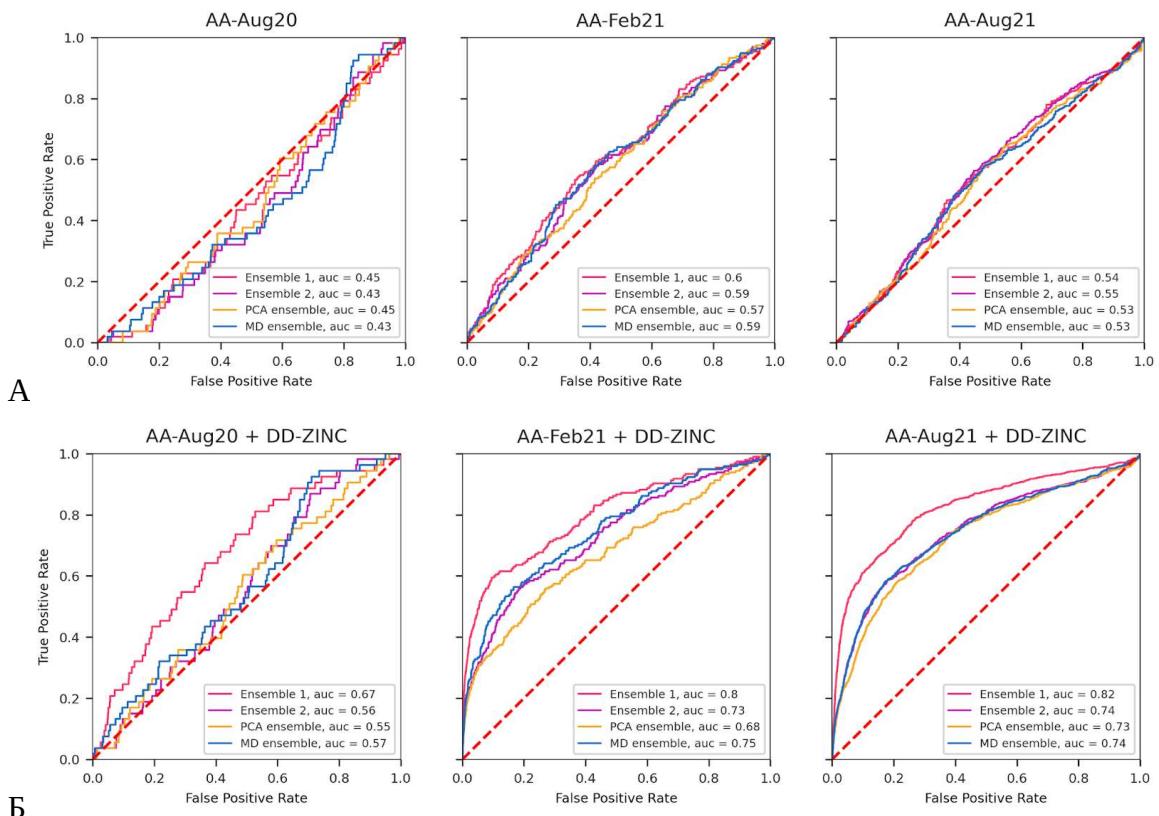


Рисунок A12. ROC-кривые для результатов докинга AA-Feb21, AA-Aug20 и AA-Aug21 (А), и тех же библиотек, обогащенных разнообразными соединениями из DD-ZINC, основанные на ранжировании по “евклидовым” оценкам докинга (Б).

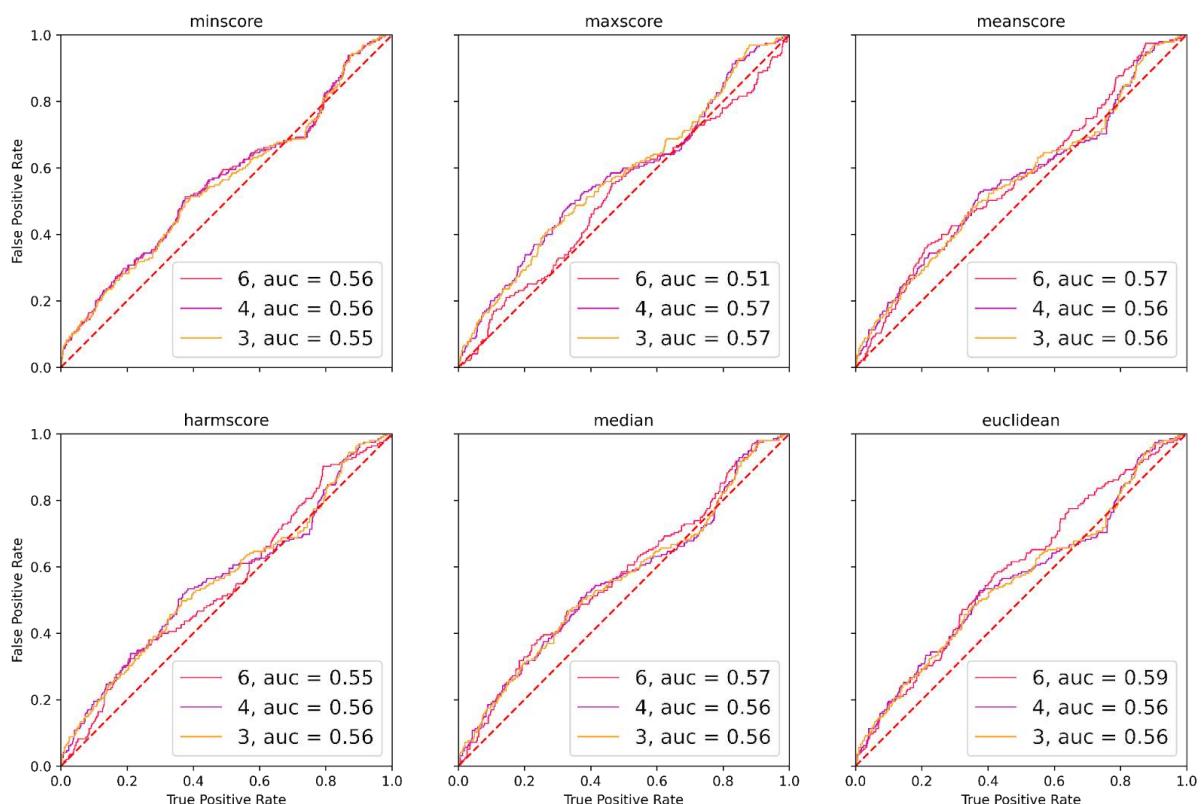


Рисунок A13. ROC-кривые для результатов докинга библиотеки AA-Feb21 в ансамбли разных размеров

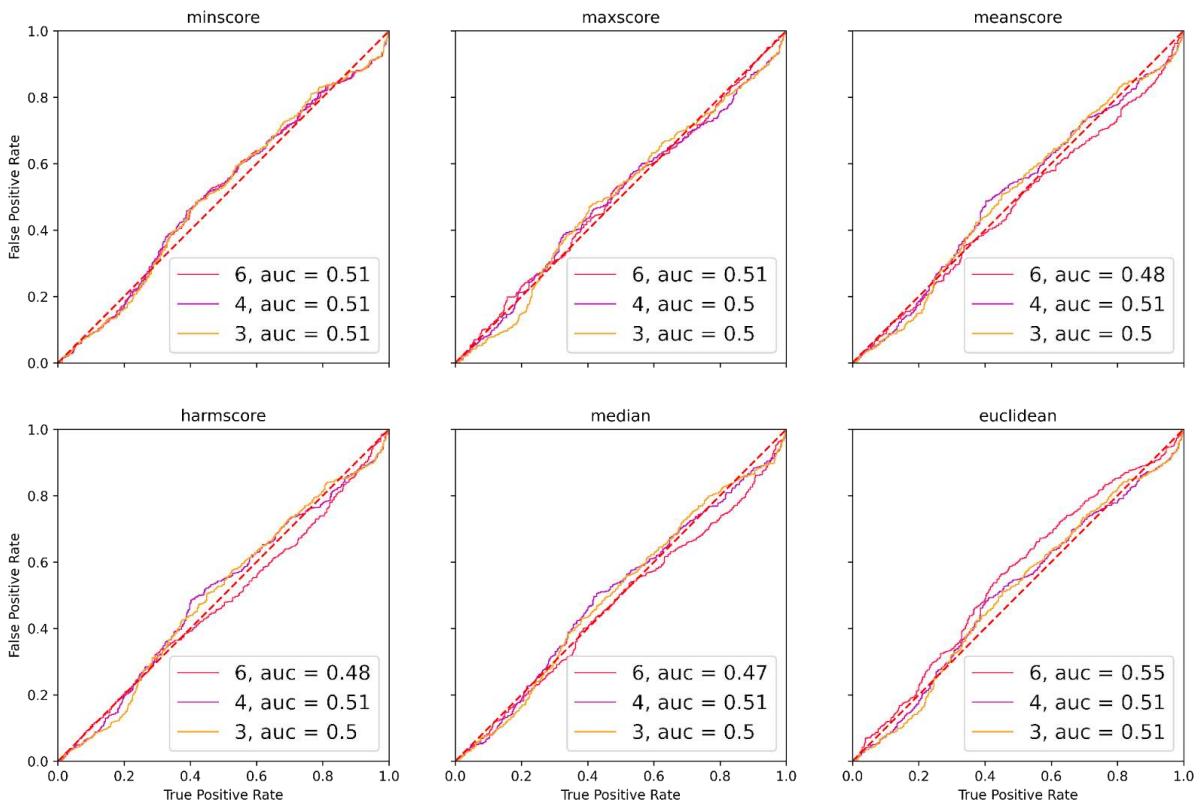


Рисунок A14. ROC-кривые для результатов докинга библиотеки AA-Aug21 в ансамблях разных размеров

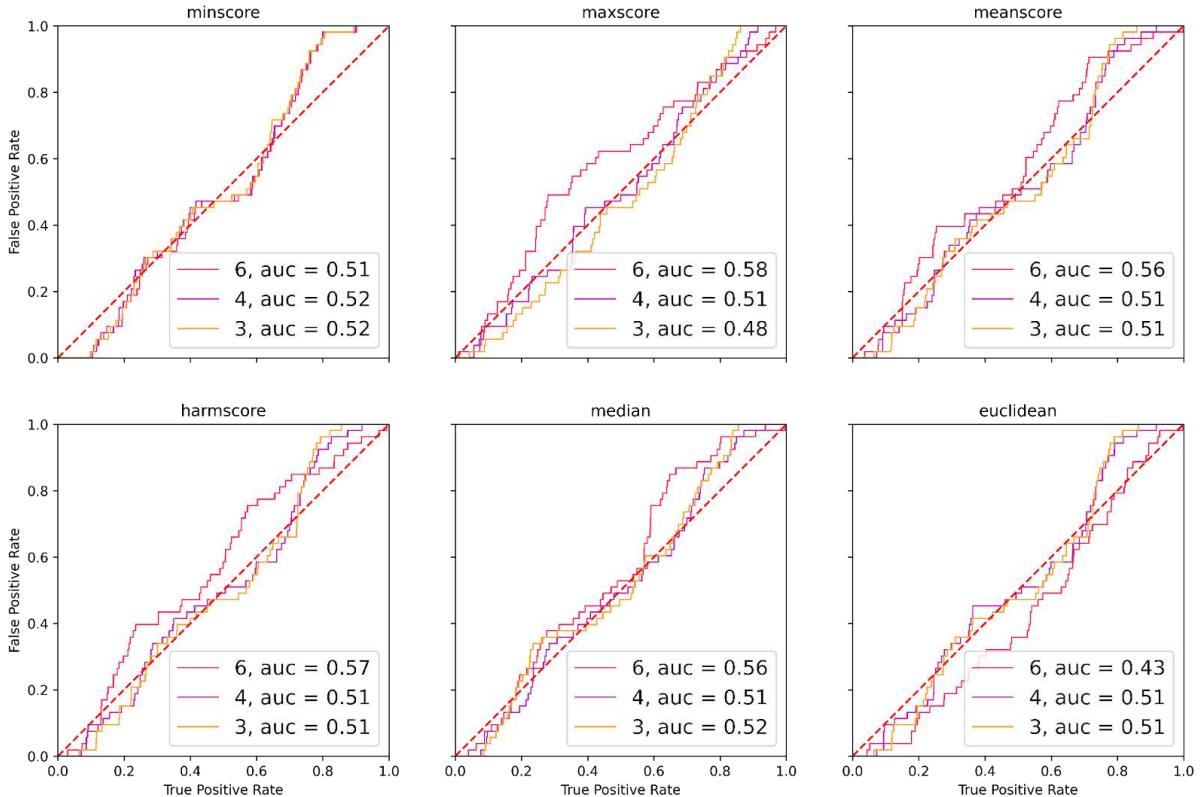


Рисунок A15. ROC-кривые для результатов докинга библиотеки AA-Aug20 в ансамблях разных размеров

## Приложение Б

### Реализация расчета фингерпринтов взаимодействия на Python3

```
from cresset import flare
from rdkit.Geometry.rdGeometry import Point3D
import urllib.request
from glob import glob

with open('concurrent_contacts.csv', "w") as output:
    output.write("pdb_id,contact_type,res1,atom1,res2,atom2,distance
\n")
    project = flare.main_window().project
for file in glob('****.pdb'):
    output = open('concurrent_contacts.csv', "a")
    output2 = open('concurrent_all_contacts.csv', "a")
    project.proteins.extend(flare.read_file(file))
    prep = flare.ProteinPrep()
    prep.proteins = project.proteins
    prep.cap_chains = True
    prep.start()
    prep.wait()
    ligand_residues = []
    for sequence in prep.proteins[0].sequences:
        if sequence.type == flare.Sequence.Type.Ligand:
            ligand_residues.extend(sequence)
    nat_seq_dict = {}
    for prot in prep.proteins:
        PDB_id = prot.title
        #Find natural amino acid sequences
        for seq_list in [prot.sequences for prot in prep.proteins]:
            #print(seq_list)
            sequences = []
            for seq in seq_list:
                if seq.type == flare.Sequence.Type.Natural:
                    sequences += [seq]
            nat_seq_dict[seq.molecule.title.lower()] = sequences
    ligands = []
    for ligand_residue in ligand_residues:
        ligand = project.ligands.append(ligand_residue)
        prep.proteins[0].residues.remove(ligand_residue)
        ligands.append(ligand)
    if len(ligands) == 0:
        print("No ligands found for {}".format(PDB_id))
```

```

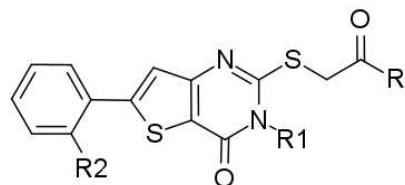
        for protein in project.proteins:
            del project.proteins[-1]
            continue
    def residues_near(what, nat_seq_dict, radius):
        residues = set()
        for al in what.atoms:
            if (al.atomic_number == 1):
                continue
            for i in
range(len(nat_seq_dict[what.protein.title.lower()]))) :
                for ap in nat_seq_dict[what.protein.title.lower()]
[i].atoms:
                    if (ap.atomic_number == 1):
                        continue
                    d = Point3D(*al.pos).Distance(Point3D(*ap.pos))
                    if (d < radius):
                        sting = str(ap.residue)+" "+str(ap.name)+" "
                        "+str(ap.atomic_number)+" "+str(al.name)+" "
                        "+str(al.atomic_number)+" "
                        "+ str(d)
                        residues.add(sting)
        return residues
    contacts1 = [residues_near(l, nat_seq_dict, 3.5) for l in
project.ligands]
    for i in range(len(contacts1)):
        string = ",".join(contacts1[i])
        string = str(PDB_id)+"," +string+"\n"
        output2.write(string)
    for lig,prot in [(lig,prot) for lig in project.ligands for prot
in project.proteins]:
        #print(prot.title, end=", ")
        for i in flare.contacts.h_bonds([lig, prot]):
            if i.strength == flare.Measurement.Strength.Strong:
                residues = [k[0].residue for k in i.points]
                if lig.residues[0] in residues:
                    string = ',' .join([prot.title, "h-bond"]+
[str(k[0].residue)+"," +str(k[0]) for k in i.points]+[str(i.value)])
                    string+="\n"
                    output.write(string)
    for ligand in project.ligands:
        del project.ligands[-1]
    for protein in project.proteins:
        del project.proteins[-1]
    output.close()
    output2.close()

```

## Приложение В

### Результаты экспериментального скрининга серии замещенных тиено[3,2-d]пиrimидинонов

Таблица В1. Результаты экспериментального скрининга ингибиторной активности против Mpro SARS-CoV-2 и против вируса SARS-CoV-2<sup>12</sup> серии замещенных тиено[3,2-d]пиrimидинонов.



Шифр	R	R1	R2	Mpro IC <sub>50</sub> (95% доверительный интервал), мкМ	SARS-CoV-2/ Vero EC <sub>50</sub> , мкМ
N3-алкилтиенопиrimидиноны (фенильная серия)					
IBS-E0680029	NHPh	CH <sub>3</sub>	H	74.3 (64.2–87.4)	>100
IBS-E0680043	N(Ph)Et	CH <sub>3</sub>	H	>100	>100
IBS-E0680114	N(Ph)Et	Et	H	>100	>100
IBS-E0680016	NHBn	CH <sub>3</sub>	H	>100	>100
IBS-E0680088	NHBn	Et	H	>100	>100
IBS-E0680026	NHCH <sub>2</sub> ) <sub>3</sub> Ph	CH <sub>3</sub>	H	>100	>100
IBS-E0680097	NH(CH <sub>2</sub> ) <sub>3</sub> Ph	Et	H	>100	>100
IBS-E0680032	NHC <sub>6</sub> H <sub>4</sub> -(4-Me)	CH <sub>3</sub>	H	≈100	>100
IBS-E0680042	NHC <sub>6</sub> H <sub>4</sub> -(4-nBu)	CH <sub>3</sub>	H	51.0 (42.0–64.5)	>100
IBS-E0680113	NHC <sub>6</sub> H <sub>4</sub> -(4-nBu)	Et	H	>100	>100
IBS-E0680065	NHC <sub>6</sub> H <sub>4</sub> -(4-CO <sub>2</sub> Me)	CH <sub>3</sub>	H	>100	>100
IBS-E0680133	NHC <sub>6</sub> H <sub>4</sub> -(4-CO <sub>2</sub> Me)	Et	H	21.2 (16.7–29.8)	>100
IBS-E0680066	NHC <sub>6</sub> H <sub>4</sub> -(4-CO <sub>2</sub> Et)	CH <sub>3</sub>	H	>100	>100
IBS-E0680134	NHC <sub>6</sub> H <sub>4</sub> -(4-CO <sub>2</sub> Et)	Et	H	>100	>100
IBS-E0680063	NHC <sub>6</sub> H <sub>4</sub> -(2-CO <sub>2</sub> Et)	CH <sub>3</sub>	H	≈100	>100
IBS-E0680132	NHC <sub>6</sub> H <sub>4</sub> -(2-CO <sub>2</sub> Et)	Et	H	>100	>100
IBS-E0680064	NHC <sub>6</sub> H <sub>4</sub> -(3-CO <sub>2</sub> Et)	CH <sub>3</sub>	H	92.6 (80.9–115)	>100

12 эксперименты по скринингу ингибиторной активности и определению IC<sub>50</sub> выполнены в Лаборатории противовирусных лекарственных средств ФГАНУ “ФНЦИРИП им. М.П. Чумакова РАН” (Институт полиомиелита) к.х.н В.И. Уваровой. Эксперименты по определению противовирусной активности выполнены в лаборатории полиомиелита и других энтеровирусов с референс-центром ВОЗ по надзору за полиомиелитом ФГАНУ “ФНЦИРИП им. М.П. Чумакова РАН” (Институт полиомиелита) д.б.н. Козловской Л.И.

Шифр	R	R1	R2	Mpro IC <sub>50</sub> (95% доверительный интервал), мкМ	SARS-CoV-2/ Vero EC <sub>50</sub> , мкМ
IBS-E0680135	NHC <sub>6</sub> H <sub>4</sub> -(2-F)	Et	H	>100	>100
IBS-E0680136	NHC <sub>6</sub> H <sub>4</sub> -(4-F)	Et	H	>100	>100
IBS-E0680070	NHC <sub>6</sub> H <sub>4</sub> -(2-Cl)	CH <sub>3</sub>	H	108 (93–147)	>100
IBS-E0680071	NHC <sub>6</sub> H <sub>4</sub> -(3-Cl)	CH <sub>3</sub>	H	>100	>100
IBS-E0680073	NHC <sub>6</sub> H <sub>4</sub> -(3-Br)	CH <sub>3</sub>	H	>100	>100
IBS-E0680074	NHC <sub>6</sub> H <sub>4</sub> -(4-Br)	CH <sub>3</sub>	H	>100	>100
IBS-E0680142	NHC <sub>6</sub> H <sub>4</sub> -(4-Br)	Et	H	>100	>100
IBS-E0680075	NHC <sub>6</sub> H <sub>4</sub> -(2-Cl-4-F)	CH <sub>3</sub>	H	>100	>100
IBS-E0680137	NHC <sub>6</sub> H <sub>4</sub> -(2-F-4-F)	Et	H	>100	>100
IBS-E0680167	NHC <sub>6</sub> H <sub>4</sub> -(4-OEt)	CH <sub>2</sub> CH=CH <sub>2</sub>	H	>100	>100
IBS-E0680171	NHC <sub>6</sub> H <sub>4</sub> -(3-Ac)	CH <sub>2</sub> CH=CH <sub>2</sub>	H	>100	>100
IBS-E0680144	NHC <sub>6</sub> H <sub>4</sub> -(2-CF <sub>3</sub> )	Et	H	>100	>100
IBS-E0680059	NHC <sub>6</sub> H <sub>4</sub> -(2-CN)	CH <sub>3</sub>	H	>100	>100
IBS-E0680128	NHC <sub>6</sub> H <sub>4</sub> -(2-CN)	Et	H	>100	>100
IBS-E0680060	NHC <sub>6</sub> H <sub>4</sub> -(3-CN)	CH <sub>3</sub>	H	66.8 (53.6–85.7)	>100
IBS-E0680129	NHC <sub>6</sub> H <sub>4</sub> -(3-CN)	Et	H	>100	>100
IBS-E0680061	NHC <sub>6</sub> H <sub>4</sub> -(4-CN)	CH <sub>3</sub>	H	24.6 (18.9–31.7)	>100
IBS-E0680047	NHC <sub>6</sub> H <sub>3</sub> -(2-MeO-4-MeO)	CH <sub>3</sub>	H	>100	>100
IBS-E0680164	NHC <sub>6</sub> H <sub>3</sub> -(2-MeO-4-MeO)	CH <sub>2</sub> CH=CH <sub>2</sub>	H	>100	>100
IBS-E0680119	NHC <sub>6</sub> H <sub>3</sub> -(2-MeO-5-MeO)	Et	H	>100	>100
IBS-E0680166	NHC <sub>6</sub> H <sub>3</sub> -(3-MeO-4-MeO)	CH <sub>2</sub> CH=CH <sub>2</sub>	H	≈100	>100
IBS-E0680055	NHC <sub>6</sub> H <sub>3</sub> -(3,4-O(CH <sub>2</sub> ) <sub>2</sub> O)	CH <sub>3</sub>	H	90 (81–107)	>100
IBS-E0680125	NHC <sub>6</sub> H <sub>3</sub> -(3,4-O(CH <sub>2</sub> ) <sub>2</sub> O)	Et	H	>100	>100
IBS-E0680169	NHC <sub>6</sub> H <sub>3</sub> -(3,4-O(CH <sub>2</sub> ) <sub>2</sub> O)	CH <sub>2</sub> CH=CH <sub>2</sub>	H	21.9 (16.1–30.3)	>100
IBS-E0680049	NHC <sub>6</sub> H <sub>3</sub> -(3-MeO-5-MeO)	CH <sub>3</sub>	H	>100	>100
IBS-E0680120	NHC <sub>6</sub> H <sub>3</sub> -(3-MeO-5-MeO)	Et	H	>100	>100
IBS-E0680165	NHC <sub>6</sub> H <sub>3</sub> -(3-MeO-5-MeO)	CH <sub>2</sub> CH=CH <sub>2</sub>	H	>100	>100
IBS-E0680051	NHC <sub>6</sub> H <sub>2</sub> -(3-MeO-4-MeO-5-MeO)	CH <sub>3</sub>	H	>100	>100
IBS-E0680085	NHC <sub>6</sub> H <sub>2</sub> -(2-MeO-4-Cl-5-Me)	CH <sub>3</sub>	H	19.7 (13.6–35.6)	>25

Шифр	R	R1	R2	Mpro IC <sub>50</sub> (95% доверительный интервал), мкМ	SARS-CoV-2/ Vero EC <sub>50</sub> , мкМ
IBS-E0680084	NHC <sub>6</sub> H <sub>2</sub> -(2-MeO-5-Cl)	CH <sub>3</sub>	H	>100	>100
IBS-E0680083	NHC <sub>6</sub> H <sub>2</sub> -(3-Cl-4-MeO)	CH <sub>3</sub>	H	>100	>100
N3-алкилтиенопиримидиноны (бензильная серия)					
IBS-E0680017	NHCH <sub>2</sub> C <sub>6</sub> H <sub>4</sub> -(4-Me)	CH <sub>3</sub>	H	>100	>100
IBS-E0680089	NHCH <sub>2</sub> C <sub>6</sub> H <sub>4</sub> -(4-Me)	Et	H	>100	>100
IBS-E0680021	NHCH <sub>2</sub> C <sub>6</sub> H <sub>4</sub> -(4-F)	CH <sub>3</sub>	H	>100	>100
IBS-E0680093	NHCH <sub>2</sub> C <sub>6</sub> H <sub>4</sub> -(4-F)	Et	H	>100	>100
IBS-E0680023	NHCH <sub>2</sub> C <sub>6</sub> H <sub>4</sub> -(4-Cl)	CH <sub>3</sub>	H	80 (60–130)	>100
IBS-E0680022	NHCH <sub>2</sub> C <sub>6</sub> H <sub>4</sub> -(2-Cl)	CH <sub>3</sub>	H	>100	>100
IBS-E0680094	NHCH <sub>2</sub> C <sub>6</sub> H <sub>4</sub> -(2-Cl)	Et	H	>100	>100
IBS-E0680018	NHCH <sub>2</sub> C <sub>6</sub> H <sub>4</sub> -(3-MeO)	CH <sub>3</sub>	H	>25	>25
IBS-E0680090	NHCH <sub>2</sub> C <sub>6</sub> H <sub>4</sub> -(3-MeO)	Et	H	>100	>100
IBS-E0680019	NHCH <sub>2</sub> C <sub>6</sub> H <sub>4</sub> -(4-MeO)	CH <sub>3</sub>	H	>100	>100
IBS-E0680091	NHCH <sub>2</sub> C <sub>6</sub> H <sub>4</sub> -(4-MeO)	Et	H	>100	>100
IBS-E0680092	NHCH <sub>2</sub> C <sub>6</sub> H <sub>3</sub> -(2-MeO-3-MeO)	Et	H	53.9 (40.2–74.2)	>100
IBS-E0680020	NHCH <sub>2</sub> C <sub>6</sub> H <sub>3</sub> -(2-MeO-3-MeO)	CH <sub>3</sub>	H	>100	>100
IBS-E0680025	NH(CH <sub>2</sub> ) <sub>2</sub> C <sub>6</sub> H <sub>3</sub> -(3-MeO-4-MeO)	CH <sub>3</sub>	H	>100	>100
N3-арилтиенопиримидиноны					
IBS-E0183441	NHPh	pEt-C <sub>6</sub> H <sub>4</sub>	OEt	>80	>80
IBS-E0183437	NHC <sub>6</sub> H <sub>4</sub> -(2-F)	pEt-C <sub>6</sub> H <sub>4</sub>	OEt	>80	>80
IBS-E0183435	NHC <sub>6</sub> H <sub>4</sub> -(3-F)	pEt-C <sub>6</sub> H <sub>4</sub>	OEt	>100	>100
IBS-E0183438	NHC <sub>6</sub> H <sub>4</sub> -(4-F)	pEt-C <sub>6</sub> H <sub>4</sub>	OEt	>100	>100
IBS-E0183443	NHC <sub>6</sub> H <sub>4</sub> -(3-Me)	pEt-C <sub>6</sub> H <sub>4</sub>	OEt	17.2 (12.8–23.3)	>80
IBS-E0183434	NHC <sub>6</sub> H <sub>3</sub> -(3-Me-5-Me)	pEt-C <sub>6</sub> H <sub>4</sub>	OEt	>80	>80
IBS-E0183439	NHC <sub>6</sub> H <sub>3</sub> -(3-Me-4-Me)	pEt-C <sub>6</sub> H <sub>4</sub>	OEt	>80	>80
IBS-E0183440	NHC <sub>6</sub> H <sub>3</sub> -(2-Et)	pEt-C <sub>6</sub> H <sub>4</sub>	OEt	>80	>80
IBS-E0183442	NHC <sub>6</sub> H <sub>3</sub> -(3-Et)	pEt-C <sub>6</sub> H <sub>4</sub>	OEt	75.8 (65.6–90.2)	>100
IBS-E0183436	NHC <sub>6</sub> H <sub>3</sub> -(4-Et)	pEt-C <sub>6</sub> H <sub>4</sub>	OEt	>80	>80

## Приложение Г

### Метрики качества моделей машинного обучения

Таблица Г1. Модели, построенные на данных CHEMBL4495582

Набор данных	Модель	ROC AUC обучение	ROC AUC std обучение	ROC AUC контроль	ROC AUC std контроль	F1 обучение	F1 std обучение	F1 контроль	F1 std контроль	Точность контроль	Точность std контроль	Полнота std контроль	Полнота std контроль
Set2	Catboost	1.00	0.00	0.59	0.03	0.83	0.03	0.05	0.03	0.63	0.43	0.03	0.02
Set2	LR	0.77	0.01	0.57	0.07	0.13	0.01	0.08	0.02	0.04	0.01	0.44	0.09
Set2	RF	0.96	0.01	0.64	0.05	0.44	0.05	0.13	0.02	0.10	0.02	0.20	0.04
Set2	SVM	0.32	0.01	0.45	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Set3	Catboost	0.93	0.01	0.62	0.06	0.24	0.04	0.01	0.02	0.11	0.31	0.00	0.01
Set3	LR	0.76	0.01	0.55	0.07	0.13	0.01	0.07	0.02	0.04	0.01	0.40	0.10
Set3	RF	0.98	0.01	0.62	0.06	0.55	0.05	0.12	0.06	0.11	0.05	0.15	0.07
Set3	SVM	0.32	0.01	0.47	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Set4	Catboost	1.00	0.00	0.67	0.04	0.89	0.03	0.10	0.06	0.64	0.34	0.05	0.03
Set4	LR	0.97	0.00	0.69	0.03	0.33	0.01	0.14	0.02	0.09	0.01	0.44	0.05
Set4	RF	0.99	0.00	0.69	0.06	0.65	0.05	0.19	0.05	0.17	0.04	0.21	0.06
Set4	SVM	0.88	0.01	0.68	0.02	0.18	0.01	0.12	0.01	0.07	0.01	0.55	0.06
Set5	Catboost	1.00	0.00	0.67	0.06	0.89	0.02	0.11	0.06	0.49	0.18	0.06	0.04
Set5	LR	0.96	0.00	0.69	0.03	0.30	0.01	0.14	0.02	0.08	0.01	0.47	0.05
Set5	RF	0.98	0.00	0.69	0.06	0.56	0.05	0.19	0.05	0.15	0.05	0.26	0.06
Set5	SVM	0.75	0.01	0.63	0.04	0.14	0.01	0.09	0.01	0.05	0.01	0.53	0.06
Set6	Catboost	0.99	0.01	0.65	0.06	0.58	0.05	0.08	0.03	0.71	0.27	0.04	0.01

Набор данных	Модель	ROC AUC обучение	ROC AUC std обучение	ROC AUC контроль	ROC AUC std контроль	F1 обучение	F1 std обучение	F1 контроль	F1 std контроль	Точность контроль	Точность std контроль	Полнота контроль	Полнота std контроль
Set6	LR	0.96	0.00	0.69	0.03	0.31	0.01	0.15	0.02	0.09	0.01	0.47	0.06
Set6	RF	0.98	0.00	0.68	0.06	0.57	0.06	0.20	0.04	0.17	0.05	0.24	0.04
Set6	SVM	0.82	0.01	0.67	0.03	0.15	0.01	0.10	0.01	0.06	0.01	0.54	0.05
Set7	Catboost	0.99	0.00	0.58	0.03	0.76	0.03	0.08	0.05	0.61	0.37	0.04	0.02
Set7	LR	0.76	0.01	0.62	0.05	0.13	0.01	0.10	0.02	0.06	0.01	0.51	0.08
Set7	RF	0.94	0.01	0.63	0.05	0.37	0.05	0.14	0.04	0.10	0.04	0.24	0.06
Set7	SVM	0.58	0.01	0.59	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Set8	Catboost	1.00	0.00	0.70	0.06	0.73	0.02	0.10	0.06	0.60	0.34	0.05	0.03
Set8	LR	0.98	0.00	0.71	0.03	0.37	0.01	0.18	0.02	0.11	0.01	0.47	0.06
Set8	RF	0.97	0.01	0.71	0.06	0.45	0.05	0.21	0.05	0.15	0.04	0.35	0.05
Set8	SVM	0.86	0.01	0.64	0.03	0.20	0.01	0.10	0.01	0.06	0.01	0.50	0.06
Set9	Catboost	1.00	0.00	0.65	0.05	0.87	0.03	0.13	0.07	0.55	0.22	0.07	0.04
Set9	LR	0.98	0.00	0.71	0.03	0.39	0.01	0.18	0.02	0.11	0.01	0.46	0.06
Set9	RF	0.97	0.01	0.70	0.06	0.48	0.05	0.22	0.04	0.17	0.04	0.33	0.05
Set9	SVM	0.91	0.01	0.68	0.03	0.21	0.01	0.12	0.01	0.07	0.01	0.54	0.07

Таблица Г2. Модели, построенные на данных AA-Feb21

Набор данных	Модель	ROC AUC обучение	ROC AUC std обучение	ROC AUC контроль	ROC AUC std контроль	F1 обучение	F1 std обучение	F1 контроль	F1 std контроль	Precision1 контроль	Precision1 std контроль	Recall1 контроль	Recall1 std контроль
Set4	LR	0.85	0.01	0.68	0.04	0.17	0.01	0.12	0.01	0.07	0.01	0.54	0.05
Set4	RF	0.99	0.00	0.69	0.06	0.67	0.05	0.19	0.04	0.18	0.04	0.21	0.05
Set4	SVM	0.80	0.01	0.68	0.04	0.10	0.01	0.08	0.01	0.04	0.01	0.75	0.07
Set5	LR	0.84	0.01	0.68	0.04	0.17	0.01	0.12	0.01	0.07	0.01	0.55	0.05
Set5	RF	0.98	0.00	0.69	0.06	0.53	0.05	0.19	0.05	0.15	0.05	0.28	0.06
Set5	SVM	0.85	0.01	0.69	0.02	0.17	0.01	0.12	0.01	0.07	0.00	0.56	0.03
Set6	LR	0.84	0.01	0.68	0.04	0.17	0.01	0.12	0.01	0.07	0.01	0.55	0.05
Set6	RF	0.98	0.00	0.68	0.06	0.57	0.06	0.20	0.04	0.17	0.05	0.24	0.04
Set6	SVM	0.82	0.01	0.67	0.03	0.15	0.01	0.10	0.01	0.06	0.01	0.54	0.05
Set7	LR	0.71	0.01	0.63	0.05	0.06	0.00	0.06	0.00	0.03	0.00	1.00	0.00
Set7	RF	0.93	0.01	0.63	0.05	0.35	0.05	0.15	0.05	0.10	0.04	0.27	0.07
Set7	SVM	0.58	0.01	0.59	0.04	0.05	0.00	0.05	0.01	0.03	0.00	0.42	0.20
Set8	LR	0.86	0.01	0.70	0.04	0.18	0.01	0.13	0.01	0.07	0.01	0.56	0.06
Set8	RF	0.96	0.01	0.71	0.06	0.42	0.05	0.21	0.04	0.14	0.03	0.37	0.04
Set8	SVM	0.89	0.01	0.71	0.02	0.19	0.01	0.13	0.01	0.07	0.01	0.57	0.07
Set9	LR	0.86	0.01	0.70	0.04	0.18	0.01	0.13	0.01	0.07	0.01	0.56	0.06
Set9	RF	0.97	0.01	0.70	0.06	0.48	0.05	0.22	0.04	0.17	0.04	0.33	0.05
Set9	SVM	0.91	0.01	0.68	0.03	0.21	0.01	0.12	0.01	0.07	0.01	0.54	0.07

## Приложение Д

## Аминокислотные последовательности белка NS1 ортофлавивирусов.

TBEV-NS1/1-351	1	.....	VGCAVDTERMELRQGEGLVWREVSWEYDNYAYYPETFGALASAIKETFEQSGCGVVPQRNLEAMWRSVTE	73
5YXA/1-177	1	.....	HHHHHHVCGSVDFSKKETRQGTGTVFVYNDV	71
5GS-B/1-349	1	.....	HHHHHHVCGSVDFSKKETRQGTGTVFVYNDV	71
5Y3/1-177	1	.....	HHHHHHVCGSVDFSKKETRQGTGTVFVYNDV	71
5GS-A/1-352	1	.....	HVCGSVDFSKKETRQGTGTVFVYNDV	74
5XBY/1-176	1	.....	YFQSNADVGCVSDFSKKETRQGTGTVFVYNDV	74
5KGK-A/1-350	1	.....	YFQSNADVGCVSDFSKKETRQGTGTVFVYNDV	80
5KGK-B/1-354	1	.....	NADVGCVSDFSKKETRQGTGTVFVYNDV	76
4O6B-A/1-319	1	.....	ADSGCVVS.....	72
4O6B-B/1-323	1	.....	DSGCVVSWNKNEKLKGSGFIFTDNVHTWTEQYKQFOPESPSKLASAIQKAHEEGICRGISVSRLENLMWKQI	72
DENV1-SN1/1-351	1	.....	SGCVINWKORELKCGSGFIFTVNHTWTEQYKQFADSPRLSAAIGKAWEEGCVGIRSATRLEN1MWKQISNE	73
4OIG/1-175	1	.....		
5O36/1-178	1	.....		
5O19/1-176	1	.....		
4OIE/1-177	1	.....		
4TPL-B/1-352	1	HHHHHSSGVLDLGTENLYFQSNADTGCACIDISROELRQGSGGVFIHNDVEAWMDRKYYPETPOGLAKI	95	
4O6C/1-338	1	.....	QKAHKEGVCGLRSVSRLEHOMWWEAVKDE	75
4O6D-B/1-335	1	.....	QKAHKEGVCGLRSVSRLEHOMWWEAVKDE	79
4O6D-A/1-338	1	.....	QKAHKEGVCGLRSVSRLEHOMWWEAVKDE	81
4OIV/1-177	1	.....		
4TPL-A/1-342	1	.....	GTENLYFQSNADTGCACIDISROELRQGSGGVFIHNDVEAWMDRKYYPETPOGLAKI	85
TBEV-NS1/1-351	74	LNALAILEGEANLTVVVDFKDPDPTDYRGVPGLLKKGKD1KVSWSHMSIWSIPEAPRREMVGTE6QSECPLERKRTGTVFTVAEFGVGL	167	
5YXA/1-177	72	LNALAILEENGVOLTVVVGSKVNPWMRGP.....	165	
5GS-B/1-349	72	LNALAILEENGVOLTVVVGSKVNPWMRGP.....	165	
5Y3/1-177	72	LNALAILEENGVOLTVVVGSKVNPWMRGP.....	165	
5GS-A/1-352	75	LNALAILEENGVOLTVVVGSKVNPWMRGP.....	168	
5XBY/1-176	81	LNALAILEENGVOLTVVVGSKVNPWMRGP.....	166	
5KGK-A/1-350	81	LNALAILEENGVOLTVVVGSKVNPWMRGP.....	166	
5KGK-B/1-354	77	LNALAILEENGVOLTVVVGSKVNPWMRGP.....	170	
4O6B-A/1-319	73	LNALIILESENEVKT1MTGD1KG1MQAGK-RSLRQP.....	136	
4O6B-B/1-323	75	LNALIILESENEVKT1MTGD1KG1MQAGK-RSLRQP.....	144	
DENV1-SN1/1-351	74	LNALIILEENDMKFTVVVGDSVGLAQKKG1MQAGK-RSLRQP.....	167	
4OIG/1-175	74	LNALAILEGEANLTVVVDFKDPDPTDYRGVPGLLKKGKD1KVSWSHMSIWSIPEAPRREMVGTE6QSECPLERKRTGTVFTVAEFGVGL	167	
5O36/1-178	74	LNALAILEENGVOLTVVVGSKVNPWMRGP.....	165	
5O19/1-176	74	LNALAILEENGVOLTVVVGSKVNPWMRGP.....	165	
4OIE/1-177	74	LNALAILEENGVOLTVVVGSKVNPWMRGP.....	165	
4TPL-B/1-352	96	LNTLLEKENGVDLSVVVEKQEGMYKSAP-KRLTAT.....	168	
4O6C/1-338	76	LNTLLEKENGVDLSVVVEKQEGMYKSAP-KRLTAT.....	154	
4O6D-B/1-335	80	LNTLLEKENGVDLSVVVEKQEGMYKSAP-KRLTAT.....	151	
4O6D-A/1-338	82	LNTLLEKENGVDLSVVVEKQEGMYKSAP-KRLTAT.....	154	
4OIV/1-177	82	LNTLLEKENGVDLSVVVEKQEGMYKSAP-KRLTAT.....	154	
4TPL-A/1-342	88	LNTLLEKENGVDLSVVVEKQEGMYKSAP-KRLTAT.....	158	
TBEV-NS1/1-351	168	LDFRQEPTTHECDTGVMAAVKNGMA1HTDQSLWWMRS-MKNDTGTYI1VELLVTDLRNSWPSHAST1TIDNAWDVDELF1FASLAGPRTWYNNRPI	261	
5YXA/1-177	1	.....	SVDCGAI1GAAANGVNGKSAHSGPSLWMMSSHEVNWTM1HTELT1-DYKECEWPLHTH1-CTGSEWNSCMFMRP1GCPVSHSNR1TREGYR	261
5GS-B/1-349	166	LKVREDESYLSCDPAVIGTAVKGKEAVHSDLGYWIES-EKNDTWRKLRAHLL1-EMKTCEWPSHTLWTDG1IEESDL11PKSLAGPLBHNNTREGYR	258	
5Y3/1-177	1	.....	SVDCGAI1GAAANGVNGKSAHSGPSLWMMSSHEVNWTM1HTELT1-DYKECEWPLHTH1-CTGSEWNSCMFMRP1GCPVSHSNR1TREGYR	258
5GS-A/1-352	169	LKVREDESYLSCDPAVIGTAVKGKEAVHSDLGYWIES-EKNDTWRKLRAHLL1-EMKTCEWPSHTLWTDG1IEESDL11PKSLAGPLBHNNTREGYR	261	
5XBY/1-176	1	.....	SVDCGAI1GAAANGVNGKSAHSGPSLWMMSSHEVNWTM1HTELT1-DYKECEWPLHTH1-CTGSEWNSCMFMRP1GCPVSHSNR1TREGYR	261
5KGK-A/1-350	169	LKVREDESYLSCDPAVIGTAVKGKEAVHSDLGYWIES-EKNDTWRKLRAHLL1-EMKTCEWPSHTLWTDG1IEESDL11PKSLAGPLBHNNTREGYR	261	
5KGK-B/1-354	171	LKVREDESYLSCDPAVIGTAVKGKEAVHSDLGYWIES-EKNDTWRKLRAHLL1-EMKTCEWPSHTLWTDG1IEESDL11PKSLAGPLBHNNTREGYR	263	
4O6B-A/1-319	139	LKLKEKDQFDCKSLKMSAAIKDNRVADMYWIES-ALNDTWRK1IEKASFI-EVKVNCWPKSHTLWSNGVLESEMI1PKNLAGPVSQHNYRPGYH	231	
4O6B-B/1-323	145	LKLKE-KODCKSLKMSAAIKDNRVADMYWIES-ALNDTWRK1IEKASFI-EVKVNCWPKSHTLWSNGVLESEMI1PKNLAGPVSQHNYRPGYH	235	
DENV1-SN1/1-351	168	LKLRLDSYTCVCDPRLMSAAIKDOSKAVADMYWIES-EKNETWKLARASFI-EVKVTCWPKSHTLWSNGVLESEMI1PKI1KYYGPI	260	
4OIG/1-175	1	.....	VCDHRLMSAAIKDOSKAVADMYWIES-EKNETWKLARASFI-EVKVTCWPKSHTLWSNGVLESEMI1PKI1KYYGPI	260
5O36/1-178	1	.....	DECDAIG1IATGAVKGHVAWSDLSWYIES-RYNDTWRK1ERAVFG-EVKVSCTWPEHTHTLGWDVSEEL1IPTI1AGPKGKHNRREGYR	84
5O19/1-176	1	.....	DECDAIG1IATGAVKGHVAWSDLSWYIES-RYNDTWRK1ERAVFG-EVKVSCTWPEHTHTLGWDVSEEL1IPTI1AGPKGKHNRREGYR	85
4OIE/1-177	1	.....	TECDSK1IMGATVAKGNLAIHSDLSWYIES-RLNDTWRK1ERAVLG-EVKVSCTWPEHTHTLGWDG1ILESDMI1PVTLAGPRSNHNRREPGYK	86
4TPL-B/1-352	169	LKVRESNTTECDSK1IATGAVKNLAIHSDLSWYIES-RLNDTWRK1ERAVLG-EVKVSCTWPEHTHTLGWDG1ILESDMI1PVTLAGPRSNHNRREPGYK	261	
4O6C/1-338	155	LKVRESNTTECDSK1IATGAVKNLAIHSDLSWYIES-RLNDTWRK1ERAVLG-EVKVSCTWPEHTHTLGWDG1ILESDMI1PVTLAGPRSNHNRREPGYK	247	
4O6D-B/1-335	152	LKVRESNTTECDSK1IATGAVKNLAIHSDLSWYIES-RLNDTWRK1ERAVLG-EVKVSCTWPEHTHTLGWDG1ILESDMI1PVTLAGPRSNHNRREPGYK	244	
4O6D-A/1-338	155	LKVRESNTTECDSK1IATGAVKNLAIHSDLSWYIES-RLNDTWRK1ERAVLG-EVKVSCTWPEHTHTLGWDG1ILESDMI1PVTLAGPRSNHNRREPGYK	247	
4OIV/1-177	1	.....	TECDSK1IATGAVKNLAIHSDLSWYIES-RLNDTWRK1ERAVLG-EVKVSCTWPEHTHTLGWDG1ILESDMI1PVTLAGPRSNHNRREPGYK	86
4TPL-A/1-342	159	LKVRESNTTECDSK1IATGAVKNLAIHSDLSWYIES-RLNDTWRK1ERAVLG-EVKVSCTWPEHTHTLGWDG1ILESDMI1PVTLAGPRSNHNRREPGYK	251	
TBEV-NS1/1-351	262	EQVKCPWVKYTP1IR1V1RECCEPGT1VT1NAK1DKRKGASVRS1TESSKG1IPEWCCRACTMPVFTFRTGDCWYAME1RPMV-HDQGGLVRSVMTA	361	
5YXA/1-177	87	VOTNQGPWMPQVPL1EV1KRECPGTVPSVWVSDN1DGRGKSTR1TDD1GOK1IPEWCCRSCTMPVFSHGSDGCPWMP1RPMKTSDSLHVRSPWTA	177	
5GS-B/1-349	259	TOMKGPWPHSEEL1IREECPGTVK1VHEET1CGRGPSSLRTTASGRV1IPEWCCRCSTMPVPSLFRAKDGCVWYGM1IRPKRPESNLVRSPWTA	349	
5Y3/1-177	87	TOMKGPWPHSEEL1IREECPGTVK1VHEET1CGRGPSSLRTTASGRV1IPEWCCRCSTMPVPSLFRAKDGCVWYGM1IRPKRPESNLVRSPWTA	177	
5GS-A/1-352	262	TOMKGPWPHSEEL1IREECPGTVK1VHEET1CGRGPSSLRTTASGRV1IPEWCCRCSTMPVPSLFRAKDGCVWYGM1IRPKRPESNLVRSPWTA	352	
5XBY/1-176	87	TOMKGPWPHSEEL1IREECPGTVK1VHEET1CGRGPSSLRTTASGRV1IPEWCCRCSTMPVPSLFRAKDGCVWYGM1IRPKRPESNLVRSPWTA	176	
5KGK-A/1-350	260	TQVKGPWPHSEEL1IREECPGTVK1VHEET1CGRGPSSLRTTASGRV1IPEWCCRCSTMPVPSLFRAKDGCVWYGM1IRPKRPESNLVRSPWTA	350	
5KGK-B/1-354	264	TQVKGPWPHSEEL1IREECPGTVK1VHEET1CGRGPSSLRTTASGRV1IPEWCCRCSTMPVPSLFRAKDGCVWYGM1IRPKRPESNLVRSPWTA	354	
4O6B-A/1-319	232	T1OTGPWPHLGK1LMDMFD1FDG1TTV1VTVEDCGRGPSSLRTTASGR1IPEWCCRCSTLPP1R1YRGEDCGCVWYGM1IRPK1KEEVLNVS1	319	
4O6B-B/1-323	236	T1OTGPWPHLGK1LMDMFD1FDG1TTV1VTVEDCGRGPSSLRTTASGR1IPEWCCRCSTLPP1R1YRGEDCGCVWYGM1IRPK1KEEVLNVS1	323	
DENV1-SN1/1-351	261	T1OTGPWPHLGK1LMDF1FDG1TTV1VTVEDCGRGPSSLRTTASGR1IPEWCCRCSTLPP1R1YRGEDCGCVWYGM1IRPK1KEEVLNVS1	361	
4OIG/1-175	85	TQTAGPWHLGK1LMDFD1FDG1TTV1VTVEDCGRGPSSLRTTASGR1IPEWCCRCSTLPP1R1YRGEDCGCVWYGM1IRPKVKEEVLNVS1	175	
5O36/1-178	86	TQTAGPWHLGK1LMDFD1FDG1TTV1VTVEDCGRGPSSLRTTASGR1IPEWCCRCSTLPP1R1YRGEDCGCVWYGM1IRPKVMDHDETTLVRSPQDAFK	178	
5O19/1-176	86	TQTAGPWHLGK1LMDFD1FDG1TTV1VTVEDCGRGPSSLRTTASGR1IPEWCCRCSTLPP1R1YRGEDCGCVWYGM1IRPKVMDHDETTLVRSPQDAFK	176	
4OIE/1-177	87	TQTAGPWHLGK1LMDFD1FDG1TTV1VTVEDCGRGPSSLRTTASGR1IPEWCCRCSTLPP1R1YRGEDCGCVWYGM1IRPKVMDHDETTLVRSPQDAFK	177	
4TPL-B/1-352	262	TQNOGPWDEGRV1E1FDYD1CPG1TTV1VTVEDCGRGPSSLRTTASGR1IPEWCCRCSTLPP1R1YRGEDCGCVWYGM1IRPKRDEKTLVQSVNA	362	
4O6C/1-338	248	TQNOGPWDEGRV1E1FDYD1CPG1TTV1VTVEDCGRGPSSLRTTASGR1IPEWCCRCSTLPP1R1YRGEDCGCVWYGM1IRPKRDEKTLVQSVNA	338	
4O6D-B/1-335	245	TQNOGPWDEGRV1E1FDYD1CPG1TTV1VTVEDCGRGPSSLRTTASGR1IPEWCCRCSTLPP1R1YRGEDCGCVWYGM1IRPKRDEKTLVQSVNA	335	
4O6D-A/1-338	248	TQNOGPWDEGRV1E1FDYD1CPG1TTV1VTVEDCGRGPSSLRTTASGR1IPEWCCRCSTLPP1R1YRGEDCGCVWYGM1IRPKRDEKTLVQSVNA	338	
4OIV/1-177	87	TQNOGPWDEGRV1E1FDYD1CPG1TTV1VTVEDCGRGPSSLRTTASGR1IPEWCCRCSTLPP1R1YRGEDCGCVWYGM1IRPKRDEKTLVQSVNA	177	
4TPL-A/1-342	252	TQNOGPWDEGRV1E1FDYD1CPG1TTV1VTVEDCGRGPSSLRTTASGR1IPEWCCRCSTLPP1R1YRGEDCGCVWYGM1IRPKRDEKTLVQSVNA	342	

Рисунок Д1. Выравнивание аминокислотных последовательностей белка NS1 ВКЭ (штамм Абсеттаров) и других ортофлавивирусов (последовательности, извлечённые из pdb-файлов), окрашенное по идентичности. Чем идентичнее остатки, тем окраска интенсивнее

Таблица Д1. Эпидемиологически значимые ортофлавивирусы

Вид вируса	Название вируса	Сокращение (рус.)	Сокращение (англ.)	GenBank ID
Вирусы, переносимые клещами				
<i>Orthoflavivirus kyasanurense</i>	Вирус Киассанурской лесной болезни		KFDV	AY323490
<i>Orthoflavivirus langatense</i>	Вирус лангат		LGTВ	AF253419
<i>Orthoflavivirus loupingi</i>	Вирус энцефаломиелита овец		LIV	Y07863
<i>Orthoflavivirus omskense</i>	Вирус Омской геморрагической лихорадки	ВОГЛ	OHFV	AY193805
<i>Orthoflavivirus powassanense</i>	Вирус Повассан		POWV	L06436
<i>Orthoflavivirus encephalitidis</i>	Вирус клещевого энцефалита – европейский подтип	ВКЭ	TBEV-Eur	U27495
	Вирус клещевого энцефалита – дальневосточный подтип	ВКЭ	TBEV-FE	X07755
	Вирус клещевого энцефалита – сибирский подтип	ВКЭ	TBEV-Sib	L40361
Вирусы, переносимые комарами				
<i>Orthoflavivirus denguei</i>	Вирус денге 1	ДЕНВ1	DENV-1	U88536
	Вирус денге 2	ДЕНВ2	DENV-2	U87411
	Вирус денге 3	ДЕНВ3	DENV-3	M93130
	Вирус денге 4	ДЕНВ4	DENV-4	AF326573
<i>Orthoflavivirus japonicum</i>	Вирус японского энцефалита	ВЯЭ	JEV	M18370
<i>Orthoflavivirus murrayense</i>	Вирус энцефалита долины Муррея		MVEV	AF161266
<i>Orthoflavivirus louisense</i>	Вирус энцефалита Сент-Луис		SLEV	DQ525916
<i>Orthoflavivirus nilense</i>	Вирус лихорадки Западного Нила	ВЛЗН	WNV	M12294
<i>Orthoflavivirus zikaense</i>	Вирус Зика	ЗИКВ	ZIKV	AY632535
<i>Orthoflavivirus flavi</i>	Вирус желтой лихорадки	ВЖЛ	YFV	X03700

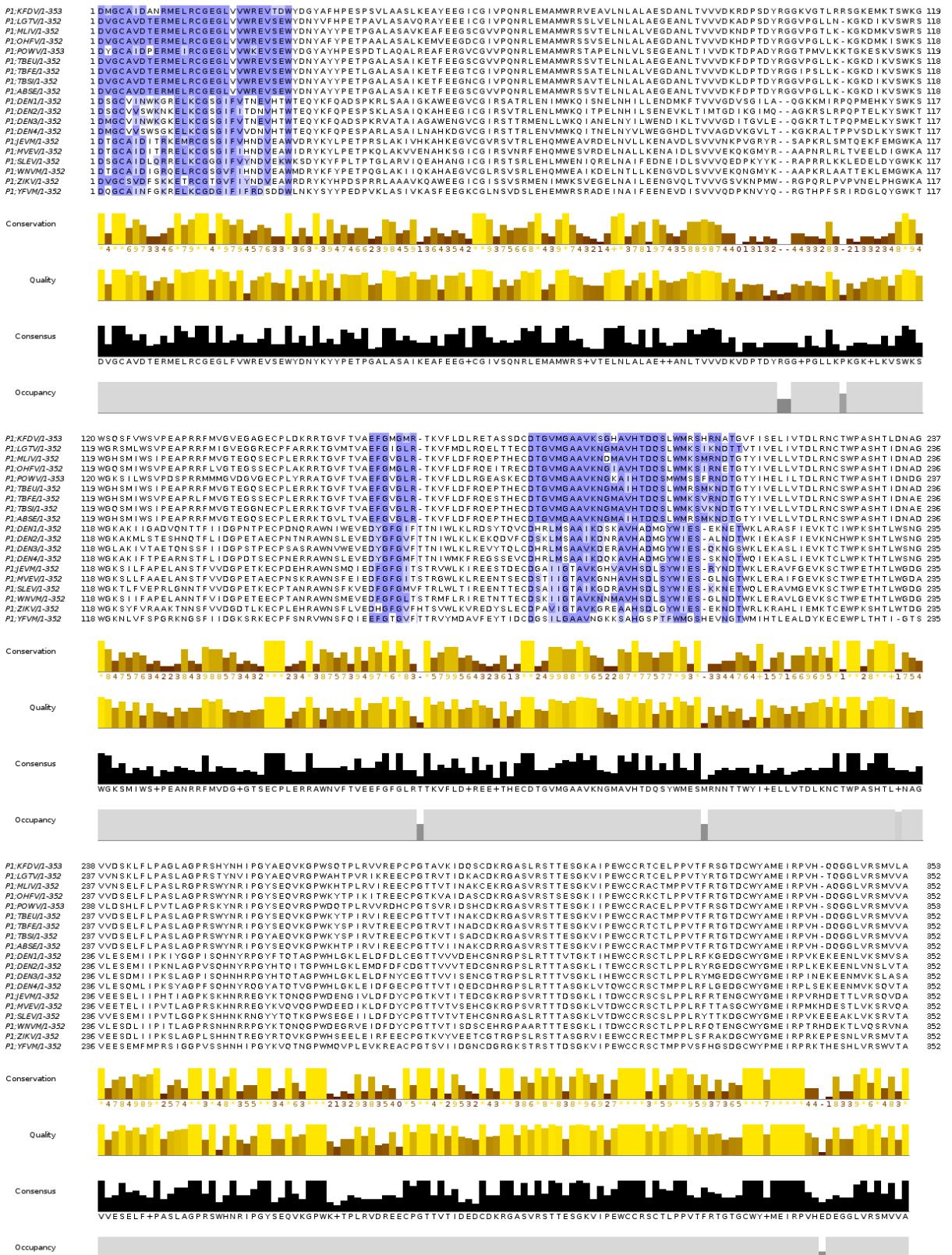


Рисунок Д2. Выравнивание аминокислотных последовательностей белка NS1 эпидемиологически значимых ортофлавивирусов. Окрашивание остатков, формирующих потенциальный сайт связывания, аналогично рисунку Д1.

## Приложение E

### Хиты виртуального скрининга потенциальных лигандов белка NS1 ортофлавивирусов

Таблица E1. Хиты виртуального скрининга потенциальных лигандов белка NS1 ортофлавивирусов, переносимых комарами

ID молекулы	SMILES	Молекулярный вес	cLogP
ZINC000342133588	C[C@H]1CN(CC#CCN2C[C@H](C)[C@H](CO)C2)C[C@H]1CO	280,41	0,911
ZINC000067804507	NC[C@H]1Cc2cc(F)cc(-c3cc4c(cc3C(N)=O)OCO4)c2O1	330,314	1,8443
ZINC000257358609	OC[C@H](O)[C@H](O)[C@H](O)[C@H]1Nc2cccc2S1	257,309	-0,7972
ZINC00000896569	NS(=O)(=O)c1cc2c(cc1Cl)NCNS2(=O)=O	297,743	-0,3795
ZINC000065407363	Cc1cccc(C)c1OC1CN(c2nc3c(c(=O)[nH]2)CCNCC3)C1	340,426	1,8501
ZINC000038777417	OCC1=C(CO)CC2=C(C1)CC(CO)=C(CO)C2	252,309	0,551
ZINC000330229468	CN(c1cc(N(C)[C@H]2CCCC[C@H]2O)ncn1)[C@H]1CCCC[C@H]1O	334,462	1,9992
ZINC00009348652	O=C(O)CNC(=O)C[C@H]1SC(NNC2=NC(=O)CS2)=NC1=O	345,359	-3,2911
ZINC000958528707	N[C@H](c1nccs1)C1(O)C[C@H]2CC[C@H](C1)S2(=O)=O	288,391	-0,497
ZINC000017112820	CC[C@H]1(C)Cc2c(sc3[nH]c(SCC(N)=O)nc(=O)c23)CO1	339,439	0,6348
ZINC00004521387	O[C@H]([C@H](O)[C@H]1COC2(CCCCC2)O1)[C@H]1COC2(CCCCC2)O1	342,43	0,9482
ZINC00000279315	Nc1ccc(c1)[C@@]13CCN4CC5=CCO[C@H]6CC(=O)N2[C@H]1[C@H]6[C@H]5C[C@H]43	349,433	0,9162
ZINC000670455453	CN(CC(=O)NC12C[C@H]3C[C@H](C1)CC(n1ncnn1)(C3)C2)C(N)=O	333,395	0,6328
ZINC000052525227	O=S1(=O)CCN([C@H](CCO)c2cccc(C(F)(F)F)c2)CC1	337,361	1,2282
ZINC000085493565	CN1C[C@]2(C)CCC[C@@]34[C@@H]2C[C@@H]([C@@H]13)[C@@]12CC[C@H](C[C@H]14)[C@@](O)(CO)[C@H]2O	347,497	1,4288
ZINC00005227699	CC(C)(CO)C1OCC2(CO1)COC(C(C)(C)CO)OC2	304,381	1,0058
ZINC00034279958	N[C@H](Cc1c[nH]c2ccc(I)cc12)C(=O)O	330,12	-1,0727
ZINC000100168449	CC(C)=C(O)[C@H](O)[C@H](O)[C@H](O)[C@@]1(O)C(=O)C1(C)C	260,285	-0,2416
ZINC000330205278	O=S(=O)(N[C@H]1CCn2ccnc2C1)c1ccc2c(c1)C(O)=NCC2	346,41	1,2286

ID молекулы	SMILES	Молекулярный вес	cLogP
ZINC000100627706	O=C1O[C@H]([C@@H](O)CO)C(O)=C1OS(=O)(=O)O	256,187	-4,0794
ZINC001226465242	Oc1c(O)c(O)c(OC2CC3(COC3)C2)c(O)c1O	270,236	0,2184
ZINC000005567009	O=C1CCN(CCCN2CCC(=O)NC2=S)C(=S)N1	300,406	-0,0308
ZINC000008728786	O=C(O)CSCc1cc(=O)n2ncsc2n1	257,294	-0,3369
ZINC000907297314	O=C(NCC#CCNC(=O)[C@@H]1[C@H]2CCC(=O)[C@H]12)[C@@H]1[C@@H]2CCC(=O)[C@H]21	328,367	0,0798
ZINC000244718786	Cc1ccc(-c2nc(CN3C[C@@H]4C[C@H](O)[C@@H](O)C[C@@H]4C3)no2)cc1	329,399	1,4617
ZINC000102635684	Cc1nc2nc(N3C(=O)[C@@H]4CC=CC[C@@H]4C3=O)nn2c(O)c1Cl	333,734	0,651
ZINC000014963550	OC1CCN([C@@H]2CCN(C34CC5CC(CC(C5)C3)C4)C[C@H]2O)CC1	334,502	1,8695
ZINC000223612342	N[C@H]1[C@H]2CCO[C@H]2[C@@H]1NS(=O)(=O)c1cc(Cl)c(Cl)s1	343,254	1,2914
ZINC000215856195	O=P(O)(O)OC[C@H]1O[C@]2(CO[P@@](=O)(O)O2)[C@H](O)[C@@H]1O	322,098	-7,4475
ZINC000330130175	C(c1nnnc2n1CCCC2)N1CCC(c2nnnc3n2CCCC3)CC1	341,461	2,6337
ZINC00006442303	C[C@H]12CCC/C(=N/N=C(/N)S)[C@@H]1C(=O)NC(=O)C2	268,34	1,4618
ZINC000025560451	CC(=O)[C@@H]1[C@@H]2[C@@H]3CO[C@@H](O3)C(=O)[C@@H]2[C@H]2C=C[C@@H]3C=CC=C[C@@H]3N21	313,352	0,1626
ZINC000413398983	Cc1[nH]c(=NCC(C)(C)CC#N)sc1S(N)(=O)=O	288,395	2,0772
ZINC000039948012	CO[C@H]1C[C@H]2/C(=N/N)[C@@H](OC(=O)OC)C[C@H]2/C1=N/N	314,297	0,4998
ZINC000058990757	N#CC[C@@H](NC(=O)C1=C[C@@H](N)[C@@H](O)[C@H](O)C1)C(N)=O	268,272	-3,2684
ZINC000330334635	CN1C(=O)Cc2cc(NC(=O)N(CCO)[C@@H]3CCCC3)ccc21	349,454	1,1046
ZINC000005807322	CSc1nc2ncc([C@@H](O)[C@@H](O)CO)nc2c(=O)[nH]1	284,295	-2,1623
ZINC000002219747	C[C@H]1CCCC2sc3ncn4nc(C(=O)NCCCO)nc4c3c21	345,426	1,6543
ZINC000001641852	CC#C[C@](O)(C(=O)OC1CCN(C)C1)c1ccccn1	288,346	1,1823
ZINC000072165275	COc1cc(Cl)cc2cc(C(=O)N3CC[C@H](CO)[C@H](O)C3)oc21	339,774	2,03
ZINC000022441635	O=C(O)[C@H]1CN(C/C=C/P(=O)(O)O)CCN1	250,19	-7,538

ID молекулы	SMILES	Молекулярный вес	cLogP
ZINC000892123001	N#Cc1ccc2c(c1)C[C@@H](NC(=O)NC/C=C\CN)C2	270,335	0,9357
ZINC00006092288	N[C@@H]1CC[C@@]2(O)[C@H]3Cc4ccc(O)c5c4[C@@]2(CCN3CC2CC2)[C@H]1O5	342,437	1,0353
ZINC000071746712	O=c1cc(CN2CCO[C@@H](C(F)(F)F)C2)[nH]c(=O)[nH]1	279,217	-0,3006
ZINC000072462920	NCCCC1C(=O)CC[C@H]2CN(c3ncnc4sccc43)CC[C@H]21	345,47	1,2164
ZINC00004632611	O=C(O)[C@H]1CC(=O)c2oc(=O)c3cc(O)c(O)c(O)c3c21	292,199	-0,673
ZINC000409380504	COc1ccc(NC(=O)[C@@H]2[C@@H](C(=O)O)[C@H]3C=C[C@@]2(C)O3)nn1	305,289	-0,6238
ZINC00015404956	O=C1Cc2cc(C3=NN=C(NC4CC4)SC3)cc3c2N1CCC3	326,423	3,2441
ZINC00033830752	O=c1cc(CCc2cccc2)oc2c1[C@H](O)[C@@H](O)[C@@H](O)[C@@H]2O	318,324	0,3017
ZINC000244899967	Cc1cccc([C@@]23NC(=O)[C@@H]4[C@H]5C[C@H]([C@H]42)[C@H](O3)[C@H]5O)c1	271,315	1,0464
ZINC000330226146	C[C@H](NCC1(O)CCS(=O)(=O)CC1)[C@@H]1OCCc2sccc21	345,483	0,4331
ZINC000309642940	N#Cc1c(N)sc2c1CCN(S(N)(=O)=O)C2	258,325	-0,3855
ZINC000328679747	O/C(=N/C[C@@H]1CCCC[C@H]1O)c1ccc2nnc(O)n2c1	276,295	0,4034
ZINC00031687741	N=C(N=C(N)N)Nc1ccc(SCC(=O)O)cc1	267,312	-0,0412
ZINC000244816997	O=C(NCC12CCCN1CCC2)[C@H]1CC(=O)N(c2cccc(O)c2)C1	343,426	1,0976
ZINC000329909867	C[C@@H]1Oc2ccc(C(=O)Cn3cnc4c(c3=O)CCC4)cc2N=C1O	339,35	0,4721
ZINC000772124687	Nc1c(Cl)cc(N2C(=O)N[C@H](CO)C2=O)cc1Cl	290,105	0,5651
ZINC000872626811	N#C[C@@H]1CN(CN2CC3(CCOCC3)[C@H](O)C2=O)CCN1C1CC1	334,418	-0,6957
ZINC000330114843	O[C@@H]1Cc2cccc2[C@@H]1Nc1nccc(N[C@@H]2CCCOC2)n1	326,399	1,7551
ZINC00004624958	CNc1nc(Cl)nc(Oc2ccc(O)nn2)n1	254,637	1,1297
ZINC000019362162	O[C@@H](CN1CCCC(C[C@H](O)C(F)(F)F)CC1)C(F)(F)F	324,264	0,9312
ZINC00004625503	O=C(O)[C@@H]1CS[C@@H](c2ccc([C@H]3N[C@H](C(=O)O)CS3)cc2)N1	340,423	-4,5476
ZINC00020540054	OCC#Cc1csc(CN2C[C@H](O)[C@@H](N3CCOCC3)C2)c1	322,428	0,5197

ID молекулы	SMILES	Молекулярный вес	cLogP
ZINC000822125131	CC(C)(C)[C@@H]1NC(=O)[C@H](Cc2cnc[nH]2)NC1=O	250,301	-0,1514
ZINC000053229721	C[C@H](CO)NC(=O)[C@@H]1C[C@@H]2c3cccc4[nH]cc(c43)C[C@H]2N(C)C1	327,427	1,6393
ZINC001236908437	C[C@H](N)CN(C)Cc1cc2c(cc1O)OCCNC2	265,356	-0,4033
ZINC000004992871	Cc1ccc(NC(=O)[C@@H]2CC(=O)N=C(NC(=N)N)S2)cc1Cl	339,806	0,6214
ZINC000330320595	CC(C)c1n[nH]c(CN2CCN(C3C[C@H]4CC[C@@H](C3)N4C)CC2)n1	332,494	0,9541
ZINC000045796127	CSc1ncc2c(n1)O[C@@H]1CNC[C@@H]1NC2=O	252,297	-0,053
ZINC001772820148	O[C@H]1[C@@H](O)[C@H]2c3cc4c(cc3[C@]3(O)CN5CCC[C@]25[C@@H]13)OCO4	317,34	1,0252
ZINC000527836098	N#Cc1cn(C2CN(C=O)N[C@H]3CCO[C@H]3C3CC3)C2)(=O)nc1O	345,358	-0,9008
ZINC000072143422	OC[C@H]1CN(c2cc(-c3cccc3F)nc3ccnn32)C[C@H]1CO	342,373	0,9409
ZINC000330575377	C[C@H](C(=O)NC(=N)O)N1CC[C@@](O)(Cc2cccc(Cl)c2)C1	325,795	1,0252
ZINC000067541660	CCNC(=O)[C@@H]1C[C@@H](N)CN1Cc1ccc2cccc(Cl)c2n1	332,834	1,208
ZINC000006650283	CC(C)c1ccc2c(c1)[C@]1(NC(N)=Nc3nc4cccc4n31)C(=O)N2	346,393	2,7113
ZINC000330039541	Cc1cc(-c2nc3sc4c(c3c(O)n2)CCS(=O)(=O)C4)no1	337,379	1,4505
ZINC000079280485	C#Cc1cccc(NC(=O)CNS(N)(=O)=O)c1	253,281	-0,1687
ZINC000858172382	CN(C)c1cc(F)c(CN[C@@H]2CO[C@H]2O)cc1F	272,294	0,3116
ZINC000104390239	N=C(N)N/N=C\ccc(O)c(S(=O)(=O)O)c1	258,257	-1,2563
ZINC000006296265	CC(=O)OC(OC(C)=O)[C@@]12C=C[C@@](C)(O1)[C@H]1C(=O)NC(=O)[C@@H]12	309,273	-1,2949
ZINC000330291110	O=C(NC[C@@]1(O)[C@@H]2[C@@H]3C[C@H]4[C@@H]1C[C@@H]1C[C@@H]4[C@@H]3[C@@H]21)c1ccnc(C2CC2)n1	337,422	1,4721
ZINC000284232839	Cc1nc2n(n1)CCN(CC#CCN1CCn3nc(C)nc3C1)C2	326,407	0,0528
ZINC000216393001	O[C@H]1CO[C@H]1N1CCCN(Cc2ccc(-n3ccnc3)cc2)CC1	342,441	0,5231
ZINC000001848555	O=S(=O)(O)c1cc(O)c2c(Cl)cc(S(=O)(=O)O)cc2c1	338,744	-1,3443
ZINC000009810510	Cc1[nH]c(=O)[nH]c(=O)c1CCC(=O)Nc1ccc2c3c(cccc31)CC2	349,389	3,0302

ID молекулы	SMILES	Молекулярный вес	cLogP
ZINC000006620309	O=C(O)CCSc1nc(O)nc2[nH]c(O)nc21	256,242	-0,235
ZINC000329404651	C[C@H](O)CN1CCN(/C(O)=N/[C@@H]2CCCS(=O)(=O)C2)[C@H](C)C1	333,451	-0,2295
ZINC000001894000	N[C@@](C(=O)O)(c1ccc(P(=O)(O)O)cc1)C1CC1	271,208	-4,8949
ZINC000009694454	CCCC1[nH]nc(O)c1[C@]1(O)C(=O)NC(=O)N=C1O	268,228	-1,5343
ZINC000020720824	Cc1nn2c(NC(=O)CSC3=NCCS3)csc2nc1=O	341,439	1,0669
ZINC000013136395	Oc1ncnc(O)c1C(c1ccc(F)cc1)c1c(O)ncnc1O	330,274	1,3541
ZINC001772711399	OC[C@@]12CNC[C@@H]1CN(Cc1ccn3nccc3c1)C2	272,351	-0,6211
ZINC000065405961	Nc1nc(NCCO)cc(N2C[C@H]3CC[C@@H](C2)N3)n1	264,332	-0,5003
ZINC000007811979	C[C@@H]1C[C@H](C)CN(CC(=O)Nc2sc3c(c2C(N)=O)CCC3)C1	335,471	2,2867
ZINC000016385168	Nc1ccc(Cl)cc1N/N=C1/NNC(=O)C1=O	253,649	0,0967
ZINC000257356591	COC(=O)C[C@@H]1[C@H]2[C@H](O)[C@H](O)[C@H]1[C@@](C)(C(=O)O)[C@H]2C(=O)O	302,278	-1,556
ZINC000057124747	CNC(=O)NNS(=O)(=O)c1cc2c(cc1F)NC(=O)CC2	316,312	-1,4696
ZINC000299765575	CO[C@]12CCN(Cc3nc4ccsc4c(O)n3)[C@H]1C[C@@H](O)CC2	335,427	1,1042
ZINC000065372420	COc1ccc([C@H]2NC(CO)(CO)[C@H]3CN(C)C[C@H]32)cc1C	306,404	0,4947
ZINC000007757820	O=C(NNC(=O)C1CCCC1)c1ccc2c(c1)nc1n(c2=O)CCC1	340,382	0,7632
ZINC000097637841	CN1CCN(c2cc(-c3ccc4c(c3)OC(F)(F)O4)nc(N)n2)CC1	349,34	3,4424
ZINC000000440425	Cc1ccc(NNC2=NC(CO)(CO)COC2=O)cc1	279,295	-0,4201
ZINC000108225329	NC[C@@]1(COc2ccc(C(F)(F)F)cn2)C[C@@H](O)[C@@H](O)C1	306,283	0,4996
ZINC000218453347	CC(=O)N1CC(Nc2cc(C(=O)O)cc(C(=O)O)c2)=CC1=O	304,257	0,0033
ZINC000007106673	NC(=O)NC(=O)COC(=O)c1csc(-c2cccc(Cl)c2)n1	339,758	1,4056
ZINC000097346828	Cc1c2cccc(F)c2[nH]c1C(=O)N1CCN(CCCO)C[C@@H](O)C1	349,405	1,4053
ZINC000016618208	Nn1c(SCc2nc(-c3ccsc3)no2)nnc1C(F)(F)F	348,333	1,9968

ID молекулы	SMILES	Молекулярный вес	cLogP
ZINC000043199179	OC1C2(CCCO2)C(O)C2(CCCO2)C(O)C12CCCO2	300,35	-0,9558
ZINC000013534655	C=C1C[C@]23C[C@H]1CC[C@H]2[C@@]12C=C[C@H](O)[C@@](C)(C(=O)O1)[C@@H]2[C@@H]3C(=O)O	330,379	0,9158
ZINC000330519080	O=C(NC[C@@H]1CCN(CC(F)(F)F)C1)C1(O)CCSCC1	326,382	0,6626
ZINC000015758341	Cc1cc(C)n(CC(=O)NNC(=O)c2cc3cc(F)ccc3s2)n1	346,385	0,9923
ZINC000012533881	C[C@H](C(=O)N1C[C@H](C)C[C@H](C)C1)N1CCN(c2cccc2O)CC1	345,485	2,4734
ZINC00004063870	Nc1nnc(Sc2ccc(Nc3cc(=O)[nH]c(=O)[nH]3)cc2)s1	334,383	1,2928
ZINC000330030034	Cc1nc2cncc(N3CCN(C[C@H](O)C4CCCCC4)CC3)n2n1	344,461	0,5462
ZINC000077505943	O=C(O)c1cc2c(nc1N1CC[C@](O)(CN3CCCCC3)C1)CCC2	345,441	0,8009
ZINC00003590790	Nc1ccn([C@@H]2CS[C@H](COP(=O)(O)O)O2)c(=O)n1	309,238	-4,2511
ZINC001235500388	OCc1ncn2c1CN(Cc1c(O)cc(O)cc1O)CC2	291,306	-0,0386
ZINC00031793749	Nc1nccc2c1ncn2[C@@H]1C=C(CO)[C@H](O)[C@H]1O	262,268	-0,8991
ZINC00005624960	CC(C)(C)N1C(=O)NC(=O)/C(=C/NC(C)(CO)CO)C1=O	299,326	-1,4388
ZINC000067562651	O=C(c1ccc([C@@H]2CCCO2)s1)N1Cc2cnn(CCO)c2C1	333,411	-0,0497
ZINC000067843403	CC(C)(C)c1ncc2c(n1)CN(c1nccc(N3CC[C@H](O)C3)n1)C2	340,43	0,9068
ZINC000013222066	NC(=O)C(=CNC(=S)c1ccncc1)C(N)=O	250,281	-1,3321
ZINC000005807390	O=C(O)C[C@H](/N=c1\[nH]c2nc[nH]c2c(=O)[nH]1)C(=O)O	267,2	-2,3525
ZINC001602247550	O=C(O)c1cc(C(=O)NCc2cc(O)ccc2O)on1	278,219	-0,1715
ZINC000330221469	O=C(NCCO)c1ccc(N2CCOC3(C2)CCCCC3)nn1	334,418	0,5201
ZINC000329947366	N[C@H]1CCc2ncc(NC(=O)c3ccc4c(c3)CCNCC4)cc2C1	336,438	1,9584
ZINC000072284228	NC(=S)c1nc(-c2cccc(C(N)=S)c2)[nH]c1C(N)=S	321,452	0,9264
ZINC000328897239	CN(C)c1cc(N/C(O)=N/[C@H]2CCCN=C2O)ncn1	278,315	-0,1106
ZINC000104089204	NC(=S)N/N=C1\c2ccc(O)nc2C(=O)C=C1O	264,265	-0,0625

ID молекулы	SMILES	Молекулярный вес	cLogP
ZINC00004940568	O=c1c2cnn(CCO)c2ncn1Cc1ccc(Cl)c(Cl)c1	339,181	1,187
ZINC00091620289	CSCC(=O)N1CC[C@H]2(O)CCN(Cc3cccnc3)C[C@H]2C1	335,471	0,6719
ZINC000330049127	CN(Cc1nccc(N)n1)[C@H]1CCN(C(=O)[C@H]2CCC(F)(F)C2)C1	339,389	0,64
ZINC000247841320	COC(=O)[C@]12[C@H]3[C@@H]4[C@H]5[C@@H]1[C@@H]1C=CC[C@H]1[C@@H]5[C@]21[C@H]4[C@H]2[C@H](CC[C@H]32)O[C@@H]1O	340,418	1,5806
ZINC00000399862	O=c1[nH]/c(=C\c2cccc(O)c2)c(=O)[nH]/c1=C/c1cccc(O)c1	322,319	1,5812
ZINC00072434424	O=C(C1CCCC1)N1CCC(Nc2ncnc3c2CCNCC3)CC1	343,473	1,9
ZINC00014199324	NC(=O)c1ccc(S(=O)(=O)N[C@@H]2CC[C@H]3CCCC[C@H]3C2)cc1	336,455	2,3574
ZINC001220825770	Cc1cc(C)c(CNC(=O)[C@@H]2C[C@H]3CNC[C@H]3N2)c(O)c1	289,378	0,5001
ZINC00072134680	O=C(c1cncc(N2C[C@H]3[C@@H](C2)C2CCC3CC2)n1)N1CCOCC1	342,441	1,4033
ZINC00491096653	C#CCN=c1[nH]c(=N[C@H]2CNC(=O)C2)[nH]c(=N[C@H]2CNC(=O)C2)[nH]1	330,351	-1,7176
ZINC00000521301	O=c1/c(=C/c2cccs2)sc2n1CN(c1cccc(F)c1)CN=2	345,421	3,4222
ZINC000862582506	O=C(/C=C/c1ccc(O)c(O)c1)N1C[C@H](O)[C@@H](O)C1	265,264	0,0895
ZINC00426546266	OCCN=c1[nH]c([C@@H]2CCOC2)cc(=NCC(F)(F)F)[nH]1	306,287	0,089
ZINC00003080955	O=C(O)[C@H](O)[C@@H](O)C(=O)Nc1ccc(Cl)cc1Cl	294,089	0,1642
ZINC00016953495	O=C(NO)c1c(F)c(F)c(C(=O)NO)c(F)c1F	268,122	0,004200 1
ZINC001506430897	Cc1nc(N)nc(N2CC[C@H](O)[C@@H](O)C2)c1Cl	258,708	0,2727
ZINC001205358690	N[C@@H]1CN(Cc2cc3c4c(c2)CCCN4CCC3)C[C@H]1O	287,406	0,9353
ZINC000330649840	C[C@H](c1cn(-c2ccccc2)nn1)N1CCC[C@H](S(N)(=O)=O)C1	335,431	0,3193
ZINC00019877995	CO[C@H]1O[C@@H]2CO[C@H](c3cccc3)O[C@]2(O)[C@H](O)[C@H]1O	298,29	-0,8042
ZINC000216448693	Cc1cc(C)c(CN2CCN(C)CC2)cc1CN1C[C@H]2CNC[C@H]2C1	342,529	1,7425
ZINC000565707690	CN1CCN(C(=O)C(=O)NCC#Cc2cccc(F)c2)[C@@H](CN)C1	332,378	-0,1121

ID молекулы	SMILES	Молекулярный вес	cLogP
ZINC000097237395	O=C(Nc1n[nH]c(Cc2ccc(F)cc2)n1)c1ccc2nnncn2c1	337,317	1,3049
ZINC000670433709	NC(=O)c1ncn([C@@@H]2OC[C@H](O)[C@@H](O)[C@@H]2O)c1C(N)=O	286,243	-3,1983
ZINC000257212371	NC(=O)N[C@@H]1C[C@@H](Oc2ccnc2)[C@H](O)[C@H]1N1CCOCC1	322,364	-1,1768
ZINC000021380768	C[C@H]1CN(CC(=O)Nc2ccc3c(c2)OC2(CCCC2)O3)CCO1	332,399	2,0687
ZINC000091270366	Cc1cc(C)c(C(=O)NCCc2nc(C)c3c(n2)CCCC3)c(=O)[nH]1	340,426	1,6455
ZINC000091757499	Cc1nc(CN2CCC3(CCC(=O)N([C@@H]4CCN(C)C4)C3)CC2)co1	346,473	1,1472
ZINC000718192660	C#C[C@H](CO)Nc1ncnc2cc(C(=O)O)ccc21	257,248	0,3566
ZINC000019993584	C[C@@H]1CCc2c(sc3nc(SC[C@H](O)CO)[nH]c(=O)c32)C1	326,44	1,67
ZINC000426589002	Cc1cc(=N[C@@H]2CCN(Cc3ccn3)C[C@H]2O)nc(N)[nH]1	320,42	-0,8387
ZINC000299771480	CN(C)[C@@H]1CCCN(Cc2cc(O)c3cc(C(N)=O)ccc3n2)CC1	342,441	1,2372
ZINC000031808989	Cc1nn(-c2ncnc3[nH]cnc32)c2c1[C@@H](c1ccnc1)CC(=O)N2	346,353	0,7312
ZINC000075792337	NS(=O)(=O)NCc1ccc(N2CCSCC2)cc1	287,407	0,4317
ZINC000104002065	O=C1NC(=O)[C@H]2[C@H]1[C@@H]1[C@H]2[C@H]2[C@H]3O[C@@H]3[C@H]1[C@H]1C(=O)NC(=O)[C@@H]12	288,258	-2,312
ZINC000072165956	O[C@H]1[C@H](n2cc(-c3cncn3)nn2)c2cccc2C12CCNCC2	348,409	0,9805
ZINC000330179756	OC1(CN2C[C@H]3CCC[C@@]3(c3nc(-c4cnccn4)no3)C2)CCC1	341,414	0,6412
ZINC000067920255	CN1CCC(Cc2noc(C[C@H]3NC(=O)c4cccc43)n2)CC1	326,399	1,9611
ZINC000082097658	Cn1c2ccccc2c2cc(CC(=O)N3CC[C@@H](N)[C@H](O)C3)ccc21	337,422	1,8831
ZINC000329851409	CC1CCC(N2CCN(S(=O)(=O)NC3CCCC3)CC2)CC1	329,507	2,3646
ZINC000272665881	O=C1CN(CC/N=C/c2ccc(O)c(O)c2O)CCN1	279,295	-0,3587
ZINC000100284158	COC(=O)CN/C=C1\C(=O)NN=C1C(F)(F)F	251,164	-0,9
ZINC00004665029	CCOC(=O)[C@@]12C(=O)C(=O)N[C@]1(c1ccccc1)[C@@H]1C=CCC[C@H]12	325,363	1,5562

ID молекулы	SMILES	Молекулярный вес	cLogP
ZINC00000281025	CC[C@H](Sc1nc2c(ncnc2N)[nH]1)C(=O)O	253,285	0,2834
ZINC000299791284	Cc1cc2c(nc(CN3CCN(c4nonc4C)CC3)nc2O)s1	346,414	1,9102
ZINC00053006806	CN[C@H]1[C@H](O)[C@H]2O[C@@H]3O[C@H](C)CC(=O)[C@]3(O)O[C@@H]2[C@@H](NC)[C@H]1O	332,352	-3,1536
ZINC00071278598	CN1C[C@H](NC(=O)c2ccco2)C[C@H]1C(=O)N[C@H]1CCCCNC1=O	348,402	-0,5277
ZINC000330329209	Cc1ccc(CN2C(=O)CC[C@H]2C(=O)N2[C@H]3CC[C@H]2[C@H](CO)C3)cc1	342,437	1,6096
ZINC00023506959	CN(C)C(=O)c1cc(-c2cnn(CCO)c2)nc2cc(F)ccc21	328,346	0,9235
ZINC000253495540	Cc1ccccc1NC(=O)[C@@H]1CC(=O)N=C2N(N)c3ccccc3N21	335,366	0,9735
ZINC00043198994	O=C(O)[C@@H]1CC2=C(C[C@@H]1C(=O)O)C[C@H](C(=O)O)[C@H](C(=O)O)C2	312,273	-0,9802
ZINC00067727172	COc1ccccc2c1OCC(CN1CC[C@](N)(C(=O)O)C1)=C2	304,345	-2,3632
ZINC000995446774	NC(=O)C(=O)N1CCN(C(=O)C#CC2CC2)[C@H]2CS(=O)(=O)C[C@H]21	339,371	-1,9106
ZINC000117713062	Cn1cc(-c2nc(N[C@@H]3CCCC[C@@H]3N)c(F)c3c2C(=O)NC3)cn1	344,393	-0,2131
ZINC001041335897	N#Cc1cnc(N2C[C@@H](O)[C@H](NC(=O)[C@H]3CC[C@@H](C(N)=O)C3)C2)cn1	344,374	-1,7067
ZINC00019537538	CC1=C(C(=O)OC2C3CC4CC(C3)CC2C4)[C@@H](CC(=O)O)C(=O)N1	333,383	1,3236
ZINC00003984185	O=C(O)[C@@H]1Cc2[nH]cnc2[C@@H](c2ccccc2O)N1	259,264	-2,0093
ZINC000330046193	C[C@H](CNC(=O)c1cc2c(nc1O)CCCC2=O)N1CCCCC1=O	345,398	1,3891
ZINC00072474189	Cc1nc2c(c(=O)[nH]1)CN(C(=O)c1cccc(CCC(C)(C)O)c1)C2	341,41	1,7514
ZINC000330343491	NC(=O)c1ncn([C@H]2CCCN(Cc3ncc(Cl)s3)C2)n1	326,811	0,4231
ZINC000330612084	OC[C@H]1CN(Cc2cn(-c3cccc(Cl)c3Cl)nn2)CCO1	343,213	0,7299
ZINC00001609918	NC(=O)c1cc(C(N)=O)c(C(N)=O)cc1C(N)=O	250,213	-1,9912
ZINC00091934781	O=C(O)[C@H]1CC(=O)NC12CCN(C/C=C/c1ccc(F)cc1)CC2	332,374	0,4838
ZINC000534638301	N[C@@H]1CN(S(=O)(=O)c2ccc(OC(F)F)cc2)C[C@H]1O	308,304	-0,4082
ZINC00299788393	Cc1nc2c(c(N3C[C@H](CC(=O)O)[C@H](c4cn(C)nn4)C3)n1)CCC2	342,402	0,9947

ID молекулы	SMILES	Молекулярный вес	cLogP
ZINC000253471491	Cc1cn([C@@H]2C[C@@H](N=[N+]=[N-])[C@@H](CO)O2)c(=O)nc1N1CCCCC1	334,379	0,2402
ZINC00005274421	COC(=O)[C@]1(O)CC(C)=NN1c1nc(-c2ccccc(O)c2)cs1	333,367	2,478
ZINC00006668315	Cc1cc2c(Cc3n[nH]cc3-c3cccs3)n[nH]c(=O)c2c(=O)[nH]1	339,378	0,7865
ZINC00006686730	CC1CCC(=NNC(=O)c2ccc(Cl)c(S(N)(=O)=O)c2)CC1	343,834	2,5916
ZINC00012858874	c1nc2sc3c(c2c(N2CCN(C[C@H]4CCCO4)CC2)n1)CCC3	344,482	2,4584
ZINC000328746844	CC1(C)SC[C@H]1/N=C(/O)N1CCN(CCS(N)(=O)=O)CC1	336,48	0,345

Таблица E2. Хиты виртуального скрининга потенциальных лигандов белка NS1 ортофлавивирусов, переносимых клещами.

ID молекулы	SMILES	Молекулярный вес	cLogP
ZINC000342133588	C[C@H]1CN(CC#CCN2C[C@@H](C)[C@H](CO)C2)C[C@H]1CO	280.41	0.911
ZINC000067804507	NC[C@@H]1Cc2cc(F)cc(-c3cc4c(cc3C(N)=O)OCO4)c2O1	330.314	1.8443
ZINC000330229468	CN(c1cc(N(C)[C@@H]2CCCC[C@H]2O)ncn1)[C@@H]1CCCC[C@H]1O	334.462	1.9992
ZINC000000000747	O=C1NOC[C@H]1/N=C/c1ccc(/C=N/[C@H]2CONC2=O)cc1	302.289	-0.8806
ZINC000009348652	O=C(O)CNC(=O)C[C@H]1SC(NNC2=NC(=O)CS2)=NC1=O	345.359	-3.2911
ZINC000958528707	N[C@H](c1nccs1)C1(O)C[C@H]2CC[C@@H](C1)S2(=O)=O	288.391	-0.497
ZINC000017112820	CC[C@@]1(C)Cc2c(sc3[nH]c(SCC(N)=O)nc(=O)c23)CO1	339.439	0.6348
ZINC000004521387	O[C@@H]([C@H](O)[C@H]1COC2(CCCCCC2)O1)[C@H]1COC2(CCCCCC2)O1	342.43	0.9482
ZINC000085493565	CN1C[C@]2(C)CCC[C@@]34[C@@H]2C[C@@H]([C@H]13)[C@@]12CC[C@@H](C[C@H]14)[C@@](O)(CO)[C@H]2O	347.497	1.4288
ZINC000021542192	N#CCOC(=O)Cn1cnc2scc(-c3cccs3)c2c1=O	331.375	1.6804
ZINC000330205278	O=S(=O)(N[C@H]1CCn2ccnc2C1)c1ccc2c(c1)C(O)=NCC2	346.41	1.2286
ZINC001226465242	Oc1c(O)c(O)c(OC2CC3(COC3)C2)c(O)c1O	270.236	0.2184

ID молекулы	SMILES	Молекулярный вес	cLogP
ZINC000450767552	CC1[C@@H](C(=O)N2CC(O)(CC#N)C2)[C@@H]1C(=O)N1CC(O)(CC#N)C1	332.359	-1.0885
ZINC000907297314	O=C(NCC#CCNC(=O)[C@@H]1[C@H]2CCC(=O)[C@H]12)[C@@H]1[C@H]2CCC(=O)[C@H]21	328.367	0.0798
ZINC000244718786	Cc1ccc(-c2nc(CN3C[C@@H]4C[C@H](O)[C@@H](O)C[C@@H]4C3)no2)cc1	329.399	1.4617
ZINC000329893505	Cc1oc2nc(CN3CCO[C@@]4(CCO[C@@H]4C)C3)nc(N)c2c1C	332.403	0.9024
ZINC000014963550	OC1CCN([C@@H]2CCN(C34CC5CC(CC(C5)C3)C4)C[C@H]2O)CC1	334.502	1.8695
ZINC000223612342	N[C@H]1[C@H]2CCO[C@H]2[C@H]1NS(=O)(=O)c1cc(Cl)c(Cl)s1	343.254	1.2914
ZINC000958500969	O=C(NC[C@@H]1CN(CC#Cc2cccc2)C[C@H]1O)c1nc[nH]n1	325.371	0.3695
ZINC000330130175	C(c1nnc2n1CCCC2)N1CCC(c2nnc3n2CCCC3)CC1	341.461	2.6337
ZINC000006442303	C[C@@]12CCC/C(=N/N=C(/N)S)[C@@H]1C(=O)NC(=O)C2	268.34	1.4618
ZINC000299759929	COC(=O)[C@@H]1C[C@@H](NC(=O)c2[nH]c3ccc(Cl)cc3c2C)CN1	335.79	1.5242
ZINC000072410008	COc1ccc2[nH]nc(CN3CCC[C@](O)(CO)C3)c2c1	291.35	0.1292
ZINC000004331330	CCNC(=S)N1CCN(C(=S)c2ccc(N(C)C)cc2)CC1	336.527	2.6263
ZINC000058990757	N#CC[C@@H](NC(=O)C1=C[C@@H](N)[C@@H](O)[C@H](O)C1)C(N)=O	268.272	-3.2684
ZINC000016318621	COCCNC(=S)NNC(=O)c1cc2c(s1)CCCCCC2	341.499	2.5152
ZINC000003452436	Cn1nc(C(=O)Nc2nc(-c3ccc(F)c(F)c3)cs2)ccc1=O	348.332	1.8961
ZINC000079190593	CC12CN3CC(C)(CN(C1)C3c1ccc(O)c3ncccc13)C2O	325.411	1.6219
ZINC000005807322	CSc1nc2ncc([C@@H](O)[C@@H](O)CO)nc2c(=O)[nH]1	284.295	-2.1623
ZINC000002219747	C[C@@H]1CCCC2sc3ncn4nc(C(=O)NCCCO)nc4c3c21	345.426	1.6543
ZINC001246201770	c1cc(CN[C@H]2CCNC3(CNC3)C2)cc2c1OCC2	273.379	0.854
ZINC000096231420	O=S1(=O)CC[C@H](c2nn3c(nnc3-c3ccsc3)s2)C1	326.424	0.7215
ZINC000072165275	COc1cc(Cl)cc2cc(C(=O)N3CC[C@H](CO)[C@H](O)C3)oc21	339.774	2.03
ZINC000008524769	CC(C)CCNC(=S)N/N=C1/c2cc3c(cc2N=C1O)OCCO3	348.426	2.0428

ID молекулы	SMILES	Молекулярный вес	cLogP
ZINC000004909307	CN(C(=O)CN1CCC(NC(=O)C(C)(C)C)CC1)C1CCCCC1	337.506	2.3001
ZINC000892123001	N#Cc1ccc2c(c1)C[C@@H](NC(=O)NC/C=C\CN)C2	270.335	0.9357
ZINC000826385401	N#Cc1ccc(NC(=O)c2nc(O)nc(O)c2N)cc1-c1nn[nH]n1	339.274	-0.2479
ZINC000008980649	Nc1nc(SCc2nc3scc(-c4cccc4)c3c(=O)[nH]2)n[nH]1	346.394	1.0087
ZINC000072462920	NCCCN1C(=O)CC[C@H]2CN(c3ncnc4sccc43)CC[C@H]21	345.47	1.2164
ZINC000520425064	O=C1NCC[C@]12CCCN2CC#CCN1CCC[C@H]12CCNC2=O	330.43	-0.0332
ZINC000015404956	O=C1Cc2cc(C3=NN=C(NC4CC4)SC3)cc3c2N1CCC3	326.423	3.2441
ZINC000072412327	C=CCN1C[C@@H]2(C(=O)O)CN(c3nc(N)nc4[nH]ncc43)C[C@H]2C1=O	343.346	-1.5194
ZINC000033830752	O=c1cc(CCc2ccccc2)oc2c1[C@H](O)[C@@H](O)[C@@H](O)[C@@H]2O	318.324	0.3017
ZINC000017379409	S=c1[nH]cnc2c1ncn2CCn1cnc2c1nc[nH]c2=S	330.399	-1.2864
ZINC000330226146	C[C@H](NCC1(O)CCS(=O)(=O)CC1)[C@@H]1OCCc2sccc21	345.483	0.4331
ZINC000003157297	N#Cc1cc2c([nH]c1=O)CC/C(=C/NCCO)C2=O	259.264	-1.1983
ZINC000426428382	O[C@@H]1CNC[C@H]1NC1CCN(c2ccc(F)c(F)c2)CC1	297.348	0.8499
ZINC000244816997	O=C(NCC12CCCN1CCC2)[C@H]1CC(=O)N(c2cccc(O)c2)C1	343.426	1.0976
ZINC000761226094	COC(=O)CNC(=O)[C@@H](C)OC(=O)C#Cc1ccc2c(c1)OCO2	333.295	0.9448
ZINC000006483425	COc1c(O)cc(C[C@H]2(O)Oc3cc(O)cc(O)c3C2=O)cc1O	334.279	1.2375
ZINC000270823378	C[C@H]1CN=C(N2CCN(C(=O)[C@H](CO)NC(N)=O)CC2)S1	315.397	-1.2302
ZINC000330114843	O[C@@H]1Cc2ccccc2[C@@H]1Nc1nccc(N[C@@H]2CCCOC2)n1	326.399	1.7551
ZINC00075171424	NC(=O)Cc1csc(NC(=O)[C@@H]2CSCCC(=O)N2)n1	314.389	-0.7613
ZINC000004625503	O=C(O)[C@@H]1CS[C@@H](c2ccc([C@H]3N[C@H](C(=O)O)CS3)cc2)N1	340.423	-4.5476
ZINC000219646651	Cc1nc(N2CCN(C(CO)CO)C[C@H](O)C2)sc1C	301.41	0.2894
ZINC00020540054	OCC#Cc1csc(CN2C[C@H](O)[C@@H](N3CCOCC3)C2)c1	322.428	0.5197

ID молекулы	SMILES	Молекулярный вес	cLogP
ZINC000053229721	C[C@@H](CO)NC(=O)[C@@H]1C[C@@H]2c3cccc4[nH]cc(c43)C[C@H]2N(C)C1	327.427	1.6393
ZINC00004992871	Cc1ccc(NC(=O)[C@@H]2CC(=O)N=C(NC(=N)N)S2)cc1Cl	339.806	0.6214
ZINC000330320595	CC(C)c1n[nH]c(CN2CCN(C3C[C@H]4CC[C@@H](C3)N4C)CC2)n1	332.494	0.9541
ZINC000072143422	OC[C@@H]1CN(c2cc(-c3cccc3F)nc3ccnn32)C[C@H]1CO	342.373	0.9409
ZINC000330575377	C[C@H](C(=O)NC(=N)O)N1CC[C@@](O)(Cc2cccc(Cl)c2)C1	325.795	1.0252
ZINC001341846663	CC(C)(O)[C@H](N)C(=O)NCc1c(O)ccc2c1CCCC2	292.378	0.7318
ZINC000017071439	COC(=O)c1c(C)c(C(=O)N(C)C)sc1NC(=S)NC1CC1	341.455	2.0371
ZINC000014315775	Cc1noc2nc(C3CC3)cc(C(=O)NNC(=S)NC(C)C)c12	333.415	1.637
ZINC000004291883	Cc1ccc2c(c1)[C@@H]1c3sc(=O)[nH]c3S[C@H](C(=O)O)[C@@H]1C(=O)O2	349.386	2.14
ZINC000004462695	COC(=O)c1ccc(COC(=O)c2ccc3c(c2)NC(=O)CS3)o1	347.346	0.6304
ZINC000009552872	CC(C)(C)NC(=O)NC(=O)CSc1nc2cc(Cl)cnc2[nH]1	341.822	1.9728
ZINC000330291110	O=C(NC[C@@]1(O)[C@@H]2[C@@H]3C[C@H]4[C@@H]1[C@H]1C[C@@H]4[C@@H]3[C@@H]21)c1ccnc(C2CC2)n1	337.422	1.4721
ZINC000284232839	Cc1nc2n(n1)CCN(CC#CCN1CCn3nc(C)nc3C1)C2	326.407	0.0528
ZINC000216393001	O[C@H]1COC[C@@H]1N1CCCN(Cc2ccc(-n3ccnc3)cc2)CC1	342.441	0.5231
ZINC00001848555	O=S(=O)(O)c1cc(O)c2c(Cl)cc(S(=O)(=O)O)cc2c1	338.744	-1.3443
ZINC000329404651	C[C@H](O)CN1CCN(/C(O)=N/[C@@H]2CCCS(=O)(=O)C2)[C@H](C)C1	333.451	-0.2295
ZINC000043199181	COc1ccc(OCC23OC4C(O)C(O2)C(O)C(O3)C4O)cc1	326.3	-0.3516
ZINC000020720824	Cc1nn2c(NC(=O)CSC3=NCCS3)csc2nc1=O	341.439	1.0669
ZINC000067587305	CCNc1ncc(C(=O)N2CCC(O)(c3ccc(C)cn3)CC2)cn1	341.414	1.4865
ZINC000087492684	CC1(C)c2nc(Cl)c3nc([C@H]4O[C@H](CO)[C@@H](O)[C@H]4O)nc-3n21	326.739	-2.2347
ZINC000097823334	Cc1[nH]c(=O)c(C(=O)NC(C)(C)c2nnc(N)s2)c(C)c1Cl	341.822	0.7386
ZINC00007811979	C[C@@H]1C[C@H](C)CN(CC(=O)Nc2sc3c(c2C(N)=O)CCC3)C1	335.471	2.2867

ID молекулы	SMILES	Молекулярный вес	cLogP
ZINC000257356591	COC(=O)C[C@@H]1[C@H]2[C@H](O)[C@H](O)[C@H]1[C@@](C)(C(=O)O)[C@H]2C(=O)O	302.278	-1.556
ZINC00004996188	O=C1c2ccccc2[C@@]2(O)Nc3nc4nonc4nc3N[C@]12O	312.245	0.0493
ZINC000330196785	CC(C)[C@@H]1C[C@H](N[C@@H]2CCO[C@@]3(CCS(=O)(=O)C3)C2)CS1	333.515	0.9545
ZINC000299765575	CO[C@]12CCN(Cc3nc4ccsc4c(O)n3)[C@H]1C[C@@H](O)CC2	335.427	1.1042
ZINC000065372420	COc1ccc([C@H]2NC(CO)(CO)[C@H]3CN(C)C[C@H]32)cc1C	306.404	0.4947
ZINC00007757820	O=C(NNC(=O)C1CCCC1)c1ccc2c(c1)nc1n(c2=O)CCC1	340.382	0.7632
ZINC000019815466	[O-][n+]1ccccc1[C@H](O)CN1CCN(c2cccc(Cl)c2)CC1	333.818	1.3997
ZINC000108225329	NC[C@@]1(COc2ccc(C(F)(F)F)cn2)C[C@@H](O)[C@@H](O)C1	306.283	0.4996
ZINC000007106673	NC(=O)NC(=O)COC(=O)c1csc(-c2cccc(Cl)c2)n1	339.758	1.4056
ZINC000329437543	CN(/C(O)=N/CCN1CCN=C(O)C1)C1CC(O)C1	270.332	-1.2921
ZINC000097346828	Cc1c2cccc(F)c2[nH]c1C(=O)N1CCN(CCCO)C[C@@H](O)C1	349.405	1.4053
ZINC000095529020	COc1c(F)c(F)c(CN2CC(=O)N3CCC[C@H]3C2=O)c(F)c1F	346.279	1.0914
ZINC000299779274	Cc1n[nH]c(CN2C[C@H]3COCC[C@@]3(O)[C@@H]3COCC[C@@H]32)c1Cl	341.838	0.1298
ZINC000063856626	O=C(O)[C@]12CNC[C@H]1CN(/C(O)=N/S(=O)(=O)c1ccccc1)C2	339.371	-2.3364
ZINC000015758341	Cc1cc(C)n(CC(=O)NNC(=O)c2cc3cc(F)ccc3s2)n1	346.385	0.9923
ZINC000012533881	C[C@H](C(=O)N1C[C@H](C)C[C@H](C)C1)N1CCN(c2cccc2O)CC1	345.485	2.4734
ZINC00004063870	Nc1nncc(Sc2ccc(Nc3cc(=O)[nH]c(=O)[nH]3)cc2)s1	334.383	1.2928
ZINC000072159246	CCn1c2ncc(C(=O)OC)cc2nc1N1C[C@@H]2[C@H]3CC[C@H](O3)[C@@H]2C1	342.398	1.4831
ZINC000067872490	O=C([C@H]1CCNC1)N1CCOc2c(O)cc(-c3ccnc3)cc2C1	339.394	1.1952
ZINC000067562651	O=C(c1ccc([C@@H]2CCCC2)s1)N1Cc2cnn(CCO)c2C1	333.411	-0.0497
ZINC000240104277	Cc1nc(C(=O)NCCn2c(C)nc3sccc3c2=O)cs1	334.423	1.1777
ZINC000330221469	O=C(NCCO)c1ccc(N2CCOC3(C2)CCCCC3)nn1	334.418	0.5201

ID молекулы	SMILES	Молекулярный вес	cLogP
ZINC000330285894	Cc1nc2c(n1[C@H]1CCN(C(=O)C[C@@H]3N=C(O)N=C3O)C1)CCCC2	345.402	0.4537
ZINC000299759775	COc1cc(OC)c(C(=O)N2CCC3(CCC(=O)N3)CC2)cc1OC	348.398	1.4277
ZINC000329947366	N[C@H]1CCc2ncc(NC(=O)c3ccc4c(c3)CCNCC4)cc2C1	336.438	1.9584
ZINC000100646803	O=C(COC(=O)[C@H]1[C@@H]2C[C@H]3OC(=O)[C@H]1[C@H]3C2)Nc1ccc(F)cc1	333.314	0.8401
ZINC000072284228	NC(=S)c1nc(-c2cccc(C(N)=S)c2)[nH]c1C(N)=S	321.452	0.9264
ZINC000101447247	COC(=O)/C=C/C(=O)/N=C(\N)NC(=N)Nc1ccccc1O	305.293	-0.2269
ZINC000330265700	O=S1(=O)NC(CN2CCOC[C@@H](C3CCC3)C2)=Nc2ccccc21	349.454	0.9423
ZINC000004940568	O=c1c2cnn(CCO)c2ncn1Cc1ccc(Cl)c(Cl)c1	339.181	1.187
ZINC000091986750	CN1CCN(Cc2ccc(-c3cc[nH]n3)o2)C[C@]12CCNC(=O)CC2	343.43	0.1082
ZINC000056768883	C[C@H](Cc1ccc(O)cc1)NC(=O)NNC(=O)C(=O)NC1CC1	320.348	-0.4721
ZINC00000399862	O=c1[nH]/c(=C\c2cccc(O)c2)c(=O)[nH]/c1=C/c1cccc(O)c1	322.319	1.5812
ZINC000329206133	C[C@H]1C[C@H](N2CCOCC2)CN1/C(O)=N/C[C@@H]1CN(C)CCN1C	339.482	-0.0278
ZINC001220825770	Cc1cc(C)c(CNC(=O)[C@@H]2C[C@H]3CNC[C@H]3N2)c(O)c1	289.378	0.5001
ZINC000225745297	Cc1noc(-c2ccn3c(c2)nn3CNC(=O)C2CCOCC2)nn1	342.358	-0.1425
ZINC000000113474	Cc1cccc(N2C(=O)c3sc4nc(=O)c(C)nn4c3C2=O)c1C	340.362	0.9815
ZINC000000521301	O=c1/c(=C/c2cccs2)sc2n1CN(c1cccc(F)c1)CN=2	345.421	3.4222
ZINC000016591147	O=C(CN1CCC[C@@H](C(F)(F)F)C1)NC(=O)NCc1ccco1	333.309	1.2272
ZINC000003080955	O=C(O)[C@H](O)[C@@H](O)C(=O)Nc1ccc(Cl)cc1Cl	294.089	0.1642
ZINC000072423511	Cc1ccc(N2CCn3nc(CN4CCSCC4)cc3C2)nn1	330.459	0.3605
ZINC000019799224	Nc1nnc2nc(O)c([C@@H](C(=O)O)c3ccc(Cl)cc3)nn12	320.695	-0.2748
ZINC000003048339	COC(=O)[C@@H]1SC[C@H]2C(=O)N(CCc3cccc3)CC(=O)N21	335.383	0.0998
ZINC001205358690	N[C@@H]1CN(Cc2cc3c4c(c2)CCCN4CCC3)C[C@H]1O	287.406	0.9353

ID молекулы	SMILES	Молекулярный вес	cLogP
ZINC000330649840	C[C@H](c1cn(-c2cccc2)nn1)N1CCC[C@@H](S(N)(=O)=O)C1	335.431	0.3193
ZINC00032070283	Cc1cc2c(nc1C)oc1c2ncn(CC(=O)N2CCOCC2)c1=O	342.354	0.6625
ZINC001148585901	O=C(NCCN1CCCCO1)c1cc(O)c(O)c(O)c1	282.295	-0.0109
ZINC000216448693	Cc1cc(C)c(CN2CCN(C)CC2)cc1CN1C[C@H]2CNC[C@H]2C1	342.529	1.7425
ZINC000565707690	CN1CCN(C(=O)C(=O)NCC#Cc2cccc(F)c2)[C@@H](CN)C1	332.378	-0.1121
ZINC00097237395	O=C(Nc1n[nH]c(Cc2ccc(F)cc2)n1)c1ccc2nncn2c1	337.317	1.3049
ZINC000172997272	NC(=O)Cn1cnc(-c2nc3cc4cccc4cc3c(=O)[nH]2)n1	320.311	-0.1881
ZINC00021380768	C[C@H]1CN(CC(=O)Nc2ccc3c(c2)OC2(CCCC2)O3)CCO1	332.399	2.0687
ZINC00091270366	Cc1cc(C)c(C(=O)NCCc2nc(C)c3c(n2)CCCC3)c(=O)[nH]1	340.426	1.6455
ZINC00091757499	Cc1nc(CN2CCC3(CCC(=O)N([C@@H]4CCN(C)C4)C3)CC2)co1	346.473	1.1472
ZINC000216537616	COCC(=O)N1CCC2(CC1)CN(C[C@H]1C[C@H]3C=C[C@@H]1C3)CCN2C	347.501	1.18
ZINC00091985985	CC(=O)N[C@@H]1CN(c2nc(C(N)=O)cs2)C[C@H]1c1ccc(C)o1	334.399	1.4065
ZINC00019993584	C[C@@H]1CCc2c(sc3nc(SC[C@@H](O)CO)[nH]c(=O)c32)C1	326.44	1.67
ZINC00008062130	CCOC(=O)[C@@H]1CCCN(Cn2c(=S)nc3nc(C)cc(C)n32)C1	349.458	0.3387
ZINC00002654612	O=C(O)c1cccc(NS(=O)(=O)c2ccc3nnn(O)c3c2)c1	334.311	0.4772
ZINC000597511720	Cc1cn(-c2ccc(C#N)cn2)nc1NC(=O)C1=NN2CCCN=C2N1	349.357	0.4703
ZINC00014900677	Nc1nccnc1C(=O)N1CCN(Cc2cccc3ccnc32)CC1	348.409	0.9037
ZINC00072165956	O[C@H]1[C@H](n2cc(-c3cncn3)nn2)c2cccc2C12CCNCC2	348.409	0.9805
ZINC00036366864	NC(=O)CSc1sc2c([nH]c(=O)cc2O)c1C(N)=O	299.33	-0.9764
ZINC000330179756	OC1(CN2C[C@H]3CCCC[C@@@]3(c3nc(-c4cnccn4)no3)C2)CCC1	341.414	0.6412
ZINC000794526092	CC(=NNC1=NC(=O)C(=O)N1)c1cc(S(N)(=O)=O)c(C)cc1C	337.359	-0.3289
ZINC00067920255	CN1CCC(Cc2noc(C[C@@H]3NC(=O)c4cccc43)n2)CC1	326.399	1.9611

ID молекулы	SMILES	Молекулярный вес	cLogP
ZINC000082097658	Cn1c2ccccc2c2cc(CC(=O)N3CC[C@@H](N)[C@H](O)C3)ccc21	337.422	1.8831
ZINC000272665881	O=C1CN(CC/N=C/C/c2ccc(O)c(O)c2O)CCN1	279.295	-0.3587
ZINC000016708199	CCOc1ccc2nc(NC(=O)C3=NN(C)C(=O)CC3)sc2c1	332.383	1.7331
ZINC000330501805	CN1CCC(NC(=O)C(=O)Nc2nc(C(F)(F)F)cs2)CC1	336.337	1.0694
ZINC000067725414	O=C1OCCN1Cc1nc([C@@]2(c3ccc(F)cc3)CCOC2)no1	333.318	1.0497
ZINC000065408548	Cc1nnc(-c2nc3c([nH]2)C[C@H](c2ccc(F)cc2)CNC3=O)o1	327.318	1.7935
ZINC000299791284	Cc1cc2c(nc(CN3CCN(c4nonc4C)CC3)nc2O)s1	346.414	1.9102
ZINC000053006806	CN[C@H]1[C@H](O)[C@H]2O[C@@H]3O[C@H](C)CC(=O)[C@]3(O)O[C@@H]2[C@@H](NC)[C@H]1O	332.352	-3.1536
ZINC000330160088	O=C(NC[C@H]1CS(=O)(=O)CCO1)c1ccc(F)c2ccccc12	337.37	1.1854
ZINC000007542856	O=C(COC(=O)[C@H]1COc2ccc(Cl)cc2C1)NC1CCCC1	337.802	2.5517
ZINC000043198994	O=C(O)[C@@H]1CC2=C(C[C@@H]1C(=O)O)C[C@H](C(=O)O)[C@H](C(=O)O)C2	312.273	-0.9802
ZINC000067727172	COc1cccc2c1OCC(CN1CC[C@](N)(C(=O)O)C1)=C2	304.345	-2.3632
ZINC001041335897	N#Cc1cnc(N2C[C@@H](O)[C@H](NC(=O)[C@H]3CC[C@@H](C(N)=O)C3)C2)cn1	344.374	-1.7067
ZINC000330046193	C[C@H](CNC(=O)c1cc2c(nc1O)CCCC2=O)N1CCCCC1=O	345.398	1.3891
ZINC000010854608	COc1ccc2c(c1)/C(=N/NC(=O)c1n[nH]c3c1CCCC3)C(O)=N2	339.354	0.9371
ZINC000329095037	CC(=O)N1CCCC[C@](O)(/C(O)=N/CCN2C(=O)CCCC2=O)C1	325.364	-0.9495
ZINC000884178958	CC(C)NC(=O)CN1CCN(C(=O)C2CC3(C2)CNC[C@@H]3C#N)CC1	347.461	-0.2921
ZINC000330612084	OC[C@H]1CN(Cc2cn(-c3cccc(Cl)c3Cl)nn2)CCO1	343.213	0.7299
ZINC00002886231	CSc1nsc(SCC(=O)NC2=NC(=O)[C@H](C)S2)n1	334.468	1.5955
ZINC000091934781	O=C(O)[C@H]1CC(=O)NC12CCN(C/C=C/c1ccc(F)cc1)CC2	332.374	0.4838
ZINC000253471491	Cc1cn([C@@H]2C[C@@H](N=[N+]=[N-])[C@@H](CO)O2)c(=O)nc1N1CCCCC1	334.379	0.2402
ZINC000215573505	CCn1ccc(C(=O)NC2CCN(C3CCSCC3)CC2)cc1=O	349.497	0.7119

ID молекулы	SMILES	Молекулярный вес	cLogP
ZINC000005274421	COC(=O)[C@]1(O)CC(C)=NN1c1nc(-c2cccc(O)c2)cs1	333.367	2.478
ZINC000009358206	O=C(Nc1cc(=O)[nH]c2nc(-c3cccc3)nn21)c1cccs1	327.323	1.4905
ZINC000006686730	CC1CCC(=NNC(=O)c2ccc(Cl)c(S(N)(=O)=O)c2)CC1	343.834	2.5916
ZINC000020432555	O=C(O)[C@H]1CSCN1S(=O)(=O)c1cc(Cl)cc(Cl)c1	342.222	1.6654
ZINC00012858874	c1nc2sc3c(c2c(N2CCN(C[C@H]4CCCCO4)CC2)n1)CCC3	344.482	2.4584
ZINC000328746844	CC1(C)SC[C@H]1/N=C(/O)N1CCN(CCS(N)(=O)=O)CC1	336.48	0.345
ZINC001170316557	O=C(CNc1cc2c(nn1)CNC2)c1ccc(O)c(O)c1	286.29	-1.109
ZINC00220276412	N[C@H]1C[C@H](c2cc(N3CCN(C4Cc5cccc5C4)CC3)ncn2)C1	349.48	2.4612
ZINC001772592027	CC(C)(O)[C@H](O)C[C@]1(O)C[C@H](O)[C@]23C[C@H]2C(C)(C)O[C@H]3[C@H]1O	316.392	-0.439
ZINC000002668814	Cc1sc2c(c1C)c(=O)nc1scc(CC(=O)NCCO)n12	337.423	1.4739
ZINC000009613968	CC(=O)Nc1nc(COC(=O)c2nc3nc(C)cc(C)n3n2)cs1	346.37	0.6587
ZINC000225666997	CS(=O)(=O)N1[C@H]2CC[C@H]1CC(NC(=O)c1c[nH]c3cccc13)C2	347.438	1.6137
ZINC000222544460	Nc1nc(N2CCN(C(=O)[C@H]3CCC[C@H](O)C3)CC2)cc(=O)[nH]1	321.38	-0.4952
ZINC000063857164	O[C@H]1(CN2CCC(Nc3cccc3)CC2)CCCN2CCCC[C@H]21	345.489	1.3581
ZINC000409189390	Cc1nc(S/C(=C/C2C(=O)NC(=O)NC2=O)C(=O)O)n[nH]1	311.277	-1.556
ZINC000655057289	Cc1c(C(=O)NCC#CCO)sc2nc(C)nc(N)c12	290.346	0.9248
ZINC000082127518	CS(=O)(=O)c1cnn2cc(CN3CCCC[C@H](N)C3)cnc12	323.42	-1.5877
ZINC000012679198	COc1cc(C(=O)Nc2cccc(C(N)=O)c2)cc2c1O[C@H](C)C2	326.351	2.3749
ZINC000096136436	CCc1nnc(CNC(=O)Nc2ccc3c(c2)O[C@H](C)C(=O)N3C)o1	345.358	0.7763
ZINC000032104634	O=C(c1nc(COc2ccc3c(c2)NC(=O)CC3)no1)N1CCCC1	342.354	1.0209
ZINC000095525018	Cc1nnsc1CN(C)C(=O)CN1CC2(CC1=O)CCCCC2	336.459	1.602
ZINC000330209970	O=S(=O)(N[C@H]1CCCC2cn(C3CCCC3)nc21)C1COC1	325.432	1.3784

ID молекулы	SMILES	Молекулярный вес	cLogP
ZINC000011428323	C[C@H](OC(=O)c1ccc2ncsc2c1)C(=O)Nc1ncccn1	328.351	1.485
ZINC000010125794	Cc1ccc2c(c1)S[C@@H](C(=O)NNC(=O)c1cccc1O)C2	328.391	1.374
ZINC000012113434	COc1cc([C@H]2NC(=O)NC3=C2C(=O)N(CCO)C3)ccc1O	319.316	-0.3576
ZINC000329630852	CO[C@@H]1COC[C@H]1n1cc(C/N=C(/O)N2CCN=C(O)[C@@H]2C)nn1	338.367	-1.5932
ZINC000097370441	O=C(O)c1nc[nH]c1C(=O)N1CCC([C@H](O)c2cccs2)CC1	335.383	0.7589
ZINC00004132375	NS(=O)(=O)c1ccc2c(c1)[C@@H]1C=CC[C@@H]1[C@H](C(=O)O)N2	294.33	0.2112
ZINC000072171153	CC(C)c1ccc(C(=O)NC[C@]2(O)CCCNCC2)c(=O)n1C	321.419	0.3092
ZINC000879743578	COC(=O)[C@H]1C[C@@H](O)CN1CC#CCN1C[C@H](O)C[C@@H]1C(=O)OC	340.375	-1.0644
ZINC000023130755	NC[C@@H]1CN(C2CCN(c3cccc(-n4ccn4)c3)CC2)CCO1	341.457	-0.0012999
ZINC000244724584	CCn1c(CN2C[C@H]3CC[C@@H](C2)[C@H]3O)nnc1[C@H]1C[C@@H](N)C1	305.424	-0.0506
ZINC00006016888	Cc1cc(O)nc(-n2cc(C(=O)c3cc(Cl)ccc3O)cn2)n1	330.73	1.6964