

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М.В. ЛОМОНОСОВА

На правах рукописи

Фомина Анастасия Дмитриевна

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

1.4.16. Медицинская химия

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата химических наук

Москва – 2025

Диссертация подготовлена на кафедре медицинской химии и тонкого органического синтеза химического факультета МГУ имени М.В.Ломоносова и в лаборатории противовирусных лекарственных средств ФГАНУ «Федеральный научный центр исследования и разработки иммунологических препаратов им. М.П. Чумакова РАН» (Институт полиомиелита)

Научные руководители: **Палюлин Владимир Александрович**
кандидат химических наук, старший научный сотрудник

Осолодкин Дмитрий Иванович
кандидат химических наук

Официальные оппоненты: **Свитанько Игорь Валентинович**
доктор химических наук,
ФГБУН Институт органической химии им. Н.Д. Зелинского
Российской академии наук, лаборатория молекулярного
моделирования и направленного синтеза, заведующий лабораторией

Лагунин Алексей Александрович
доктор биологических наук, профессор РАН,
ФГАОУ ВО «Российский Национальный Исследовательский
Медицинский Университет им. Н.И. Пирогова» Министерства
здравоохранения Российской Федерации, Институт биомедицины
(МБФ), заведующий кафедрой биоинформатики

Хренова Мария Григорьевна
доктор физико-математических наук, профессор РАН
ФГБОУ ВО «Московский государственный университет имени
М.В.Ломоносова», химический факультет, профессор кафедры
физической химии

Защита состоится « 17 » октября 2025 г. в 13 часов 00 минут на заседании диссертационного совета МГУ.014.7 Московского государственного университета имени М.В.Ломоносова по адресу: 119991, г. Москва, ГСП-1, Ленинские горы, д. 1, стр. 3, МГУ имени М.В.Ломоносова, химический факультет, аудитория 446.

E-mail: 7422990@mail.ru (e-mail совета)

С диссертацией можно ознакомиться в отделе диссертаций научной библиотеки МГУ имени М.В.Ломоносова (Ломоносовский просп., д. 27) и на портале <https://dissovet.msu.ru/dissertation/3513>

Автореферат разослан «__» сентября 2025 г.

Ученый секретарь
диссертационного совета МГУ.014.7,
кандидат химических наук

Н.А. Синикова

Общая характеристика работы

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 и генерации при помощи нейронных сетей 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 мишенях, 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. 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.

Степень разработанности темы

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, 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 стандартизированных 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 ортофлавириусов. 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 ортофлавириусов методом систематического ансамблевого докинга и формирование фокусированной библиотеки для экспериментальных исследований.

Объекты исследования: базы данных низкомолекулярных химических соединений и кристаллографические данные о строении белков.

Предмет исследования: стратегия построения ансамблей для виртуального скрининга

лигандов вирусных белков методом ансамблевого докинга.

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*]пиримидиноны, 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 ортофлавириусов 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 ортофлавириусов 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 ингибирования 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 ортофлавиринов разделением ансамблей по переносчикам вируса является эффективной стратегией при виртуальном скрининге потенциальных лигандов.

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 стендовых 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».

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 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 [1], где автором проделана работа по виртуальному скринингу новых хемотипов потенциальных противокоронавирусных соединений, которая легла в основу экспериментального исследования.

Публикации По результатам работы опубликовано 5 работ, в том числе 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), а также 1 свидетельство о регистрации в государственном реестре программ для ЭВМ РФ и 1 свидетельство о регистрации в государственном реестре баз данных РФ, приравненные к таковым.

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.

1. Zakharova M.Y., Kaliberda E.N., Kurbatskaia I.N., Smirnov I.V., Knorre V.D., Gabibov A.G., Kuznetsova A.A., Bulygin A.A., Fedorova O.S., Kuznetsov N.A., Uvarova V.I., Fomina A.D., Kozlovskaya L.I., Osolodkin D.I., Ishmukhametov A.A., Egorov A.M., Varnek A. Pre-Steady-State Kinetics of the SARS-CoV-2 Main Protease as a Powerful Tool for Antiviral Drug Discovery // *Frontiers in Pharmacology*. – 2021. – Vol. 12. – P. 773198. – DOI 10.3389/fphar.2021.773198. – EDN UCTNDK. JIF - 5.6, 1.5 п.л., доля вклада автора 10%

Основное содержание работы

Во введении обоснована актуальность темы диссертационной работы, сформулированы цель и задачи, описана научная новизна, практическая и теоретическая значимость работы. В главе 1 представлен обзор научной литературы по разработке ингибиторов главной протеазы коронавируса и белка NS1 ортофлавиринов и методам структурно-обоснованного виртуального скрининга. В последующих главах приведены методы, использованные в работе, и обсуждение полученных результатов.

Во второй главе описана **разработка метода отбора ансамбля структур Mpro SARS-CoV-2 для ансамблевого докинга**. Структуры главной протеазы SARS-CoV-2 извлекали из

базы данных PDB и стандартизовали для сравнения и отбора ансамблей структур. Далее исключали мутантные, окисленные и незрелые формы и структуры с недостаточным разрешением, а также структуры с некорректной формой димеризации (рис. 1).

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, 306 из которых были прошли фильтрацию.

Структуры Mpro пространственно выравнивали по набору C α -атомов цепи A, присутствующих в каждой структуре. Для всех пар структур рассчитывали среднеквадратичное отклонение (СКО) между неводородными атомами остатков, формирующих активный центр фермента.

На основании распределения значений СКО отобрали 1% самых разнообразных пар структур. 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 избыточными. Такие избыточные пары были отфильтрованы по порогу подобия, чтобы сохранить одну репрезентативную пару из набора сходных пар. В *Ансамбль 1* вошли 7 структур, а в *Ансамбль 2* — 6 структур.

В качестве контроля были использованы два распространенных подхода к формированию ансамблей: анализ главных компонент (PCA, principal component analysis) и моделирование молекулярной динамики (МД). PCA провели для координат тяжелых атомов того же набора



Рисунок 1. Порядок стандартизации структур и отбора ансамбля.

остатков активного сайта, который использовался для расчета СКО. Данные в приведенных координатах двух первых главных компонент были кластеризованы с использованием метода k-средних. *Ансамбль PCA* был составлен из 4 структур, соответствующих точкам, ближайшим к центроидам кластеров.

Моделирование молекулярной динамики (МД) — наиболее распространенный источник белковых структур для ансамблевого докинга. Чтобы сравнить ансамбли, мы использовали пять конформаций, соответствующих 2, 4, 6, 8 и 10 мкс, извлеченных из 10-мкс траектории МД структуры 6LU7, опубликованной другими авторами (*Ансамбль МД*). На тепловой карте попарных СКО между неводородными атомами остатков активного центра структур всех ансамблей (рис. 2) явственно видно, что конформации из траектории МД значительно разнообразнее, чем кристаллографические структуры, извлеченные из PDB. Хотя высокое разнообразие может быть полезным для обеспечения разнообразия результатов виртуального скрининга, биологическое значение таких конформаций может быть ограничено. 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.

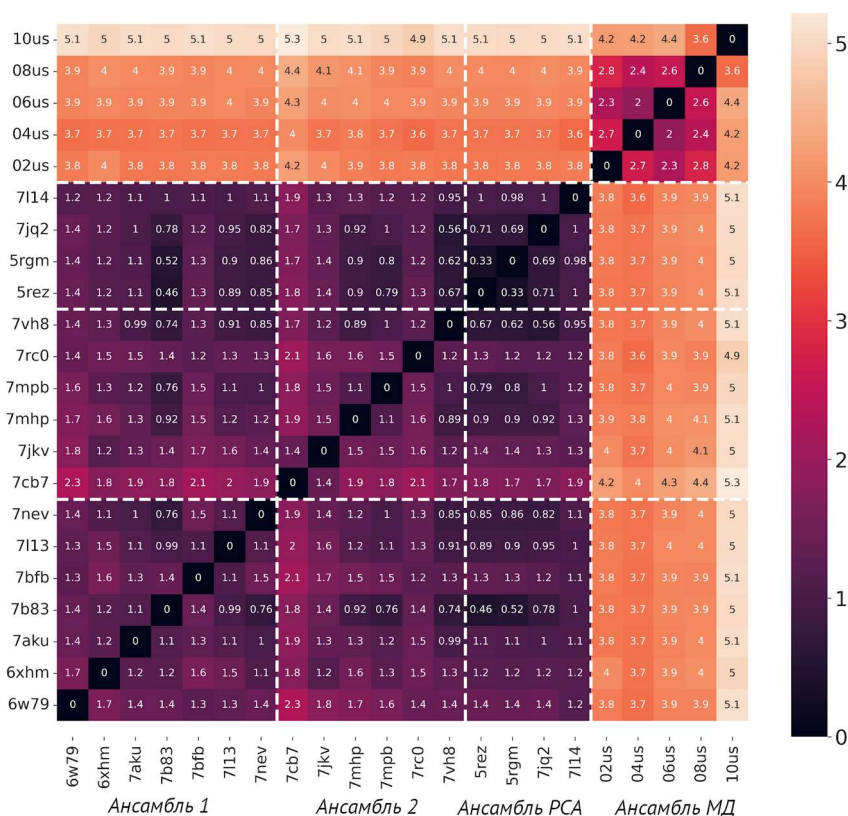


Рисунок 2. Тепловая карта попарных СКО (Å) между координатами тяжелых атомов остатков активного центра структур *Ансамбля 1*, *Ансамбля 2*, *Ансамбля PCA* и *Ансамбля МД*.

В третьей главе описана **подготовка библиотек низкомолекулярных соединений для виртуального скрининга**. Для сравнения результатов виртуального скрининга при помощи

различных ансамблей была собрана аннотированная библиотека соединений, для которых в период с 07.2020 по 03.2021, в статьях, индексированных в PubMed и препринтах, была опубликована экспериментально определенная активность против Mpro SARS-CoV-2 (*AA-Feb21*). Соединения, тестированные против конкретных ферментов (не Mpro), или имеющие известную мишень или механизм действия (не Mpro), исключали. Проанализировали 286 статей и препринтов, из которых извлекли 2181 результат определения ингибиторной активности по отношению к протеазе Mpro и/или к репродукции различных коронавирусов. Для дальнейшего использования в библиотеке были оставлены только соединения, для которых была определена концентрация полуингибирования (IC_{50}) для Mpro хотя бы одного из бетакоронавирусов (SARS-CoV-2, SARS-CoV, MERS-CoV). В итоговой библиотеке 788 соединений, 195 из которых имеют $IC_{50} < 10$ мкМ и отнесены к классу активных (44.7%).

Также были использованы две опубликованные ранее аннотированные библиотеки, собранные другими авторами из статей, опубликованных до 30.08.2020 (*AA-Aug20*: 53 активных (6.5%), 763 неактивных) и в период с января 2020 по август 2021 (*AA-Aug21*: 754 активных (62.6%), 451 неактивных). Химическое пространство трех библиотек сравнили методом двумерного t-SNE (t-distributed stochastic neighbor embedding). Три библиотеки слабо перекрываются и различаются по представленным в них хемотипам (рис. 3).

Библиотека *AA-Aug20* обладает наибольшим разнообразием, причем активные и неактивные соединения имеют одинаковые распределения индексов Танимото. В других библиотеках наблюдается большее разнообразие неактивных соединений, что связано с высоким процентом активных соединений и их меньшим разнообразием. Это отражает тенденцию публиковать данные о сериях активных соединений, часто полученных в результате целевых программ поиска ингибиторов Mpro, которые не были завершены к середине 2020 года. Поэтому *AA-Aug20* включает больше результатов репозиционирования и скрининга природных соединений, что приводит к более разнообразным результатам.

В дополнение к аннотированным библиотекам была сформирована выборка из 20000 разнообразных лекарственноподобных соединений из базы данных ZINC, итеративно отобранная по химическому разнообразию (*DD-ZINC*). Ее использовали в качестве класса потенциально неактивных соединений при сравнении результатов виртуального скрининга различными ансамблями и для обогащения выборок неактивными соединениями при построении моделей машинного обучения.

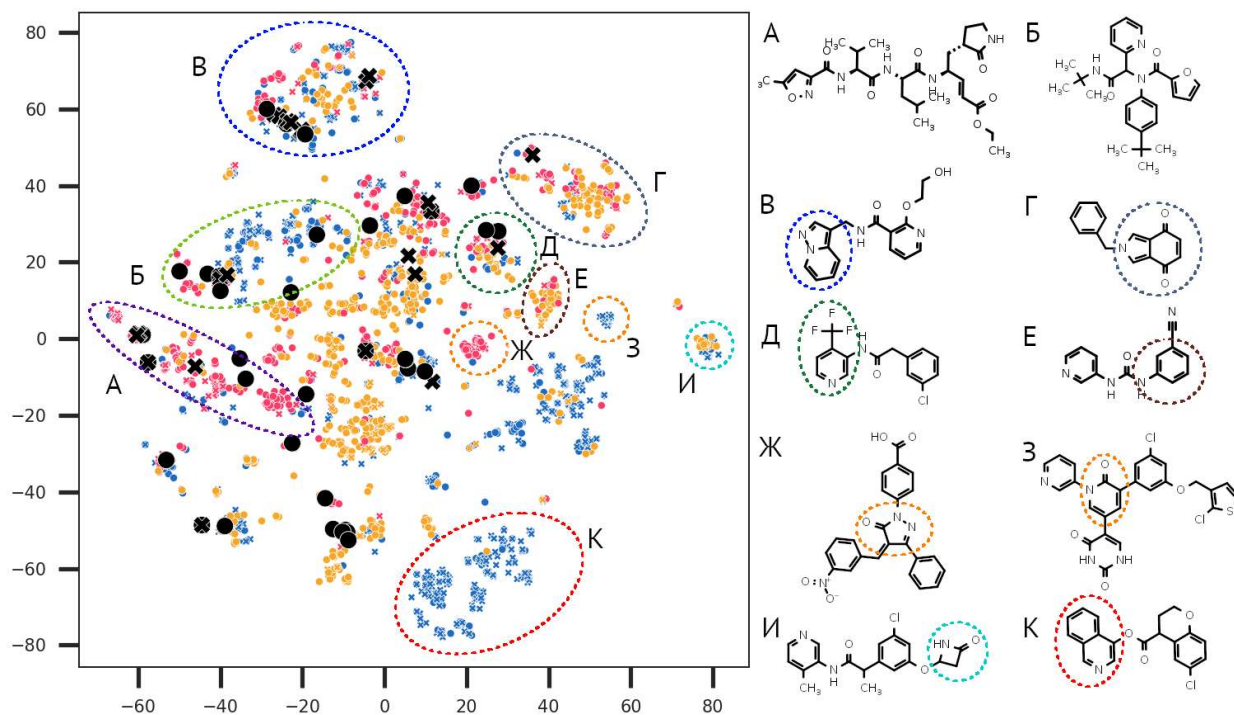


Рисунок 3. Карта t-SNE для библиотек соединений *AA-Aug20* (желтый), *AA-Feb21* (розовый) и *AA-Aug21* (синий). Маленькие точки соответствуют соединениям, присутствующим в одной библиотеке, большие черные точки представляют соединения из двух наборов данных. Активные соединения показаны крестиками, неактивные — кругами.

В четвертой главе представлена **разработка метода систематического ансамблевого докинга**. Все молекулы библиотек *AA-Aug20*, *AA-Feb21*, *AA-Aug21* и *DD-ZINC* были докированы во все структуры каждого ансамбля при помощи программы DOCK6.9. Ранжирование результатов докинга основывалось на консенсусных оценках. Несколько функций от индивидуальных оценок докинга сравнивали, чтобы найти наиболее мощную. Использовали минимальную, максимальную, среднюю и медианную оценку по ансамблю, «евклидову». Затем рассчитали распределение оценок докинга (рис. 4) для каждой библиотеки и каждого ансамбля. Если между левыми плечами распределений оценок активных и неактивных соединений есть разрыв, ансамбль может различать активные и неактивные соединения, давая большую по модулю оценку активным. Исследовали способность ансамблевого докинга отделять истинно активные молекулы от истинно неактивных в аннотированных библиотеках и от случайных соединений разнообразной библиотеки. Несмотря на то, что распределения для активных и неактивных соединений значительно различаются только для библиотеки *AA-Feb21*, распределения для всех библиотек имеют плечи в диапазоне более низких оценок по сравнению с распределением оценок докинга для набора данных *DD-ZINC*. Это означает, что ансамбли могут быть использованы в сценарии виртуального скрининга.

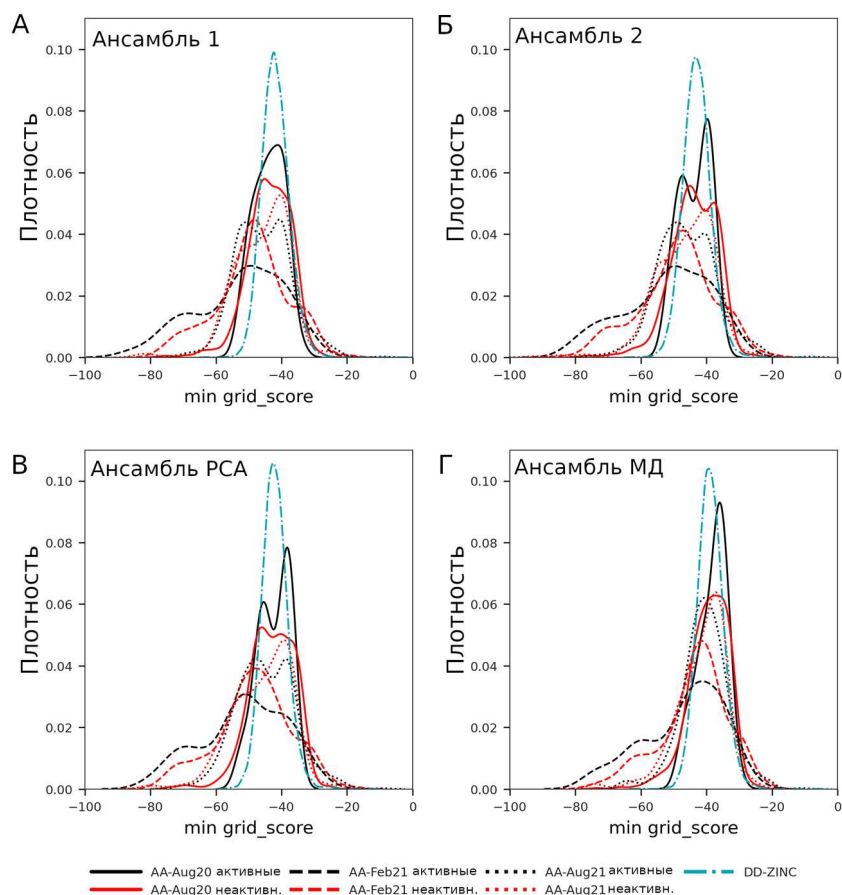


Рисунок 4. Распределения минимальных оценок докинга (*min grid_score*) для активных (черный) и неактивных (красный) соединений из библиотек *AA-Aug20*, *AA-Feb21*, *AA-Aug21* и *DD-ZINC* (голубой), докированных в *Ансамбль 1* (А), *Ансамбль 2* (Б), *Ансамбль PCA* (В) и *Ансамбль МД* (Г).

Обогащение результатов сравнили методом ROC, построив ROC-кривые и рассчитав площадь под ними (AUC) (рис. 5). 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. Судя по сравнению ROC-кривых, качество ранжирования соединений разными ансамблями примерно одинаково. Однако для библиотек, дополненных неактивными соединениями, наблюдаются бóльшие значения ROC AUC и углы наклона начального участка кривой, а различия между ансамблями становятся более заметны (рис. 5Б). Для дальнейшей работы были выбраны *Ансамбль 1* и *Ансамбль 2*, позволяющие наиболее полно отделить истинно активные соединения из аннотированных библиотек от неактивных соединений, а также отражающие конформационную подвижность аминокислотных остатков, составляющих активный центр фермента. *Ансамбль PCA* обладает менее воспроизводимой на разных библиотеках ранжирующей способностью, а *Ансамбль МД* обладает наиболее низкой точностью ранжирования среди всех ансамблей.

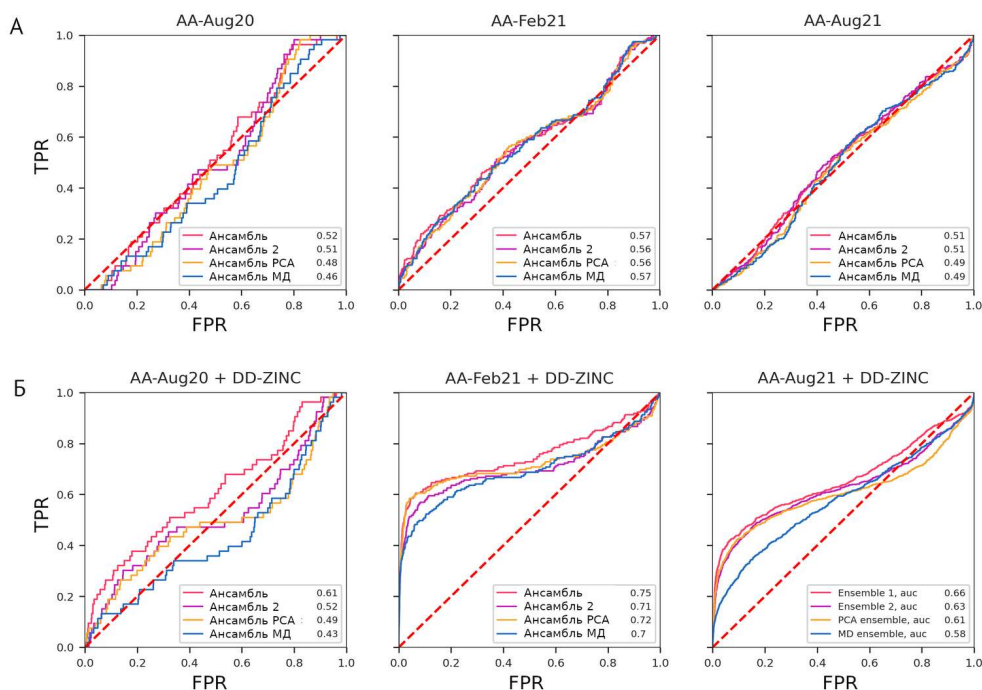
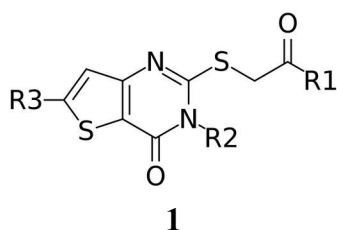


Рисунок 5. ROC-кривые для минимальной оценки результатов докинга библиотек *AA-Aug20*, *AA-Feb21* и *AA-Aug21* без обогащения (А) и с обогащением (Б) неактивными соединениями. FPR и TPR — true positive rate (доля истинноположительных результатов) и false positive rate (доля ложноположительных результатов) соответственно.

Чтобы подтвердить возможность использования методики систематического ансамблевого докинга для различных белковых мишеней, автоматизированная процедура отбора ансамбля была без изменений применена для структур циклин-зависимой киназы CDK2. Этот фермент хорошо изучен, для него доступно много структур в PDB и специально разработана аннотированная библиотека низкомолекулярных соединений DUD-e, содержащая активные и неактивные соединения, часто оказывающиеся в хит-листах при виртуальном скрининге в соответствующую мишень. Отбор ансамбля проводили среди структур, содержащих ингибитор в активном центре (при подготовке структур все лиганды удаляли), поскольку связывание циклина или его аналогов в активном центре вызывает конформационные изменения в структуре фермента. На основании разнообразия в ансамбль были выбраны 4 структуры. Рассчитанное по результатам докинга значение AUC, равное 0.65, было одинаковым как для консенсусной оценки по ансамблю, так и для докинга в структуру 1CKP, использованную авторами выборки DUD-e, в то время как индивидуальные AUC для структур ансамбля лежали в интервале от 0.57 до 0.61. Таким образом, процедура систематического докинга позволяет выбрать ансамбль структур, репрезентативных относительно конформационного пространства белка и обеспечивающих метрики качества виртуального скрининга по крайней мере на уровне одиночной структуры и нивелировать фактор случайности при выборе структуры для виртуального скрининга.

Чтобы повысить достоверность результатов виртуального скрининга, исследовали возможность введения ограничений на конформационный поиск с помощью методики Chemical Matching в DOCK6.9. Сферы в областях, где должны были формироваться необходимые водородные связи, были помечены так, чтобы в них оказывались фрагменты молекулы, содержащие акцептор водородной связи. Ограничения вводили на основе анализа контактов сокристаллизованных лигандов с активным центром Mpro SARS-CoV-2. Провели докинг библиотеки *AA-Feb21* с ограничениями и без. Распределения минимальных оценок докинга оказались почти идентичными, а при анализе ROC-кривых не было выявлено значительных различий. Значения коэффициентов корреляции рангов Спирмана между результатами докинга с ограничениями и без также свидетельствуют о высокой корреляции рангов. В дальнейшем использовали докинг без конформационных ограничений, так как ограничения не увеличивали точность виртуального скрининга и усложняли автоматизацию процесса.

В главе 5 описан **виртуальный скрининг потенциальных ингибиторов Mpro SARS-CoV-2**. В начале работы по поиску потенциальных ингибиторов главной протеазы SARS-CoV-2 был выполнен докинг соединений, ранее предложенных методом генеративного топографического картирования, в структуру 6WQF. Из 473 хитов картирования были отобраны 39, для которых оценка докинга была меньше -45. Они были разделены на группы по скаффолдам, и из каждой группы были выбраны от 2 до 4 коммерчески доступных соединений — одно с лучшей оценкой докинга и еще несколько достаточно разнообразных. Для десяти соединений определили эффективность ингибирования Mpro SARS-CoV-2 в присутствии фиксированной концентрации ингибитора, а для трех соединений, остаточная активность при скрининге которых оказалась меньше 0.5 (табл. 1), определили IC_{50} и K_i ². Оба соединения относились к ряду замещённых тиенопиримидинонов, соответствующих общей формуле 1.



С целью поиска более активных соединений и установления соотношений «структура-активность» в тиенопиримидиновой серии был проведён поиск соединений по подструктуре 1 в каталоге поставщика, в результате которого были отобраны 178 соединений. По результатам виртуального скрининга были выбраны соединения, получившие наиболее благоприятные минимальные оценки среди всех структур *Ансамбля 1*, а также ключевые для анализа соотношений «структура — активность» молекулы. Для 72 отобранных таким образом соединений была исследована ингибиторная активность по отношению к Mpro. IBS-E0680092

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

был родоначальником N³-алкильной серии с незамещённым фенилом (R₂ = H), а IBS-E0183442 — N³-арильной серии с замещённым фенилом (R₂ = OEt). Результаты определения ингибиторной активности и консенсусные оценки докинга наиболее активных соединений серии приведены в таблице 2.

Таблица 1. Результаты докинга и результаты определения ингибиторной активности для трех соединений-хитов.

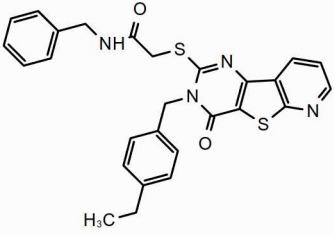
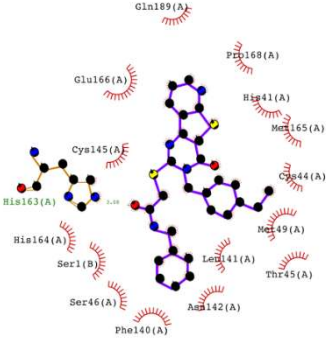
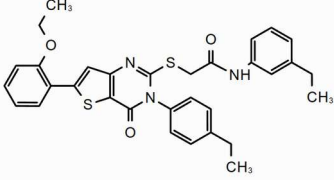
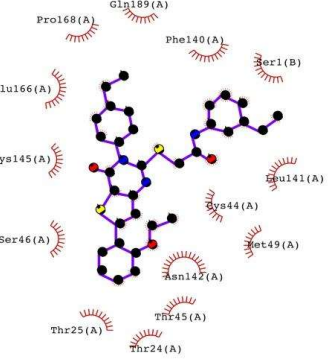
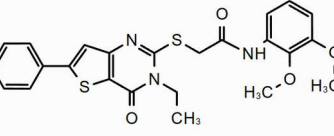
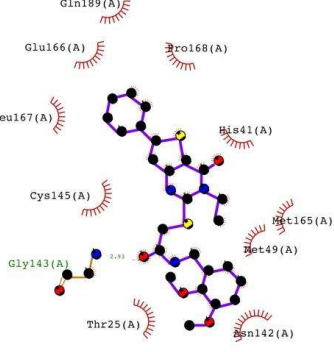
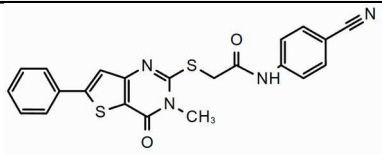
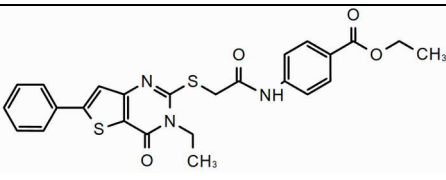
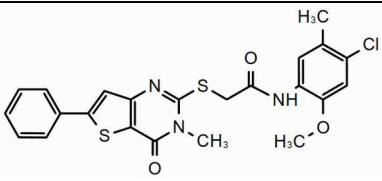
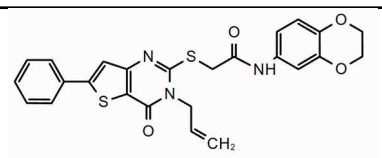
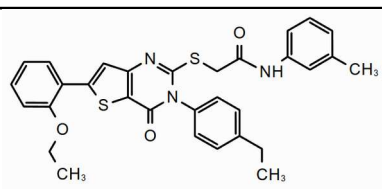
№	Шифр	Оценка докинга	IC ₅₀ , мкМ	Структура	Способ связывания по результатам докинга
2	IBS-E0474913	-57.06	38.6±0.4		
3	IBS-E0183442	-51.92	35.5±0.2		
4	IBS-E0680092	-50.91	27.9±0.2		

Таблица 2. Структуры и консенсусные оценки докинга для активных соединений расширенной серии.

№	Шифр	Структура	Минимальная оценка докинга	М _{pro} IC ₅₀ (95% ДИ*), мкМ
5	IBS-E0680061		-45.84	24.6 (18.9—31.7)
6	IBS-E0680133		-50.15	21.2 (16.7—29.8)
7	IBS-E0680085		-51.28	19.7 (13.6—35.6)
8	IBS-E0680169		-52.91	21.9 (16.1—30.3)
9	IBS-E0183443		-52.52	17.2 (12.8—23.3)

* ДИ — доверительный интервал

По итогам скрининга коммерчески доступной тиенопиримидиновой серии не было выявлено соединений, обладающих концентрацией полумаксимального ингибирования на уровне ниже 15 мкМ. При этом в наиболее исследованном ряду N³-алкилтиенопиримидинов отсутствует варьирование заместителей по положению R₃, а в положении R₂ присутствуют только 3 вида заместителей: метил, этил и аллил. Таким образом, за изменение ингибиторной активности в этом ряду преимущественно должен отвечать ароматический заместитель R₁, присоединённый по 2-сульфанильной группе. Эта серия соединений была выбрана для анализа потенциальных направлений синтетической модификации на основе результатов молекулярного докинга.

В ряду N³-арилтиенопиримидинов были представлены только гидрофобные соединения, содержащие доноры и акцепторы водородной связи исключительно в центральном фрагменте. Хотя среди них также имелись соединения с IC₅₀ на уровне 20 мкМ, недостаток направленных межмолекулярных взаимодействий и недостаточное покрытие серии затрудняют

её развитие и требуют дальнейшего исследования на базе коммерчески доступных и новых соединений.

Для N³-алкильных производных фенильного типа наблюдаются наиболее однородные результаты докинга. При этом заместитель R1 (формула 1) 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 (рис. 6). Поэтому для дальнейшего моделирования были выбраны две начальные структуры, характеризующиеся наиболее высокой активностью в серии (IC₅₀ ~20 мкМ) и содержащие акцепторы водородной связи при заместителе R1: IBS-E0680085 и IBS-E0680133. Методом молекулярного докинга провели систематический анализ допустимости различных заместителей в положении R2 тиенопиримидинового фрагмента, а также замещения фенильной группы в положении R3 (1). В соответствии с результатами докинга исходной серии, в положении R2 исследовали возможность введения небольших гидрофобных заместителей, а в положении R3 — ароматических заместителей. 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, а также расчётную липофильность. По результатам анализа предложена фокусированная библиотека для синтеза и дальнейшей экспериментальной валидации.

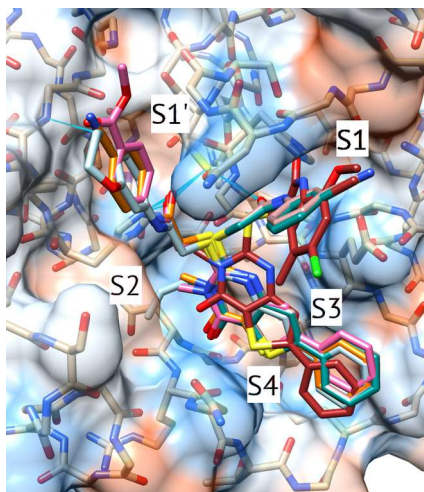


Рисунок 6. Наиболее активные соединения в активном центре главной протеазы SARS-CoV-2 (PDB id: 7NEV) в наиболее распространенной конформации среди N³-алкильных производных.

Глава 6 посвящена **прогнозированию класса активности ингибиторов Mpro SARS-CoV-2 методами машинного обучения**. На первой стадии для построения моделей были выбраны результаты высокопроизводительного скрининга *in vitro* ингибиторной активности лекарственных средств в отношении протеазы Mpro SARS-CoV-2 из базы данных ChEMBL (CHEMBL4495582). Активными считали соединения с процентом ингибирования Mpro SARS-

CoV-2 > 60%. После фильтрации и стандартизации библиотека насчитывала 6897 соединений, из которых 3.2% были активными. Библиотеку докировали в структуры *Ансамбля 2*, для каждого соединения выбирали одну наиболее высоко оцененную конформацию для каждой структуры ансамбля и на основании этих конформаций связывания рассчитывали фингерпринты взаимодействия — векторные представления контактов лиганда с каждым из аминокислотных остатков белка. Использовали два метода построения фингерпринтов: в “Flare-IFP” учитывали 4 типа контактов (“хорошие” — углерод-углерод, “плохие” — углерод-гетероатом, водородные связи и отсутствие контакта), а в “ODDT-IFP” учитывали 8 типов контактов (гидрофобные, π -стекинг, Т-стекинг, водородные связи, в которых аминокислотный остаток выступает донором или акцептором, солевые мостики, в которых аминокислотный остаток положительно заряжен либо отрицательно заряжен, а также связи с неорганическими ионами). Для векторного описания молекул использовали фингерпринты Моргана. Строили классификационные модели методами логистической регрессии, случайного леса, SVM и градиентного бустинга (CatBoost). Оптимальные параметры моделей подбирали по схеме стратифицированного скользящего контроля. Оказалось, что все модели *CHEMBL4495582* склонны к переобучению и характеризуются очень низкими значениями точности и поздним обогащением. Это делает их непригодными для виртуального скрининга — для больших библиотек такие модели будут прогнозировать слишком много ложноположительных результатов. Кроме того, в этих моделях фингерпринты взаимодействия не улучшают качество классификации по сравнению с моделями, построенным только на молекулярных фингерпринтах Моргана.

Процедуру построения моделей повторили аналогично на основе библиотеки *AA-Feb21* (рис. 7). Эта библиотека меньше по размеру и менее разнообразна, однако надежно аннотирована по активности и содержит серии соединений, специфически разрабатывавшихся в качестве ингибиторов главной протеазы коронавируса. Исходя из сравнения средних значений AUC и F1-меры, модели *AA-Feb21* характеризуются меньшим переобучением, чем модели, построенные на данных из *CHEMBL4495582*. Точность (precision) и полнота (recall) на классе активных для моделей *AA-Feb21* также превышают таковые для моделей, обученных на данных из *CHEMBL4495582*, как при оценке для контрольной выборки с соотношением классов как в полной библиотеке, так и при дополнении контрольной выборки соединениями из библиотеки *DD-ZINC*, помещенными в класс неактивных соединений. Во втором случае эффект значительно сильнее выражен, что свидетельствует о приемлемой надежности моделей в сценарии виртуального скрининга. Модели случайного леса, построенные на выборках, включающих оценки докинга, фингерпринты Моргана и фингерпринты взаимодействия, характеризовались наименьшим переобучением и самыми высокими значениями AUC и F1-меры.

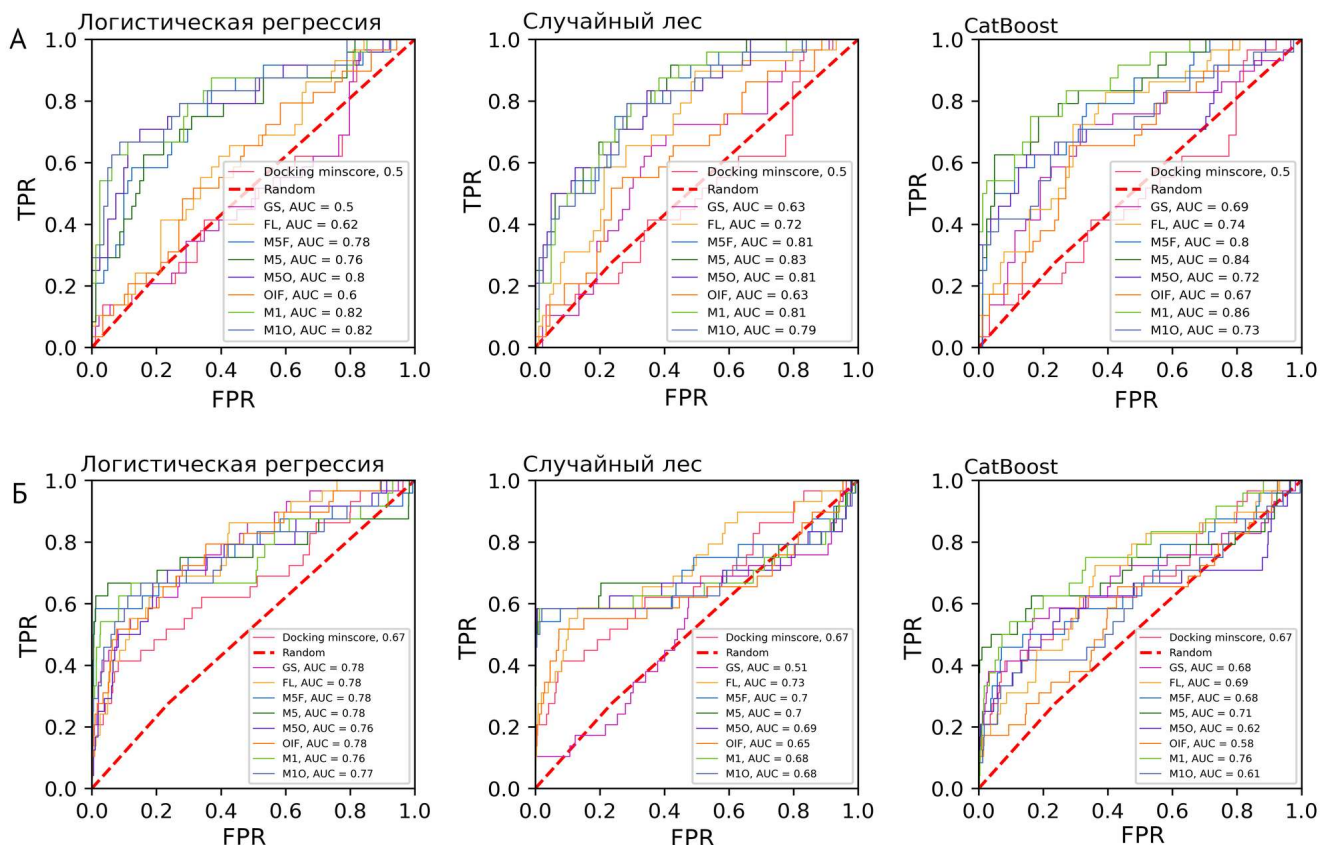


Рисунок 7. ROC-кривые прогнозов различных моделей для контрольных выборок, состоящих только из молекул библиотеки *AA-Feb21* (A), и для контрольной выборки, дополненной неактивными соединениями из *DD-ZINC* (Б). Наборы данных: GS — только индивидуальные оценки докинга в структуры ансамбля, FL — GS + Flare-IFP, M5 — GS + 512-битные fingerprints Моргана, M5F — FL + M5, OIF — GS + ODDT-IFP, M1 — GS + 1024-битные fingerprints Моргана, M1O — M1+ ODDT-IFP.

В случае моделей *AA-Feb21* включение fingerprints взаимодействия позволяет получить более высокие значения F1-меры и AUC по сравнению с моделями, построенными только на основе fingerprints Моргана, оценочных функций докинга и молекулярных дескрипторов. Для моделей *CHEMBL4495582* такой закономерности не наблюдалось, что можно связать с более высоким качеством аннотации библиотеки *AA-Feb21* и наличием в ней большего количества примеров молекул, специфически связывающихся в активном центре. Добавление fingerprints взаимодействия в модели способствует тому, что в прогнозах меньше ложноположительных результатов по сравнению с моделями, содержащими информацию только о структуре малых молекул.

Библиотека *AA-Feb21* меньше по размеру и менее разнообразна, чем результаты репозиционирования лекарственных средств *CHEMBL4495582*, кроме того, она собрана из разных источников, из-за чего данные, полученные в разных условиях, могут иметь разные распределения, нормировать которые не представляется возможным. Но в этой библиотеке

используется более надежная количественная характеристика активности соединений — IC_{50} , а не процент ингибирования при фиксированной концентрации, как при высокопроизводительном скрининге. Соединения библиотеки *AA-Feb21* обладают более высокой специфичностью к главной протеазе коронавируса и имеют значительно меньший дисбаланс классов, чем результаты высокопроизводительного скрининга лекарственных средств. С другой стороны, *CHEMBL4495582* обладает высоким разнообразием, которое должно вести к более высокой обобщающей способности модели. Однако проценты ингибирования $Mrgo$ при фиксированной концентрации соединения оказались недостаточно надежной характеристикой активности.

Точность классификационных моделей превышает точность виртуального скрининга методом ансамблевого докинга, и, более того, докинга в одну структуру. Тем не менее, целесообразность применения таких моделей для виртуального скрининга ограничена ввиду значительного количества дополнительных вычислительных процедур, необходимых для их использования, при незначительном повышении качества. Впрочем, дальнейшая разработка моделей машинного обучения, использующих информацию о связывании молекулы с целевым белком, перспективна. Преимущество фингерпринтов взаимодействия в возможности приоритизации соединений новых хемотипов, отличающихся от соединений, использованных при построении модели. Это расширяет прогностическую способность модели по отношению к “чистому” лиганд-обоснованному виртуальному скринингу, однако требует разработки аннотированных библиотек большего размера, содержащих информацию о разнообразных соединениях с надежными показателями активности (например, IC_{50}).

В седьмой главе описано **применение ансамблевого докинга для поиска лигандов белка NS1 ортофлавивирусов**. В отличие от главной протеазы коронавируса, для белка NS1 ортофлавивирусов нет большого количества опубликованных структур, причем только шесть из доступных структур имеют полную аминокислотную последовательность. Также нет структур со связанными лигандами, поэтому поиск потенциальных сайтов связывания осуществляли двумя разными расчетными методами (FTSite, DoGSiteScorer). Обоими методами были обнаружены карманы в области рядом с сайтом димеризации, содержащей аминокислотные остатки с доказанной значимостью для димеризации белка (рис. 8). Эта область была выбрана для виртуального скрининга потенциальных лигандов белка NS1.

Несмотря на малое количество доступной экспериментальной структурной информации, процедура систематического отбора ансамбля по СКО может быть применена для моделей белков NS1 различных ортофлавивирусов, построенных по гомологии. Использование для виртуального скрининга такого ансамбля, составленного из структур белка разных вирусов, могло бы способствовать приоритизации потенциальных лигандов широкого спектра действия.

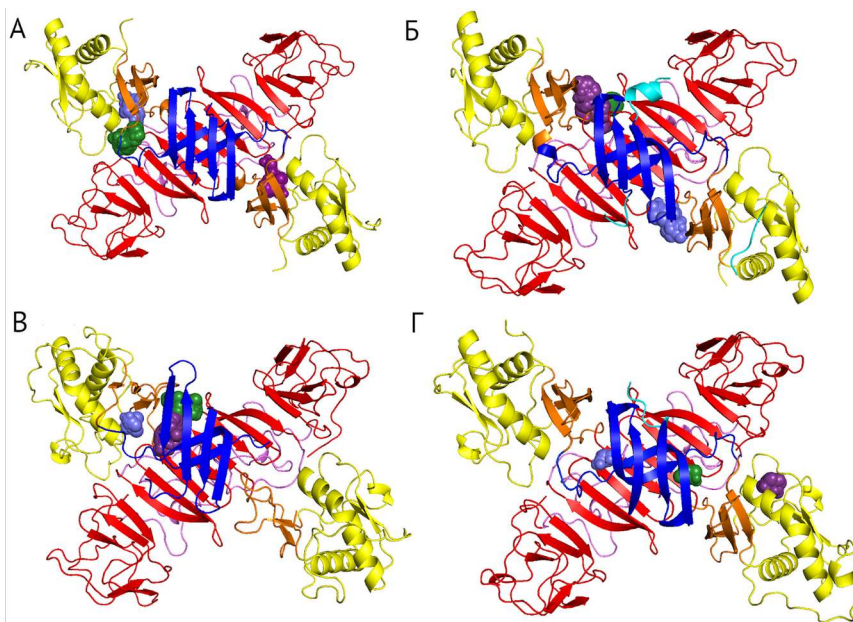


Рисунок 8. Карманы FTSite для димерных структур. Карманы отмечены по убыванию оценки фиолетовым, темно-зеленым и светло-синим. А — 4O6C (ВЛЗН), Б — 4O6D (ВЛЗН), В — 5GS6 (ЗИКВ), Г — 5K6K (ЗИКВ).

Из анализа гомологии аминокислотных последовательностей белка NS1 эпидемиологически значимых ортофлавириусов можно сделать вывод о низком перекрестном подобии (процент гомологичности от 34 до 40) белков NS1 вирусов, переносимых клещами (TBFV) и комарами (MBFV). Как следствие, доступные в PDB кристаллические структуры, принадлежащие только MBFV, могут быть недостаточно репрезентативны относительно разнообразия последовательностей белка NS1 различных ортофлавириусов, ограничивая достоверность моделей структуры NS1 TBFV, построенных по доступным шаблонам. Более того, существование двух различных групп последовательностей белка NS1, соответствующих филогенетическому делению рода *Orthoflavivirus*, может быть достаточным для возникновения селективности потенциальных лигандов относительно вирусов с разными переносчиками.

Моделирование по гомологии проводили при помощи программы Modeller 10.5. В рамках описанной ранее процедуры формирования ансамбля для всех моделей структуры белка NS1 были рассчитаны попарные СКО С α -атомов остатков, формирующих карман в области интерфейса димеризации. Отбор ансамблей по разнообразию конформаций аминокислотных остатков, формирующих потенциальный сайт связывания, не позволил сформировать репрезентативный ансамбль приемлемого размера ввиду очень высокого конформационного разнообразия сгенерированных координат. При раздельном рассмотрении моделей белка NS1 MBFV и TBFV заметно, что вторые имеют более высокое конформационное разнообразие.

В качестве альтернативы формированию ансамбля на основе СКО был использован подход, основанный на результатах докинга. При этом конформационное пространство белков

описывается неявным образом на основе корреляции рангов результатов докинга: при высокой корреляции модели считаются условно идентичными, при низкой — модели считаются принадлежащими принципиально разным ансамблям, модели же с промежуточными значениями коэффициента корреляции могут быть использованы для консенсусного ранжирования. Для сравнения ранжирующей способности моделей был проведен докинг библиотеки *DD-ZINC5000* (5000 разнообразных соединений из библиотеки *DD-ZINC*) в карман в области сайта димеризации всех сгенерированных моделей и кристаллических структур при помощи программы Flare.

Анализировали попарную корреляцию между СКО α -атомов остатков активного центра, гомологией аминокислотных последовательностей и коэффициентом корреляции рангов результатов докинга. При общем рассмотрении всех пар структур чётко выраженной зависимости между этими параметрами не наблюдается, однако попарные коэффициенты корреляции рангов не превышают 0.739. Следовательно, имеющиеся модели конформационно различны и могут быть использованы для консенсусного ранжирования. Например, в качестве ансамбля можно выбрать тройку моделей с наибольшей суммой коэффициентов корреляции рангов. В этом ансамбле (рис. 9А) представлены TBFV и MBFV, и он может быть использован для поиска лигандов широкого спектра действия. Аналогичным образом могут быть составлены ансамбли из четырёх и более структур.

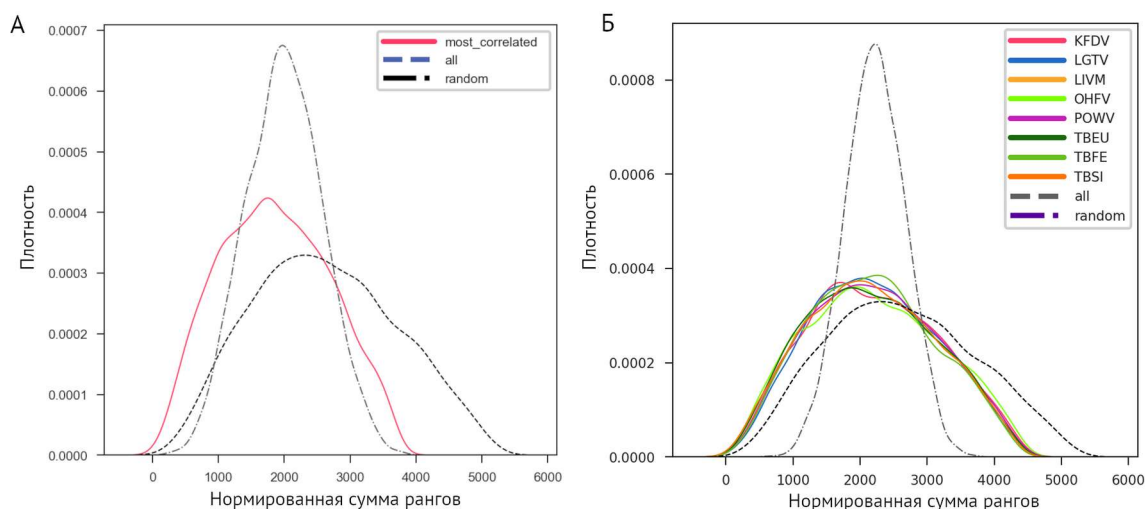


Рисунок 9. Распределение суммы рангов для результатов докинга библиотеки *DD-ZINC5000*. (А) — нормированная сумма по ансамблю, составленному из модели NS1 ДЕНВ4 по шаблону 4O6C, модели NS1 ВЯЭ по шаблону 5K6K, модели NS1 вируса шотландского энцефаломиелита овец по шаблону 5GS6. (Б) — нормированная сумма по моделям, принадлежащим белку NS1 одного вируса (для вирусов, распространяемых клещами).

При рассмотрении распределений оценок докинга не выявлено корреляции между оценками, типами шаблонных структур и аминокислотными последовательностями. Это согла-

суется с другими результатами, подтверждающими высокий вклад стохастической генерации координат при моделировании по гомологии. В такой ситуации закономерно применять в качестве консенсусной оценки сумму рангов по ансамблю. Анализ распределения суммы рангов провели для ансамблей моделей на основе общего шаблона или общей аминокислотной последовательности с целью выявления возможных закономерностей, связанных с различием последовательностей белка NS1 как индивидуальных вирусов, так и сгруппированных по переносчикам. Для сравнения были построены распределения суммы случайно сгенерированных рангов и суммы рангов по всем моделям. Все рассмотренные ансамбли моделей имели более широкие распределения, чем сумма случайных рангов (рис. 9Б), что демонстрирует их способность приоритизировать молекулы лучше случайных моделей. Также у ансамблей, состоящих из 7 моделей NS1 одного вируса по разным шаблонам, плечо в области малых сумм рангов расположено левее, чем у распределения суммы рангов по всем моделям. Это означает, что в результатах докинга в такие ансамбли консенсусно приоритизируются одни и те же молекулы. При этом для разделенных по переносчику ансамблей число молекул с нормированной на размер ансамбля суммой рангов ≤ 500 (потенциальные хиты виртуального скрининга) больше, чем для ансамблей, включающих белки NS1 всех исследованных вирусов. Разделение TBFV и MBFV при построении ансамблей следует считать эффективной стратегией при отборе низкомолекулярных соединений для экспериментального исследования. Так были сформированы два ансамбля (табл. 3).

Таблица 3. Ансамбли моделей белка NS1 TBFV и MBFV.

<i>MB3</i>		<i>TB3</i>	
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

Библиотека *DD-ZINC5000* была докирована в структуры ансамблей. Ансамбль *TB3* приоритизирует соединения с большей молекулярной массой, чем ансамбль *MB3*, что соответствует большему объему карманов у моделей структур белка NS1 TBFV. Хиты виртуального скрининга характеризуются большей липофильностью и меньшей расчетной растворимостью, чем в целом по выборке. Сравнение распределений индекса молекулярной формы позволило сделать вывод, что ансамбли приоритизируют удлиненные линейные моле-

кулы, что связано с продолговатой формой потенциального сайта связывания и расположением в нем остатков, способных участвовать в формировании водородных связей.

На основе результатов виртуального скрининга была сформулирована фармакофорная гипотеза (рис. 10) на основании частоты контактов лигандов и остатков, формирующих потенциальный сайт связывания. Все остатки, формирующие наибольшее количество контактов в ансамбле *MB3*, находятся на периферии кармана. Это связано с его меньшим объемом и более закрытой конформацией петли β -бочки. При этом закрытое положение петли поддерживается водородными связями между остатками β -бочки цепи В (Asp2, Asp8, Phe21) и остатком Lys190А. В моделях ансамбля *TB3* карман пребывает в более открытой конформации, фиксируемой только водородной связью между остатками Lys190А и Asp2В. Этим также можно объяснить смещение распределения молекулярной массы хитов виртуального скрининга в ансамбле *MB3* в область меньших значений и их менее линейную форму.

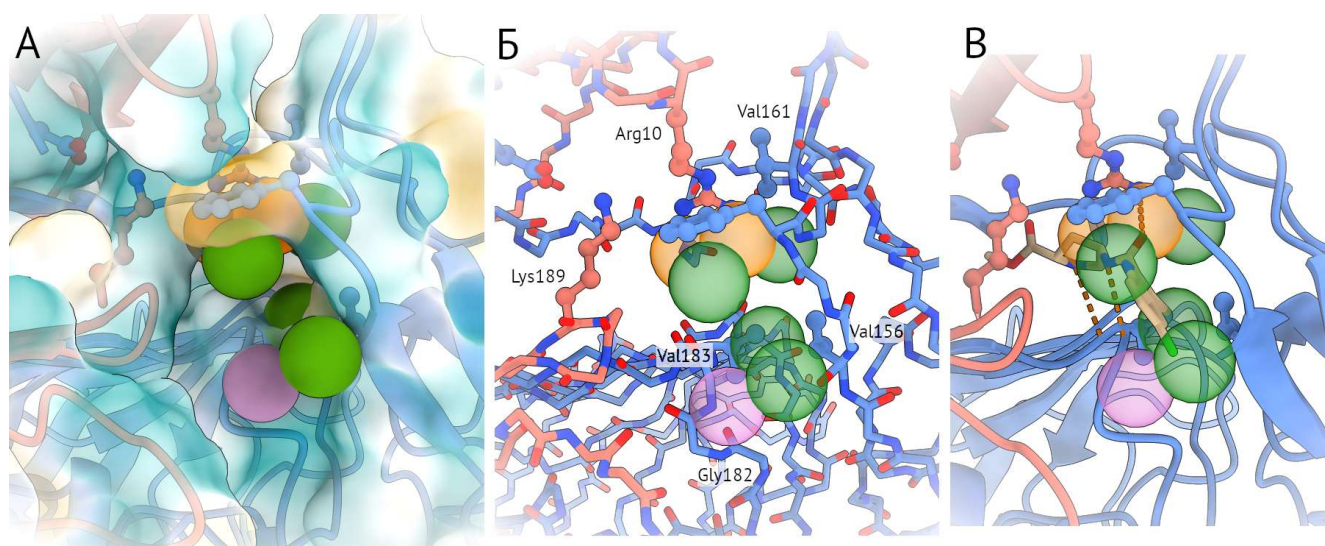


Рисунок 10. Фармакофорная гипотеза для белка NS1 ортофлавивирусов, переносимых клещами. Зеленые сферы — гидрофобные взаимодействия, оранжевые — акцепторы водородной связи, розовые — доноры водородной связи. А — проекция с отображением поверхности, окрашенной по гидрофобности, Б — проекция с указанием ключевых остатков и атомов основной цепи, В — проекция с лигандом в активном центре. Цепь А окрашена голубым, цепь В — светло-красным.

Из хитов виртуального скрининга лигандов белка NS1 ортофлавивирусов, переносимых комарами и ортофлавивирусов, переносимых клещами, сформированы фокусированные библиотеки, однако для дальнейшей разработки таких соединений требуется создание методов определения эффективности связывания малых молекул с белком NS1, которые можно будет использовать для экспериментального скрининга.

ЗАКЛЮЧЕНИЕ

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*]пиримидиноны. 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 работ, опубликованных по теме диссертации

Научные статьи, опубликованные в журналах Scopus, Web of Science, RSCI,
рекомендованных для защиты в диссертационном совете МГУ

по специальности 1.4.16. Медицинская химия

1. **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. – Vol. 43, No. 8. – P. e202300279. DOI 10.1002/minf.202300279. – EDN APAGEM. Импакт-фактор **3,1 (JIF)**, 2.54 п.л., доля вклада автора 70%;
2. **Фомина А.Д.**, Палюлин В.А., Осолодкин Д.И. Моделирование структуры белка NS1 ортофлавириусов по гомологии для виртуального скрининга потенциальных лигандов // *Биомедицинская химия*. – 2024. – Т. 70, № 6. – С. 456-468. – DOI 10.18097/PBMC20247006456. – EDN PSYRQY. Импакт-фактор **0,941 (РИНЦ)**, 1.5 п.л., доля вклада автора 90%;
[перевод **Fomina A.D.**, Palyulin V.A., Osolodkin D.I. Homology modeling of the orthoflavivirus NS1 protein for virtual screening of potential ligands // *Biomeditsinskaya Khimiya*. – 2024 – Vol. 70, № 6. – P. 456-468. Импакт-фактор **0,217 (SJR)**];
3. Zakharova M.Y., Kaliberda E.N., Kurbatskaia I.N., Smirnov I.V., Knorre V.D., Gabibov A.G., Kuznetsova A.A., Bulygin A.A., Fedorova O.S., Kuznetsov N.A., Uvarova V.I., **Fomina A.D.**, Kozlovskaya L.I., Osolodkin D.I., Ishmukhametov A.A., Egorov A.M., Varnek A. Pre-Steady-State Kinetics of the SARS-CoV-2 Main Protease as a Powerful Tool for Antiviral Drug Discovery // *Frontiers in Pharmacology*. – 2021. – Vol. 12. – P. 773198. – DOI 10.3389/fphar.2021.773198. – EDN UCTNDK. Импакт-фактор **4,8 (JIF)**, 1.5 п.л., доля вклада автора 10%³.
4. **Фомина А.Д.**, Осолодкин Д.И. Программный комплекс для систематического ансамблевого докинга // Свидетельство о государственной регистрации программы для ЭВМ № 2024684465 Российская Федерация. – Номер заявки 2024684027. – Дата регистрации 17.10.24. – Дата публикации 17.10.24, Бюл. № 10. – EDN CCPFAG. доля вклада автора 90%.
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 Российская Федерация. – Номер заявки 2024624786. – Дата регистрации 29.10.2024. – Дата публикации 15.11.2024, Бюл. № 11. – EDN LPGAGO. доля вклада автора 90%.

3 Автором проделана работа по виртуальному скринингу новых хемотипов потенциальных противокоронавирусных соединений, которая легла в основу экспериментального исследования.