

©Коллектив авторов

МОДЕЛИРОВАНИЕ СТРУКТУРЫ БЕЛКА NS1 ОРТОФЛАВИВИРУСОВ ПО ГОМОЛОГИИ ДЛЯ ВИРТУАЛЬНОГО СКРИНИНГА ПОТЕНЦИАЛЬНЫХ ЛИГАНДОВ

А.Д. Фомина^{1,2}, В.А. Палюлин², Д.И. Осолодкин^{1,3*}

¹Федеральный научный центр исследований и разработки иммунобиологических препаратов имени М.П. Чумакова РАН (Институт полиомиелита), 108819, поселение Московский, посёлок Института полиомиелита, домовладение 8, корп. 1, Москва; *эл. почта: osolodkin_di@chumakovs.su

²Химический факультет Московского государственного университета имени М.В. Ломоносова, Москва

³Институт трансляционной медицины и биотехнологии, Первый московский государственный медицинский университет имени И.М. Сеченова Минздрава России, Москва

Белок NS1 ортофлавиринов является относительно малоизученной мишенью для разработки противоортофлавириновых препаратов широкого спектра действия. В настоящее время не опубликованы структуры белка NS1 ортофлавиринов, переносимых клещами, однако эти структуры можно моделировать по гомологии, генерируя таким образом большое количество структурной информации. Нами было проведено моделирование по гомологии структур белка NS1 эпидемиологически значимых ортофлавиринов и анализ возможности использования этих моделей для виртуального скрининга методом ансамблевого докинга. Были показаны ограничения метода и важность разделения моделей на основании организма-переносчика при отборе ансамбля.

Ключевые слова: ортофлавириновы; белок NS1; ансамблевый докинг; моделирование по гомологии

DOI: 10.18097/PBMC20247006456

ВВЕДЕНИЕ

Род *Orthoflavivirus* включает оболочечные вирусы, переносимые клещами и комарами [1] и вызывающие опасные заболевания, такие как лихорадка денге, лихорадка Зика и лихорадка Западного Нила. В России и странах СНГ наибольшее среди ортофлавиринов распространение имеет вирус клещевого энцефалита (ВКЭ). Заболеваемость клещевым энцефалитом в 2020 году в мире составила до 0,9 случаев на 100 тыс. человек, что выше показателей 0,7 случаев в 2019 году и 0,4 случаев в 2015 году [2]. В России самый высокий зарегистрированный уровень заболеваемости был отмечен в 1996 году (7,0 случаев на 100 тыс. населения), в 2020 году он составил 0,66 случаев на 100 тыс. человек [3]. Кроме того, в связи с расширением ареала обитания комаров, на территории России регистрируют случаи заболевания лихорадкой Западного Нила [4]. К 2024 году не существует разрешённых к медицинскому применению противовирусных препаратов прямого действия против инфекций, вызванных флавириновыми. Для профилактики некоторых инфекций, вызванных ортофлавириновыми (ВКЭ [5], вирусом денге (ДЕНВ) [6], вирусом жёлтой лихорадки (ВЖЛ) [7], вирусом японского энцефалита (ВЯЭ) [8]) разработаны вакцины, однако они не могут быть использованы для лечения заболеваний после заражения, не обеспечивают полной защиты от инфекции, а их разработка и широкое применение осложнены тем, что для некоторых флавиринов (ДЕНВ, вируса Зика (ЗИКВ)) наблюдается антителозависимое усиление инфекции [9, 10]. Например, было показано, что вакцинация против ВКЭ вызывает при последующей вакцинации против ВЖЛ выработку малоэффективных кросс-реактивных антител [11].

Геном ортофлавиринов представлен одноцепочечной РНК положительной полярности длиной 10–11 тыс. оснований [12]; в ходе её трансляции образуется полипротеин, расщепляемый вирусными и клеточными протеазами на три структурных и семь неструктурных белков (рис. 1А). Среди них особую роль играет неструктурный белок NS1, в острой форме протекания заболеваний вызывающий нарушение целостности кровеносных сосудов за счёт взаимодействия с клетками эндотелия [13, 14] и разрушения гликокаликса [15]. Одновременно с этим белок NS1 участвует в подавлении иммунного ответа, связываясь с белками комплемента на поверхности заражённой клетки [16, 17]. Также белок NS1 задействован в репликативном цикле вируса, стабилизируя инвагинации мембран эндоплазматического ретикулума и размещённый в них репликативный комплекс [18, 19]. Малые молекулы, взаимодействующие с белком NS1, могут быть использованы в качестве соединений-лидеров для разработки противоортофлавириновых препаратов. Предлагаются такие механизмы действия, как модулирование связывания белка NS1 с мембранами или другими белками [20] или ингибирование его димеризации [21–23]. Однако комплексы белка NS1 ортофлавиринов с малыми молекулами не были описаны экспериментально, что затрудняет применение классических методов структурно-обоснованного дизайна лигандов.

Тем не менее, структура апо-формы белка NS1 переносимых комарами ортофлавиринов была исследована методами рентгеноструктурного анализа и криоэлектронной микроскопии (табл. 1, Дополнительные материалы, табл. ST1). Согласно данным из PDB, белок NS1 ортофлавиринов

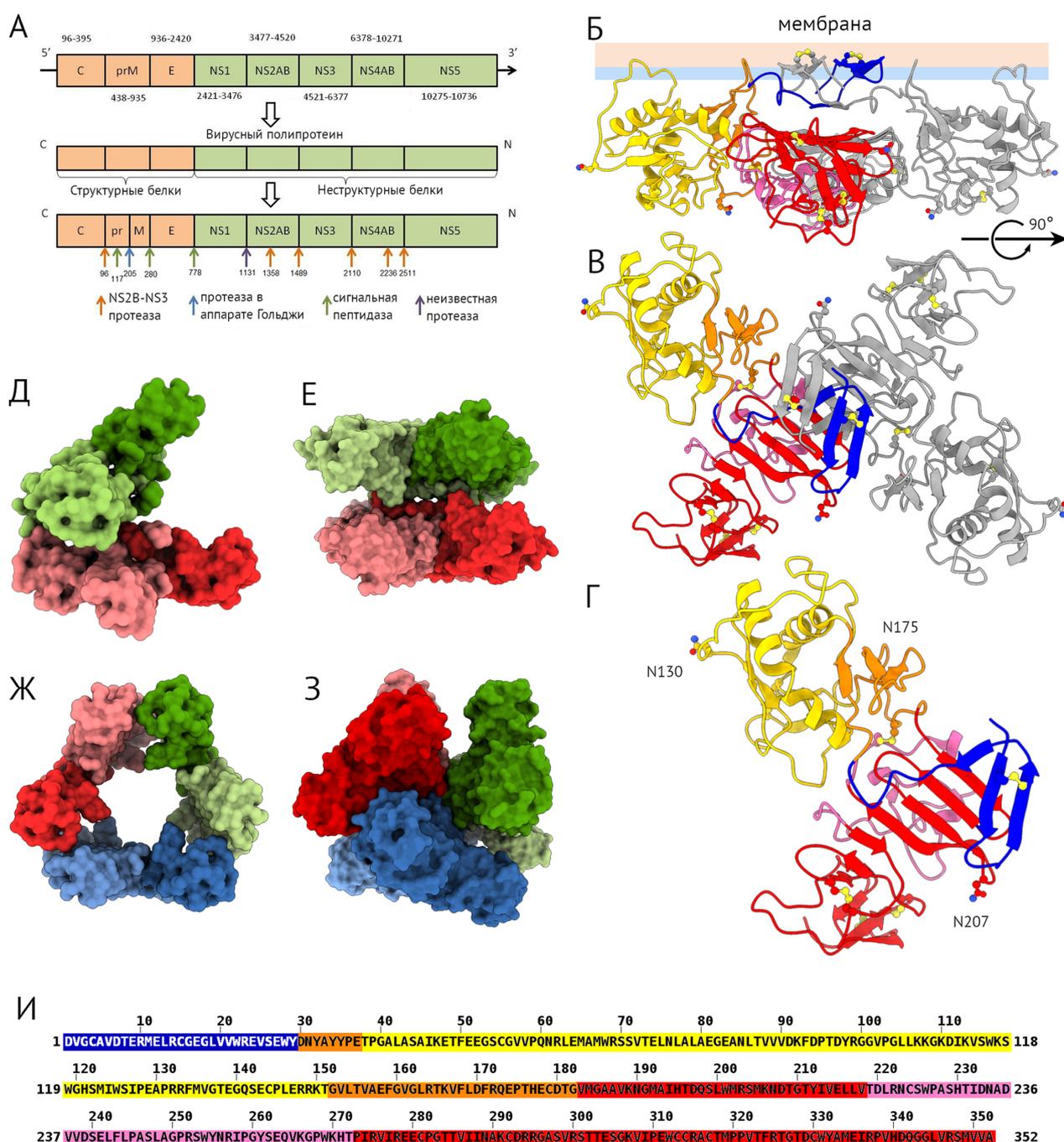


Рисунок 1. Генетические и структурные основы функционирования белка NS1 ортофлавириусов. (А) – Схема генома и полипротеина ортофлавириусов, нумерация приведена для последовательностей вируса денге серотипа 1 (GenBank ID: U88536). (Б–Г) – структура NS1, установленная методом криоэлектронной микроскопии (PDB ID 7WUS [24]): (Б, В) – в форме димера, (Г) – в форме мономера. Цветовое кодирование: β-бочка (а.о. 1–29) – синий, домен-коннектор (а.о. 30–37, 152–182) – оранжевый, α-спиральное крыло (а.о. 38–151) – желтый, красным – β-лестница (а.о. 183–352), розовый – “спагетти-петля” (а.о. 219–272). Отмечены сайты гликозилирования – Asn130 и Asn270 (Asn175 гликозилирован у вируса желтой лихорадки), дисульфидные связи показаны в виде шаростержневых моделей. (Д–З) – мультимерные формы белка NS1: (Д) – “мягкий” тетрамер (PDB ID 7WUU [24]), (Е) – “стабильный” тетрамер (PDB ID: 7WUT [24]), (Ж) – секретируемый гексамер “голова-к-голове” (PDB ID: 7WUV [24]), (З) – гексамер “сторона-к-стороне” (PDB ID: 8WBE [25]). (И) – последовательность белка NS1 ВКЭ, штамм Абсеттаров (GenBank KU885457.1), окрашенная по доменам белка. Цветной вариант рисунка представлен в электронной версии статьи на сайте журнала.

МОДЕЛИРОВАНИЕ ПО ГОМОЛОГИИ БЕЛКА NS1 ОРТОФЛАВИВИРУСОВ

Таблица 1. Полные структуры белка NS1 в PDB

PDB ID	Вирус	Год	Метод	Разрешение, Å	Олигомерное состояние	Ссылка
4O6D	ВЛЗН	2014	Рентгеноструктурный анализ	2,59	Димер	[28]
4O6C	ВЛЗН	2014	Рентгеноструктурный анализ	2,75	Димер	[28]
4TPL	ВЛЗН	2014	Рентгеноструктурный анализ	2,90	Димер	[32]
4O6B	ДЕНВ2	2014	Рентгеноструктурный анализ	3,00	Димер	[28]
7WUS	ДЕНВ2	2022	Криоэлектронная микроскопия	3,40	Димер	[24]
7WUT	ДЕНВ2	2022	Криоэлектронная микроскопия	3,50	Тетрамер	[24]
7WUU	ДЕНВ2	2022	Криоэлектронная микроскопия	8,00	Тетрамер	[24]
7WUV	ДЕНВ2	2022	Криоэлектронная микроскопия	8,30	Тетрамер	[24]
7WUR	ДЕНВ2	2022	Криоэлектронная микроскопия	3,50	Комплекс димера с антителами	[24]
5K6K	ЗИКВ	2016	Рентгеноструктурный анализ	1,89	Димер	[33]
5GS6	ЗИКВ	2016	Рентгеноструктурный анализ	2,85	Димер	[34]

представляет собой гликопротеин, состоящий из 352 аминокислотных остатков (а.о.). Молекулярная масса мономера составляет 48–50 кДа в зависимости от аминокислотной последовательности и состояния гликозилирования. На разных стадиях инфекции в заражённом организме преобладают различные мультимерные состояния белка NS1: на начальном этапе белок NS1 в основном представлен в виде экспонированного в цитоплазму димера на поверхности мембраны ЭПР или во внеклеточное пространство на поверхности клеточной мембраны (рис. 1Б,В). В дальнейшем белок секретируется в кровь в виде тетрамера, существующего в двух формах — “мягкой” (рис. 1Д) и “стабильной” (рис. 1Е), — или циклического гексамера (рис. 1Ж), внутренняя полость которого заполнена липидами [24, 26]. Кроме того, для рекомбинантных белков NS1 наблюдали гексамер с плотной упаковкой субъединиц (рис. 1З) [25]. В структуре мономера выделяют четыре домена: β -бочка (β -barrel), домен-коннектор (connector), α -спиральное крыло (α -wing) и β -лестница (β -ladder) (рис. 1Г) [27]. При этом поверхности β -бочки и α -спирального крыла гидрофобны и ориентированы внутрь мембран или липидного ядра гексамера, а “спагетти-петля” (spaghetti-loop, часть β -лестницы) ориентирована наружу [28]. Стабилизацию мультимеров, а также взаимодействие с компонентами системы комплемента обеспечивают сложные гликаны, связанные с Asn130 и Asn207 (а также Asn175 у белка NS1 ВЖЛ) [28–31].

Доступность структурных данных позволяет использовать докинг в качестве инструмента поиска потенциальных лигандов, а достаточное количество структур позволяет использовать ансамблевый докинг, в котором учитывается конформационная подвижность мишени за счёт докинга в набор разных структур одного белка, что обогащает результаты виртуального скрининга [35]. Ранее мы разработали методику систематического ансамблевого докинга, позволяющую использовать всю доступную структурную информацию о белке для структурно-обоснованного виртуального скрининга малых органических молекул [36]. Поскольку до настоящего

времени отсутствует структурная информация как о белке NS1 ВКЭ, так и о взаимодействии малых органических молекул с белком NS1 ортофлавириусов в целом, мы провели моделирование по гомологии белка NS1 ВКЭ и анализ структуры белка NS1 для идентификации потенциальных сайтов связывания низкомолекулярных соединений. На основе полученного набора моделей мы оценили применимость нашей методологии систематического ансамблевого докинга к неоднородным по аминокислотной последовательности наборам структур белка.

МЕТОДИКА

Анализ данных о структуре белка NS1 флавириусов

Извлечение структурной информации. Структуры белка NS1 флавириусов (Дополнительные материалы, табл. ST1) извлекали из БД PDB [37] путём анализа результатов поиска по запросу Full Text = “flavivirus NS1 protein” AND Parent Scientific Name (typically superkingdom or clade) = “Riboviria”. Для извлечения некорректно аннотированных структур, которые не идентифицируются поиском среди белков вирусов рода *Orthoflavivirus*, была использована программа BLASTP [38], в качестве запроса для которой фигурировала последовательность белка NS1 вируса денге серотипа 1 (GenBank ID AAB70695). Таким образом, в выборку были добавлены 2 структуры (PDB ID: 5O19, 5O36). Проверка стереохимической корректности структур была выполнена при помощи веб-сервиса ProCheck [39].

Пространственное выравнивание структур. Пространственное выравнивание проводили в программе VMD 1.9.3 [40]. Для выравнивания димерных структур объединяли цепи в файлах PDB в одну, затем проводили выравнивание по всем C_{α} -атомам объединённой цепи, а затем разделяли цепи обратно при помощи скрипта на Python 3.9.

После выравнивания рассчитывали значения среднеквадратичного отклонения (RMSD) C_{α} -атомов для всех пар структур. Визуализацию попарных RMSD осуществляли методом тепловых карт, реализованным

на Python 3.9 с использованием библиотек Pandas 1.4.1 [41], NumPy 1.21.6 [42], Matplotlib 3.5.1 [43] и seaborn 0.11.2 [44] аналогично процедуре, описанной в [36].

Поиск потенциальных сайтов связывания малых молекул

Поиск карманов — полостей в белке, в которых может происходить специфическое связывание малых молекул, — проводили для мономера и димера наиболее различных полных структур белка (4O6C, 4O6D, 5K6K, 5GS6), очищенных от сокристаллизованных молекул растворителя, при помощи двух программ, в основе которых лежат принципиально различные методы идентификации потенциальных сайтов связывания малых молекул: FTSite (FTMap) [45–47] и ProteinsPlus (DoGSiteScorer) [48, 49]. Обе программы использовались через веб-интерфейс. Результаты работы программ представляют собой наборы точек или сфер, определяющих потенциальные карманы. Для структур 4O6D и 5GS6 поиск карманов проводили до и после удаления полигистидинового якоря, однако результаты совпали, поэтому далее его наличие не учитывали.

В программе FTSite реализован энергетический метод поиска карманов. В ходе поиска проводится жёсткий докинг 16 малых молекулярных зондов, результаты докинга кластеризуют для каждого зонда, и объединённые кластеры ранжируются на основе общего количества взаимодействий между белком и всеми зондами кластера.

Алгоритм программы ProteinsPlus (DoGSiteScorer) относится к геометрическим алгоритмам и основан на методах молекулярной формы. Вокруг молекулы белка генерируется сеточное поле, и точки сетки размечаются как занятые, если они лежат на расстоянии ван-дер-ваальсового радиуса от какого-либо атома белка, и как свободные в противном случае. На основании этой разметки проводится выбор карманов, которые достаточно велики, чтобы вместить хотя бы один тяжёлый атом, и их ранжирование по объёму. В случае анализа димеров мы выбирали три кармана наибольшего объёма и симметричные им.

Выравнивание аминокислотных последовательностей и моделирование по гомологии

Выравнивание аминокислотных последовательностей проводили с помощью веб-приложения MUSCLE [50, 51]. В качестве базовой использовали последовательность NS1 ВКЭ штамма Абсеттаров (GenBank KU885457.1), а также последовательности, извлечённые из pdb-файлов, и последовательности белков NS1 эпидемиологически значимых ортофлавириусов (Дополнительные материалы, табл. ST3).

Моделирование по гомологии проводили при помощи программы Modeller 10.5 [52]. На основе каждого индивидуального шаблона генерировали 100 моделей, также генерировали 300 моделей по всем шаблонам одновременно

с функциями DOPE и GA341 в качестве оценки. Моделирование проводили с неявно заданными атомами водорода (класс automodel). Использовали “медленный” алгоритм оптимизации с максимальным количеством итераций, равным 300. Оптимизацию повторяли 2 раза. Поскольку при отборе итоговой модели по оценке DOPE оптимальные модели имели очень большой разброс попарных среднеквадратичных отклонений (RMSD), итоговую модель для сравнения и докинга выбирали на основании наименьшего среднеквадратичного отклонения C_α-атомов от структуры 7WUS, так как это единственная кристаллическая структура, не содержащая неразрешённых участков (все структуры, смоделированные в работе, представлены в Дополнительных материалах).

Построение выборки разнообразных соединений

Выборку 5000 разнообразных соединений отбирали из ранее сформированной библиотеки 20000 различных веществ [36] по следующим критериям: молекулярная масса от 200 Да до 400 Да, липофильность (clogP [53]) между -5 и 5 включительно, число доноров водородной связи не более 5, число акцепторов водородной связи не более 10, формальный заряд от -2 до 2, число вращаемых связей не более 8, число тяжёлых атомов от 15 до 50. Фильтрацию прошло 9341 соединение, из которых выбрали по разнообразию 5000 методом MaxMinPicker [36].

Ансамблевый докинг

Конвейер докинга был реализован при помощи ruflare [54] — библиотеки на языке Python 3, позволяющей исполнять скрипты для программ Flare 5.0.0, Cresset [55, 56]. Подготовку белков (добавление атомов водорода, расстановка зарядов при pH 7,4) осуществляли при помощи скрипта proteinprep.py. Докинг в подготовленную структуру белка проводили при помощи скрипта docking.py для каждой структуры белка отдельно. Генерацию сетки для докинга проводили на расстоянии 6 Å от молекулы ZINC000734046780 (Дополнительные материалы, рис. SF1), размещённой вручную при помощи UCSF Chimera в кармане между β-бочкой, β-лестницей и α-спиральным крылом. Докинг проводили с параметром качества “normal”, что соответствует стандартному алгоритму докинга, реализованному в программе Lead Finder [57]. Сохраняли 5 наиболее высоко оцененных по Rank score (оценка, оптимизированная для корректного энергетического ранжирования результатов докинга конкретной молекулы) результатов докинга для каждого соединения. Обработку результатов докинга проводили при помощи скриптов на Python 3.9. Оценки считывали из sdf-файлов и для конформеров, наиболее высоко оцененных по Rank score, проводили ранжирование на основании уменьшения модуля оценки VS score (оценка, оптимизированная для корректного ранжирования активных молекул относительно неактивных в процессе виртуального скрининга).

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Анализ данных о структуре белка NS1 флавивирусов

Поскольку для белка NS1 ортофлавивирусов не описаны сайты связывания малых молекул, необходимо было провести сравнительный анализ доступных структур (табл. 1) и определить потенциальные сайты связывания расчётными методами. Только 11 структур содержат полную аминокислотную последовательность, причём две из них не обладают достаточным разрешением (7WUU, 7WUV), одна (7WUT) представляет собой “стабильный тетрамер” и содержит нетипичную укладку β-бочки, а ещё одна является структурой комплекса белка NS1 с антителами (7WUR [24]). Остальные депонированные структуры представляют собой только часть белка NS1 с β-лестницей и спагетти-петлёй (указаны в Дополнительных материалах, табл. ST1 как C-домен). Структуры характеризуются высоким стереохимическим качеством: лишь отдельные остатки расположены в неблагоприятных областях карт Рамачандрана, что позволяет сделать вывод об общей корректности данных (Дополнительные материалы, табл. ST2).

Пространственное выравнивание мономеров и димеров по RMSD C_α-атомов проводили для полных структур с целью сравнительного анализа, выявления конформационных различий, а также определения степени симметричности мономеров в индивидуальных структурах (рис. 2, Дополнительные материалы, рис. SF2). Для количественного анализа результатов выравнивания были построены тепловые карты попарных RMSD (рис. 2).

Все структуры обладают высоким сходством, что подтверждается низкими значениями попарного RMSD: в большинстве случаев ≤1,4 Å. Такие значения RMSD характерны для структур одинаковых белков и объясняются разрешением структур [58] и общей подвижностью полипептидных цепей.

При выравнивании полных структур наиболее выраженные конформационные различия наблюдаются в области β-бочки (остатки 6–14) и в петлях α-спирального крыла. Во взаимном расположении мономеров в димере разных структур нет заметных различий. Высокие значения RMSD для структур 5GS6 и 5K6K обусловлены отклонениями в области β-бочки (5GS6-B, остатки 6–14) и α-спирального крыла (5GS6-B, A остатки 62–82) (Дополнительные материалы, рис. SF3).

Поиск потенциальных сайтов связывания малых молекул

Основываясь на анализе пространственного выравнивания, а также данных по полноте разрешения и стереохимической корректности структур, для поиска карманов на поверхности белка были выбраны наиболее различные и качественные структуры: 4O6C (ВЛЗН), 4O6D (ВЛЗН), 5K6K (ЗИКВ) и 5GS6 (ЗИКВ).

Обе программы обнаружили карман в области интерфейса димеризации между β-бочкой, доменом-коннектором, α-спиральным крылом и β-лестницей (рис. 3А), образованный остатками 1–15, 186–193, 202–207 одной цепи и 1, 16–26, 156–162, 179–185, 193–200 другой цепи (рис. 3Б). Во всех структурах карманы содержат в основном гидрофильные остатки — как доноры, так и акцепторы водородной связи. Гидрофобные боковые цепи остатков β-бочки ориентированы вовне кармана и формируют гидрофобную поверхность, которая способствует связыванию белка с мембранами (рис. 3В). Таким образом, внутри кармана оказываются доступны атомы основной цепи, также способные формировать водородные связи. Этот карман был выбран в качестве потенциального сайта связывания для дальнейшего виртуального скрининга и анализа как наиболее консервативный.

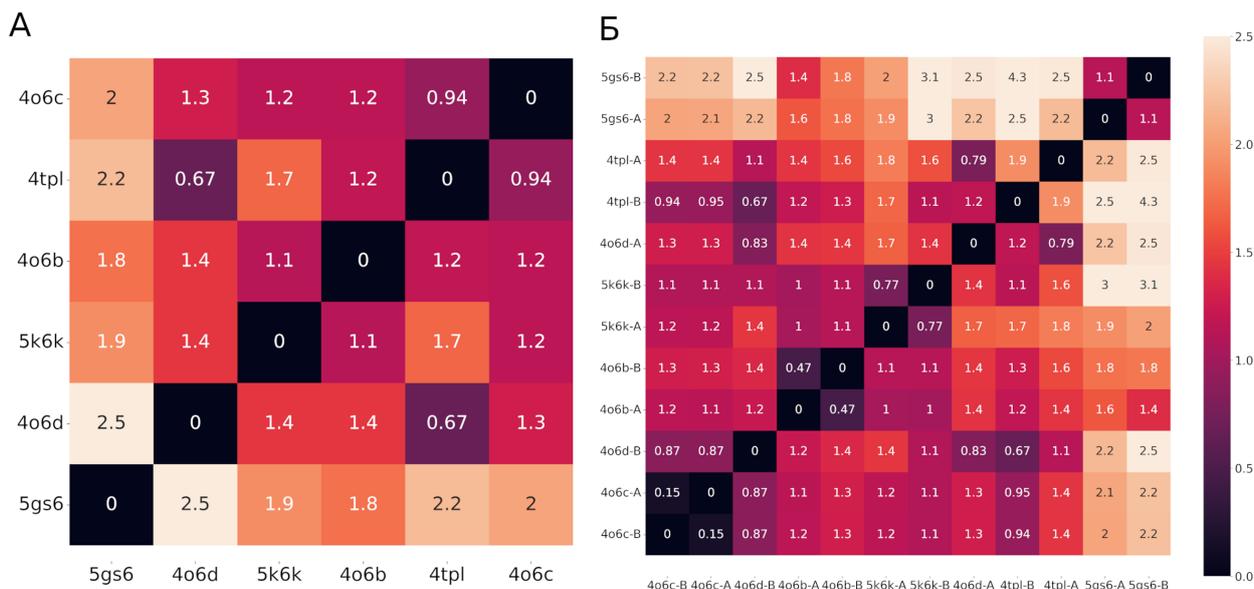


Рисунок 2. (А, Б) – Тепловые карты RMSD C_α-атомов пространственного выравнивания (А) – димеров и (Б) – мономеров полных структур NS1 флавивирусов. Цветовая шкала соответствует значениям RMSD от 0 Å до 2,5 Å. Цветной вариант рисунка представлен в электронной версии статьи на сайте журнала.

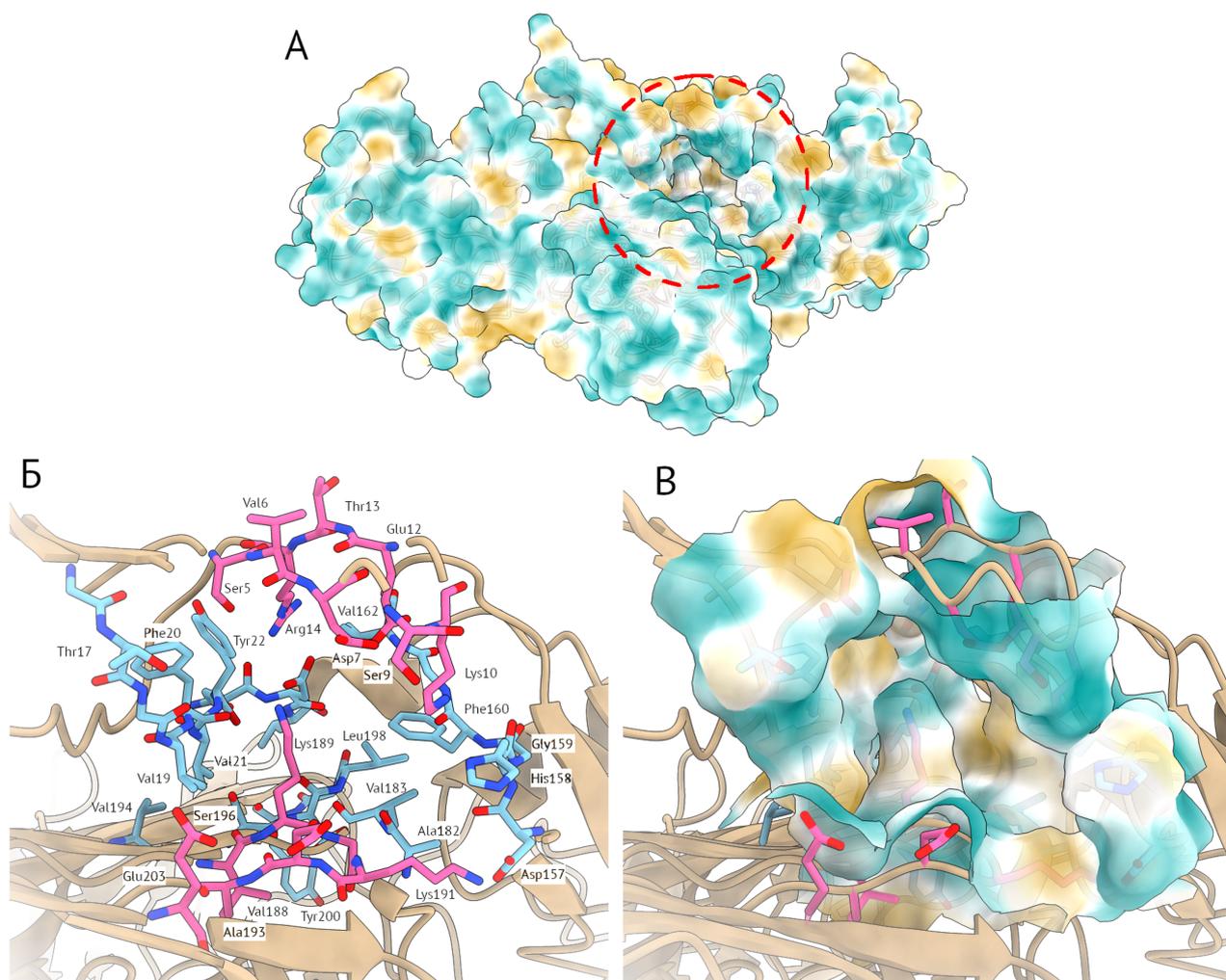


Рисунок 3. Потенциальный сайт связывания малых молекул. (А) – Расположение кармана, (Б) – остатки, формирующие карман, окрашенные в соответствии с цепью, где цепь А розовая, а цепь В – голубая, (В) – поверхность кармана, окрашенная в зависимости от гидрофобности, где жёлтый цвет отображает высокую, а голубой – низкую гидрофобность. Цветной вариант рисунка представлен в электронной версии статьи на сайте журнала.

Разнообразие аминокислотных последовательностей эпидемиологически значимых ортофлавириусов

Аминокислотные последовательности эпидемиологически значимых ортофлавириусов (Дополнительные материалы, табл. ST4) [59, 60] были получены из базы данных GenBank. Для моделирования по гомологии они были выравнены с шаблонными последовательностями, извлечёнными из файлов со структурами белка NS1, депонированных в RSCB PDB (PDB ID: 4O6B, 4O6C, 4O6D, 4TPL, 5K6K, 5GS6, 7WUS). Также было проведено отдельное выравнивание аминокислотных последовательностей белка NS1 эпидемиологически значимых ортофлавириусов с последовательностью белка NS1 ВКЭ штамма Абсеттаров (Дополнительные материалы, рис. SF4) для расчёта матрицы гомологичности последовательностей (рис. 4). На её основании можно разделить вирусы, переносимые клещами и комарами, на две группы с относительно низким перекрёстным подобием. При этом белки NS1 вирусов, переносимых клещами, обладают высоким внутрикласовым подобием,

в то время как белки NS1 вирусов, переносимых комарами, более значительно различаются между собой. Как следствие, доступные в PDB кристаллические структуры могут быть недостаточно репрезентативны применительно к разнообразию последовательностей белка NS1 ортофлавириусов. Поэтому достоверность моделей структуры NS1 ортофлавириусов, переносимых клещами, по доступным шаблонам (табл. 1) может быть ограничена. 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 структуры белка NS1 ортофлавириусов по гомологии

Модели структуры белка NS1 были построены при помощи программы Modeller 10.5 [61] методом моделирования по гомологии с использованием в качестве шаблонов всех доступных в PDB

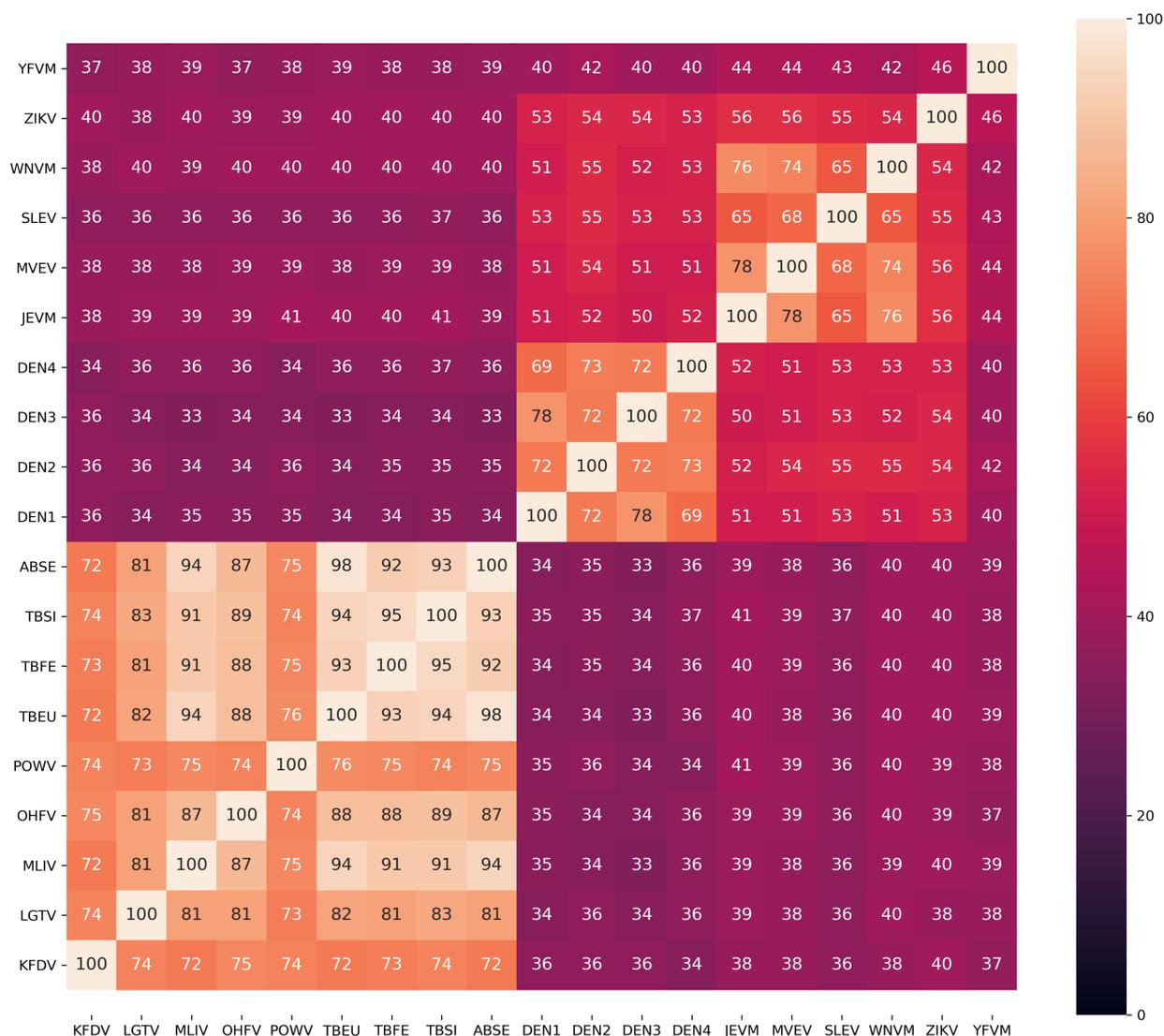


Рисунок 4. Матрица гомологичности аминокислотных последовательностей белка NS1 ортофлавириусов. Расшифровка сокращений приведена в Дополнительных материалах, таблица ST4. Цветной вариант рисунка представлен в электронной версии статьи на сайте журнала.

полных димерных структур белка NS1 (табл. 1) по отдельности, а также в комбинации (файлы со структурами представлены в Дополнительных материалах). Для дальнейшего формирования ансамбля и виртуального скрининга необходимо было для каждого шаблона и каждого вируса отобрать одну оптимальную модель. Стандартный критерий отбора основан на значении оценочной функции DOPE [62], которая носит статистический характер и приоритизирует структуры, наиболее близкие к нативному состоянию белка, на основании информации о плотности вероятности расстояний между атомами. При этом оптимальное значение DOPE в серии моделей не гарантирует того, что у соответствующей модели не будет значительного конформационного отклонения от шаблона, в частности, в подвижных областях. Поскольку для всех сгенерированных моделей значения DOPE находились в приемлемом диапазоне, а различия между ними не были значительными (Дополнительные материалы, табл. ST3), выбор модели для дальнейшего

исследования проводили не по значению DOPE, а на основе пространственного выравнивания всех моделей относительно структуры 7WUS, не содержащей неразрешённых участков. Для каждого вируса и каждого шаблона выбирали модели с наименьшим RMSD C_α-атомов относительно 7WUS.

В рамках описанной ранее процедуры формирования ансамбля [36] для всех моделей были рассчитаны попарные RMSD C_α-атомов остатков, формирующих карман в области интерфейса димеризации (V1–G15, A186–A193, E202–D207 цепи A и V1, T16–W26, D156–F162, D179–T185, V193–W200 цепи B). Отбор ансамблей по разнообразию конформаций а.о., формирующих потенциальный сайт связывания, не позволил сформировать репрезентативный ансамбль приемлемого размера ввиду очень высокого конформационного разнообразия сгенерированных координат: в широком диапазоне варьирования пороговых значений сходства и разнообразия структур в ансамбль отбирались всего две модели.

На тепловой карте RMSD аминокислотных остатков кармана (Дополнительные материалы, рис. SF5) выделяются модели, построенные по гомологии со структурой 4TPL, имеющие особенно высокие средние значения RMSD (Дополнительные материалы, рис. SF5, выделены оранжевым). Эти модели значительно отличаются от остальных, что негативно сказывается на отборе ансамбля по конформационному разнообразию.

При раздельном рассмотрении моделей белка NS1 вирусов, переносимых комарами (Дополнительные материалы, рис. SF6) и клещами (Дополнительные материалы, рис. SF7), заметно, что вторые имеют более высокое конформационное разнообразие. В случае моделей, построенных по одинаковым шаблонам, наибольшие конформационные отличия наблюдаются для вирусов Повассан (POWV) и киассанурской лесной болезни (KFDV), имеющих наименьшее сходство аминокислотных последовательностей как с белками NS1 других вирусов, переносимых клещами, так и с белками NS1 вирусов, переносимых комарами. Кроме того, последовательности белков NS1 этих вирусов имеют множественные неконсервативные замены в области, выбранной в качестве потенциального сайта связывания (Дополнительные материалы, рис. SF8). Это ставит под сомнение достоверность моделирования этих белков по шаблонным структурам белков NS1 ортофлавириусов, переносимых комарами.

На следующей стадии был проведён анализ главных компонент (PCA) координат C_{α} -атомов остатков, формирующих целевой карман (рис. 5).

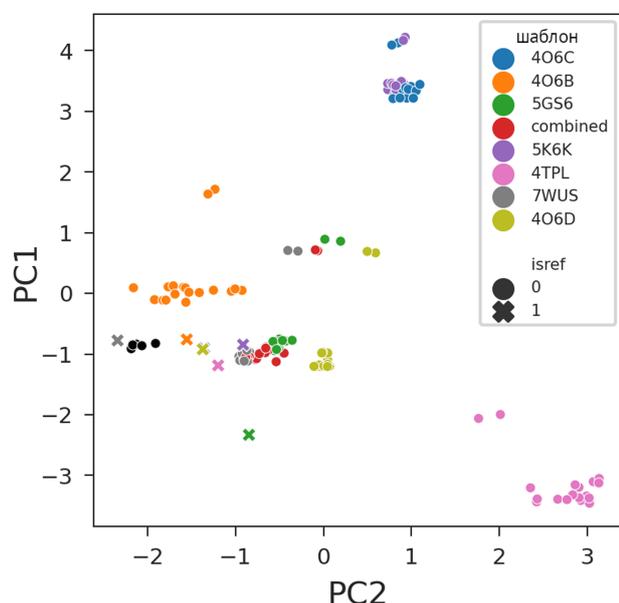


Рисунок 5. Анализ главных компонент координат C_{α} -атомов остатков, формирующих целевой карман белка NS1 ортофлавириусов моделей и шаблонных кристаллических структур. Модели окрашены по шаблону. Категория “isref” показывает, является ли структура моделью или шаблоном: модели показаны кружками, шаблоны – крестиками. Цветной вариант рисунка представлен в электронной версии статьи на сайте журнала.

Модели, построенные по шаблону 4TPL, выделяются в отдельный кластер, удалённый от основного множества моделей, что согласуется с результатами расчёта RMSD. При этом сама шаблонная структура 4TPL лежит в основном кластере рядом с остальными кристаллическими структурами (отмечены крестиками, рис. 5). Это связано с тем, что структура 4TPL содержит наибольший по длине неразрешённый участок в α -спиральном крыле (остатки 108–130) и соответствующие фрагменты структуры моделируются стохастически. В связи с этим модели по шаблону 4TPL далее не использовали. Модели по шаблонам 4O6C и 5K6K выделяются в самостоятельный кластер, однако при визуальном анализе значительных отличий от моделей, построенных по другим шаблонам, в сгенерированных структурах не наблюдается. Следует также отметить, что модели, построенные по объединённому шаблону, лежат в проекции на главные компоненты между моделями, построенными по индивидуальным шаблонам. Такие “усреднённые” модели обладают наиболее высоким сходством со всеми остальными (ср. Дополнительные материалы, рис. SF5, выделены голубым), вследствие чего при отборе ансамбля по разнообразию имеют низкий приоритет.

Модели структуры белков NS1 POWV и KFDV лежат в проекции на главные компоненты (Дополнительные материалы, рис. SF9) отдельно от основных кластеров моделей, построенных по тем же шаблонам, при этом повторяя характер расположения. Это согласуется с низкой гомологией их последовательностей с последовательностями белков NS1. Для отбора ансамблей белковых структур имеет смысл исключать из рассмотрения такие модели, поскольку они не репрезентативны относительно всего набора структур.

Удаление моделей, построенных по шаблону 4TPL, и моделей белков NS1 POWV и KFDV не способствовало улучшению результатов отбора ансамбля. Поэтому от систематического подхода к отбору ансамбля, основанного на RMSD остатков сайта связывания, было решено отказаться.

Сравнение моделей по ранжирующей способности

В качестве альтернативы формированию ансамбля на основе структуры белка был использован подход, основанный на результатах докинга. При этом конформационное пространство белков описывается неявным образом на основе корреляции рангов: при высокой корреляции модели считаются условно идентичными, при низкой корреляции 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 проведён докинг библиотеки разнообразных лекарственноподобных соединений из базы данных ZINC15 в карман в области сайта димеризации всех сгенерированных моделей и кристаллических структур при помощи программы Flare. Результаты докинга ранжировали на основании

МОДЕЛИРОВАНИЕ ПО ГОМОЛОГИИ БЕЛКА NS1 ОРТОФЛАВИВИРУСОВ

значений оценочной функции VS score. Анализировали корреляцию между попарным RMSD, гомологией аминокислотных последовательностей и коэффициентом корреляции рангов. При общем рассмотрении всех пар структур чётко выраженной зависимости между этими параметрами отмечено не было (рис. 6, Дополнительные материалы, рис. SF10, SF11), однако можно заметить, что 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 — 2,188: модель NS1 ДЕНВ4 по шаблону 4O6C, модель NS1 ВЯЭ по шаблону 5K6K, модель NS1 вируса шотландского энцефаломиелита овец по шаблону 5GS6 (рис. 7А). В этом ансамбле представлены вирусы, переносимые как клещами, так и комарами, и он может быть использован

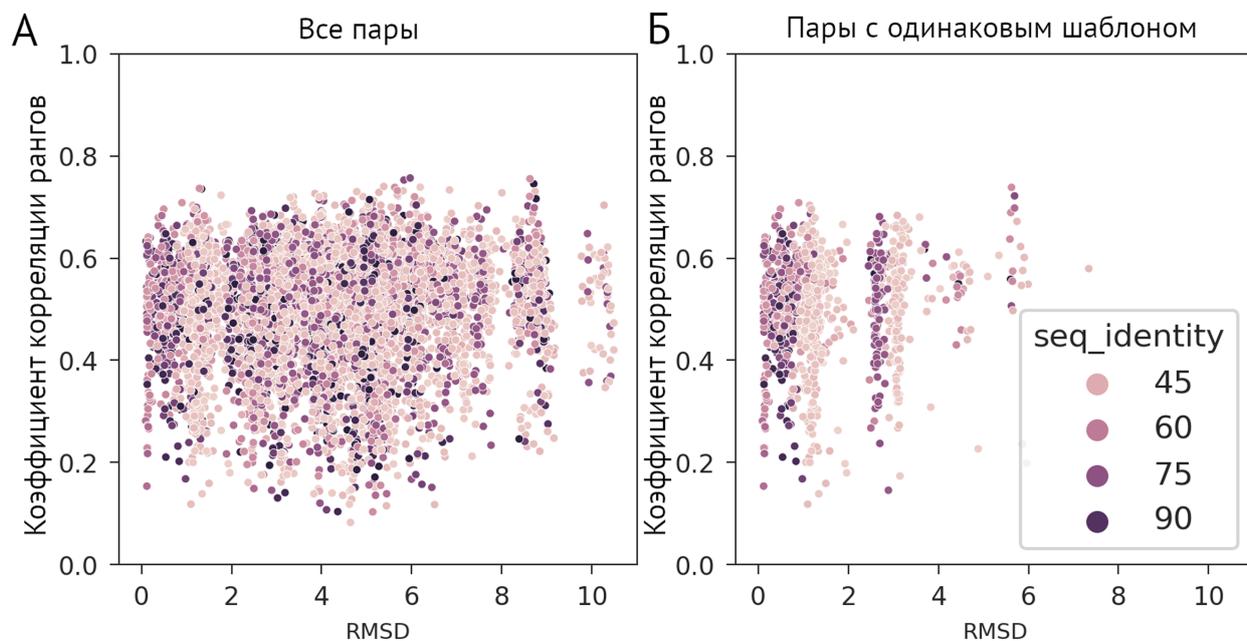


Рисунок 6. Зависимость попарного коэффициента корреляции рангов результатов докинга в структуры моделей белков NS1 флавивирусов от попарного RMSD между соответствующими структурами, окраска по проценту сходства между соответствующими последовательностями. (А) – для всех пар структур, (Б) – для пар моделей разных белков с одинаковым шаблоном. Цветной вариант рисунка представлен в электронной версии статьи на сайте журнала.

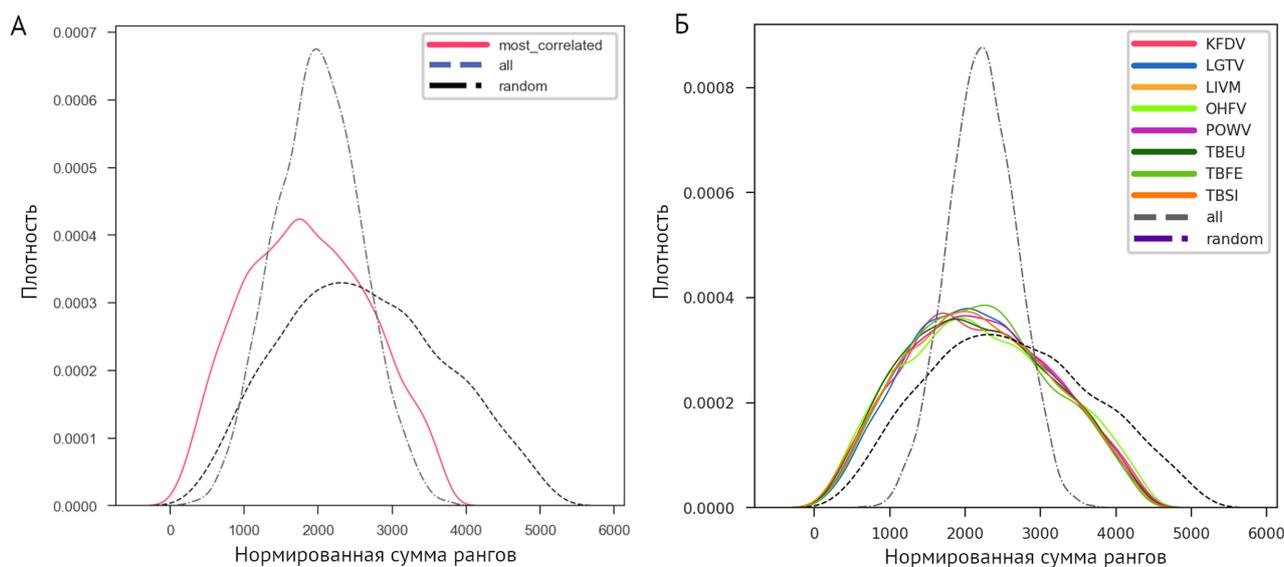


Рисунок 7. Распределение суммы рангов для результатов докинга. (А) – нормированная сумма по ансамблю, составленному из модели NS1 ДЕНВ4 по шаблону 4O6C, модели NS1 ВЯЭ по шаблону 5K6K, модели NS1 вируса шотландского энцефаломиелита овец по шаблону 5GS6. (Б) – нормированная сумма по моделям, принадлежащим белку NS1 одного вируса (для вирусов, распространяемых клещами). Цветной вариант рисунка представлен в электронной версии статьи на сайте журнала.

для поиска лигандов широкого спектра действия. Аналогичным образом могут быть составлены ансамбли из четырех и более структур.

При рассмотрении пар моделей на основе одинаковых шаблонов выделяются три группы с RMSD в диапазоне от 0 до 2, от 2 до 4 и от 4 до 6, и в рамках каждой группы структуры с большим подобием аминокислотных последовательностей ожидаемо имеют более низкие RMSD, однако на характер ранжирования структурное подобие практически не влияет (рис. 6Б).

При анализе распределений оценок докинга (Дополнительные материалы, рис. SF12–14) не было обнаружено корреляции между медианой, минимальным и максимальным значением оценки, типом шаблонной структуры и аминокислотной последовательностью. Это согласуется 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 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. Распределение нормированной суммы рангов по ансамблям моделей, построенных по одному шаблону и по одному вирусу (рис. 7Б), имеет одновершинную ассиметричную форму. 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, 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.

Распределения нормированной суммы рангов по ансамблям моделей NS1 разных вирусов по каждому шаблону в отдельности и ансамблям моделей NS1 индивидуальных вирусов по всем шаблонам сравнили при помощи критерия Колмогорова-Смирнова (Дополнительные материалы, табл. ST5). 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 разделённых по переносчику ансамблей число молекул с нормированной суммой рангов ≤ 500 (потенциальные хиты виртуального скрининга) больше, чем для ансамблей, включающих белки NS1 всех исследованных вирусов (Дополнительные материалы, табл. ST6). Таким образом, разделение вирусов, переносимых комарами и клещами,

при построении ансамблей следует считать эффективной стратегией при отборе низкомолекулярных соединений для экспериментального исследования.

ЗАКЛЮЧЕНИЕ

Белок NS1 ортофлавиринов — относительно малоизученная мишень, для которой имеется мало структурной информации. Однако она представляет интерес и как терапевтическая мишень из-за низкого сходства с человеческими белками и важности для репликации ортофлавиринов. Мы рассмотрели, насколько моделирование по гомологии применимо для генерации моделей структуры для ансамблевого докинга и как неоднородность аминокислотных последовательностей влияет на ранжирующую способность структурно-обоснованного виртуального скрининга. Отбор структур на основе ранжирования результатов докинга позволил составить ансамбли более согласованно, чем отбор на основе RMSD, из-за очень большого разброса атомных координат. Ансамбль, построенный на основе ранжирования, отражает разнообразие моделей структур белков и позволяет выбирать структуры, которые обеспечивают приемлемую степень сходства для консенсусного отбора потенциальных лигандов NS1 широкого спектра действия. Дальнейшее объединение моделей в ансамбли на основании белка-шаблона или целевого вируса при использовании суммы рангов в качестве оценки приводит к ансамблям, ранжирующим соединения лучше случайного выбора. При этом объединение по вирусу обеспечивает более согласованное ранжирование результатов. Напротив, объединение в одном ансамбле структур, белков NS1, принадлежащих вирусам с разными переносчиками, негативно сказывается на обогащении результатов виртуального скрининга потенциальными специфическими лигандами белка NS1 ортофлавиринов. Ансамбли дают возможность приоритизировать некоторые группы соединений над случайными, таким образом, они подходят для виртуального скрининга потенциальных лигандов белка NS1.

ФИНАНСИРОВАНИЕ

Работа выполнена в рамках государственного задания FNZG-2024-0005.

СОБЛЮДЕНИЕ ЭТИЧЕСКИХ СТАНДАРТОВ

Настоящая статья не содержит каких-либо исследований с участием людей или с использованием животных в качестве объектов.

КОНФЛИКТ ИНТЕРЕСОВ

Авторы заявляют об отсутствии конфликта интересов.

Дополнительные материалы доступны в электронной версии статьи на сайте журнала (pbmc.ibmc.msk.ru).

ЛИТЕРАТУРА

1. Pierson T.C., Diamond M.S. (2020) The continued threat of emerging flaviviruses. *Nat. Microbiol.*, **5**(6), 796–812. DOI: 10.1038/s41564-020-0714-0
2. Pustijanac E., Buršić M., Talapko J., Škrlec I., Meštrović T., Lišnjić D. (2023) Tick-borne encephalitis virus: A comprehensive review of transmission, pathogenesis, epidemiology, clinical manifestations, diagnosis, and prevention. *Microorganisms*, **11**(7), 1634. DOI: 10.3390/microorganisms11071634
3. Колясникова Н.М., Ишмухаметов А.А., Акимкин В.Г. (2023) Современное состояние проблемы клещевого энцефалита в России и мире. *Эпидемиология и вакцинопрофилактика*, **22**(1), 104–123. [Kolyasnikova N.M., Ishmukhametov A.A., Akimkin V.G. (2023) The current state of the problem of tick-borne encephalitis in Russia and the world. *Epidemiology and Vaccinal Prevention.*, **22**(1), 104–123.] DOI: 10.31631/2073-3046-2023-22-1-104-123
4. Путинцева Е.В., Удовиченко С.К., Никитин Д.Н., Бородай Н.В., Батурун А.А., Мачнева А.Ю., Антонов А.С., Зарубин Н.А., Топорков А.В. (2023) Лихорадка Западного Нила в Российской Федерации в 2022 г., прогноз заболеваемости на 2023 г. Проблемы особо опасных инфекций, **1**, 75–84. [Putintseva E.V., Udovichenko S.K., Nikitin D.N., Boroday N.V., Baturin A.A., Machneva A.Yu., Antonov A.S., Zarubin N.A., Toporkov A.V. (2023) West Nile fever in the Russian Federation in 2022, the incidence forecast for 2023. *Problems of Particularly Dangerous Infections*, **1**, 75–84.] DOI: 10.21055/0370-1069-2023-1-75-84
5. Ruzek D., Avšič Županc T., Borde J., Chrdele A., Eyer L., Karganova G., Kholodilov I., Knap N., Kozlovskaya L., Matveev A., Miller A.D., Osolodkin D.I., Överby A.K., Tikunova N., Tkachev S., Zajkowska J. (2019) Tick-borne encephalitis in Europe and Russia: Review of pathogenesis, clinical features, therapy, and vaccines. *Antiviral Res.*, **164**, 23–51. DOI: 10.1016/j.antiviral.2019.01.014
6. Redoni M., Yacoub S., Rivino L., Giacobbe D.R., Luzzati R., di Bella S. (2020) Dengue: Status of current and under-development vaccines. *Rev. Med. Virol.*, **30**(4), e2101. DOI: 10.1002/rmv.2101
7. Staples J.E., Monath T.P., Gershman M.D., Barrett A.D.T. (2017) Yellow fever vaccines. In: Plotkin's Vaccines, Elsevier, pp. 1181–1265. e20. DOI: 10.1016/B978-0-323-35761-6.00063-8
8. Yun S.-I., Lee Y.-M. (2014) Japanese encephalitis: The virus and vaccines. *Hum. Vaccin. Immunother.*, **10**(2), 263–279. DOI: 10.4161/hv.26902
9. Morrone S.R., Lok S.-M. (2019) Structural perspectives of antibody-dependent enhancement of infection of dengue virus. *Curr. Opin. Virol.*, **36**, 1–8. DOI: 10.1016/j.coviro.2019.02.002
10. Bardina S.V., Bunduc P., Tripathi S., Duehr J., Frere J.J., Brown J.A., Nachbagauer R., Foster G.A., Krysztof D., Tortorella D., Stramer S.L., Garcia-Sastre A., Krammer F., Lim J.K. (2017) Enhancement of Zika virus pathogenesis by preexisting anti-flavivirus immunity. *Science*, **356**(6334), 175–180. DOI: 10.1126/science.aal4365
11. Santos-Peral A., Luppa F., Goresch S., Nikolova E., Zaucha M., Lehmann L., Dahlstroem F., Karimzadeh H., Thorn-Seshold J., Winheim E., Schuster E.-M., Dobler G., Hoelscher M., Kümmerer B.M., Endres S., Schober K., Krug A.B., Pritsch M., Barba-Spaeth G., Rothenfusser S. (2024) Prior flavivirus immunity skews the yellow fever vaccine response to cross-reactive antibodies with potential to enhance dengue virus infection. *Nat. Commun.*, **15**(1), 1696. DOI: 10.1038/s41467-024-45806-x
12. Chong H.Y., Leow C.Y., Abdul Majeed A.B., Leow C.H. (2019) Flavivirus infection — A review of immunopathogenesis, immunological response, and immunodiagnosis. *Virus Res.*, **274**, 197770. DOI: 10.1016/j.virusres.2019.197770
13. Rastogi M., Singh S.K. (2020) Zika virus NS1 affects the junctional integrity of human brain microvascular endothelial cells. *Biochimie*, **176**, 52–61. DOI: 10.1016/j.biochi.2020.06.011
14. Puerta-Guardo H., Glasner D.R., Espinosa D.A., Biering S.B., Patana M., Ratnasiri K., Wang C., Beatty P.R., Harris E. (2019) Flavivirus NS1 triggers tissue-specific vascular endothelial dysfunction reflecting disease tropism. *Cell Rep.*, **26**(6), 1598–1613. e8. DOI: 10.1016/j.celrep.2019.01.036
15. Puerta-Guardo H., Glasner D.R., Harris E. (2016) Dengue virus NS1 disrupts the endothelial glycocalyx, leading to hyperpermeability. *PLOS Pathogens*, **12**(7), e1005738. DOI: 10.1371/journal.ppat.1005738
16. Avirutnan P., Fuchs A., Hauhart R.E., Somnuk P., Youn S., Diamond M.S., Atkinson J.P. (2010) Antagonism of the complement component C4 by flavivirus nonstructural protein NS1. *J. Exp. Med.*, **207**(4), 793–806. DOI: 10.1084/jem.20092545
17. Chen S., Wu Z., Wang M., Cheng A. (2017) Innate immune evasion mediated by flaviviridae non-structural proteins. *Viruses*, **9**(10), 291. DOI: 10.3390/v9100291
18. Bílý T., Palus M., Eyer L., Elsterová J., Vancová M., Růžek D. (2015) Electron tomography analysis of tick-borne encephalitis virus infection in human neurons. *Sci. Rep.*, **5**(1), 10745. DOI: 10.1038/srep10745
19. Plaszczyca A., Scaturro P., Neufeldt C.J., Cortese M., Cerikan B., Ferla S., Brancale A., Pichlmair A., Bartenschlager R. (2019) 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. *PLOS Pathogens*, **15**(5), e1007736. DOI: 10.1371/journal.ppat.1007736
20. de Silva A.M., Rey F.A., Young P.R., Hilgenfeld R., Vasudevan S.G. (2018) Viral entry and NS1 as potential antiviral drug targets. *Adv. Exp. Med. Biol.*, **1062**, 107–113. DOI: 10.1007/978-981-10-8727-1_8
21. Songprakhon P., Thaingtamtanha T., Limjindaporn T., Puttikhant C., Srisawat C., Luangaram P., Dechtawewat T., Uthaiipull C., Thongsima S., Yenchitsomanus P.-T., Malasit P., Noisakran S. (2020) Peptides targeting dengue viral nonstructural protein 1 inhibit dengue virus production. *Sci. Rep.*, **10**(1), 12933. DOI: 10.1038/s41598-020-69515-9
22. Raza S., Abbas G., Azam S.S. (2020) Screening pipeline for flavivirus based inhibitors for Zika virus NS1. *IEEE/ACM Trans. Comput. Biol. Bioinform.*, **17**(5), 1751–1761. DOI: 10.1109/TCBB.2019.2911081
23. Modhiran N., Gandhi N.S., Wimmer N., Cheung S., Stacey K., Young P.R., Ferro V., Watterson D. (2019) Dual targeting of dengue virus virions and NS1 protein with the heparan sulfate mimic PG545. *Antiviral Res.*, **168**, 121–127. DOI: 10.1016/j.antiviral.2019.05.004
24. Shu B., Ooi J.S.G., Tan A.W.K., Ng T.-S., Dejnirattisai W., Mongkolsapaya J., Fibriansah G., Shi J., Kostyuchenko V.A., Screaton G.R., Lok S.-M. (2022) CryoEM structures of the multimeric secreted NS1, a major factor for dengue hemorrhagic fever. *Nat. Commun.*, **13**(1), 6756. DOI: 10.1038/s41467-022-34415-1

25. Pan Q., Jiao H., Zhang W., Chen Q., Zhang G., Yu J., Zhao W., Hu H. (2024) The step-by-step assembly mechanism of secreted flavivirus NS1 tetramer and hexamer captured at atomic resolution. *Sci. Adv.*, **10**(18), eadm8275. DOI: 10.1126/sciadv.adm8275
26. Gutsche I., Coulibaly F., Voss J.E., Salmon J., d'Alayer J., Ermonval M., Larquet E., Charneau P., Krey T., Mégret F., Guittet E., Rey F.A., Flamand M. (2011) Secreted dengue virus nonstructural protein NS1 is an atypical barrel-shaped high-density lipoprotein. *Proc. Natl. Acad. Sci. USA*, **108**(19), 8003–8008. DOI: 10.1073/pnas.1017338108
27. Roy P., Roy S., Sengupta N. (2020) Disulfide reduction allosterically destabilizes the β -ladder subdomain assembly within the NS1 dimer of ZIKV. *Biophys. J.*, **119**(8), 1525–1537. DOI: 10.1016/j.bpj.2020.08.036
28. Akey D.L., Brown W.C., Dutta S., Konwerski J., Jose J., Jurkiw T.J., del Proposto J., Ogata C.M., Skinnotis G., Kuhn R.J., Smith J.L. (2014) Flavivirus NS1 structures reveal surfaces for associations with membranes and the immune system. *Science*, **343**(6173), 881–885. DOI: 10.1126/science.1247749
29. Zhang S., Wang X., He Y., Hu T., Guo J., Wang M., Jia R., Zhu D., Liu M., Zhao X., Yang Q., Wu Y., Zhang S., Huang J., Mao S., Ou X., Gao Q., Sun D., Liu Y., Zhang L., Chen S., Cheng A. (2021) 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. *Vet. Microbiol.*, **261**, 109215. DOI: 10.1016/j.vetmic.2021.109215
30. Somnuk P., Hauhart R.E., Atkinson J.P., Diamond M.S., Avirutnan P. (2011) N-Linked glycosylation of dengue virus NS1 protein modulates secretion, cell-surface expression, hexamer stability, and interactions with human complement. *Virology*, **413**(2), 253–264. DOI: 10.1016/j.virol.2011.02.022
31. Pryor M.J., Wright P.J. (1994) Glycosylation mutants of dengue virus NS1 protein. *J. Gen. Virol.*, **75**(Pt 5), 1183–1187. DOI: 10.1099/0022-1317-75-5-1183
32. Akey D.L., Brown W.C., Konwerski J.R., Ogata C.M., Smith J.L. (2014) Use of massively multiple merged data for low-resolution S-SAD phasing and refinement of flavivirus NS1. *Acta Cryst. D.*, **70**(10), 2719–2729. DOI: 10.1107/S1399004714017556
33. Brown W.C., Akey D.L., Konwerski J.R., Tarrasch J.T., Skinnotis G., Kuhn R.J., Smith J.L. (2016) Extended surface for membrane association in Zika virus NS1 structure. *Nat. Struct. Mol. Biol.*, **23**(9), 865–867. DOI: 10.1038/nsmb.3268
34. Xu X., Song H., Qi J., Liu Y., Wang H., Su C., Shi Y., Gao G.F. (2016) Contribution of intertwined loop to membrane association revealed by Zika virus full-length NS1 structure. *EMBO J.*, **35**(20), 2170–2178. DOI: 10.15252/embj.201695290
35. Carlson H.A., Masukawa K.M., McCammon J.A. (1999) Method for including the dynamic fluctuations of a protein in computer-aided drug design. *J. Phys. Chem. A*, **103**(49), 10213–10219. DOI: 10.1021/jp991997z
36. Fomina A.D., Uvarova V.I., Kozlovskaya L.I., Palyulin V.A., Osolodkin D.I., Ishmukhametov A.A. (2024) Ensemble docking based virtual screening of SARS-CoV-2 main protease inhibitors. *Mol. Inf.*, **43**(8), e202300279. DOI: 10.1002/minf.202300279
37. Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N., Bourne P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**(1), 235–242. DOI: 10.1093/nar/28.1.235
38. Boratyn G.M., Thierry-Mieg J., Thierry-Mieg D., Busby B., Madden T.L. (2019) Magic-BLAST, an accurate RNA-Seq aligner for long and short reads. *BMC Bioinformatics*, **20**(1), 405. DOI: 10.1186/s12859-019-2996-x
39. Laskowski R.A., MacArthur M.W., Moss D.S., Thornton J.M. (1993) PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**(2), 283–291. DOI: 10.1107/S0021889892009944
40. Humphrey W., Dalke A., Schulten K. (1996) VMD: Visual molecular dynamics. *J. Mol. Graph.*, **14**(1), 33–38. DOI: 10.1016/0263-7855(96)00018-5
41. McKinney W. (2010) Data structures for statistical computing in Python. *Proceedings of the 9th Python in Science Conference*, 56–61. DOI: 10.25080/Majora-92bf1922-00a
42. Harris C.R., Millman K.J., van der Walt S.J., Gommers R., Virtanen P., Cournapeau D., Wieser E., Taylor J., Berg S., Smith N.J., Kern R., Picus M., Hoyer S., van Kerkwijk M.H., Brett M., Haldane A., del Río J.F., Wiebe M., Peterson P., Gérard-Marchant P., Sheppard K., Reddy T., Weckesser W., Abbasi H., Gohlke C., Oliphant T.E. (2020) Array programming with NumPy. *Nature*, **585**(7825), 357–362. DOI: 10.1038/s41586-020-2649-2
43. Hunter J.D. (2007) Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.*, **9**(3), 90–95. DOI: 10.1109/MCSE.2007.55
44. Waskom M.L. (2021) Seaborn: Statistical data visualization. *J. Open Source Softw.*, **6**(60), 3021. DOI: 10.21105/joss.03021
45. Ngan C.-H., Hall D.R., Zerbe B., Grove L.E., Kozakov D., Vajda S. (2012) FTSite: High accuracy detection of ligand binding sites on unbound protein structures. *Bioinformatics*, **28**(2), 286–287. DOI: 10.1093/bioinformatics/btr651
46. Kozakov D., Grove L.E., Hall D.R., Bohnuud T., Mottarella S.E., Luo L., Xia B., Beglov D., Vajda S. (2015) The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. *Nat. Protoc.*, **10**(5), 733–755. DOI: 10.1038/nprot.2015.043
47. Retrieved January 15, 2020 from <https://ftsitesite.bu.edu/>
48. Volkamer A., Griewel A., Grombacher T., Rarey M. (2010) Analyzing the topology of active sites: On the prediction of pockets and subpockets. *J. Chem. Inf. Model.*, **50**(11), 2041–2052. DOI: 10.1021/ci100241y
49. Zentrum für Bioinformatik: Universität Hamburg — Proteins Plus Server. Retrieved January 15, 2020 from <https://proteins.plus/>
50. Edgar R.C. (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*, **32**(5), 1792–1797. DOI: 10.1093/nar/gkh340
51. Retrieved August 08, 2023 from <https://www.ebi.ac.uk/Tools/msa/muscle/>
52. Webb B., Sali A. (2016) Comparative protein structure modeling using MODELLER. *Curr. Protoc. Bioinformatics*, **54**, 5.6.1–5.6.37. DOI: 10.1002/epbi.3
53. Wildman S.A., Crippen G.M. (1999) Prediction of physicochemical parameters by atomic contributions. *J. Chem. Inf. Comput. Sci.*, **39**(5), 868–873. DOI: 10.1021/ci9903071
54. Cresset/flare-python-pyflare GitLab. GitLab. Retrieved August 20, 2023 from <https://gitlab.com/cresset/flare-python-pyflare>
55. Ligand and structure-based drug design software | Flare. Retrieved October 01, 2024 from <https://www.cresset-group.com/software/flare/>
56. Bauer M.R., Mackey M.D. (2019) Electrostatic complementarity as a fast and effective tool to optimize binding and selectivity of protein-ligand complexes. *J. Med. Chem.*, **62**(6), 3036–3050. DOI: 10.1021/acs.jmedchem.8b01925

57. *Stroganov O.V., Novikov F.N., Stroylov V.S., Kulkov V., Chilov G.G.* (2008) Lead Finder: An approach to improve accuracy of protein-ligand docking, binding energy estimation, and virtual screening. *J. Chem. Inf. Model.*, **48**(12), 2371–2385. DOI: 10.1021/ci800166p
58. *Kufareva I., Abagyan R.* (2012) Methods of protein structure comparison. *Methods Mol. Biol.*, **857**, 231–257. DOI: 10.1007/978-1-61779-588-6_10
59. Retrieved October 01, 2024 from <https://ictv.global/report/chapter/flaviviridae/flaviviridae/orthoflavivirus>
60. *Walker P.J., Siddell S.G., Lefkowitz E.J., Mushegian A.R., Adriaenssens E.M., Alfenas-Zerbini P., Dempsey D.M., Dutilh B.E., Garcia M.L., Curtis Hendrickson R., Junglen S., Krupovic M., Kuhn J.H., Lambert A.J., Lobočka M., Oksanen H.M., Orton R.J., Robertson D.L., Rubino L., Sabanadzovic S., Simmonds P., Smith D.B., Suzuki N., van Doorslaer K., Vandamme A.-M., Varsani A., Zerbini F.M.* (2022) Recent changes to virus taxonomy ratified by the international committee on taxonomy of viruses. *Arch. Virol.*, **167**(11), 2429–2440. DOI: 10.1007/s00705-022-05516-5
61. *Fiser A., Sali A.* (2003) Modeller: Generation and refinement of homology-based protein structure models. *Methods Enzymol.*, **374**, 461–491. DOI: 10.1016/S0076-6879(03)74020-8
62. *Shen M., Sali A.* (2006) Statistical potential for assessment and prediction of protein structures. *Prot. Sci.*, **15**(11), 2507–2624. DOI: 10.1110/ps.062416606

Поступила в редакцию: 16. 10. 2024.
После доработки: 28. 10. 2024.
Принята к печати: 05. 11. 2024.

HOMOLOGY MODELING OF THE ORTHOFLAVIVIRUS NS1 PROTEIN FOR VIRTUAL SCREENING OF POTENTIAL LIGANDS

A.D. Fomina^{1,2}, V.A. Palyulin², D.I. Osolodkin^{1,3}*

¹Chumakov Federal Scientific Center for Research and Development of Immune-and-Biological Products of Russian Academy of Sciences (Institute of Poliomyelitis), premises 8, bldg. 1, Village of Institute of Poliomyelitis, Settlement Moskovskiy, Moscow, 108819 Russia; *e-mail: osolodkin_di@chumakovs.su

²Department of Chemistry, Lomonosov Moscow State University, Moscow, Russia

³Institute of Translational Medicine and Biotechnology, Sechenov First Moscow State Medical University, Moscow, Russia

The orthoflavivirus NS1 protein is a relatively understudied target for the design of broad-spectrum anti-orthoflaviviral drugs. Currently, the NS1 protein structures of tick-borne orthoflaviviruses have not been published yet, but these structures can be modelled by homology, thus generating a large amount of structural data. We performed homology modelling of the NS1 protein structures of epidemiologically significant orthoflaviviruses and analysed the possibility of using these models in ensemble docking-based virtual screening. The limitations of the method and the importance of separating the models based on the vector organism when selecting an ensemble have been demonstrated.

The whole English version is available at <http://pbmc.ibmc.msk.ru>.

Key words: orthoflaviviruses; NS1 protein; ensemble docking; homology modeling

Funding. Chumakov FSC R&D IBP RAS (Institute of Poliomyelitis) fundamental research assignment FNZG-2024-0005.

Received: 16.10.2024; revised: 28.10.2024; accepted: 05.11.2024.