

ОЦЕНКА АФФИННОСТИ СВЯЗЫВАНИЯ В КОМПЛЕКСАХ ACE2–RBD S-БЕЛКА КОРОНАВИРУСОВ С ИСПОЛЬЗОВАНИЕМ СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ

© 2024 г. Е.А. Богданова*, А.В. Чернухин**, К.В. Шайтан*, В.Н. Новоселецкий***, #

*Биологический факультет Московского государственного университета имени М.В. Ломоносова, Ленинские горы, 1/12, Москва, 119991, Россия

**Факультет цифровых технологий и химического инжиниринга Российского химико-технологического университета имени Д.И. Менделеева, Миусская площадь, 9, Москва, 125047, Россия

***Биологический факультет Университета МГУ-ППИ в Шэньчжэнь, Гоцзидасюэюань, 1, Даюньсиньчэн, район Лунган, Шэньчжэнь, 518172, Китайская Народная Республика

#E-mail: novoseletsky@smbu.edu.cn

Поступила в редакцию 14.05.2024 г.

После доработки 15.05.2024 г.

Принята к публикации 22.05.2024 г.

Проведена оценка экспериментально полученных структур 48 комплексов рецептора ACE2 с RBD S-белка коронавируса SARS-CoV и SARS-CoV-2 (включая мутантные формы последнего), для которых рассчитана константа диссоциации. Для предсказания аффинности связывания использован разработанный авторами ранее нейросетевой алгоритм ProBAN, а также ряд других алгоритмов оценки свободной энергии Гиббса: Prodigy, FoldX, DFIRE и RosettaDock. Сопоставление результатов оценки показывает, что ProBAN демонстрирует лучшее качество предсказания (корреляция Пирсона – 0.56, средняя абсолютная ошибка – 0.7 ккал/моль). Полученные результаты позволяют предположить лучшее качество предсказания аффинности и для других белок-белковых комплексов. Информация об исследуемых комплексах и результаты предсказания доступны в репозитории по ссылке: https://github.com/EABogdanova/ProBAN_RBD-ACE2.

Ключевые слова: белок-белковые взаимодействия, интерфейс взаимодействия, аффинность связывания, коронавирус, RBD–ACE2, биоинформатика.

DOI: 10.31857/S0006302924050053, **EDN:** MKJZVI

Проникновение вируса SARS-CoV-2 в организм человека начинается со взаимодействия RBD-домена его S-белка с ангиотензинпревращающим ферментом 2 (ACE2, англ. angiotensin converting enzyme 2), выполняющим роль мембранного рецептора. Широкое распространение SARS-CoV-2 по сравнению с его предшественниками было связано с более высокой аффинностью связывания S-белка с ACE2 [1, 2]. В процессе накопления мутаций в S-белке, в RBD-домене, в частности, происходило возникновение новых вариантов, таких как Alpha, Beta, Omicron и др., которые характеризуются более быстрым распространением и эффективным проникновением в клетки человека. Осознание данной закономерности повлекло за собой большое количество исследований интерфейса связывания RBD–ACE2,

а также аминокислотных замен, влияющих на аффинность связывания между данными молекулами [3–5].

Физическое взаимодействие между молекулами белков имеет давнюю историю изучения многочисленными экспериментальными и вычислительными методами [6, 7], включая методы биоинформатики [8]. Одной из главных характеристик взаимодействия является константа диссоциации комплексов «белок–белок» (K_d), которая может быть выражена через энергию связывания $\Delta G = RT \ln K_d$. Создание алгоритмов аккуратной оценки этой энергии позволило бы более эффективно проводить направленный мутагенез взаимодействующих белков [9], что имеет существенное значение для создания медицинских препаратов белковой природы, включая антитела [10].

Сокращения: ACE2 – ангиотензинпревращающий фермент 2, MAE – средняя абсолютная ошибка.

В настоящее время в банке данных белковых структур PDB содержится более сотни структур комплексов ACE2 человека (hACE2) и RBD SARS-CoV-2, причем как нативного ACE2 и исходной формы RBD, так и их многочисленных мутантных форм, а также комплексов hACE2–RBD и SARS-CoV. Столь обширные структурные данные позволяют провести анализ молекулярных взаимодействий между RBD и hACE2 и выявить типичные и уникальные контакты между остатками [11–14]. Кроме того, этот уникальный набор высококонсервативных структур комплексов позволяет тестировать различные методы оценки белок-белковых взаимодействий.

В данной работе мы провели тестирование ряда ранее предложенных алгоритмов оценки энергии взаимодействия в комплексах «белок–белок» (Prodigy [15], FoldX [16], DFIRE [17], RosettaDock [18]), а также разработанного нами алгоритма ProBAN [19, 20]. Сопоставление обобщающих характеристик этих алгоритмов показывает значительно большую точность алгоритма ProBAN (коэффициент корреляции Пирсона равен 0.56, средняя абсолютная ошибка (англ. MAE – mean absolute error) равна 0.68 ккал/моль), что делает его перспективным инструментом для оценки аффинности связывания в белок-белковых комплексах.

МЕТОДЫ

Для анализа аффинности связывания были отобраны комплексы RBD–ACE2 с известными пространственными структурами и значениями константы диссоциации. Итоговый набор данных содержит 48 комплексов ACE2 и RBD вирусов SARS-CoV (3 структуры) и SARS-CoV-2 (45 структур), включая исходный вариант, а также различные штаммы и инженерные варианты.

Предсказание аффинности связывания осуществлялось алгоритмом ProBAN [20], основанным на сверточной нейронной сети. Кратко, основные шаги алгоритма следующие:

Шаг 1. Методом логистической регрессии определяется оптимальная разделяющая плоскость (рис. 1а) между взаимодействующими молекулами белков. Для лучшего разделения атомов двух молекул в процессе обучения алгоритма для координат подбираются коэффициенты (для моделей машинного обучения эти параметры называются весами) и добавляется общее смещение. Из полученных параметров можно составить уравнение разделяющей плоскости следующего вида:

$$D = W_1x + W_2y + W_3z + \beta,$$

где W_1 , W_2 , W_3 – коэффициенты, β – смещение.

Далее производится расчет центров масс по обе стороны в области на расстоянии 10 Å от разделяю-

щей плоскости. Точка пересечения вектора, объединяющего центры масс, и разделяющей плоскости определяет центр будущей решетки. Высота решетки задается нормалью к плоскости с ограничением ± 20 Å от центра решетки. Длина решетки определяется по направлению наибольшего разброса атомов в области, ограниченной высотой. Ограничение по длине (± 40 Å от центра) подбиралось эмпирически из соображения, что для большей части комплексов, используемых для обучения модели, данной длины ограничительной ячейки будет достаточно для локализации всего интерфейса связывания. Ширина определяется по вектору, ортогональному двум уже заданным с границей, идентичной длине (± 40 Å от центра). На основе полученных векторов строится итоговая решетка размером $81 \times 81 \times 41$ с шагом 1 Å.

Шаг 2. Для дальнейшего рассмотрения происходит отбор таких атомов, которые удалены не более чем на 10 Å от ближайшего атома взаимодействующей цепи, и располагаются в пределах решетки (рис. 1б).

Шаг 3. Для каждой ячейки размером $1 \times 1 \times 1$ Å, определяются атомы, лежащие внутри нее, и свойства атомов становятся свойствами ячейки. Набор и число свойств ячейки (каналы в терминах машинного обучения) подбирались на основе их влияния на продуктивность и качество обучения нейронной сети. Всего было использовано 10 различных свойств (гидрофобность, заряд, полярность, способность образовывать водородные связи и др.), которые вносят наибольший вклад в предсказание константы диссоциации.

Шаг 4. Полученный массив размера $10 \times 81 \times 81 \times 41$ подается на вход сверточной нейронной сети (рис. 1в), на выходе которой генерируется значение pK_d ($pK_d = -\log(K_d)$).

Для последующего сравнения с другими алгоритмами осуществляется расчет энергии Гиббса по следующей формуле:

$$\Delta G = RT \ln(K_d),$$

где T – температура (300 K), R – универсальная газовая постоянная (1.99 ккал/(моль·K)), K_d – константа диссоциации (M).

Данная модель была обучена и протестирована на наборе белок-белковых комплексов с известными значениями константы диссоциации в базе данных PDBBind v.2020 [21]. Исследуемые комплексы RBD–ACE2 не содержатся в данной версии базы данных, что позволяет проводить предсказания на обученной модели без возможности «утечки» данных из обучающей выборки.

Для сравнительной оценки алгоритма ProBAN нами было отобрано несколько широко используемых и доступных методов предсказания ΔG связывания белок-белковых комплексов на основе пространственной структуры: Prodigy (пред-

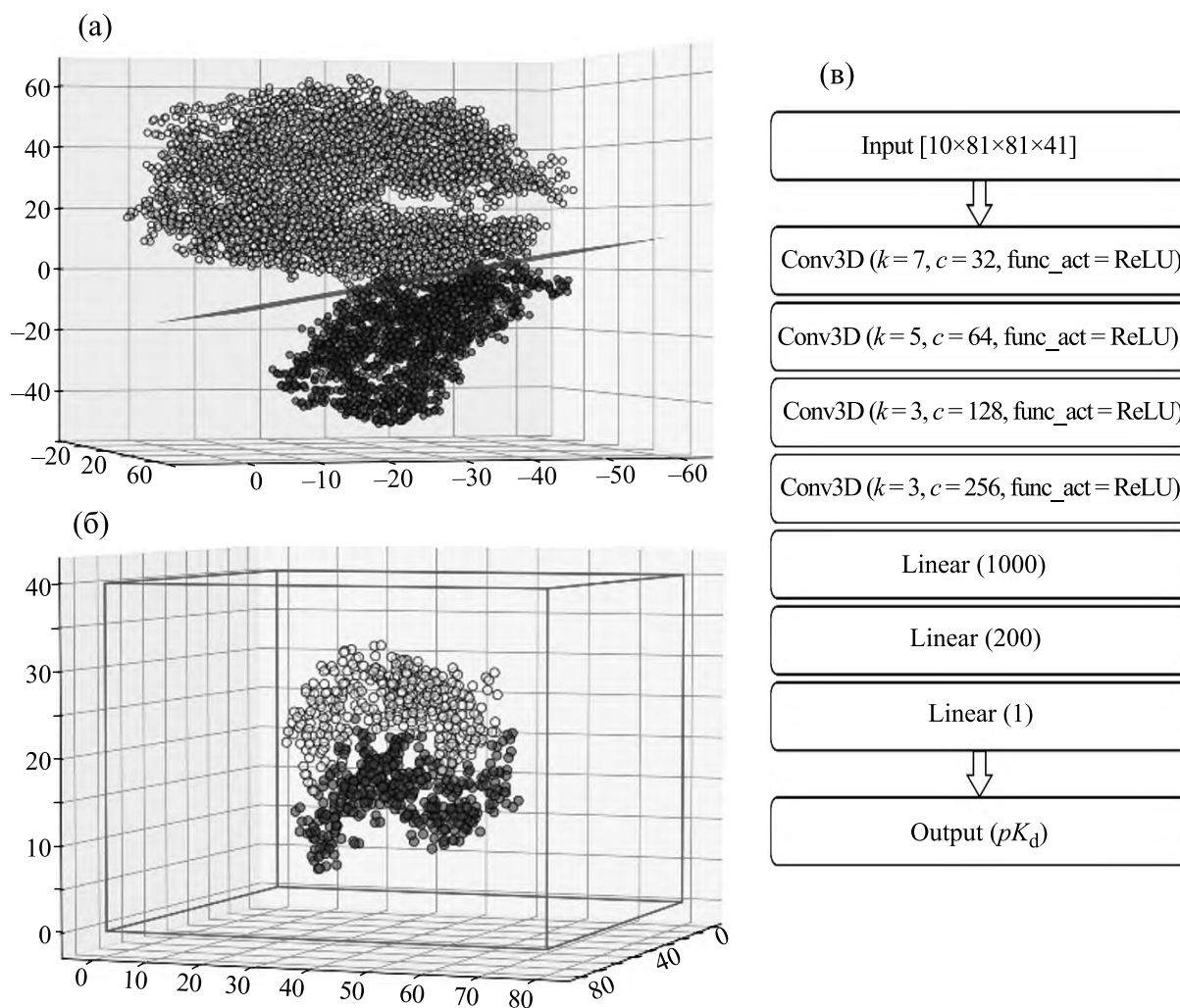


Рис. 1. Основные этапы в работе алгоритма ProBAN: (а) – точечная диаграмма атомов комплекса ACE2-RBD (6lzg) с рассчитанной разделяющей плоскостью; (б) – точечная диаграмма атомов, отобранных для дальнейшего анализа, в ограничивающей ячейке; (в) – архитектура нейронной сети ProBAN. На вход подается четырехмерный массив, включающий в себя информацию о белок-белковом комплексе. Данные проходят последовательно через 4 сверточных слоя (Conv3D) с уменьшением размера ядра свертки (k), увеличением количества каналов (c) и функцией активации (func_act) ReLU, затем проходят через 3 полносвязных слоя (Linear) с уменьшением количества каналов, формируя в итоге значение pK_d .

сказывает также K_d) [15], FoldX [16], DFIRE [17] и RosettaDock [18]. Предсказание с использованием Prodigy осуществляли на специальном онлайн-сервисе [22], для FoldX использовали одноименную программу FoldX Suite 4.0. Расчет с использованием остальных методов осуществляли на онлайн-сервисе CCharPPI [23, 24]. В качестве метрик качества использовали корреляцию Пирсона и среднюю абсолютную ошибку (англ. MAE).

Анализ полученных результатов, а также предсказание алгоритмом ProBAN осуществлялось с использованием Python дистрибутива Anaconda.

Анализ межмолекулярных взаимодействий в комплексах белок-белок был выполнен с помощью программы PLATINUM [25]. Визуализация

комплексов и оптимизация сети водородных связей выполнена с помощью программ Maestro (Schrödinger, LLC) и VMD [26].

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Анализ интерфейсов. Рассматриваемые в настоящей работе комплексы RBD–ACE2 образованы рецептор-связывающим доменом (receptor binding domain, RBD) S-белка коронавируса SARS-CoV и SARS-CoV-2 и молекулой ангиотензин-превращающего фермента 2 (рис. 2а). В непосредственном контакте RBD–ACE2 принимают участие 26 остатков со стороны RBD (позиции 403, 417, 439, 446, 449, 453, 455, 458, 475–477, 484–487, 489, 490, 493, 496, 498, 500–503, 505 согласно последовательности UniProt P0DTC2) и

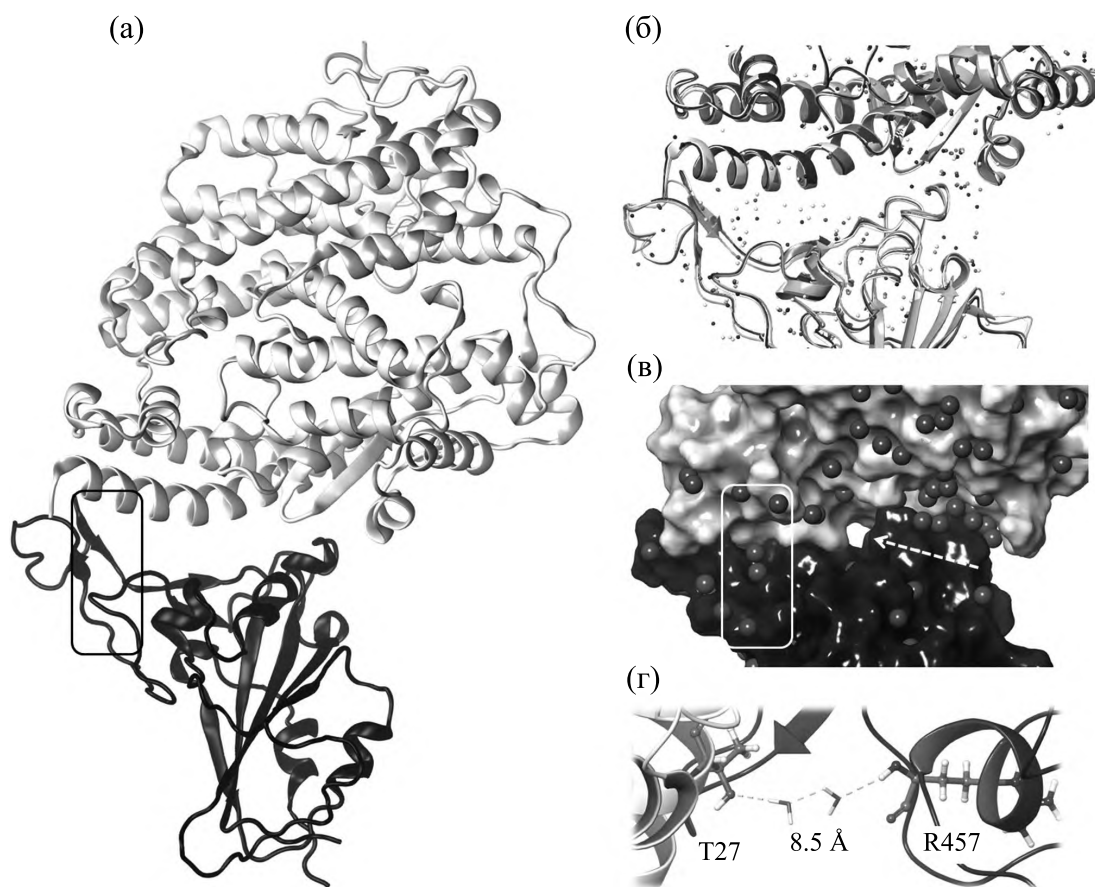


Рис. 2. (а) — Общий вид комплекса ACE2 (показан светлым) и RBD (показан темным) (PDB-код 6lzg) в ленточном представлении. Рамка соответствует области расположения остатков T27 (ACE2) и R457 (RBD). (б) — Совмещение интерфейсов взаимодействия ряда структур комплекса ACE2–RBD: 6lzg (показана самым темным), 7ekh (показана темным), 7lo4 (показана светлым) и 8df5 (показана самым светлым). Молекулы кристаллизационной воды показаны шариками соответствующих цветов (масштаб не соблюден). (в) — Интерфейс взаимодействия ACE2–RBD (PDB-код 6lzg). Молекулярные поверхности субъединиц, соответствующих ACE2 и RBD, показаны светлым и темным, соответственно. Атомы кислорода, соответствующие молекулам воды, показаны промежуточным серым. Рамка соответствует области расположения остатков T27 (ACE2) и R457 (RBD). Стрелка указывает на полость. (г) — Две молекулы воды и цепочка водородных связей, обеспечивающие взаимодействие остатка T27 ACE2 (слева) и остатка R457 RBD (справа). Расстояние между соответствующими атомами белка 8.5 Å.

22 остатка со стороны ACE2 (позиции 19, 24, 27, 28, 30, 31, 34, 35, 37, 38, 41, 42, 79, 82, 83, 329, 330, 352–355, 357 согласно последовательности UniProt Q9BYF1). Взаимная ориентация взаимодействующих белков и конформация интерфейса весьма консервативны (рис. 2б): совмещение структур комплексов по Ca-атомам указанных 48 остатков дает среднеквадратичное отклонение <1 Å для всех рассмотренных нами структур. Поверхность непосредственного контакта имеет сложную форму, но может быть заключена в параллелепипед с размерами $45 \times 15 \times 15$ Å. Анализ межмолекулярных взаимодействий показывает, что основными из них являются гидрофобные контакты и водородные связи (данные не приведены).

В первом приближении интерфейс взаимодействия RBD–ACE2 состоит из двух макрообластей

плотного примыкания молекул и полости между ними (рис. 2в). Будем для определенности называть эти макрообласти большой (рис. 2в, слева) и малой (рис. 2в, справа), поскольку число вовлеченных в их образование аминокислотных остатков со стороны RBD составляет 15 и 11, а со стороны ACE2 — 11 и 11 соответственно.

Межмолекулярные взаимодействия в комплексах RBD–ACE2 подробно освещаются практически в каждой работе, посвященной той или иной расшифрованной структуре [27–30]. Считается, что наиболее значимыми для связывания являются три области (hot spot), первая из которых соответствует малой макрообласти, а две другие — большой макрообласти. Итак, первая область локализована вокруг остатка Lys353 ACE2 и характеризуется взаимодействиями, образованными остатками Lys353, Asp38, Tyr41,

Gln42, Leu45 и Asn330 со стороны ACE2 и остатками Thr500, Asn501, Gln498 и Tyr505 со стороны RBD. Мутации в указанных позициях значимо влияют на аффинность связывания. Так, например, выявлено, что замена аспарагина на тирозин в 501 позиции RBD у альфа штамма SARS-CoV-2 значительно увеличивает аффинность [11, 12].

Вторая область ассоциирована с остатком Lys31 ACE2 и характеризуется контактами, образованными остатками Leu455, Glu484, Lys417 со стороны RBD и Asp30 и Lys31 со стороны ACE2 [29]. Также в мутантах дикого штамма выявлен остаток Gln493, который усиливает взаимодействие в данной области.

Третья область также расположена вблизи N-конца ACE2, с которым взаимодействует короткая подвижная петля RBD (рис. 2а, слева). Эта петля охватывает спираль ACE2 (остатки 21–52) с почти противоположной стороны от основного интерфейса, что обеспечивает более обширную область взаимодействий [31] и стабильное относительное расположение RBD и ACE2. Критически важными для взаимодействия в этой области являются следующие остатки RBD: Ala475, Gly476 и Phe486, первые два взаимодействуют с Ser19, а третий – с Met82 и Leu79 [32]. При этом ранее было показано, что мутации в позиции Gly476, а также Ala475 негативно сказываются на аффинности связывания [13].

Обращает на себя внимание полость, расположенная между макрообластями (рис. 2в). В естественных условиях она, очевидно, заполнена молекулами воды и ионами, однако в известных кристаллографических структурах молекулы кристаллизационной воды в этой полости отсутствуют. Мы полагаем, что ключевой причиной этого является высокая подвижность молекул воды в этой области, вызванная несоответствием гидрофобных свойств поверхностей молекул ACE2 и RBD в этой области пространства. Рассмотрение свойств поверхностей показывает, что поверхность ACE2 между остатками Lys353 и Lys31 обладает гидрофильными свойствами, в то время как соответствующая ей поверхность RBD между пятнами контактов этих остатков обладает гидрофобными свойствами (результаты не приведены). Это наблюдение хорошо соотносится с ранее опубликованными результатами анализа структур комплексов «белок–белок», полученных с высоким разрешением, который выявил более стабильное состояние молекул кристаллизационной воды вблизи полярных незаряженных остатков белка по сравнению с заряженными или неполярными остатками [33].

Несмотря на влияние молекул воды и ионов, расположенных на интерфейсе взаимодействия или в его окрестности, на организацию белок-белковых комплексов [34], при описании экспе-

риментальных структур комплексов RBD–ACE2 этим молекулам практически не уделяется внимание. Между тем структуры, полученные методом рентгеноструктурного анализа с высоким разрешением (как правило, 2.5 Å или лучше), содержат большое число молекул кристаллизационной воды. Так, в структурах 6lzg, 7ekh, 7lo4 и 8df5 в непосредственной близости от ACE2 или RBD содержится 322, 250, 132 и 163 молекулы воды соответственно, а структура 8df5 содержит еще и один ион хлора. Большая часть этих молекул расположена в карманах на поверхности белка hACE2, однако заметное число находится и в окрестности интерфейса взаимодействия этого белка с RBD (рис. 2в). Интересно отметить, что множества молекул воды, находящиеся в этих структурах, пересекаются не полностью, что, с одной стороны, позволяет выявить наиболее консервативные сайты связывания воды, а с другой стороны, гипотетически, позволяет создать молекулярную модель интерфейса, содержащую в себе все возможные молекулы воды, включая подвижные молекулы воды, расположенные в вышеупомянутой полости. Однако создание такой модели лежит за рамками данной работы.

Для выявления молекул кристаллизационной воды, которые могут опосредовать белок-белковое взаимодействие, нами было выполнено добавление атомов водорода и оптимизация сети водородных связей для рассматриваемых структур высокого разрешения (6lzg, 7ekh, 7lo4 и 8df5) в программе Maestro (Schrodinger LLC, США), причем положение тяжелых атомов не подвергалось изменению. Собственно выявление молекул воды выполняли визуально. Итоговая информация представлена в табл. 1. Видно, что для рассмотренных структур высокого разрешения характерно наличие как минимум нескольких цепочек водородных связей. Рассмотрение таких цепочек показывает, что во взаимодействии белков в комплексе играют роль не только остатки, непосредственно образующие нековалентные взаимодействия, но и остатки, образующие такие взаимодействия посредством молекул воды. Это предположение находит подтверждение и в литературе [35]. В частности, показано, что поверхность взаимодействующих белков, примыкающих к интерфейсу их взаимодействия, но не вовлеченная в него непосредственно, обогащена полярными атомами (т.е. атомами N и O). Очевидно, полярные атомы, расположенные недостаточно близко для образования непосредственной водородной связи, могут сделать это, образуя связь посредством молекул воды (рис. 2г).

Однако прямой учет подобных связей в белок-белковых комплексах по ряду причин является затруднительным. Прежде всего, для такого учета необходимо наличие молекул воды в явном виде и в достаточном количестве, что далеко не всегда

Таблица 1. Наблюдаемые в структурах высокого разрешения цепочки водородных связей с участием молекул воды (остаток ACE2 – молекулы воды – остаток RBD)

Код структуры, (разрешение, Å)	Непрямое взаимодействие посредством молекул воды		
	одной	двух	трех
6lzg (2.5)	K31-w-E484 (4.2) A386-w-Y505 (5.4)	T27-w-w-R457 (8.5) Q325-w-w-P499 (8.3) G354-w-w-V503 (5.4) G354-w-w-G504 (6.9)	G354-w-w-w-D405 (9.0)
7ekh (2.4)	T27-w-A475 (5.9) Y83-w-N487 (4.4)	T324-w-w-T500 (7.4) G354-w-w-G504 (6.9)	—
7lo4 (2.5)	—	G354-w-w-V503 (5.1) G326-w-w-V503 (7.0) Q325-w-w-V503 (7.3)	—
8df5 (2.7)	T27-w-A475 (5.6) D30-w-N417 (5.6) H34-w-N417 (4.2)	E35-w-w-F490 (7.6) G354-w-w-G504 (6.8)	E35-w-w-w-G485 (10.5)

Примечание. В скобках дано расстояние (в Å) между соответствующими атомами указанных остатков.

наблюдается даже для структур высокого разрешения. Это делает необходимым обращение к методам молекулярного моделирования для создания и оптимизации водного окружения интерфейса. Эти же методы должны быть использованы и для учета тепловых колебаний молекул как белков, так и воды. Далее, даже будучи выявленной, такая цепочка водородных связей сложна в интерпретации с точки зрения собственной энергии и вклада в энергию взаимодействия белков. Наконец, зачастую ставится задача быстрой оценки энергии взаимодействия в комплексе (например, в задачах белок-белкового докинга), что вообще не позволяет рассматривать молекулы воды в явном виде.

Для обхода этих затруднений и одновременно учета не прямых (опосредованных молекулами воды или как-то иначе) взаимодействий мы разработали алгоритм, использующий сверточную нейронную сеть.

Нейросетевая модель оценки связывания. Рассмотрение комплексов RBD–ACE2 с известными пространственной структурой и значением K_d выявило 48 комплексов. Большая часть этих комплексов имеют значение pK_d от 7 до 9 (рис. 3), в то время как низкая аффинность связывания ($pK_d < 7$) характерна для 5 комплексов, два из которых образованы RBD вируса SARS-CoV (3sci, 6cs2), а три оставшихся содержат RBD SARS-CoV-2 с мутациями F486L (7eke (ACE2 человека), 7wa1 (ACE2 норки)) или Y453F (7w8s (ACE2 норки)). Комплексы с самой высокой аффинностью связывания содержат hACE2 и RBD SARS-CoV-2 с мутациями D614G, N501Y, E484K, K417N (7sy4, 7sy8) или RBD SARS-CoV-2 Omicron BA.2.75

(8asy), BQ.1.1 (8if2) и рекомбинантный вариант XBB.1 (8iov).

Для анализа предсказания характеристик связывания в белок-белковых комплексах алгоритмом ProBAN было проведено его сравнение с предсказаниями, полученными веб-сервисом Prodigy. Данный метод осуществляет оценку аффинности связывания функцией, основанной на межмолекулярных контактах и признаках непосредственно на интерфейсе и полученных из анализа поверхности, не относящейся к интерфейсу взаимодействия. Метрики, полученные в результате оценки данного алгоритма, находятся в табл. 2. Хорошо видно, что ProBAN показывает более высокое качество предсказания по сравнению с Prodigy. Мы полагаем, что причиной этого

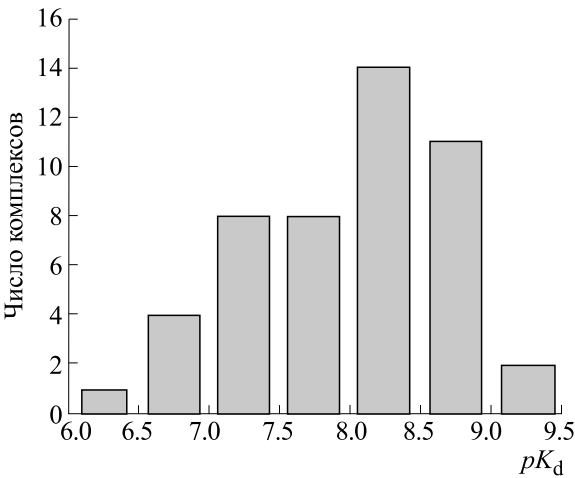


Рис. 3. Распределение комплексов RBD–ACE2 из исследуемого набора по значениям pK_d .

Таблица 2. Метрики качества предсказания аффинности связывания для комплексов ACE2–RBD для отобранных алгоритмов

Алгоритм	Коэффициент корр. Пирсона для ΔG	p -value*	MAE для ΔG (ккал/моль)	MAE для pK_d
ProBAN	0.56	3.3e-05	0.7	0.5
Prodigy	−0.38	7.2e-03	1.2	0.9
FoldX	0.41	4e-03	8.1	—
DFIRE (все комплексы)	−0.04	0.74	12.3	—
DFIRE (без 7u0n)	0.14	0.36	9.5	—
RosettaDock	−0.11	0.46	5	—

Примечание. * — Расчет p -value осуществлялся с использованием Z-преобразования Фишера.

является использование как более полной информации о взаимодействиях между атомами, так и большего порогового значения расстояния между атомами (10 Å), которое классифицирует пары атомов на взаимодействующие и нет. Используемое в Prodigy аналогичное пороговое значение расстояния между атомами (5.5 Å), по-видимому, отсеивает часть важных атомов, вносящих вклад в связывание.

Большое число алгоритмов, используемых для оценки аффинности связывания в белок-белковых комплексах, предсказывают не значение константы диссоциации, а свободную энергию Гиббса связывания. Для оценки работы данных алгоритмов (FoldX, DFIRE, RosettaDock) на исследуемом наборе данных из полученных нами значений K_d мы рассчитали ΔG и сравнили с ΔG полученной с использованием данных алгоритмов (рис. 4). Рассчитанные значения метрик качества для разных алгоритмов представлены в табл. 2, причем значения p -value, близкие к нулю, говорят о высокой достоверности рассчитанных коэффициентов корреляции. Видно, что ProBAN оказывается наиболее эффективным среди всех проанализированных алгоритмов. На втором месте по размеру средней ошибки находится Prodigy, однако корреляция принимает отрицательное значение, что свидетельствует о неспособности алгоритма оценивать влияние мутаций на направление изменения аффинности связывания в изучаемом наборе данных. Таким образом, учитывая рассчитанное значение MAE, используемая в Prodigy оценочная функция, может использоваться для оценки ΔG с погрешностью в 1.2 ккал/моль. В свою очередь, для определения вклада мутаций в аффинность относительно нативной структуры RBD–ACE2 более успешно может быть использован FoldX, который по значе-

нию корреляции (0.41) на исследуемом наборе данных находится на втором месте после ProBAN. Данный вывод согласуется с более ранними исследованиями по предсказанию аффинности связывания в комплексах RBD–ACE2 [36]. Остальные алгоритмы (DFIRE, RosettaDock) оказались менее успешными в предсказании свободной энергии Гиббса для изучаемых белок-белковых комплексов. Также стоит заметить, что для одного из комплексов (7u0n) DFIRE предсказал значение $\Delta G = 137.6$ ккал/моль, что явно является выбросом и свидетельствует о возможной нестабильности алгоритма. Для более объективной оценки работы данного метода метрики качества были пересчитаны для набора из 47 комплексов без 7u0n, что позволило их улучшить, но статистически значимой корреляции также не удалось достичь (табл. 2), как и для RosettaDock.

В результате предсказания константы диссоциации моделью ProBAN удалось получить значение корреляции Пирсона между экспериментальными и рассчитанными значениями pK_d , равное 0.56, и MAE = 0.5 (рис. 5). Также отдельно для каждого комплекса рассчитывалась абсолютная ошибка (англ. AE — Average Error) по формуле:

$$AE = |\text{True_pKd} - \text{Pred_pKd}|,$$

где True_pKd — значение pK_d , рассчитанное из экспериментально полученной K_d для комплекса, Pred_pKd — значение pK_d , предсказанное алгоритмом.

Среди с комплексов с наибольшей абсолютной ошибкой (больше 1) подавляющая часть имела разрешение хуже 3 Å (6cs2, 7wk6, 7tex) и один имел разрешение 2.85 Å (8asy). Наблюдаемая закономерность свидетельствует о негативном вкладе нечетко разрешенного положения атомов в качество предсказания аффинности связыва-

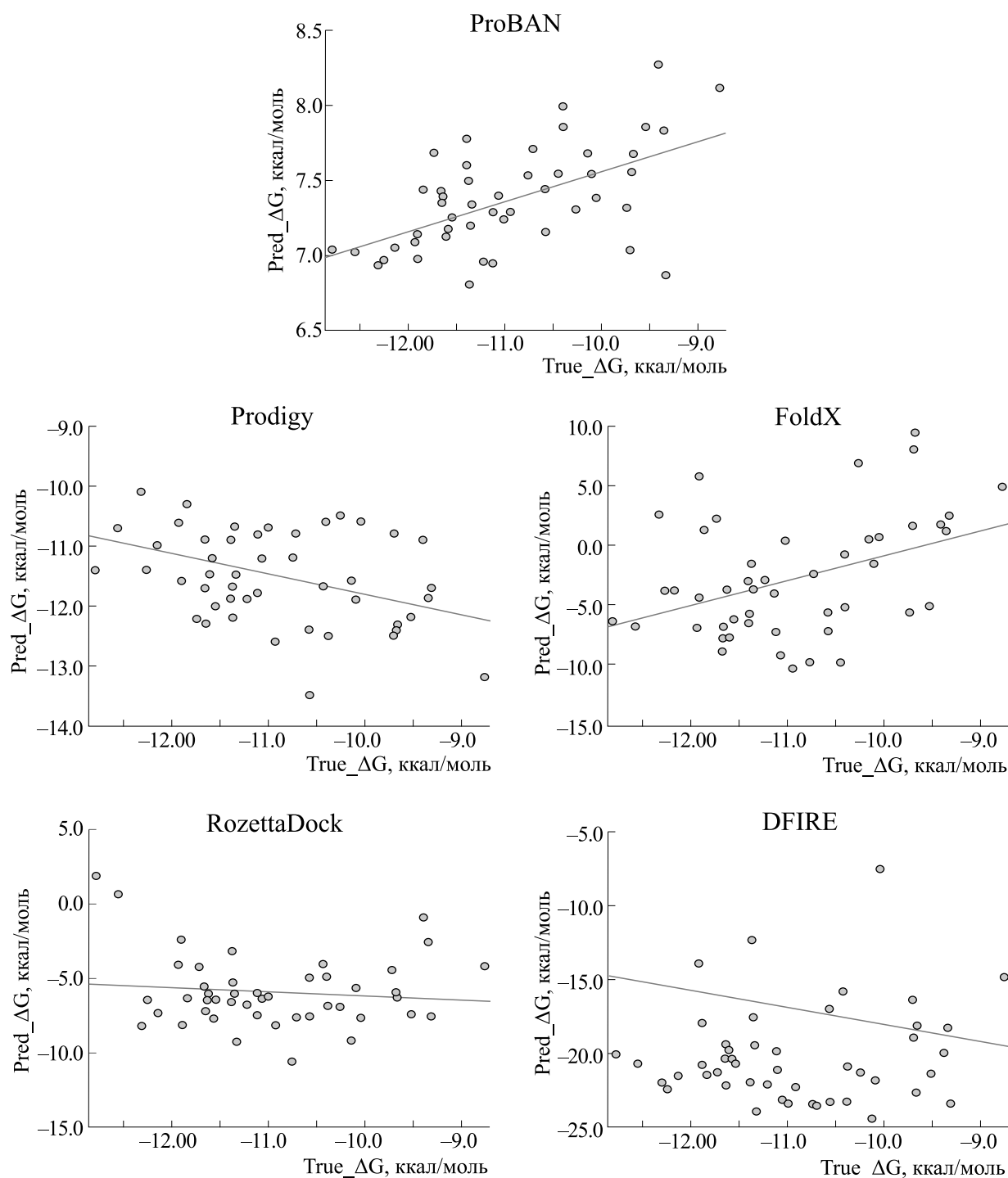


Рис. 4. Результаты предсказаний ΔG для комплексов RBD-ACE2 моделей: ProBAN, Prodigy, FoldX, DFIRE (было исключено предсказание для комплекса 7u0n) и RosettaDock.

ния, так как искажается информация об меж-атомных расстояниях, играющих ключевую роль во взаимодействии между белковыми молекулами. Так, совмещение структуры комплексов bcs2 и 3sci по Ca-атомам остатков, образующих интерфейс RBD-ACE2, показывает, что, несмотря на

незначительное отклонение в положении основной цепи (среднеквадратичное отклонение равно $0.8 \text{ \AA}$), наблюдаются значительные отличия в ориентации боковых цепей многих соседних остатков. В частности, рассмотрение пары остатков R395(RBD)–R559(ACE2) дает расстояние между

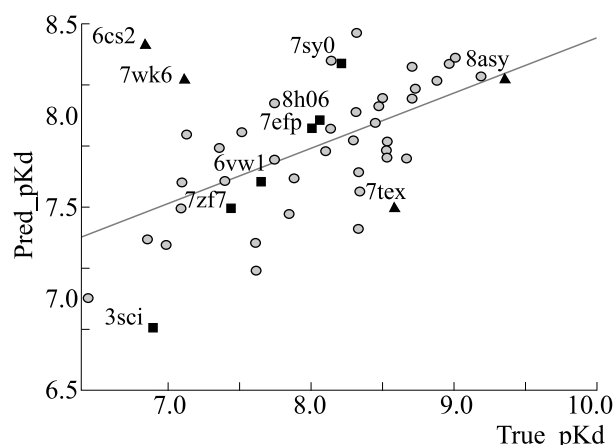


Рис. 5. Результаты предсказаний pK_d для комплексов RBD–ACE2 алгоритмом ProBAN. Треугольниками отмечены комплексы с абсолютной ошибкой больше 1, квадратами — с ошибкой меньше 0.1; добавлены PDB-коды в качестве ярлыков. Серыми кружками отмечены комплексы с абсолютной ошибкой меньше 1, но больше 0.1).

атомами CZ равное 8.9 и 14.3 Å соответственно. Таким образом, согласно нашему представлению, эти остатки аргинина вносят вклад в связывание при рассмотрении структуры 6cs2 и не вносят его при рассмотрении структуры 3sci. Различное положение боковых цепей других остатков интерфейса или соседних с ними так же имеет большое значение, однако подробное рассмотрение подобных различий лежит за рамками данной работы. При этом стоит отметить, что, не считая комплексы 6cs2, 7wk6, для которых структуры получены с низким разрешением (4.4 и 3.7 Å соответственно), наилучшие предсказания характерны для комплексов с более низкой аффинностью связывания ($pK_d < 8$), что ранее было замечено в работе, посвященной оценке других алгоритмов [36]. Данная закономерность может быть связана с тем, что мутации, дестабилизирующие интерфейс связывания, приводят к более крупным конформационным перестройкам, которые более эффективно могут учитываться предсказательными алгоритмами.

ЗАКЛЮЧЕНИЕ

Собран набор данных, содержащий структуры комплексов ACE2–RBD с известными константами связывания, который может использоваться для изучения факторов, влияющих на связывание между RBD и ACE2, а также для обучения и тестирования алгоритмов, осуществляющих предсказание аффинности связывания. Использование нейросетевых методов, в частности, ProBAN, для оценки константы диссоциации и свободной энергии Гиббса возможно не только между нативными RBD и ACE2, но и между другими

мутантными формами. Полученные метрики качества свидетельствуют о недостаточной стабильности предсказаний для комплексов с высокой аффинностью связывания, что указывает на необходимость улучшения предикторов при увеличении набора доступных данных для анализа. Собранный набор данных, а также результаты предсказания выложены в онлайн-репозитории, доступном по ссылке: https://github.com/EABogdanova/ProBAN_RBD-ACE2.

ИСТОЧНИКИ ФИНАНСИРОВАНИЯ

Работа выполнена при финансовой поддержке Российского научного фонда, грант № 19-74-30003.

КОНФЛИКТ ИНТЕРЕСОВ

Авторы заявляют об отсутствии конфликта интересов, связанных с изложенными в статье данными.

СОБЛЮДЕНИЕ ЭТИЧЕСКИХ СТАНДАРТОВ

Настоящая статья не содержит описания собственных исследований с участием людей или животных в качестве объектов.

СПИСОК ЛИТЕРАТУРЫ

1. Choudhary S., Sreenivasulu K., Mitra P., Misra S., and Sharma P. Role of genetic variants and gene expression in the susceptibility and severity of COVID-19. *Ann. Lab. Med.*, **41** (2), 129–138 (2021). DOI: 10.3343/alm.2021.41.2.129
2. Ali A. and Vijayan R. Dynamics of the ACE2-SARS-CoV-2/SARS-CoV spike protein interface reveal unique mechanisms. *Sci. Rep.*, **10** (1), 14214 (2020). DOI: 10.1038/s41598-020-71188-3
3. Vogel M., Augusto G., Chang X., Liu X., Speiser D., Mohsen M. O., and Bachmann M. F. Molecular definition of severe acute respiratory syndrome coronavirus 2 receptor-binding domain mutations: Receptor affinity versus neutralization of receptor interaction. *Allergy*, **77** (1), 143–149 (2022). DOI: 10.1111/all.15002
4. Bozdaganyan M. E., Shaitan K. V., Kirpichnikov M. P., Sokolova O. S., and Orekhov P. S. Computational analysis of mutations in the receptor-binding domain of SARS-CoV-2 spike and their effects on antibody binding. *Viruses*, **14** (2), 295 (2022). DOI: 10.3390/v14020295
5. Lan J., He X., and Ren Y. Structural insights into the SARS-CoV-2 Omicron RBD-ACE2 interaction. *Cell Res.*, **32**, 593–595 (2022). DOI: 10.1038/s41422-022-00644-8

6. Chothia C. and Janin J. Principles of protein–protein recognition. *Nature*, **256**, 705–708 (1975). DOI: 10.1038/256705a0
7. Archakov A. I., Govorun V. M., Dubanov A. V., Ivanov Y. D., Veselovsky A. V., Lewi P., and Janssen P. Protein-protein interactions as a target for drugs in proteomics. *Proteomics*, **3** (4), 380–391 (2003). DOI: 10.1002/pmic.200390053
8. Shi T. L., Li Y. X., Cai Y. D., and Chou K. C. Computational Methods for protein–protein interaction and their application. *Curr. Prot. Peptide Sci.*, **6**, 443–449 (2005). DOI: 10.2174/138920305774329313
9. Vangone A. and Bonvin A. M. Contacts-based prediction of binding affinity in protein-protein complexes. *eLife*, **4**, e07454 (2015). DOI: 10.7554/eLife.07454
10. Purisima E. O., Corbeil C. R., Gaudreault F., Wei W., Deprez C., and Sulea T. Solvated interaction energy: from small-molecule to antibody drug design. *Front. Mol. Biosci.*, **7** (10), 1210576 (2023). DOI: 10.3389/fmolb.2023.1210576
11. Jawad B., Adhikari P., Podgornik R., and Ching W. Y. Key interacting residues between RBD of SARS-CoV-2 and ACE2 receptor: combination of molecular dynamics simulation and density functional calculation. *J. Chem. Inf. Model.*, **61** (9), 4425–4441 (2021). DOI: 10.1021/acs.jcim.1c00560
12. Salleh M. Z., Derrick J. P., and Deris Z. Z. Structural evaluation of the spike glycoprotein variants on SARS-CoV-2 transmission and immune evasion. *Int. J. Mol. Sci.*, **22** (14), 7425 (2021). DOI: 10.3390/ijms22147425
13. Yang Y., Zhang Y., Qu Y., Zhang C., Liu X. W., Zhao M., Mu Y., and Li W. Key residues of the receptor binding domain in the spike protein of SARS-CoV-2 mediating the interactions with ACE2: a molecular dynamics study. *Nanoscale*, **13** (20), 9364–9370 (2021). DOI: 10.1039/d1nr01672e
14. Zhang Z. B., Xia Y. L., Shen J. X., Du W. W., Fu Y. X., and Liu S. Q. Mechanistic origin of different binding affinities of SARS-CoV and SARS-CoV-2 spike RBDs to human ACE2. *Cells*, **11** (8), 1274 (2022). DOI: 10.3390/cells11081274
15. Xue L. C., Rodrigues J. P., Kastrius P. L., Bonvin A. M., and Vangone A. PRODIGY: a web server for predicting the binding affinity of protein-protein complexes. *Bioinformatics*, **32** (23), 3676–3678 (2016). DOI: 10.1093/bioinformatics/btw514
16. Schymkowitz J., Borg J., Stricher F., Nys R., Rousseau F., and Serrano L. The FoldX web server: an online force field. *Nucl. Acids Res.*, **33**, W382–W388 (2005). DOI: 10.1093/nar/gki387
17. Zhang C., Liu S., and Zhou Y. Accurate and efficient loop selections by the DFIRE-based all-atom statistical potential. *Prot. Sci.*, **13** (2), 391–399 (2004). DOI: 10.1110/ps.03411904
18. Lyskov S., and Gray J. J. The RosettaDock server for local protein-protein docking. *Nucl. Acids Res.*, **36**, W233–W238 (2008). DOI: 10.1093/nar/gkn216
19. Bogdanova E. A., Novoseletsky V. N., and Shaitan K. V. Binding affinity prediction in protein-protein complexes using convolutional neural network. In *Advances in Neural Computation, Machine Learning, and Cognitive Research VII (NEUROINFORMATICS 2023)*. Ed. by B. Kryzhanovsky, V. Dunin-Barkowski, et al. (Springer, Cham., 2023), pp. 389–397. DOI: 10.1007/978-3-031-44865-2_42
20. Bogdanova E. A. and Novoseletsky V. N. ProBAN: neural network algorithm for predicting binding affinity in protein-protein complexes. *Proteins*, **92** (9), 1127–1136 (2024). DOI: 10.1002/prot.26700
21. Wang R., Fang X., Lu Y., Yang C. Y., and Wang, S. The PDBbind database: methodologies and updates. *J. Med. Chem.*, **48** (12), 4111–4119 (2005). DOI: 10.1021/jm048957q
22. Moal I. H., Jiménez-García B., and Fernández-Recio J. CCharPPI web server: computational characterization of protein-protein interactions from structure. *Bioinformatics*, **31**(1), 123–125 (2015). DOI: 10.1093/bioinformatics/btu594
23. Pyrkov T. V., Chugunov A. O., Krylov N. A., Nolde D. E., and Efremov R. G. PLATINUM: a web tool for analysis of hydrophobic/hydrophilic organization of biomolecular complexes. *Bioinformatics*, **25** (9), 1201–1202 (2009). DOI: 10.1093/bioinformatics/btp111
24. Humphrey W., Dalke A., and Schulten K. VMD: visual molecular dynamics. *J. Mol. Graphics*, **14** (1), 33–28 (1996). DOI: 10.1016/0263-7855(96)00018-5
25. Su C., He J., Han P., Bai B., Li D., Cao J., Tian M., Hu Y., Zheng A., Niu S., Chen Q., Rong X., Zhang Y., Li W., Qi J., Zhao X., Yang M., Wang Q., and Gao G. F. Molecular basis of mink ACE2 binding to SARS-CoV-2 and its mink-derived variants. *J. Virology*, **96** (17), e0081422 (2022). DOI: 10.1128/jvi.00814-22
26. Han P., Su C., Zhang Y., Bai C., Zheng A., Qiao C., Wang Q., Niu S., Chen Q., Zhang Y., Li W., Liao H., Li J., Zhang Z., Cho H., Yang M., Rong X., Hu Y., Huang N., Yan J., Wang Q., Zhao X., Gao G. F., and Qi J. Molecular insights into receptor binding of recent emerging SARS-CoV-2 variants. *Nature Commun.*, **12** (1), 6103 (2021). DOI: 10.1038/s41467-021-26401-w
27. Wang Q., Zhang Y., Wu L., Niu S., Song C., Zhang Z., Lu G., Qiao C., Hu Y., Yuen K. Y., Wang Q., Zhou H., Yan J., and Qi J. Structural and functional basis of SARS-CoV-2 entry by using human ACE2. *Cell*, **181** (4), 894–904 (2020). DOI: 10.1016/j.cell.2020.03.045
28. Starr T. N., Greaney A. J., Hannon W. W., Loes A. N., Hauser K., et al. Shifting mutational constraints in the SARS-CoV-2 receptor-binding domain during viral evolution. *Science (New York)*, **377** (6604), 420–424 (2022). DOI: 10.1126/science.abo7896

29. Geng Q., Shi K., Ye G., Zhang W., Aihara H., and Li F. Structural basis for human receptor recognition by SARS-CoV-2 Omicron variant BA.1. *J. Virology*, **96** (8), e0024922 (2022). DOI: 10.1128/jvi.00249-22
30. Nelson-Sathi S., Umasankar P. K., Sreekumar E., Nair R. R., Joseph I., Nori S. R. C., Philip J. S., Prasad R., Navyasree K. V., Ramesh S., Pillai H., Ghosh S., Santosh Kumar T. R., and Pillai M. R. Mutational landscape and *in silico* structure models of SARS-CoV-2 spike receptor binding domain reveal key molecular determinants for virus-host interaction. *BMC Mol. Cell Biol.*, **23** (1), 2 (2022). DOI: 10.1186/s12860-021-00403-4
31. Kastriitis P. L., Rodrigues J. P., Folkers G. E., Boelens R., and Bonvin, A. M. Proteins feel more than they see: Fine-tuning of binding affinity by properties of the non-interacting surface. *J. Mol. Biol.*, **426** (14), 2632–2652 (2014). DOI: 10.1016/j.jmb.2014.04.017
32. Reichmann D., Phillip Y., Carmi A., and Schreiber G. On the contribution of water-mediated interactions to protein-complex stability. *Biochemistry*, **47** (3), 1051–1060 (2008). DOI: 10.1021/bi7019639
33. Schweke H., Mucchielli M. H., Sacquin-Mora S., Bei W., and Lopes A. Protein interaction energy landscapes are shaped by functional and also non-functional partners. *J. Mol. Biol.*, **432** (4), 1183–1198 (2020). DOI: 10.1016/j.jmb.2019.12.047
34. Ozden B., Şamiloğlu E., Özsan A., Erguven M., Yükrük C., Koşaca M., Oktayoğlu M., Menteş M., Arslan N., Karakulah G., Barlas A. B., Savaş B., and Karaca E. Benchmarking the accuracy of structure-based binding affinity predictors on Spike-ACE2 deep mutational interaction set. *Proteins*, **92** (4), 529–539 (2024). DOI: 10.1002/prot.26645

Assessment of Binding Affinity in the Complexes of CoV-S-Protein's RBD and the ACE2 Using Convolutional Neural Networks

E.A. Bogdanova*, A.V. Chernukhin, K.V. Shaitan*, and V.N. Novoseletsky*****

*Faculty of Biology, Lomonosov Moscow State University, Leninskie Gory 1/12, Moscow, 119991 Russia

**Faculty of Digital Technologies, and Chemical Engineering, D.I. Mendeleev University of Chemical Technology of Russia, Miusskaya pl. 9, Moscow, 125047 Russia

***Faculty of Biology, Shenzhen MSU-BIT University, International University Park Road, 1, Dayun New Town, Longgang District, Shenzhen, Guangdong Province, 518172, People's Republic of China

The experimentally obtained structures of 48 complexes of the ACE2 receptor with the S protein's RBD of the coronaviruses SARS-CoV and SARS-CoV-2 (including mutant forms of the latter) were assessed and the dissociation constant was calculated for them. Prediction of binding affinity was carried out using ProBAN, a neural network algorithm, previously developed by the authors, and a number of other algorithms for estimating the Gibbs free energy such as Prodigy, FoldX, DFIRE and RosettaDock. A comparison of the evaluation results shows that ProBAN has the best prediction quality (Pearson correlation – 0.56, MAE – 0.7 kcal/mol) of all the analyzed algorithms. The results obtained suggest better quality of affinity prediction for other protein-protein complexes. Information about the complexes under study and prediction results are available in the repository at the link: https://github.com/EABogdanova/ProBAN_RBD-ACE2.

Keywords: protein-protein interactions, interaction interface, binding affinity, coronavirus, RBD–ACE2, bioinformatics