

ОЦЕНКА КАЧЕСТВА ПРОГНОЗА АНТИ-SARS-CoV-2 АКТИВНОСТИ
С ПОМОЩЬЮ ВЕБ-СЕРВИСА D3Targets-2019-nCoV

Н.С. Ионов*, П.В. Погодин, В.В. Поройков

Научно-исследовательский институт биомедицинской химии имени В.Н. Ореховича,
119121 Москва, ул. Погодинская, 10; *e-mail: ionov.nikita.serg@gmail.com

Проведена оценка веб-сервиса D3Targets-2019-nCoV, позволяющего на основании структурного сходства и молекулярного докинга предсказывать взаимодействие химических соединений с белками вируса SARS-CoV-2 и белками человека, вовлеченными в патогенез COVID-19. Качество прогноза оценено как сбалансированная точность, которая была рассчитана по результатам прогноза для структур химических соединений из сформированной нами тестовой выборки. В тестовую выборку вошло 35 активных и 59 неактивных молекул, включающих в себя соединения с установленным отсутствием активности в отношении выбранных мишеней и соединения, активные по отношению к мишеням SARS-CoV-2, не представленным в базе данных CoViLigands. Авторами анализируемого веб-сервиса не указаны пороги значений оценки сходства и оценочной функции докинга, используя которые можно было бы достоверно разделить соединения на активные и неактивные по отношению к белкам-мишеням. Поэтому нами была проведена оценка сбалансированной точности прогностических методов D3Targets-2019-nCoV при различных порогах отсека активных веществ от неактивных. С использованием сформированной нами выборки установлено, что наибольшее значение сбалансированной точности (0.59) достигается при выборе активных молекул по результатам оценки 2D сходства (порог отсека равен 46%). Оценка 3D сходства не позволила достичь значений сбалансированной точности, превышающих 0.5. Показано, что при использовании рекомендуемой авторами интегральной оценки сходства 2Dx3D, максимальное значение сбалансированной точности (0.57) достигается при пороге, равном 31%. При расчете сбалансированной точности для результатов молекулярного докинга показано, что она не превышает 0.51. На примере препарата тидеглусиб показано, что значения оценочной функции при докинге к двум белкам-мишеням, активность в отношении которых установлена в эксперименте (3CLpro и GSK3B), существенно не отличаются от значений оценочной функции докинга к остальным 44 белкам-мишеням, активность в отношении которых не подтверждена экспериментально.

Ключевые слова: SARS-CoV-2; оценка молекулярного подобия; молекулярный докинг; поиск новых антикоронавирусных соединений; тестирование качества веб-сервиса

DOI: 10.18097/BMCRM00140

ВВЕДЕНИЕ

COVID-19 – передающаяся воздушно-капельным путем острая респираторная инфекция, которую вызывает представитель семейства коронавирусов, SARS-CoV-2 (2019-nCoV) [1]. По данным Всемирной организации здравоохранения, на 25 октября 2020 года в мире было выявлено около 44.8 млн. случаев заражения COVID-19 и почти 1.15 млн. летальных исходов, связанных с коронавирусной инфекцией [2]. Российская Федерация занимает 4 место в мире по количеству заболевших с 1.51 млн. подтвержденных случаев при 17 803 955 проведенных тестах, 1.14 млн. на текущий момент разрешились выздоровлением, а 25 050 – летальным исходом [3].

По данным ReDO project, на момент написания публикации, в мире проводится 1618 клинических исследований, свыше 64% которых направлено на репозиционирование разрешенных к медицинскому применению фармакологических субстанций [4]. Такое положение, в частности, связано с тем, что репозиционирование лекарств требует меньших финансовых затрат; при этом само исследование занимает существенно более короткое время в сравнении с традиционными подходами, направленными на поиск оригинальных препаратов («first-in-the-class drugs») [5]. В настоящее время наибольшее количество исследований посвящено антибиотику азитромицину (65) и иммунодепрессанту

тоцилизумабу (57). С целью поиска новых химических соединений с антикоронавирусной активностью проводится около 550 клинических испытаний [6].

Препараты, которые в настоящее время применяются для экспериментальной (off-the-label) терапии COVID-19, характеризуются умеренной эффективностью, которая зависит от стадии и тяжести течения заболевания, а также широким набором побочных эффектов, что ограничивает возможность их применения [7]. Так, результаты недавнего рандомизированного исследования эффективности ремдезивира [8] на 1062 пациентах показали, что по сравнению с плацебо данный препарат позволил добиться более быстрого выздоровления (на 5 дней), снижения смертности (на 5.2 % и 3.8% при оценке на 15 и 29 дни соответственно), при этом серьезные побочные эффекты наблюдались у 24.6% пациентов (впрочем, при приеме плацебо они наблюдались у 31.6% пациентов). Более того, согласно препринту группы исследователей ВОЗ [9], в котором представлены результаты рандомизированного исследования эффективности целого ряда противовирусных препаратов, применяемых в терапии COVID-19, ни один из существующих вариантов терапии не позволяет добиться существенного снижения смертности, времени госпитализации и потребности в искусственной вентиляции легких, что было показано на 11266 пациентах, включенных в выборку случайным образом в госпиталях разных стран [8]. Таким образом, для дальнейшей



разработки высокоэффективных терапевтических препаратов актуальным является поиск и создание новых фармакологических веществ, активных в отношении белков-мишеней вируса SARS-CoV-2 или белков человека, которые участвуют в патогенезе COVID-19.

С целью выявления в обширных библиотеках химических соединений, перспективных для дальнейшего тестирования на антикоронавирусную активность *in vitro* и *in vivo*, широко применяют методы компьютерного конструирования лекарств. При проведении поискового запроса по ключевым словосочетаниям «COVID-19» и «Virtual screening» в Google Scholar установлено, что за десять месяцев с начала 2020 года, опубликовано свыше 42 тыс. работ, посвященных виртуальному скринингу с использованием оценки структурного сходства, методов машинного обучения и молекулярного докинга.

Первым в мире свободно доступным в сети Интернет веб-ресурсом, позволяющим прогнозировать взаимодействие химических соединений с белками-мишенями, стала платформа D3Targets-2019-nCoV [10]. В её основе лежит поиск по молекулярному подобию (D3Similarity) среди 604 соединений из базы данных (БД) CoViLigands, которая была составлена разработчиками с учетом информации о химических соединениях, активных по отношению к представителям семейства коронавирусов SARS, MERS и SARS-CoV-2, и молекулярный докинг (D3Docking) к набору из 47 отобранных авторами белков-мишеней. При этом в оригинальной публикации, посвященной описанию модуля D3Similarity [10], количество соединений, содержащихся в БД CoViLigands, составляет 470.

В оригинальной публикации [10], авторами приведено описание результатов тестирования модуля D3Similarity на примере двух мишеней - 3CLpro и PLPro. В ходе тестирования проводился расчёт оценок сходства между каждым активным соединением и остальными соединениями БД CoViLigands. Исследователями оценивался процент верных результатов прогноза, попавших в топ 10 при ранжировании по разным оценкам сходства (2D, 3D и 2Dx3D). Установлено, что наибольший процент предсказанных верных совпадений с известными белками-мишенями наблюдается при использовании интегральной оценки 2Dx3D. Однако, как для модуля D3Similarity, так и для модуля D3Docking, независимая оценка качества прогноза, учитывающая как правильные, так и ошибочные результаты прогноза, отсутствует. В связи с этим, целью настоящей работы стало независимое проведение оценки возможности применения веб-ресурса D3Targets-2019-nCoV для отбора соединений, обладающих антикоронавирусной активностью, т.е. потенциальных фармакологических веществ для терапии SARS-CoV-2/COVID-19.

МАТЕРИАЛЫ И МЕТОДЫ

D3Targets-2019-nCoV

Тестируемый веб-сервис предоставляет пользователям возможность проводить виртуальный скрининг с использованием двух подходов: оценки сходства структуры лигандов и молекулярного докинга к ряду белков-мишеней. Данные подходы реализованы в вычислительных модулях «D3Similarity» и «D3Docking» соответственно.

Для предварительной обработки структур, как вводимых

пользователем, так и содержащихся в БД, используется программа OpenBabel [11] и библиотека RDKit [12]; в результате обработки структуры химических соединений транслируются в 3D формат MOL2 и оптимизируются [13].

D3Similarity

D3Similarity – вычислительный модуль платформы D3Targets-2019-nCoV для расчёта оценок сходства (2D, 3D и 2Dx3D) между загруженными пользователем структурами и структурами, содержащимися в БД CoViLigands [14].

В модуле «D3Similarity» реализованы две функции: TargetPrediction и VirtualScreening.

Функция TargetPrediction принимает на вход структуру одного химического соединения, представленную в формате MOL или SDF и, выводит пользователю оценку сходства со структурами соединений, содержащихся в БД CoViLigands. Пользователь получает таблицу, содержащую 20 наивысших оценок сходства. Наивысшие оценки сходства выбираются среди произведений оценок 2D и 3D сходства, так как, согласно результатам авторов [15], именно этот интегральный параметр показал наилучшую предсказательную способность.

Функция VirtualScreening принимает на вход до 99 структур, однако оценка сходства проводится только с соединениями, представленными в БД CoViLigands как активные по отношению к выбранным пользователем белкам-мишеням (максимум двум).

D3Docking

Докинг проводится в установленные авторами потенциальные карманы связывания объёмом более 200 Å³ с помощью программы AutoDock Vina [16] с последующей симуляцией молекулярной динамики моделей лиганд-белковых комплексов с применением разработанной исследователями компьютерной программы vsREMD [17]. Детальное описание подготовки структур для докинга и перечень используемых авторами компьютерных программ представлены в оригинальной публикации авторов [10].

Как и D3Similarity, модуль D3Docking включает две функции – VirtualScreening и TargetPrediction, которые позволяют проводить докинг для нескольких химических соединений (максимум 99) с одним или двумя лиганд-связывающими сайтами белка-мишени, или молекулярный докинг одного соединения со всеми доступными потенциальными карманами связывания.

Тестовая выборка

Тестовая выборка для проведения прогноза была составлена нами на основе информации, извлеченной из доступных источников: БД ChEMBL 27 и Coronavirus Antiviral Research Database [18, 19]. Всего в тестовую выборку вошли структуры 94 химических соединений. В тестовой выборке представлены как структуры, для которых в условиях эксперимента *in vitro* было продемонстрировано наличие антикоронавирусной активности по отношению к SARS-CoV-2 с указанием предполагаемых белков-мишеней, так и структуры соединений, для которых либо было показано отсутствие антикоронавирусной активности в эксперименте, либо соответствующее тестирование не

Таблица 1. Характеристика соединений тестовой выборки.

Белок-мишень	Сокращение	Организм	Количество протестированных соединений активно / неактивно	Общее количество отрицательных примеров
Основная протеаза коронавируса	<i>3CLpro</i>	SARS-CoV-2	26/16	59
Папаин-подобная протеаза	<i>PLPro</i>	SARS-CoV-2	3/7	59
Мембрано-связанная сериновая протеаза	<i>TMPRSS2</i>	Человек	3/5	59
РНК-зависимая РНК-полимераза	<i>RdRp</i>	SARS-CoV-2	3/0	59

проводилось. Количественные характеристики тестовой выборки с разбивкой по белкам-мишеням приведены в таблице 1.

В таблице 1 к отрицательным примерам отнесены соединения, для которых было экспериментально установлено отсутствие активности по отношению к рассматриваемым в нашей работе белкам-мишеням (16 - к 3CLpro, 7 - к PLPro, 5 - к RdRp), а также соединения, ингибирующие репликацию вируса в клеточной культуре путем воздействия на иные мишени, не представленные в БД CoViLigands (например, на ангиотензин-конвертирующий фермент (ACE2)).

Структуры соединений с установленными экспериментально видами активности рассматривались в качестве положительных примеров; остальные – в качестве отрицательных примеров. Под истинно положительным (ИП) результатом прогноза понимается случай, когда прогнозируемая веб-ресурсом D3Targets-2019-nCoV для анализируемого вещества активность подтверждена экспериментально; под ложноотрицательным (ЛО) результатом прогноза понимается случай, когда активность установлена экспериментально, но не прогнозируется. В качестве истинно отрицательных (ИО) результатов нами принимались случаи, когда для неактивных соединений мишень не была предсказана; ложноположительными (ЛП) результатами считались случаи, когда для соединения прогнозировалась мишень, взаимодействие с которой экспериментально не установлено.

Известно, что в большинстве случаев соединения с высоким структурным сходством проявляют сходные виды активности [20]. Поскольку ни для одного из веществ сформированной нами тестовой выборки не было исследовано взаимодействие со всеми анализируемыми в работе мишенями, мы оценили структурное сходство внутри группы соединений, активных по отношению к конкретной мишени, и их структурное сходство с соединениями, для которых взаимодействие данной мишени не установлено. Оценку сходства проводили между химическими структурами, представленными в виде наборов двух типов подструктурных дескрипторов – MNA [21] и QNA [22].

MNA-дескрипторы (англ. Multilevel Neighbourhoods of Atoms) – представление молекулы, в котором явно указаны, согласно валентностям и зарядам атомов, все атомы водорода и не учитываются типы связей. Построение MNA-дескрипторов для каждого атома молекулы происходит рекурсивно и основано на следующем алгоритме:

- MNA-дескриптор нулевого уровня – метка A самого атома;

- MNA-дескриптор любого следующего уровня – условное обозначение структурного фрагмента $A(D_1D_2...D_i...)$, где D_i – MNA-дескриптор предыдущего уровня для i -го непосредственного соседа данного атома с меткой A . Дескрипторы соседей $D_1D_2...D_i...$ записываются в однозначном лексикографическом порядке.

QNA-дескрипторы (англ. Quantitative Neighbourhoods of Atoms) – рассчитываются с использованием значений потенциала ионизации (ИП) и сродства к электрону (ЕА) для каждого атома молекулы. Расчёт QNA-дескрипторов производится по формулам (1-3).

$$P_i = B_i \sum_k \left(\text{Exp} \left(-\frac{1}{2} C \right) \right)_{ik} B_k \quad (1),$$

$$Q_i = B_i \sum_k \left(\text{Exp} \left(-\frac{1}{2} C \right) \right)_{ik} B_k A_k \quad (2),$$

$$A_k = \frac{1}{2} (IP_k + EA_k), B_k = (IP_k - EA_k)^{\frac{1}{2}} \quad (3),$$

где i, k — номера атомов в молекуле, C – матрица смежности молекулы.

Методы оценки структурного сходства на основе MNA и QNA дескрипторов детально описаны в работе [23].

Оценка точности прогноза взаимодействия анализируемых соединений с белками-мишенями COVID-19 с использованием D3Targets-2019-nCoV.

На основе экспериментально установленных данных о взаимодействии веществ тестовой выборки с молекулярными мишенями мы рассчитали сбалансированную точность прогноза (СТ) по формуле (4):

$$СТ = \frac{1}{2} \left(\frac{ИП}{ИП + ЛО} + \frac{ИО}{ИО + ЛП} \right) \quad (4),$$

где ИП – количество истинно положительных; ЛО – количество ложноотрицательных результатов прогноза; ИО – количество истинно отрицательных результатов; ЛП – количество ложноположительных результатов.

Важно заметить, что расчёт СТ проводили по совокупности полученных результатов, т. е. по всем мишеням; значения СТ для каждой из мишеней рассчитаны не были, поскольку количество активных соединений по отношению к отдельным мишеням невелико.

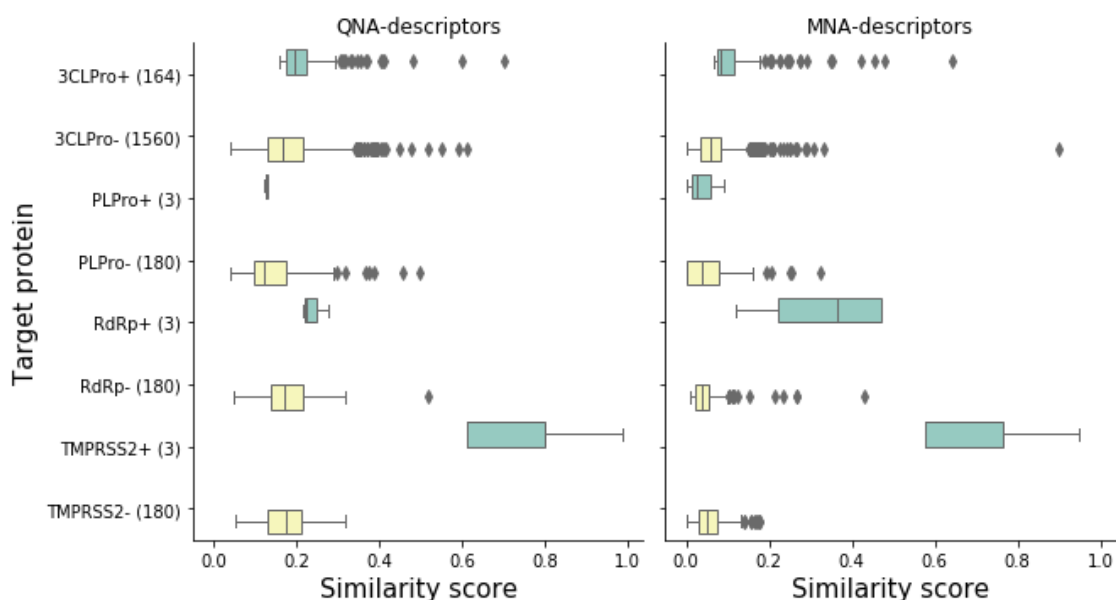


Рисунок 1. Оценки сходства (Similarity score) группы активных соединений между собой и с отрицательными примерами на основе MNA- и QNA-дескрипторов. 3CLpro, PLPro, RdRp, TMPRSS2 – аббревиатуры соответствующих белков-мишеней, «+» – указание на группу попарных оценок сходства, полученных между соединениями активными по отношению к мишени, «-» – указание на группу попарных оценок сходства между соединениями активными по отношению к мишени и соединениями для которых данная активность отсутствует или не исследована (отрицательные примеры). В скобках указано число попарных оценок сходства.

Отметим, что авторы ресурса не приводят пороги значений сходства и оценочной функции докинга для чёткого деления анализируемых химических соединений на активные/неактивные по отношению к белкам-мишеням. Поэтому, в процессе работы, мы использовали результаты оценки сходства и расчёта оценочной функции докинга, полученные для соединений тестовой выборки и белков-мишеней, активность в отношении которых подтверждена в условиях тестирования *in vitro*. Это соответствует подходу, использованному авторами веб-ресурса [10].

РЕЗУЛЬТАТЫ

Сравнение оценок сходства соединений тестовой выборки

Оценки сходства для активных и отрицательных примеров соединений и их сопоставление приведены на рисунке 1. Распределения оценок сходства для каждой из групп представлены в виде прямоугольников. Левая граница прямоугольника указывает на значение нижнего квантиля, правая – значение верхнего квантиля. Границы, выходящие на пределы прямоугольника, соответствуют минимальному и максимальному значению из совокупности оценок сходства внутри каждой из групп. Выбросы обозначены в виде отдельных точек. Вертикальная линия, локализованная внутри прямоугольника, указывает на значение медианы.

Как видно из приведённых на рисунке 1 данных, оценки сходства внутри группы соединений, активных по отношению к TMPRSS2 (мембрано-связанная сериновая протеаза, единственная мишень, представляющая белок человека), существенно выше, чем оценки сходства активных соединений с отрицательными примерами для TMPRSS2 (3 протестированных соединения и 59 отрицательных примеров). Такие результаты получены как при оценке сходства по MNA-дескрипторам (среднее значение для

активных соединений равно 0.70 ± 0.18 , среднее значение оценок сходства отрицательных примеров с активными соединениями равно 0.06 ± 0.04), так и при оценке сходства по QNA-дескрипторам (0.74 ± 0.18 для группы активных соединений, 0.17 ± 0.06 для оценки сходства между активными соединениями и отрицательными примерами). Для группы активных соединений и оценки сходства отрицательных примеров и активными соединениями в отношении РНК-зависимой РНК-полимеразы RdRp получен аналогичный результат. Среднее значение полученных оценок сходства по MNA-дескрипторам для группы активных соединений равно 0.33 ± 0.15 , по QNA-дескрипторам – 0.24 ± 0.03 . Для оценки сходства между отрицательными примерами и активными соединениями среднее значение оценки сходства по MNA-дескрипторам равно 0.05 ± 0.04 , по QNA-дескрипторам – 0.18 ± 0.06 .

В анализируемой нами тестовой выборке только три соединения обладают активностью по отношению к PLpro. Значение оценки сходства по MNA-дескрипторам составляет 0.04 ± 0.03 , по QNA-дескрипторам составляет 0.13 ± 0.01 . Среднее значение оценки сходства между отрицательными примерами и соединениями, активными в отношении PLpro, по MNA-дескрипторам составляет 0.05 ± 0.04 , по QNA-дескрипторам составляет 0.14 ± 0.07 .

Средняя оценка сходства для соединений, активных по отношению к 3CLpro, по MNA-дескрипторам равна 0.12 ± 0.08 , по QNA-дескрипторам равна 0.22 ± 0.07 . Для отрицательных примеров и соединений, активных по отношению к 3CLpro, средние оценки составляют 0.06 ± 0.05 по MNA-дескрипторам и 0.18 ± 0.07 по QNA-дескрипторам.

Таким образом, значимые структурные различия наблюдаются для TMPRSS2, и RdRp. В случае PLpro значимых различий по сходству активных соединений друг с другом и с отрицательными примерами не наблюдается, а для 3CLpro активные соединения также незначительно отличаются от отрицательных примеров.

Таблица 2. Результаты оценки сходства функцией TargetPrediction, реализованной в вычислительном модуле D3Similarity, на примере препарата тидеглусиб.

Rank	Mol ID	Similarity	3D Similarity (%)	2D Similarity (%)	Target Name/ MoA	UniProt ID/ ProteinID	Activity
1	ICV265	78.90	78.90	100	virus: 3C-like protease	P0DTD1	SARS-CoV-2 3CL (IC ₅₀ =1.55 μM)
2	ICV616	18.66	75.47	24.72	Unknown		SARS (IC ₅₀ =14.5 μM) Vero
3	ICV459	17.58	75.39	23.31	Unknown		MERS (EC ₅₀ =6.212 μM) Vero E6; SARS (EC ₅₀ =1.274 μM); Vero E6
4	ICV431	17.22	64.83	26.56	human: Histamine H1 receptor	P35367	MERS (EC ₅₀ =4.884 μM) Vero E6; SARS (EC ₅₀ =5.59 1μM) Vero E6
5	ICV432	16.94	72.66	23.31	human: Histamine N-ethyltransferase	P50135	MERS-CoV (EC ₅₀ =2.1 μM)
6	ICV341	16.77	67.09	25	virus: Papain-like protease	P0DTD1	SARS PL (IC ₅₀ =4.5 μM)
7	ICV340	16.72	68.50	24.40	virus: Papain-like protease	P0DTD1	SARS PL (IC ₅₀ =7 μM)
8	ICV189	16.68	79.90	20.88	virus: 3C-like protease	P0DTD1	SARS 3CL (IC ₅₀ =0.03 μM)
9	ICV310	16.68	71.21	23.42	virus: Papain-like protease	P0DTD1	SARS PL (IC ₅₀ =2.7 μM)
10	ICV197	16.67	73.95	22.54	virus: 3C-like protease	P0DTD1	SARS 3CL (IC ₅₀ =18.2 μM)
11	ICV305	16.38	85.55	19.15	virus: Papain-like protease	P0DTD1	SARS PL (IC ₅₀ =46.1 μM)
12	ICV344	16.37	78.95	20.73	virus: Papain-like protease	P0DTD1	SARS PL (IC ₅₀ =26.3 μM)
13	ICV274	16.33	76.75	21.28	virus: Papain-like protease	P0DTD1	MERS PL (IC ₅₀ =20 μM)
14	ICV328	16.12	70.29	22.93	virus: Papain-like protease	P0DTD1	SARS PL (IC ₅₀ =1.9 μM)
15	ICV487	16.01	69.87	22.92	Unknown		MERS (EC ₅₀ =7.416 μM) Vero E6; SARS (EC ₅₀ =15.553 μM) Vero E6
16	ICV329	15.62	77.07	20.27	virus: Papain-like protease	P0DTD1	SARS PL (IC ₅₀ =0.47 μM)
17	ICV193	15.56	80.71	19.28	virus: 3C-like protease	P0DTD1	SARS 3CL (IC ₅₀ =11.2 μM)
18	ICV342	15.54	62.16	25	virus: Papain-like protease	P0DTD1	SARS PL (IC ₅₀ =6.8 μM)
19	ICV327	15.49	67.56	22.93	virus: Papain-like protease	P0DTD1	SARS PL (IC ₅₀ =18 μM)
20	ICV457	15.36	76.30	20.13	Unknown		MERS (EC ₅₀ =6.275 μM) Vero E6; SARS (EC ₅₀ =6.538 μM) Vero E6

Примечание. «Rank» - ранг соединений по степени сходства со структурами из БД ресурса (от самой высокой оценки сходства – «1», к самой низкой); «Mol ID» - идентификатор соединения, представленного в базе данных CoViLigands, с которым сравнивалось исследуемое соединение; «3D Similarity» – оценка сходства трёхмерных структур; «2D Similarity» – оценка сходства двумерных структур; «Similarity» - произведение оценок двумерного и трёхмерного сходства; «Target name/MoA» - название белка-мишени или механизма действия; «Activity» - данные об активности соединения из CoViLigands.

Таблица 3. Оценка сходства функцией VirtualScreening, реализованной в вычислительном модуле D3Similarity, для 3CLpro и PLpro.

Compound name/ID	Similarity	Similarity 3D (%)	Similarity 2D (%)	Ligands	Target name
GS-441524	26.21	65.40	40.08	ICV199	3CLpro
6-thioguanine	39.18	88.67	44.19	ICV46	3CLpro
Beta-D-N4-Hydroxycytidine	20.32	81.84	24.83	ICV46	3CLpro
C25H22FN4O4	69.61	69.61	100	ICV273	3CLpro
CHEMBL1809250	23.41	66.95	34.97	ICV42	3CLpro
CHEMBL1809251	22.69	66.52	34.10	ICV272	3CLpro
CHEMBL1809259	30.15	71.45	42.20	ICV272	3CLpro
GC376	55.42	55.42	100	ICV21	3CLpro
GRL0617	27.06	75.67	35.77	ICV190	3CLpro
CHEMBL1388469	19.71	72.98	27	ICV206	3CLpro
Remdesivir	16.34	62.09	26.32	ICV202	3CLpro
Tideglusib	74.6	74.60	100	ICV265	3CLpro
GS-441524	23.11	62.60	36.92	ICV322	PLpro
6-thioguanine	91.12	91.12	100	ICV350	PLpro
Beta-D-N4-Hydroxycytidine	14.93	74.64	20	ICV350	PLpro
C25H22FN4O4	33.83	56.89	59.46	ICV343	PLpro
CHEMBL1809250	24.82	68.76	36.09	ICV330	PLpro
CHEMBL1809251	23.08	66.16	34.88	ICV326	PLpro
CHEMBL1809259	22.51	65.33	34.46	ICV314	PLpro
GC376	19.27	64.50	29.88	ICV327	PLpro
GRL0617	85.39	85.39	100	ICV357	PLpro
CHEMBL1388469	17.28	67.02	25.78	ICV275	PLpro
Remdesivir	15.28	71.37	21.40	ICV287	PLpro
Tideglusib	18.69	79.81	23.42	ICV310	PLpro

Примечание. Столбец с названием с выбранной мишени (в данной случае – 3C-like protease и papain-like protease), в данной столбце содержатся данные «Similarity»; «Similarity 3D» – оценка сходства трехмерных структур; «Similarity 2D» – оценка сходства двумерных структур; «Ligands» – идентификатор соединения, представленного в базе данных CoViLigands, с которым сравнивалось исследуемое соединение. В таблице 2 наименование столбца «Mol ID» было заменено нами, на «Compound name/ID», в котором указаны названия или идентификаторы соединений тестовой выборки. В изначально полученных результатах данные столбца «Mol ID» имели вид MOL_n, где n – порядковый номер структуры в загружаемом SDF. Жирным шрифтом выделены соединения, характеризующиеся экспериментально установленной активностью в отношении соответствующих белков-мишеней.

Оценка применения методов подобия, реализованных в модуле D3Similarity

Возможности использования функции TargetPrediction модуля D3Similarity мы рассмотрели на примере прогноза для структурной формулы препарата тидеглусиб.

Тидеглусиб – селективный ингибитор киназы гликогенсинтазы 3 бета (GSK3-beta), который в настоящее время проходит клинические исследования как препарат для лечения болезни Альцгеймера [24]. В исследованиях *in vitro* было показано, что тидеглусиб является ингибитором основной протеазы вируса SARS-CoV-2 (3C-like protease, 3CLpro) [25]. Результаты оценки его сходства со структурными формулами веществ, содержащихся в БД CoViLigands, приведены в таблице 2.

Таблицы 2-5, содержащие результаты расчёта оценочной функции докинга и оценки сходства, приведены в соответствии с представлением на тестируемом веб-сервисе. Название столбца «Target ID» заменено на «UniProt ID/ProteinID».

Как видно из приведенных в таблице 2 данных, тидеглусиб содержится в базе данных CoViLigands (ICV265),

и оценка 2D сходства позволяет его идентифицировать (сходство равно 100%). Ранее было отмечено, что структуры, представленные в БД CoViLigands, и структуры, загруженные пользователем, подвергаются идентичной предобработке. При этом, однако, оценка 3D сходства между двумя структурами тидеглусиба составляет только 78.90%. Было также обнаружено сходство с другими ингибиторами основной протеазы (ICV189, ICV197 и ICV193); для них оценки 2D сходства составляют 20.88%, 22.54% и 19.28%, а 3D сходства 79.90%, 73.95% и 80.71% соответственно. Помимо вышеупомянутых 4-х ингибиторов основной протеазы SARS-CoV-2, в БД CoViLigands содержатся 266 соединений с аналогичной активностью, оценка сходства с которыми для тидеглусиба ниже, чем для соединений, вошедших в список Top 20 (табл. 2).

Необходимо отметить, что в Top 20 вошли мишени, о взаимодействии с которыми препарата тидеглусиб сведения отсутствуют (Histamine H1 receptor, Histamine N-methyltransferase и Papain-like protease), а также 3 неизвестные мишени (в табл. 2 Unknown). Опираясь на сведения, представленные в ячейке «Activity», можно сделать вывод о том, что Unknown, вероятно, указывает на

Таблица 4. Примеры превышение полученных оценок 3D сходства для различных соединений в сравнении с оценками 3D сходства для одинаковых соединений.

Оценка 3D сходства для одинаковых соединений	Mol ID	Similarity	3D Similarity	2D Similarity	Target Name/ MoA	Target ID	Activity	Compound name
61.50%	ICV10	34.98	73.29%	47.73%	virus: 3C-like protease	P0DTD1	SARS 3CL (IC ₅₀ =63 µM)	C25H22FN4O4
78.90%	ICV193	15.56	80.71%	19.28%	virus: 3C-like protease	P0DTD1	SARS 3CL (IC ₅₀ =11.2 µM)	Tideglusib
78.90%	ICV305	16.38	85.55%	19.15%	virus: Papain-like protease	P0DTD1	SARS PL (IC ₅₀ =46.1 µM)	Tideglusib
60.58%	ICV139	13.42	68.83%	19.50%	virus: 3C-like protease	P0DTD1	SARS 3CL (IC ₅₀ =5.52 µM)	Remdesivir
60.58%	ICV118	14.63	68.64%	21.32%	virus: 3C-like protease	P0DTD1	SARS 3CL (K _i =6.7 µM)	Remdesivir
60.58%	ICV412	15.83	67.92%	23.30%	human: Calcium channel	-	MERS (IC ₅₀ =2.36 µM) Vero	Remdesivir
60.58%	ICV428	15.2	63.10%	24.09%	human: Eukaryotic initiation factor 4A-I	P60842	MERS (EC ₅₀ =3 nM) MRC-5 cell; HCoV-229E (EC ₅₀ =1.3 nM) MRC-5 cell	Remdesivir
60.58%	ICV392	18.23	62.17%	29.32%	virus: Helicase	P0DTD1	SARS Helicase (IC ₅₀ =5.4 µM)	Remdesivir

Примечание. «Mol ID» – идентификатор соединения, представленного в базе данных CoViLigands, с которым сравнивалось исследуемое соединение; «3D Similarity» – оценка сходства трёхмерных структур; «2D Similarity» – оценка сходства двумерных структур; «Similarity» – произведение оценок двумерного и трёхмерного сходства; «Target name/MoA» – название белка-мишени или механизма действия; «Activity» – данные об активности соединения из CoViLigands.

наличие антикоронавирусной активности, установленной при тестировании в культуре клеток без идентификации молекулярной мишени.

Результаты работы функции VirtualScreening модуля D3Similarity были получены с использованием SD файла (SDF, Structure Data File), содержащего все структуры тестовой выборки (табл. 1). Оценки сходства, полученные с использованием функции VirtualScreening, приведены в таблице 3.

Как видно из приведенных в таблице 3 данных, для всех загруженных нами соединений, активных по отношению к 3CLpro и PLpro, оценки 2D сходства составляют 100%; следовательно, их структуры представлены в БД CoViLigands. Как уже говорилось выше, загруженные для оценки сходства и представленные в БД CoViLigands соединения проходят одинаковые процедуры предобработки и приводятся к единому формату. Из таблицы 3, однако, видно, что оценки 3D сходства для идентичных соединений не достигают 100%. Также было установлено, что в полученных результатах многократно встречаются случаи, когда оценка 3D сходства между идентичными соединениями превышает оценку 3D сходства соединений самих с собой. Примеры

таких случаев представлены в таблице 4. В связи с этим возникает вопрос о корректности методов оценки 3D сходства между анализируемыми и представленными в БД структурами.

Функции TargetPrediction и VirtualScreening модуля D3Similarity основаны на идентичных алгоритмах; однако TargetPrediction позволяет получить результаты для интерпретации сразу по многим мишеням; поэтому в дальнейшей работе по оценке качества прогноза, мы использовали именно функцию TargetPrediction. Оценки сходства, полученные с использованием функции TargetPrediction для каждой структуры тестовой выборки, приведены в дополнительных материалах к статье.

Поскольку авторами работы не указан порог значений сходства для разделения соединений на активные и неактивные, нами была произведена оценка сбалансированной точности при разных порогах отсека для оценок сходства двумерных и трехмерных структур. При этом результаты прогноза, превышающие установленный порог, считались положительными, а результаты, находящиеся ниже порога – отрицательными. Результаты оценки приведены на рисунках 2 и 3.

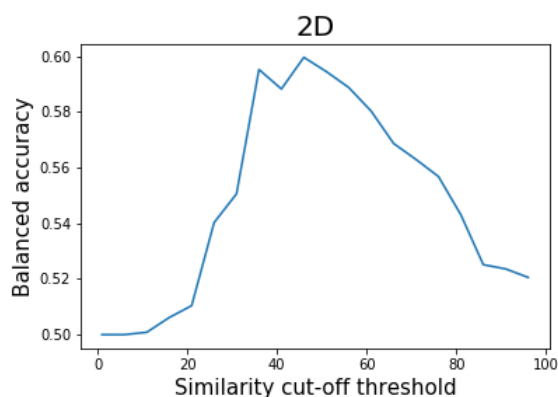


Рисунок 2. Изменение значения сбалансированной точности (Balanced accuracy), в зависимости от выбранного порога сходства (Similarity cut-off threshold), рассчитанное с использованием полученных оценок сходства по 2D методу.

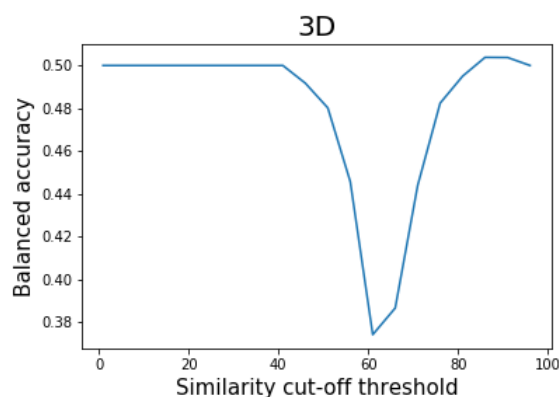


Рисунок 3. Изменение значения сбалансированной точности (Balanced accuracy), в зависимости от выбранного порога сходства (Similarity cut-off threshold), рассчитанное с использованием полученных оценок сходства по 3D методу.

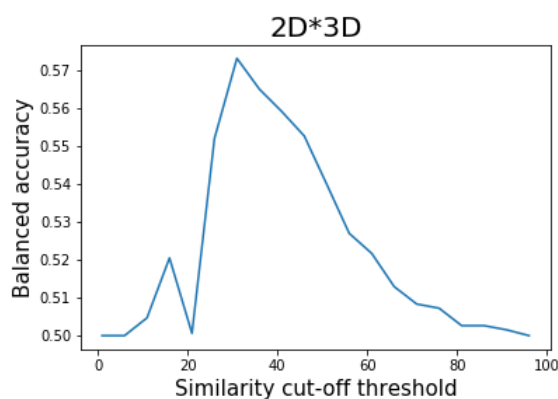


Рисунок 4. Изменение значения сбалансированной точности (Balanced accuracy), в зависимости от выбранного порога сходства (Similarity cut-off threshold), рассчитанное с использованием полученных значений сходства по 2Dx3D.

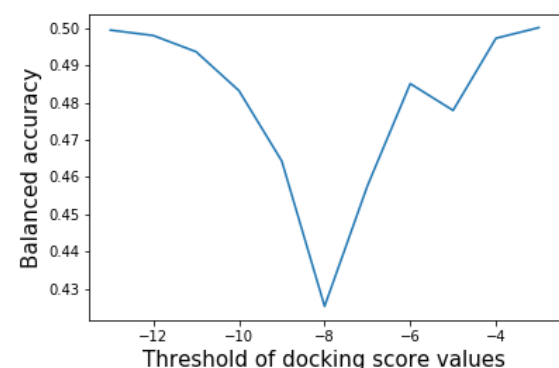


Рисунок 5. Изменение сбалансированной точности (Balanced accuracy) молекулярного докинга, в зависимости от изменения порога по значениям оценочной функции (Threshold of docking score values).

Как видно из приведенных на рисунках 2 и 3 данных, результаты расчёта сбалансированной точности для 2D и 3D методов существенно отличаются. Наибольшее значение сбалансированной точности (0.59) для 2D методов наблюдается при пороге отсечения по сходству равному 46%. Для 3D методов значения сбалансированной точности не превышает 0.51. Минимальное значение сбалансированной точности для 3D методов достигается при пороге отсечения по сходству 61 (0.37), что соответствует значениям чувствительности 0.53 и специфичности 0.22.

Параметром, который рекомендуется авторами работы для оценки сходства двух структур является Similarity (2Dx3D) [15]. По мнению авторов, эта операция, помогает сгладить влияние недостатков отдельных методов, тем самым повышая качество итоговых оценок. Результаты расчёта сбалансированной точности для данного параметра приведены на рисунке 4.

Как видно из представленных на рисунке 4 данных, максимальное значение сбалансированной точности (0.57) достигается при пороге отсечения по сходству, равному 31%, что ниже, чем таковая для 2D методов.

Как описано выше, при проведении оценки качества прогноза авторами работы [10] учитывались только положительные примеры результатов. Однако с

использованием отрицательных примеров нами показано, что применение интегральной оценки сходства (2Dx3D) на использованной нами тестовой выборке не является информативным, а с учётом приведенных выше замечаний о некорректности оценки сходства 3D методами, может вызвать затруднения при интерпретации полученных результатов у пользователей ресурса.

Оценка результатов молекулярного докинга, полученных с помощью D3Docking

Для демонстрации результатов работы функции VirtualScreening модуля D3Docking был использован SD файл, в который были включены все структуры тестовой выборки (табл. 1). Результаты докинга соединений тестовой выборки к папаин-подобной протеазе и главной протеазе SARS-CoV-2 представлены в таблицах 5 и 6 соответственно.

Как видно из представленных в таблице 5 данных, для соединения GRL0617 значение оценочной функции докинга к папаин-подобной протеазе (PLpro) равно -10.49, что является наилучшим результатом среди всех оценок. Для соединения 6-тиогауин значение оценочной функции составило -6.06. Такое значение является наихудшим среди всех расчетных значений оценочной функции докинга соединений тестовой выборки к PLpro.

Таблица 5. Результаты молекулярного докинга соединений тестовой выборки к папаин подобной протеазе (PLP) с применением функции VirtualScreening модуля D3Docking.

Название соединения	Название белка-мишени	Значение оценочной функции докинга	Установленный карман связывания лиганда с белком-мишенью
GRL0617	PLP/PLpro-Monomer	-10.49	Pocket1
C25H22FN4O4	PLP/PLpro-Monomer	-8.44	Pocket4
Remdesivir	PLP/PLpro-Monomer	-8.06	Pocket2
CHEMBL1388469	PLP/PLpro-Monomer	-8.05	Pocket1
CHEMBL1809259	PLP/PLpro-Monomer	-7.92	Pocket1
GC-376	PLP/PLpro-Monomer	-7.9	Pocket4
GS-441524	PLP/PLpro-Monomer	-7.89	Pocket4
Tideglusib	PLP/PLpro-Monomer	-7.87	Pocket4
CHEMBL1809250	PLP/PLpro-Monomer	-7.79	Pocket1
CHEMBL1809251	PLP/PLpro-Monomer	-7.33	Pocket1
Beta-D-N4-Hydroxycytidine	PLP/PLpro-Monomer	-7.3	Pocket4
6-thioguanine	PLP/PLpro-Monomer	-6.06	Pocket4

Таблица 6. Результаты молекулярного докинга соединений тестовой выборки к основной протеазе (3CLpro) с применением функции VirtualScreening модуля D3Docking

Название соединения	Название белка-мишени	Значение оценочной функции докинга	Установленный карман связывания лиганда с белком-мишенью
GRL0617	3CLpro/Mpro-Dimer	-8.21086	Pocket2
C25H22FN4O4	3CLpro/Mpro-Dimer	-9.90902	Pocket1
Remdesivir	3CLpro/Mpro-Dimer	-9.49245	Pocket2
CHEMBL1388469	3CLpro/Mpro-Dimer	-7.64298	Pocket2
CHEMBL1809259	3CLpro/Mpro-Dimer	-9.55744	Pocket1
GC-376	3CLpro/Mpro-Dimer	-8.14954	Pocket2
GS-441524	3CLpro/Mpro-Dimer	-8.16574	Pocket1
Tideglusib	3CLpro/Mpro-Dimer	-8.25794	Pocket1
CHEMBL1809250	3CLpro/Mpro-Dimer	-8.21429	Pocket2
CHEMBL1809251	3CLpro/Mpro-Dimer	-8.25598	Pocket2
Beta-D-N4-Hydroxycytidine	3CLpro/Mpro-Dimer	-7.38478	Pocket2
6-thioguanine	3CLpro/Mpro-Dimer	-5.33091	Pocket1

Ранее мы отмечали, что активными по отношению к основной протеазе являются соединения GC-376, Тидеглусиб, C25H22FN4O4 и CHEMBL1388469. Значения оценочной функции для данных соединений составляют -8.14954, -8.25794, -9.90902, -7.64298 соответственно. Среднее значение оценочной функции, рассчитанное для соединений, потенциально неактивных в отношении основной протеазы, равно -8.48 ± 0.8 . Таким образом, C25H22FN4O4 имеет значение оценочной функции, существенно превышающее среднее значение, а для остальных соединений значение оценочной функции ниже среднего. Только для двух соединений (ремдесивир и CHEMBL1809259) расчетные значения оценочной функции существенно превышают средние значения оценочной функции для неактивных соединений – -9.49 и -9.56 соответственно.

Результаты, полученные с использованием функции TargetPrediction модуля D3Docking, для препарата тидеглусиб представлены в таблице 7.

Порог значений оценочной функции докинга, опираясь на который, согласно прогнозу, можно было бы разделить соединения на активные/неактивные по отношению к белкам-мишеням, авторами не указан. В настоящей работе в качестве такого порога нами использованы значения оценочной функции, полученные при докинге лигандов к белкам-мишеням, активность в отношении которых подтверждена в условиях *in vitro*.

Расчет среднего значения оценочной функции для 45 белков-мишеней, в отношении которых активность тидеглусиба не подтверждена экспериментально, на основе представленных в таблице 5 данных даёт величину

Таблица 7. Результаты докинга соединения тидеглусиб к белкам-мишеням, представленным в базе данных D3Docking (функция TargetPrediction)

Target Full Name	Target Abbreviation	State	Score	Uniprot id / Protein Id	Pocket	Template PDB ID	Organism	Model Source	Rank
Dihydroorotate dehydrogenase	DHODH		-10.65	Q02127	Pocket2	3u2o	human	PDB	1
RAC-alpha serine/threonine-protein kinase	AKT1		-10.57	P31749	Pocket1	3o96	human	PDB	3
cGMP-specific 3',5'-cyclic phosphodiesterase	PDE5		-10.21	O76074	Pocket1	2h44	human	PDB	13
ORF1ab 5325-5925 Helicase	Helicase	Dimer	-9.62	QHD43415.1	Pocket4	6jyt	SARS-CoV-2	Swiss-model	2
Eukaryotic initiation factor 4A-I	eIF4A		-9.61	P60842	Pocket2	5zc9	human	PDB	4
Non-structural protein 2	Nsp2		-9.6	QHD43415.1	Pocket4	-	SARS-CoV-2	Robetta	5
Caveolin-2	CAV2		-9.57	P51636	Pocket1	-	human	Robetta	2
ORF1ab polyprotein 5325-5925 Helicase	Helicase	Monomer	-9.56	QHD43415.1	Pocket2	6jyt	SARS-CoV-2	Swiss-model	2
Spike protein	S protein	Closed	-9.51	QHD43416.1	Pocket2	6vxx	SARS-CoV-2	PDB	1
Nucleocapsid phosphoprotein	N protein	N terminal	-9.46	QHD43423.2	Pocket1	2ofz	SARS-CoV-2	Swiss-model	3
Complement Component 5	C5	Dimer	-9.38	P01031	Pocket8	3cu7	human	PDB	2
Complement Component 5	C5	Monomer	-9.33	P01031	Pocket1	3cu7	human	PDB	4
ORF1ab polyprotein 6453-6798 Uridylate-specific endoribonuclease	NendoU		-9.33	QHD43415.1	Pocket1	2h85	SARS-CoV-2	Swiss-model	4
Spike protein	S protein	Open	-9.31	QHD43416.1	Pocket4	5x5b	SARS-CoV-2	Swiss-model	2
Non-structural protein 6	Nsp6		-9.3	QHD43415.1	Pocket1	-	SARS-CoV-2	Robetta	1
ADP ribose phosphatase	ADRP	Dimer	-9.17	QHD43415.1	Pocket3	6w02	SARS-CoV-2	PDB	4
ADP ribose phosphatase	ADRP	Monomer	-9.14	QHD43415.1	Pocket1	6w02	SARS-CoV-2	PDB	5
ORF1ab polyprotein 5926-6452 Guanine-N7 methyltransferase	N7-MTase		-9.13	QHD43415.1	Pocket1	5nfy	SARS-CoV-2	Swiss-model	9
ORF1ab polyprotein 819-2763 Papain-like proteinase	PLP/PLpro	Dimer	-9.13	QHD43415.1	Pocket2	5y3e	SARS-CoV-2	Swiss-model	7
Mitogen-activated protein kinase 8(c-Jun N-terminal kinase 1	JNK1		-9.07	P45983	Pocket1	2g01	human	Swiss-model	2
Glycogen synthase kinase-3 beta	GSK3β		-8.84	P49841	Pocket1	1j1b	human	PDB	4
RAC-beta serine/threonine-protein kinase	AKT2		-8.8	P31751	Pocket1	3d0e	human	PDB	3
eterogeneous nuclear ribonucleoprotein A1	HNRNPA1		-8.66	P09651	Pocket1	1u1q	human	PDB	6
SARS-COV-2 chimeric receptor-binding domain+Angiotensin converting enzyme 2	SARS-CoV-2 chimeric RB-D+ACE2		-8.65	/	Pocket1	6vw1	human	PDB	2
Uridylate-specific endoribonuclease	NendoU	Dimer	-8.65	QHD43415.1	Pocket2	6w01	SARS-CoV-2	PDB	6

Таблица 7. Продолжение.

Uridylate-specific endoribonuclease	NendoU	Monomer	-8.64	QHD43415.1	Pocket1	6w01	SARS-CoV-2	PDB	4
Spike protein	S protein	Open	-8.61	QHD43416.1	Pocket2	6vyb	SARS-CoV-2	PDB	7
Nucleocapsid phosphoprotein	N protein	C terminal	-8.6	QHD43423.2	Pocket1	2gib	SARS-CoV-2	Swiss-model	3
Papain-like proteinase	PLP/PLpro	Trimer	-8.59	QHD43415.1	Pocket1	6w9c	SARS-CoV-2	PDB	6
Non-structural protein 4	Nsp4		-8.53	QHD43415.1	Pocket1	-	SARS-CoV-2	Robetta	9
ORF1ab polypeptide 4393-5324 RNA-dependent RNA polymerase	RdRp		-8.48	QHD43415.1	Pocket2	6nur	SARS-CoV-2	Swiss-model	3
ORF1ab polypeptide 4393-5324 RNA-dependent RNA polymerase	RdRp	+Mg	-8.46	QHD43415.1	Pocket2	6nur	SARS-CoV-2	Swiss-model	3
Envelope protein	E protein	Pentamer	-8.45	QHD43418.1	Pocket1	5x29	SARS-CoV-2	Swiss-model	5
Angiotensin converting enzyme 2	ACE2		-8.45	P59594	Pocket1	1r42	human	PDB	5
ORF1ab polypeptide 6799-7096 2'-O-methyltransferase	2'-O-Mtase		-8.42	QHD43415.1	Pocket1	2xyr	SARS-CoV-2	Swiss-model	3
ORF1ab polypeptide 3264-3569 3C-like proteinase	3CLpro/Mpro	Dimer	-8.31	QHD43415.1	Pocket1	2z9j	SARS-CoV-2	Swiss-model	7
Complement Component 3	C3		-8.31	P01024	Pocket2	2a73	human	PDB	3
Transmembrane protease serine 2	TMPRSS2		-8.29	O15393	Pocket1	-	human	Robetta	5
Spike protein S2 subunit	S protein-S2 subunit		-8.28	/	Pocket1	6lxt	SARS-CoV-2	PDB	3
Tyrosine-protein kinase ABL2	ABL2		-8.27	P42684	Pocket2	3hmi	human	PDB	5
3C-like proteinase	3CLpro/Mpro	Dimer	-8.26	QHD43415.1	Pocket1	6y2g	SARS-CoV-2	PDB	6
RNA-dependent RNA polymerase	RdRp	+RNA and Mg	-8.23	QHD43415.1	Pocket6	7bv2	SARS-CoV-2	PDB	27
Nucleocapsid phosphoprotein	N protein	N terminal +Tetramer	-8.17	QHD43423.2	Pocket1	6vyo	SARS-CoV-2	PDB	11
AP2-associated protein kinase 1	AAK1		-8.16	Q2M2I8	Pocket1	5te0	human	PDB	6
Furin	Furin		-8.11	P09958	Pocket1	5jxg	human	PDB	7
Disintegrin and metalloproteinase domain-containing protein 17	ADAM17/TACE	+Zn	-8.05	P78536	Pocket1	2i47	human		7
ORF1ab polypeptide 3264-3569 3C-like proteinase	3CLpro/Mpro	Monomer	-8.01	QHD43415.1	Pocket1	1z1i	SARS-CoV-2	Swiss-model	4
Disintegrin and metalloproteinase domain-containing protein 17	ADAM17/TACE		-8.01	P78536	Pocket1	2i47	human	Swiss-model	8
Spike protein	S protein	Closed	-8	QHD43416.1	Pocket6	5x58	SARS-CoV-2	Swiss-model	5
ORF1ab polypeptide 819-2763 Papain-like proteinase	PLP/PLpro	Monomer	-7.87	QHD43415.1	Pocket4	3e9s	SARS-CoV-2	Swiss-model	6
RNA-dependent RNA polymerase	RdRp		-7.8	QHD43415.1	Pocket3	6m71	SARS-CoV-2	PDB	4
3C-like proteinase	3CLpro/Mpro	MD	-7.73	QHD43415.1	Pocket2	1z1i	SARS-CoV-2	Swiss-model	4

Таблица 7. Продолжение.

Papain-like protease	PLP/P/Iprio	Monomer	-7.73	QHD43415.1	Pocket5	6w9c	SARS-CoV-2	PDB	3
3C-like proteinase	3CLpro/Mpro	MD	-7.7	QHD43415.1	Pocket2	1z1i	SARS-CoV-2	Swiss-model	11
Mitogen-activated protein kinase 10/c-Jun N-terminal kinase 3	JNK3		-7.66	P53779	Pocket1	1jnk	human	PDB	18
3C-like proteinase	3CLpro/Mpro	MD	-7.65	QHD43415.1	Pocket1	1z1i	SARS-CoV-2	Swiss-model	3
Calnexin	Calnexin		-7.64	P27824	Pocket4	1jhn	human	Swiss-model	3
3C-like proteinase	3CLpro/Mpro	Monomer	-7.56	QHD43415.1	Pocket3	5r82	SARS-CoV-2	PDB	2
Spike protein-heptad repeat 1	S-HR1		-7.49	QHD43416.1	Pocket3	5zvm	SARS-CoV-2	Swiss-model	6
Cathepsin L	CTSL		-7.46	P07711	Pocket2	2xu3	human	PDB	5
3C-like proteinase	3CLpro/Mpro	MD	-7.35	QHD43415.1	Pocket1	1z1i	SARS-CoV-2	Swiss-model	7
Mitogen-activated protein kinase 9/c-Jun N-terminal kinase 2	JNK2		-7.33	P45984	Pocket2	3npc	human	Swiss-model	4
Non-structural protein 9	Nsp9	Dimer	-7.32	QHD43415.1	Pocket2	6w4b	SARS-CoV-2	PDB	6
Calnexin	Calnexin	+Ca	-7.25	P27823	Pocket4	1jhn	human		6
ORF1ab polyprotein 3943-4140 Non-structural protein 8	Nsp8		-7.22	QHD43415.1	Pocket1	2ahm	SARS-CoV-2	Swiss-model	6
RAC-gamma serine/threonine-protein kinase	AKT3		-7.2	Q9Y243	Pocket1	2x18	human	PDB	7
Nucleocapsid phosphoprotein	N protein	N terminal +Monomer	-7.16	QHD43423.2	Pocket3	6vyo	SARS-CoV-2	PDB	3
ORF1ab polyprotein 3860-3942 Non-structural protein 7	Nsp7		-7.15	QHD43415.1	Pocket1	1ysy	SARS-CoV-2	Swiss-model	2
Cyclophilin A	CypA		-7.02	P62937	Pocket1	4n1m	human	PDB	7
ORF1ab polyprotein 4254-4392 Non-structural protein 10	Nsp10		-6.93	QHD43415.1	Pocket1	5nfy	SARS-CoV-2	Swiss-model	10
ORF1ab polyprotein 4141-4253 Non-structural protein 9	Nsp9	Dimer	-6.77	QHD43415.1	Pocket4	1qz8	SARS-CoV-2	Swiss-model	4
ORF1ab polyprotein 4141-4253 Non-structural protein 9	Nsp9	Monomer	-6.75	QHD43415.1	Pocket2	1uw7	SARS-CoV-2	Swiss-model	14
Non-structural protein 9	Nsp9	Monomer	-6.4	QHD43415.1	Pocket2	6w4b	SARS-CoV-2	PDB	10
ORF1ab polyprotein 1-180 Host translation inhibitor nsp1	Nsp1		-6.25	QHD43415.1	Pocket2	2hsx	SARS-CoV-2	Swiss-model	7
C type lectin domain family 4 member M	CLEC4M/ CD209L/L- SIGN		-6.2	Q9H2X3	Pocket1	1xar	human	PDB	10
Envelope protein	E protein	Monomer	-5.91	QHD43418.1	Pocket1	2mm4	SARS-CoV-2	Swiss-model	9
ORF7a protein	ORF7a		-5.67	QHD43421.1	Pocket1	1yo4	SARS-CoV-2	Swiss-model	12

Примечание. «Target Full Name» - полное название белка-мишени; «Score» - расчетные значения оценочной функции; «Target Abbreviation» - сокращенные наименования мишеней; «State» - форма белка-мишени; «Uniprot id / Protein Id» - идентификаторы UniProt ID или Protein ID; «Pocket» - установленный карман связывания; «Template PDB ID» - код мишени в PDB; «Organism» - организм; «Model Source» - ресурс с помощью которого была получена трехмерная структура; «Rank» - ранг мишени. Названия белков-мишеней, активность тителгусиба в отношении которых подтверждена экспериментально, выделены жирным шрифтом.

-8.24±1.00. Вышеуказанные расчетные значения оценочной функции для двух белков-мишеней (GSK3-beta и 3CLpro), в отношении которых для препарата тидеглусиб была установлена активность, близки к данному порогу, либо ниже его.

Результаты расчёта оценочной функции докинга препарата тидеглусиб с бета киназой гликоген-синтазы 3 (GSK3-beta) равны -8.84, а с главной протеазой (3CLpro) варьируют в диапазоне от -8.31 до -7.35 в зависимости от выбранной трехмерной структуры белка-мишени. Наилучшие значения оценочной функции получены для дигидрооратат дегидрогеназы (DHODH) – -10.65, взаимодействие с которой препарата тидеглусиб экспериментально не установлено.

Для 17 белков-мишеней значения оценочной функции докинга препарата тидеглусиб не уступают таковым, полученным для главной протеазы (-8.84).

Результаты расчёта оценочной функции докинга тидеглусиба к белкам-мишеням DHODH, AKT1 и PDE5 равны -10.65, -10.57 и -10.21 соответственно. Полученные значения оценочной функции являются наилучшими даже среди тех белков-мишеней, активность тидеглусиба в отношении которых подтверждена в условиях тестирования *in vitro*. Вероятно, такого рода информация может быть использована для дальнейшего экспериментального исследования действия данного препарата на эти белки-мишени.

Оценка сбалансированной точности используемых авторами подходов рассчитывалась на основе значений оценочной функции докинга, полученных с использованием функции TargetPrediction для каждой структуры тестовой выборки. Результаты приведены в дополнительных материалах к статье.

Изменения сбалансированной точности в зависимости от выбранного порога отсеивания по значению оценочной функции приведены на рисунке 5.

Как видно из приведенных на рисунке 5 данных, значения сбалансированной точности не превышают 0.5. Минимальное значение сбалансированной точности достигается при пороге значений оценочной функции, равном -8 (0.42), что соответствует значениям чувствительности 0.24 и специфичности 0.60.

ЗАКЛЮЧЕНИЕ

В связи с остротой проблемы пандемии коронавирусной инфекции SARS-CoV-2/COVID-19 и отсутствием достаточно эффективных и безопасных терапевтических средств, поиск новых фармакологических веществ с антикоронавирусной активностью крайне актуален [1]. Первым свободно доступным в сети Интернет веб-ресурсом, который позволяет проводить оценку сходства и молекулярный докинг для выявления соединений, потенциально пригодных для терапии COVID-19, стала платформа D3Targets-2019-nCoV.

В данной работе мы оценили качество прогноза с помощью данной платформы для 94 веществ тестовой выборки, содержащей 35 активных соединений и 59 соединений, которые включают в себя соединения с установленным отсутствием активности в отношении выбранных мишеней и соединения, активные по отношению к мишеням, не представленным в БД CoViLigands. Нами была проведена оценка достоверности прогноза с использованием параметра, рекомендованного авторами, а

именно произведением оценок двумерного и трехмерного сходства (2Dx3D). В ходе оценки качества прогноза показано, что максимум сбалансированной точности (0.57) достигается при пороге отсеивания по сходству, равному 31%. При расчёте сбалансированной точности молекулярного докинга показано, что её значения не превышают 0.5. Помимо этого, было установлено, что максимальное значение сбалансированной точности 2D метода оценки сходства (0.59) наблюдаются при пороге отсеивания по сходству, равном 46%. Для метода оценки 3D сходства максимальное значение сбалансированной точности не превышает 0.51. При анализе результатов, полученных при работе с модулем D3Similarity установлено, что для идентичных структур оценка 2D сходства составляет 100%, однако, при этом оценка 3D сходства ниже. Принимая во внимание тот факт, что загруженные нами для анализа и представленные в БД CoViLigands структуры соединений, проходят одинаковую предобработку, возникает вопрос о корректности выбранных авторами веб-сервиса методов для анализа сходства трехмерных структур.

Исходя из значений сбалансированной точности для всего спектра прогнозируемых мишеней, полученных на сформированной нами тестовой выборке, и установленной несогласованности результатов расчёта структурного сходства можно сделать вывод об ограниченной полезности прогноза веб-сервиса D3Target-2019-nCoV. Таким образом, данный ресурс требует серьезной доработки, включающей в себя в том числе и усовершенствование используемых алгоритмов, в частности методов оценки 3D сходства.

ФИНАНСИРОВАНИЕ РАБОТЫ

Работа выполнена при поддержке проекта РФФИ № 20-04-60285.

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

К данной статье приложены дополнительные материалы, свободно доступные на сайте журнала (<http://dx.doi.org/10.18097/BMCRM00140>).

ЛИТЕРАТУРА

1. *Sohrabi, C., Alsafi, Z., O'Neill, N., Khan, M., Kerwan, A., Al-Jabir, A., Iosifidis, C., & Agha, R.* (2020) World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). *International Journal of Surgery*. DOI: 10.1016/j.ijssu.2020.02.034
2. Coronavirus disease 2019 (COVID-19). Situation report – 134. Retrieved October 25, 2020, from <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/>
3. Coronavirus – simptoms, priznaki, obshchaya informaciya, otvety na voprosy — Minzdrav Rossii. Retrieved October 25, 2020, from <https://covid19.rosminzdrav.ru/>
4. Covid19db | ReDO Project. Retrieved June 9, 2020, from <http://www.redo-project.org/covid19db/>

5. Poroikov, V., Druzhilovskiy, D. Drug Repositioning: New Opportunities for Older Drugs. In: *In Silico Drug Design*, 1st Edition. Repurposing Techniques and Methodologies. Chapter 1. Editors: Kunal Roy. Elsevier, Academic Press, 2019, p.3-17. <https://www.elsevier.com/books/in-silico-drug-design/roy/978-0-12-816125-8>
6. Liu, C., Zhou, Q., Li, Y., Garner, L. V., Watkins, S. P., Carter, L. J., ... & Al-baiu, D. (2020). Research and development on therapeutic agents and vaccines for COVID-19 and related human coronavirus diseases. DOI: 10.1021/acscentsci.0c00272
7. Vremennye metodicheskie rekomendacii. Retrieved June 3, 2020, from https://static-0.rosminzdrav.ru/system/attachments/attaches/000/050/584/original/03062020_%D0%9CR_COVID-19_v7.pdf
8. Beigel, J. H., Tomashek, K. M., Dodd, L. E., Mehta, A. K., Zingman, B. S., Kalil, A. C., ... & de Castilla, D. L. (2020). Remdesivir for the treatment of Covid-19. *New England Journal of Medicine*. DOI: 10.1056/NEJMoa2007764
9. Pan, H., Peto, R., Karim, Q. A., Alejandria, M., Restrepo, A. M. H., Garcia, C. H., ... & Reddy, S. (2020). Repurposed antiviral drugs for COVID-19; interim WHO SOLIDARITY trial results. medRxiv. DOI: 10.1101/2020.10.15.20209817
10. Shi, Y., Zhang, X., Mu, K., Peng, C., Zhu, Z., Wang, X., Yang, Y., Xu, Z., & others. (2020). D3Targets-2019-nCoV: A Web Server to Identify Potential Targets for Antivirals Against 2019-nCoV. DOI:10.1016/j.apbs.2020.04.006
11. Open Babel. Retrieved June 12, 2020, from http://openbabel.org/wiki/Main_Page
12. Getting Started with the RDKit in Python — The RDKit 2020.03.1 documentation. Retrieved June 12, 2020, from <https://www.rdkit.org/docs/Getting-StartedInPython.html#mmff1>
13. Halgren, T. A. (1996). Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *Journal of Computational Chemistry*, 17(5–6), 490–519. DOI:10.1002/(SICI)1096-987X(199604)17:5/6<490::AID-JCC1>3.0.CO;2-P
14. D3Targets-2019-nCoV (CoViLigands database). Retrieved June 26, 2020, from <https://www.d3pharma.com/D3Targets-2019-nCoV/CoViLigands/2019-nCoV.php>
15. Zhu, Z., Wang, X., Yang, Y., Zhang, X., Mu, K., Shi, Y., Peng, C., Xu, Z., & others. (2020). D3Similarity: A ligand-based approach for predicting drug targets and for virtual screening of active compounds against COVID-19. DOI: 10.26434/chemrxiv.11959323.v1
16. Trott, O., & Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry*, 31(2), 455-461. DOI: 10.1002/jcc.21334
17. Wang, J., Peng, C., Yu, Y., Chen, Z., Xu, Z., Cai, T., ... & Zhu, W. (2020). Exploring conformational change of adenylate kinase by replica exchange molecular dynamic simulation. *Biophysical Journal*, 118(5), 1009-1018. DOI: 10.1016/j.bpj.2020.01.001
18. Gaulton, A., Hersey, A., Nowotka Michalando Bento, A. P., Chambers, J., Mendez, D., Mutowo, P., Atkinson, F., Bellis, L. J., Cibrián-Uhalte, E., & others. (2016). The ChEMBL database in 2017. *Nucleic Acids Research*, 45(D1), D945–D954. DOI: 10.1093/nar/gkw1074
19. Stanford Coronavirus Antiviral Research Database. Retrieved July 03, 2020, from <https://covdb.stanford.edu/search/?compound=EIDD-2801>
20. Wermuth, C.G. Similarity in drugs: reflections on analogue design. *Drug Discov. Today*, 2006, 11(7-8), 348-354. DOI: 10.1016/j.drudis.2006.02.006
21. Filimonov, D. A., Druzhilovskiy, D. S., Lagunin, A. A., Glorizova, T. A., Rudik, A. V., Dmitriev, A. V., ... & Poroikov, V. V. (2018). Computer-aided prediction of biological activity spectra for chemical compounds: opportunities and limitations. *Biomedical Chemistry: Research and Methods*, 1(1), e00004. DOI: 10.18097/bmcrm00004
22. Lagunin, A., Zakharov, A., Filimonov, D., & Poroikov, V. (2011). QSAR modelling of rat acute toxicity on the basis of PASS prediction. *Molecular informatics*, 30(23), 241-250. DOI: 10.1002/minf.201000151
23. Druzhilovskiy, D.S., Stolbov, L.A., Savosina, P.I., Pogodin, P.V., Filimonov, D.A., Veselovsky, A.V., Tarasova, N.I., Nicklaus, M.C., Poroikov, V.V. (2020). Computational approaches to identify a hidden pharmacological potential in large chemical libraries. *Supercomputing Frontiers and Innovations*, 7(3), 57-76. DOI: 10.14529/jsfi200306
24. Xie, J., Liang, R., Wang, Y., Huang, J., Cao, X., Niu, B. (2020). Progress in Target Drug Molecules for Alzheimer's Disease. *Current Topics in Medicinal Chemistry*, 20(1), 4–36. DOI: 10.2174/1568026619666191203113745
25. Jin, Z., Du, X., Xu, Y., Deng, Y., Liu, M., Zhao, Y., Zhang, B., Li, X., Zhang, L., Duan, Y., & others. (2020). Structure of Mpro from COVID-19 virus and discovery of its inhibitors. *bioRxiv*. DOI:10.1101/2020.02.26.

Поступила: 30.10.2020
 После доработки: 27.11.2020
 Принята к публикации: 10.12.2020

ASSESSING THE PREDICTION QUALITY OF THE ANTI-SARS-CoV-2 ACTIVITY USING THE D3Targets-2019-nCoV WEB SERVICE

N.S. Ionov*, P.V. Pogodin, V.V. Poroikov

Institute of Biomedical Chemistry,
 10 Pogodinskaya Street, Moscow, 119121 Russia; * e-mail: ionov.nikita.serg@gmail.com

The D3Targets-2019-nCoV web service predicting the interaction of chemical compounds with SARS-CoV-2 virus proteins and human proteins involved in the pathogenesis of COVID-19 by structural similarity and molecular docking was evaluated. The quality of the prediction was assessed as a balanced accuracy, which was calculated based on the results of the prediction for the structures of chemical compounds from the test set we compiled. The test set consisted of 35 active and 59 inactive molecules, including compounds with the experimentally confirmed absence of activity against the selected targets and compounds active against SARS-CoV-2 targets, not presented in the CoViLigands database. The authors of the analyzed web service did not indicate the thresholds for the values of the similarity score and the docking scoring function, using which it would be possible to reliably divide the compounds into active and inactive with respect to target proteins. Therefore, we assessed the balanced accuracy of the predictive methods D3Targets-2019-nCoV at various thresholds for cutting off active substances from inactive ones. Using our test set it was found that the highest value of balanced accuracy (0.59) was achieved when choosing active molecules based on the results of 2D similarity assessment (cutoff threshold was 46%). Assessment of 3D similarity did not allow achieving balanced accuracy values exceeding 0.5. It is shown that using the 2Dx3D integral similarity assessment recommended by the authors, the maximum value of the balanced accuracy 0.57 was achieved at a threshold of 31%. The calculated balanced accuracy for molecular docking results does not exceed 0.51. On the case study for the tideglusib, it was shown that the values of the scoring function for two target proteins, the activity against which was confirmed in the experiment (3CLpro and GSK3B), do not differ significantly from the values of the scoring function for the remaining 44 targets were not confirmed.

Key words: SARS-CoV-2; estimation of molecular similarity; molecular docking; finding of new anticoronaviral agents; estimating of the web-service quality

FUNDING

The work was supported by the Russian Foundation for Basic Research (project 20-04-60285).

Received: 30.10.2020, revised: 27.11.2020, accepted: 10.12.2020