Биомедицинская химия, 2023 том 69, вып. 5, с. 322–327.

КРАТКОЕ СООБЩЕНИЕ

ПРЕДСКАЗАНИЕ ИНГИБИРОВАНИЯ ГЛАВНОЙ ПРОТЕАЗЫ SARS-CoV-2 НА МОДЕЛЯХ КОМПЛЕКСОВ ИНГИБИТОР-ФЕРМЕНТ

Я.О. Иванова*, В.С. Скворцов

Научно-исследовательский институт биомедицинской химии имени В.Н. Ореховича,
119121, Москва, ул. Погодинская, 10; *эл. почта: yana.emris@gmail.com

Проанализирован набор уравнений линейной регрессии, предсказывающих величину IC_{50} для 180 конкурентных ингибиторов главной протеазы SARS-CoV-2. Проведена симуляция молекулярной динамики комплексов фермент-ингибитор, либо имеющих известную кристаллическую структуру, либо промоделированных методом молекулярного докинга с наложенным ограничением на отбор конечных поз по сходству со структурными аналогами. В качестве независимых переменных использовали величины энергетических вкладов, полученных при расчёте двумя вариантами метода MMPBSA (MMGBSA), изменения свободной энергии комплекса, и ряд физико-химических характеристик ингибиторов. При обучении для подвыборок, полученных из различных источников, использовали индикаторные переменные, чтобы нивелировать имеющиеся систематические отклонения целевой величины. Качество предсказания оценивали по процедуре скользящего контроля методом выбрасывания по одному и по 20% выборки. Средняя ошибка при предсказании величины $lg(IC_{50})$ составила 0,45 логарифмической единицы при общей ширине диапазона значений 3,71. Рассмотрена зависимость устойчивости предсказания от вариативности комплекса в процедуре молекулярной динамики.

Ключевые слова: SARS-CoV-2; главная протеаза; конкурентные ингибиторы; QSAR

DOI: 10.18097/PBMC20236905322

ВВЕДЕНИЕ

Данная работа является продолжением работы, опубликованной нами ранее [1], в которой была показана принципиальная важность контроля положения лиганда в месте связывания при процедуре докирования. Ранее был использован неконтролируемый докинг, когда отбор “лучшего” положения проводили исключительно на основе оценочной функции, а контроль положения лиганда осуществляли постфактум. В настоящее время в различных процедурах докирования заложена возможность контроля положения лиганда по аналогии с уже известной. Это позволяет избежать пропуска данных, если решение для группы близких по структуре молекул не совпадало с данными, полученными из анализа кристаллографических структур или общей тенденцией для большинства членов группы. В предыдущей работе таких решений было 70 из 146 [1]. В настоящее исследование был также добавлен дополнительный набор данных [2] по конкурентным ингибиторам главной протеазы SARS-CoV-2 (M^{pro} SARS-CoV-2) с известными значениями IC_{50} (концентрация, вызывающая 50% ингибирование), а также внесён ряд дополнений и модификаций в процедуру подготовки и расчёта данных с использованием пакета Amber 19 [3]. Как и ранее, главной задачей работы был анализ простых линейных уравнений, основанных на использовании данных об известных или смоделированных комплексах ингибиторов с главной протеазой SARS-CoV-2.

МЕТОДИКА

В работе была использована выборка данных для 146 соединений — конкурентных ингибиторов M^{pro} SARS-CoV-2 (набор данных S1), отобранная в предыдущей работе [1]. Для 34 соединений известна структура комплекса с ферментом, доступная в Protein Data Bank [4]. К данной выборке было добавлено 35 соединений из работы [2], причём, для одного из них (L014, идентификатор соединения в Дополнительных материалах), ранее включённого в выборку S1, есть данные о кристаллической структуре комплекса фермент-ингибитор (PDB ID: 7LMF). Данные о IC_{50} всех добавленных соединений были получены с использованием того же субстрата HiyteFluor-488ESATLQSGLRKAK-(QXL)-NH_2 (набор S2). При этом диапазон значений величины $lg(IC_{50})$ для суммарной выборки составил 3,71 логарифмических единиц (л.е.; от 1,26 до 4,97; рис. 1). Для набора S2 диапазон значений более узкий (2,69 л.е.). Полный набор данные о структурах, распределение по структурным группам, величинах IC_{50}, литературных ссылки и др. представлен в Дополнительных материалах к этой статье.

Начальное моделирование комплексов фермент-ингибитор выполняли, как и в работе [1], с использованием программы Schrodinger [5]. При этом (в отличие от [1]) для каждой из структурных групп, полученных ранее, в качестве прототипа была определена молекула ингибитора с известной кристаллической структурой комплекса.

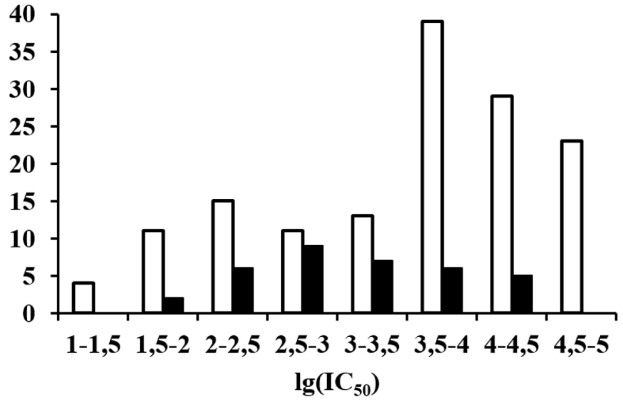

Рисунок 1. Распределение величины $lg(IC_{50})$ для выборки соединений с известной ингибиторной активностью по отношению к M^{pro} SARS-CoV-2. Белые столбцы – набор S1, чёрные – S2.

Было наложено ограничение на отбор только структур, имеющих сходное расположение в процессе докирования. Соответственно, в качестве структуры белка (участок связывания) для докирования соединений из каждой группы использовали 3D структуру комплекса фермент-ингибитор одного из ингибиторов группы, взятого из PDB. Если для данной группы соединений кристаллические структуры комплексов фермент-ингибитор неизвестны, то в качестве приоритетной структуры комплекса использовали совпадающие варианты расположения ингибиторов в участке связывания, максимально представленные в группе, и как участок связывания использовали вариант структуры M^{pro} из комплекса с соединением L039 (PDB ID: 7N44). Все использованные структуры M^{pro} были предварительно выровнены в пространстве между собой. Общую структуру комплексов оптимизировали с использованием поля сил OPLS3e [6]. Данную процедуру проводили для всех комплексов, включая структуры, полученные из PDB. Как и в работе [1], для каждого комплекса рассчитывали изменения свободной энергии с использованием метода MMGBSA (модель сольватации VSGB). Набор из 7 покомпонентных значений, включающих изменение энергии кулоновского взаимодействия, энергии ковалентных взаимодействий в лиганде и рецепторе, энергии ван-дер-ваальсовых взаимодействий, неполярного вклада в энергию сольватации по площади поверхности, электростатического вклада в энергию сольватации на основе обобщённой модели Борна, вклада водородного связывания и вклада, связанного с липофильностью, использовали в качестве независимых переменных (группа E1) для создания уравнений линейной регрессии. В отличие от работы [1], в качестве следующего этапа оптимизации структуры комплекса использовали дополнительно набор последовательных процедур симуляции молекулярной динамики (по схеме, описанной нами ранее [2]) с использованием пакета программ Amber19 [3]. Последний этап симуляции молекулярной динамики служил основой для расчёта энергетических вкладов в изменение свободной энергии комплексов методом MMPBSA [8]. Набор этих параметров (E2) также использовали в качестве независимых переменных. Данный набор включал: (i) изменение величин электростатического взаимодействия и ван-дер-ваальсовых взаимодействий; (ii) гидрофобный и сольватационный вклады, рассчитанные методом Poisson-Boltzmann; (iii) аналогичные вклады, рассчитанные методом Generalized Born; (iv) трансляционный, ротационный и колебательный энтропийные вклады [8].

В качестве целевой при создании предсказательных уравнений использовали величину $lg(IC_{50})$, при этом саму величину IC_{50} выражали в нМ. Кроме энергетических параметров, характеризующих свойства комплекса, для каждого из лигандов средствами программы Sybyl-X [9] был рассчитан набор из 6 параметров (P), характеризующих свойства самого лиганда (молекулярный вес, общий и полярный объём, площадь общей и полярной поверхности, число связей, по которым возможно вращение). Эти параметры также использовали как независимые переменные. В процессе подбора уравнений линейной регрессии варьировали как набор независимых переменных, так и состав выборки наблюдений. Качество моделей, как и в работе [1], оценивали по результатам процедуры скользящего контроля методом выкидывания по одному (“leave-one-out cross-validation”); кроме того, дополнительно использовали тест на выкидывание 20% значений. Для этого данные сортировали по значению IC_{50}, а затем отбирали каждый n-ый элемент, сформировав 5 выборок, имеющих сходные диапазоны значений IC_{50}.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Вариации положения ингибиторов в месте связывания M^{pro} достаточно велики даже при выравнивании между собой структур комплексов из PDB (рис. 2А). Однако выбранная процедура докирования позволяет получить сходное расположение ингибиторов в пределах своей структурной группы (рис. 2Б,В).

Статистические характеристики всех моделей (уравнений) предсказания величины $lg(IC_{50})$, обсуждаемые далее, представлены в таблице. Во всех случаях возможность выбросов наблюдений не рассматривали, несмотря на то, что для части структур, взятых из PDB, наблюдали существенные отклонения при обучении (до 2,22 л.е.). В качестве первого варианта варьировали все 3 группы переменных (P, E1 и E2) по выборке (S1, 145 наблюдений), использованной в работе [1]. Данные для соединения L014, входящего также в группу добавленных ингибиторов (выборка S2, 35 наблюдений), из группы S1 удалили. Ни одной из трёх групп переменных по отдельности недостаточно, чтобы построить уравнения с приемлемыми характеристиками (см. уравнения № 1, 2 и 3 таблицы). Тем не менее, процедура случайного смешивания целевого значения демонстрирует, что и в этом варианте уравнения имеют слабую предсказательную способность. Её нельзя считать значимой,

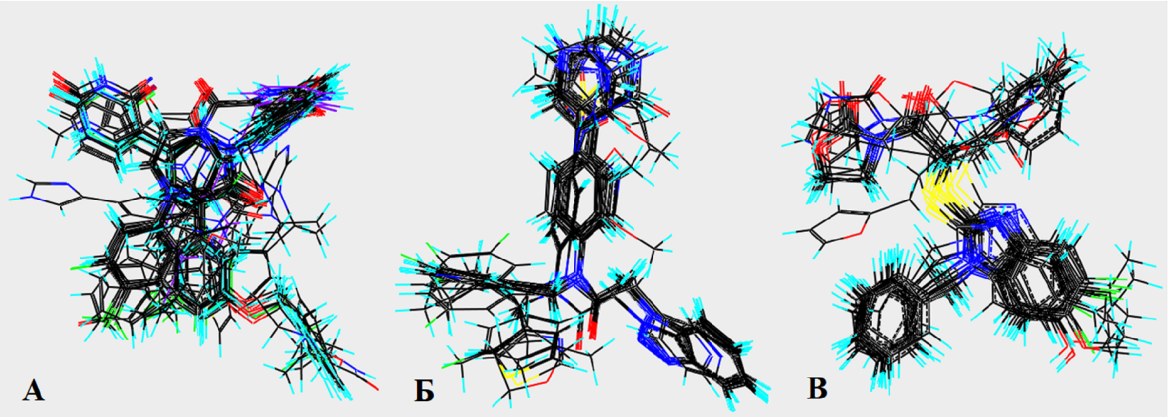

Рисунок 2. Расположение ингибиторов, для которых известна структура комплексов, полученное выравниванием структур из PDB (**А**); выравнивание для наборов ингибиторов из работ [2] (**Б**) и [12] (**В**), полученное в результате докирования к структуре протеазы M^{pro} SARS-CoV-2. Ориентация выравниваний **А**, **Б** и **В** не совпадает.

так как критерием значимости обычно считают величину $Q^2 > 0{,}6$ в процедуре скользящего контроля методом выбрасывания по одному. Значимую предсказательную способность имеют уравнения, построенные для объединённых наборов переменных (№ 4, 5 и 6). В то же время, если использовать набор S2 в качестве тестовой выборки, то результаты не удовлетворяют критериям значимости. Однако, набор S2 имеет более узкий диапазон целевых значений, и уравнения линейной регрессии для того же набора независимых переменных (уравнения № 7, 8 и 9 таблицы) имеют слабые характеристики, даже если не обращать внимание на тот факт, что при таком числе переменных и эти параметры спорные (см. характеристики при случайном смешивании целевой функции). Кроме того, предсказанные значения $\lg(IC_{50})$ для набора S2 имеют выраженную систематическую погрешность (рис. 3А). Величина IC_{50}, полученная в разных лабораториях на одном и том же наборе соединений, может отличаться даже при использовании одной и той же экспериментальной методики. Наилучший вариант, если различные наборы данных имеют некоторое количество одинаковых соединений в выборках, и тогда данные можно выровнять [10]. В данном случае для нашей выборки, собранной по литературным источникам, таких пересечений нет. Выходом может быть использование при обучении индикаторных переменных, отмечающих группы данных, полученные из одного источника. По сути, это смещает целевое значение на некоторую фиксированную величину. Из общих соображений, чем меньше таких индикаторных переменных, тем меньше вероятность переобучения. Мы проанализировали выборку S1, и выявили 2 набора данных, собранных из работ [11, 12], для которых актуально введение индикаторной переменной (пример на рис. 3Б). Эти 2 индикаторные переменные обозначены в таблице как набор параметров I1 (I2 — индикаторная переменная для набора S2). При использовании I1 в комбинации с наборами данных P, E1 и/или E2 характеристики уравнений улучшаются (№ 10, 11 и 12). Однако, из-за того, что диапазон целевых значений

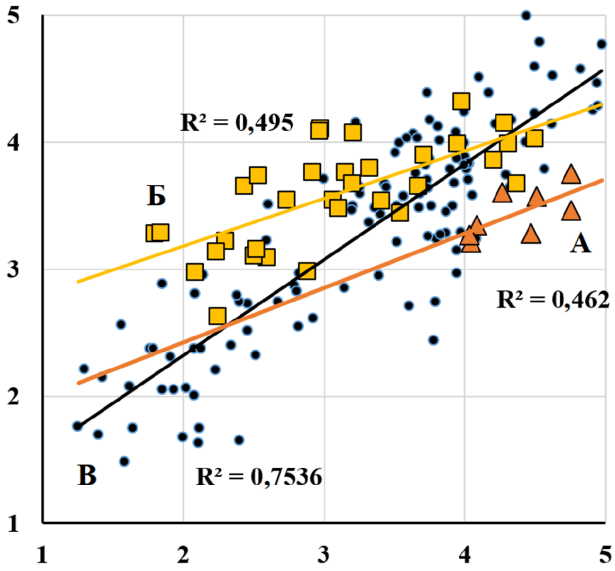

Рисунок 3. Сравнение экспериментально определённых и предсказанных величин $\lg(IC_{50})$ для наборов данных из работ [11] (**А**, треугольник), [2] (**Б**, квадрат) и обучающей выборки (**В**). Предсказание выполнено по модифицированному уравнению 6 (таблица), пересчитанному после удаления из выборки наблюдений из работ [11] и [12]. По оси абсцисс экспериментально определённые величины $\lg(IC_{50})$. По оси ординат предсказанные.

в тестовой выборке S2 уже чем в обучающей S1, и в этом случае разброс предсказанных значений достаточно велик. Более того, так как объединённая выборка S1+S2 даже с использованием индикаторной переменной I2 даёт немного худший результат (№ 13, 14 и 15), то причина, вероятнее всего, в “зашумлённости” выборки S2. Тем не менее, общие параметры уравнений удовлетворяют критерию значимости предсказательной силы. В качестве дополнительного теста качества приведены данные для процедуры перекрёстного контроля с выкидыванием по 20% выборки (№ 14.1–14.5). В среднем R^2 при тестировании составил 0,65, а средняя ошибка 0,45 л.е. (общий диапазон значений 3,71 л.е.).

Таблица. Параметры уравнений предсказания величины $\lg(IC_{50})$, полученные при обучении, и результаты тестирования

№	Число наблюдений (обучение)	Набор переменных	Число переменных + постоянная, лучшая модель (всего)	R^2_L	ME_L	$MaxE_L$	R^2_{rand}	ME_{rand}	Q^2_{loo}	ME_{loo}	$MaxE_{loo}$	Число наблюдений (тестирование)	R^2_{test}	ME_{test}	$MaxE_{test}$
1	145	P	6(7)	0,49	0,57	2,22	0,05	0,78	0,44	0,60	2,30	35	0,01	0,73	1,86
2	145	E1	4(7)	0,51	0,56	1,94	0,02	0,78	0,48	0,57	2,04	35	0,24	1,42	3,73
3	145	E2	8(10)	0,49	0,57	1,59	0,06	0,77	0,43	0,61	1,85	35	0,07	0,62	1,67
4	145	P+E1	7(13)	0,71	0,41	1,45	0,05	0,77	0,67	0,43	1,57	35	0,32	1,80	3,70
5	145	P+E2	8(16)	0,73	0,39	1,46	0,07	0,78	0,71	0,41	1,50	35	0,37	0,76	1,99
6	145	P+E1+E2	11(22)	0,77	0,37	1,39	0,08	0,77	0,73	0,40	1,48	35	0,41	1,20	2,49
7	35	P+E1	7(13)	0,62	0,36	0,87	0,18	0,53	0,46	0,45	1,07	—	—	—	—
8	35	P+E2	5(16)	0,61	0,36	1,16	0,19	0,53	0,51	0,42	1,26	—	—	—	—
9	35	P+E1+E2	11(22)	0,75	0,28	0,93	0,42	0,45	0,56	0,40	1,09	—	—	—	—
10	145	P+E1+I1	6(15)	0,76	0,38	1,47	0,04	0,79	0,74	0,39	1,49	35	0,27	0,75	2,17
11	145	P+E2+I1	8(18)	0,76	0,38	1,23	0,09	0,76	0,74	0,40	1,25	35	0,19	0,55	1,58
12	145	P+E1+E2+I1	9(24)	0,80	0,36	1,18	0,05	0,78	0,77	0,38	1,25	35	0,35	0,59	1,73
13	180	P+E1+I1+I2	7(16)	0,67	0,44	1,70	0,06	0,76	0,64	0,46	1,75	—	—	—	—
14	180	P+E2+I1+I2	7(19)	0,71	0,41	1,37	0,07	0,77	0,69	0,42	1,40	—	—	—	—
15	180	P+E1+E2+I1+I2	8(25)	0,73	0,41	1,22	0,80	0,76	0,70	0,43	1,33	—	—	—	—
14.1	144	P+E2+I1+I2	11(19)	0,73	0,39	1,48	0,07	0,76	0,69	0,42	1,56	36	0,70	0,42	1,22
14.2	144	P+E2+I1+I2	8(19)	0,73	0,40	1,12	0,05	0,77	0,70	0,42	1,16	36	0,64	0,45	1,47
14.3	144	P+E2+I1+I2	8(19)	0,72	0,39	1,45	0,05	0,78	0,69	0,41	1,49	36	0,69	0,46	1,23
14.4	144	P+E2+I1+I2	8(19)	0,71	0,41	1,31	0,04	0,78	0,68	0,43	1,37	36	0,68	0,46	1,28
14.5	144	P+E2+I1+I2	9(19)	0,73	0,40	1,18	0,09	0,76	0,70	0,42	1,22	36	0,57	0,46	1,56

Примечание: R^2_L – R^2 обучения; ME_L – средняя ошибка обучения; $MaxE_L$ – максимальная ошибка обучения; R^2_{rand} – усреднённый R^2 в процедуре обучения по выборке со смешанным целевым значением; ME_{rand} – усреднённая средняя ошибка обучения по выборке со смешанным целевым значением; Q^2_{loo} – Q^2 модели в процедуре скользящего контроля; ME_{loo} – средняя ошибка для метода скользящего контроля; $MaxE_{loo}$ – максимальная ошибка для метода скользящего контроля; R^2_{test} – R^2 тестирования; ME_{test} – средняя ошибка тестирования; $MaxE_{test}$ – максимальная ошибка тестирования.

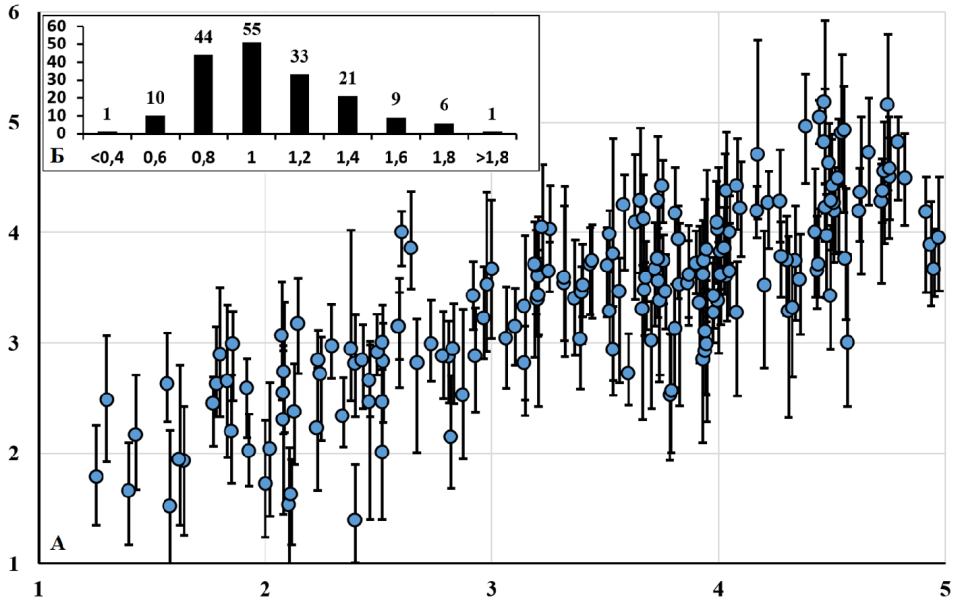

Рисунок 4. А. Сравнение экспериментально определённых и предсказанных величин $lg(IC_{50})$ полного набора данных с учётом вариантов комплексов, полученных в ходе симуляции молекулярной динамики (15 для каждого комплекса ингибитор/фермент). Предсказание выполнено по группе моделей 14.1-14.5, для каждого ингибитора использовали уравнение, в обучении которого данный ингибитор не использовался. Для каждого ингибитора представлено среднее значение и диапазон от минимального до максимального. По оси абсцисс экспериментально определённые величины $lg(IC_{50})$. По оси ординат предсказанные. **Б.** Распределение диапазона предсказаний для каждого из ингибиторов.

Набор независимых переменных, использованный для уравнения 14, включает в себя переменные из групп P и E2. При отборе лучших из общего числа переменных в уравнениях всегда присутствует набор I1, площадь полярной поверхности и полярный объём из набора P, из набора E2 — величины электростатического и ван-дер-ваальсовых взаимодействий и сольватационный вклад, рассчитанный методом Poisson-Boltzmann. В зависимости от подвыборки в уравнениях используется иногда I2, общий объём или молекулярный вес из P, гидрофобный вклад, рассчитанный Generalized Born, трансляционный, ротационный и колебательный энтропийные вклады из E2.

Поскольку некоторые из переменных коррелируют между собой, то в различных вариантах они могут подменять друг друга. Переменная I2 была использована в общем уравнении 14 и в двух из пяти тестовых (14.1 и 14.5). Можно предположить, что выборка S2 неоднородна, но других доказательств этого не обнаружено.

Группы переменных E1 и E2 по сути характеризуют одни и те же параметры комплексов, но при использовании комбинации E1+E2, конечные уравнения из набора E1 включали, как правило, только величину изменения энергии ковалентных взаимодействий в лиганде и рецепторе, аналога которой нет в E2.

Параметры из набора E2 получены усреднением соответствующего набора данных, рассчитанного с регулярным шагом на траектории симуляции молекулярной динамики. Предсказание величины $lg(IC_{50})$ можно рассчитывать для варианта комплекса на каждом таком шаге (рис. 4). При этом разброс предсказанных значений можно трактовать как дополнительную меру качества (аналог разброса данных при экспериментальных измерениях). Чем больше разброс предсказанных значений, тем сомнительнее результат предсказания.

В заключение отметим, что использование ограничения по сходному пространственному положению вариантов докирования для структурно близких соединений и добавление индикаторных переменных, связанных и с источником данных, даёт возможность создать набор уравнений, адекватно предсказывающий величину IC_{50} для ингибиторов главной протеазы SARS-CoV-2.

ФИНАНСИРОВАНИЕ

Работа выполнена в рамках Программы фундаментальных научных исследований в Российской Федерации на долгосрочный период (2021–2030 годы) (№ 122030100170-5).

СОБЛЮДЕНИЕ ЭТИЧЕСКИХ СТАНДАРТОВ

Данная работа не содержит каких-либо исследований с использованием людей и животных в качестве объектов исследования.

КОНФЛИКТ ИНТЕРЕСОВ

Авторы заявляют об отсутствии конфликта интересов.

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

ЛИТЕРАТУРА

1. *Иванова Я.О., Воронина А.И., Скворцов В.С.* (2022) Предсказание ингибирования главной протеазы SARS-CoV-2 с учётом фильтрации данных о положении лигандов. Биомедицинская химия, **68**(6), 444-458. [*Ivanova Ya.O., Voronina A.I., Skvortsov V.S.* (2022) The prediction of SARS-CoV-2 main protease inhibition with filtering by position of ligand. Biomeditsinskaya Khimiya, **68**(6), 444-458.] DOI: 10.18097/PBMC20226806444
2. *Han S.H., Goins C.M., Arya T., Shin W.-J., Maw J., Hooper A., Sonawane D.P., Porter M.R., Bannister B.E., Crouch R.D., Lindsey A.A., Lakatos G., Martinez S.R., Alvarado J., Akers W.S., Wang N.S., Jung J.U., Macdonald J.D., Stauffer S.R.* (2022) Structure-based optimization of ML300-derived, noncovalent inhibitors targeting the severe acute respiratory syndrome coronavirus 3CL protease (SARS-CoV-2 3CL(pro)). J. Med. Chem., **65**(4), 2880-2904. DOI: 10.1021/acs.jmedchem.1c00598
3. *Case D.A., Aktulga H.M., Belfon K., Ben-Shalom I.Y., Berryman J.T., Brozell S.R., Cerutti D.S., Cheatham T.E. III, Cisneros G.A., Cruzeiro V.W.D., Darden T.A., Forouzesh N., Giambaeu G., Giese T., Gilson M.K., Gohlke H., Goetz A.W., Harris J., Izadi S., Izmailov S.A., Kasavajhala K., Kaymak M.C. et al* (2023) Amber 2023, University of California, San Francisco.
4. *Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N., Bourne P.E.* (2000) The protein data bank. Nucleic Acids Res., **28**, 235-242. DOI: 10.1093/nar/28.1.235
5. Schrodinger (Schrodinger, LLC, New York, NY). Retrieved September 02, 2023 from https://www.schrodinger.com/
6. *Harder E., Damm W., Maple J., Wu C., Reboul M., Xiang J.Y., Wang L., Lupyan D., Dahlgren M.K., Knight J.L., Kaus J.W., Cerutti D.S., Krilov G, Jorgensen W.L., Abel R., Friesner R.A.* (2015) OPLS3: A force field providing broad coverage of drug-like small molecules and proteins. J. Chem. Theory Comput., **12(**1), 281-296. DOI: 10.1021/acs.jctc.5b00864
7. *Mikurova A.V., Skvortsov V.S., Grigoryev V.V.* (2020) Generalized predictive model of estimation of inhibition of muscarinic receptors M1-M5. Biomedical Chemistry: Research and Methods, **3**(3), e00129. DOI: 10.18097/bmcrm00129
8. *Massova I., Kollman P.A.* (2000) Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. Perspectives Drug Discovery Design, **18**, 113-135. DOI: 10.1023/A:1008763014207
9. SYBYL-X, Tripos, St. Louis, MO, USA.
10. *Mikurova A.V., Skvortsov V.S.* (2018) Prediction of progestin affinity for the human progesterone receptor based on corrected RBA data. Biomedical Chemistry: Research and Methods, **1**(4), e00080. DOI: 10.18097/BMCRM00080
11. *Gentile F., Fernandez M., Ban F., Ton A.-T., Mslati H., Perez C.F., Leblanc E., Yaacoub J.C., Gleave J., Stern A., Wong B., Jean F., Strynadka N., Cherkasov A.* (2021) Automated discovery of noncovalent inhibitors of SARS-CoV-2 main protease by consensus deep docking of 40 billion small molecules. Chemical Science, **12**(48), 15960-15974. DOI: 10.1039/d1sc05579h
12. *Deodato D., Asad N., Dore T.M.* (2022) Discovery of 2-thiobenzimidazoles as noncovalent inhibitors of SARS-CoV-2 main protease. Bioorganic Med. Chem. Lett., **72**, 128867. DOI: 10.1016/j.bmcl.2022.128867

Поступила в редакцию: 22. 09. 2023.
После доработки: 18. 10. 2023.
Принята к печати: 20. 10. 2023.

THE PREDICTION OF MAIN PROTEASE SARS-CoV-2 INHIBITION BASED ON MODELS OF ENZYME-INHIBITOR COMPLEXES

Ya.O. Ivanova*, V.S. Skvortsov

Institute of Biomedical Chemistry,
10 Pogodinskaya str., Moscow, 119121 Russia; *e-mail: yana.emris@gmail.com

A set of linear regression equations predicting the IC_{50} values for SARS-CoV-2 main protease inhibitors was analyzed. For 180 competitive inhibitors, we have simulated the molecular dynamics of enzyme-inhibitor complexes with known structures or modeled using molecular docking. In the docking procedure, the selection of final poses was restricted by similarity to known structural analogs. The values of the energy contributions obtained by means of calculation of the free energy change of the enzyme-inhibitor complex performed by two variants of the MMPBSA (MMGBSA) method and a number of physicochemical characteristics of the inhibitors were used as independent variables. During the learning process, indicator variables were used for inhibitor subsets obtained from various literature sources to compensate the existing systematic deviations from the target value. A leave one out and leave 20% out cross validation procedures were used to evaluate the prediction quality. For the total logarithmic range width of 3.71, the mean error in predicting the $lg(IC_{50})$ value was 0.45 log units. The stability of the prediction depending on the variability of the complex in molecular dynamics was investigated.

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

Key words: SARS-CoV-2; main protease; competitive inhibitors; QSAR

Funding. The work was performed within the framework of the Program for Basic Research in the Russian Federation for a long-term period (2021–2030) (No. 122030100170-5).

Received: 22.09.2023; revised: 18.10.2023; accepted: 20.10.2023.