

©Коллектив авторов

ПРЕДСКАЗАНИЕ ИНГИБИРОВАНИЯ ГЛАВНОЙ ПРОТЕАЗЫ SARS-CoV-2 С УЧЁТОМ ФИЛЬТРАЦИИ ДАННЫХ О ПОЛОЖЕНИИ ЛИГАНДОВ

Я.О. Иванова*, А.И. Воронина, В.С. Сковрцов

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

В работе анализируется набор уравнений, адекватно предсказывающий величину IC_{50} для ингибиторов главной протеазы SARS-CoV-2. Обучающая выборка получена с использованием фильтрации по независимым от предсказания целевой величины критериям. В её состав вошло 76 соединений, тестовая выборка составила 9 соединений. В качестве независимых переменных использовали величины энергетических вкладов, полученных при расчёте методом MMGBSA изменения свободной энергии комплекса, и ряд характеристик физико-химических свойств ингибиторов. Результаты нашего исследования свидетельствуют о том, что достаточно использовать всего 7 независимых переменных без потери качества предсказания ($Q^2 = 0,79$; $R^2_{\text{предсказания}} = 0,89$). Максимальная ошибка при этом не превышает 0,92 единицы $\lg(IC_{50})$ при полном диапазоне наблюдаемых величин от 1,26 до 4,95.

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

DOI: 10.18097/PBMC20226806444

ВВЕДЕНИЕ

Коронавирусы представляют собой семейство одноцепочечных (+)РНК-содержащих вирусов, относящихся к порядку *Nidovirales* и семейству *Coronaviridae*, в котором выделяют: *Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus*, *Deltacoronavirus* [1]. Способные заражать человека коронавирусы относятся к родам *Alphacoronavirus* и *Betacoronavirus*, в то время как вирусы родов *Gammacoronavirus* и *Deltacoronavirus* заражают в основном птиц [2]. Геном коронавирусов составляет ~26-32 тысяч нуклеотидов, что является одним из самых больших геномов среди РНК-содержащих вирусов [3].

Вирусы SARS-CoV, MERS-CoV, SARS-CoV-2 получили эпидемиологическое распространение в человеческой популяции. Более того, 11 марта 2020 г вспышка SARS-CoV-2 была объявлена Всемирной организацией здравоохранения пандемией. Появление новых штаммов вируса SARS-CoV-2 приводит к потере эффективности лекарственных соединений, поэтому разработка новых и анализ уже существующих ингибиторов является актуальной задачей.

Генетически SARS-CoV-2 схож с коронавирусом летучих мышей BatCoV RaTG13 (96% идентичной последовательности) и с SARS-CoV (80% идентичной последовательности) [4]. Целями для воздействия фармакологически активных соединений являются S-белок, РНК-зависимая РНК-полимераза [5], хеликаза [6], папаин-подобная протеаза [7] и главная протеаза (3CL^{pro} или M^{pro}). Последняя является привлекательной мишенью для разработки новых лекарственных соединений, так как данная протеаза консервативна и структурно схожа с главными протеазами других коронавирусов [8]. Последовательность аминокислот показывает ~99% идентичности с главной протеазой BatCoV RaTG13, ~96% идентичности с главной протеазой SARS-CoV и 50% идентичности с MERS-CoV.

В структуре главной протеазы выделяют N-конец, и два химотрипсин-подобных домена I и II, глобулярный домен III, состоящий из пяти спиралей с С-концом и длинной петлей, соединяющей домены II и III. Место связывания с субстратом состоит из двух глубоко расположенных сайтов S1 и S2, гидрофобного сайта S4 и нескольких открытых растворителей сайтов S3, S5 и S1' [9]. Сайт S1 состоит из остатков аминокислот Phe140/Leu141/Asn142/Gly143/Ser144/His163/Glu166/His172, сайт S2 — из остатков His41/Cys44/Met49/Pro52/Tyr54 и алкильной части боковых цепей Asp187/Gln189. Остатки аминокислот Met165/Pro168/Gln192 и Gln189/Thr190/Ala191 составляют основу сайта S4, а боковые цепи Thr25/Leu27/His41/Cys145/H164 относятся к сайту S1' (рис. 1). При этом все эти остатки консервативны в 12 M^{pro} разных коронавирусов.

В зависимости от механизма действия разделяют ковалентные и нековалентные (конкурентные) ингибиторы главной протеазы SARS-CoV-2. Ковалентные ингибиторы содержат активную группу, которая способна вступать в химическую реакцию с местом связывания главной протеазы и блокировать её. К таким ингибиторам относятся соединения, содержащие, например, акцепторы Михаэля [18-19], альдегидную группу [10], нитрильную группу [11], α -кетоамидную группу [12] и др. К достоинствам данного вида ингибиторов относят высокую селективность и активность, но в то же время ковалентные ингибиторы обладают рядом существенных недостатков, такими как высокая токсичность [13]. Механизм действия нековалентных ингибиторов основан на взаимодействии фармакологически активного соединения с местом связывания при помощи гидрофобных взаимодействий, водородных связей, электростатических и ван-дер-ваальсовых сил. И хотя нековалентные ингибиторы менее активные, чем ковалентные,

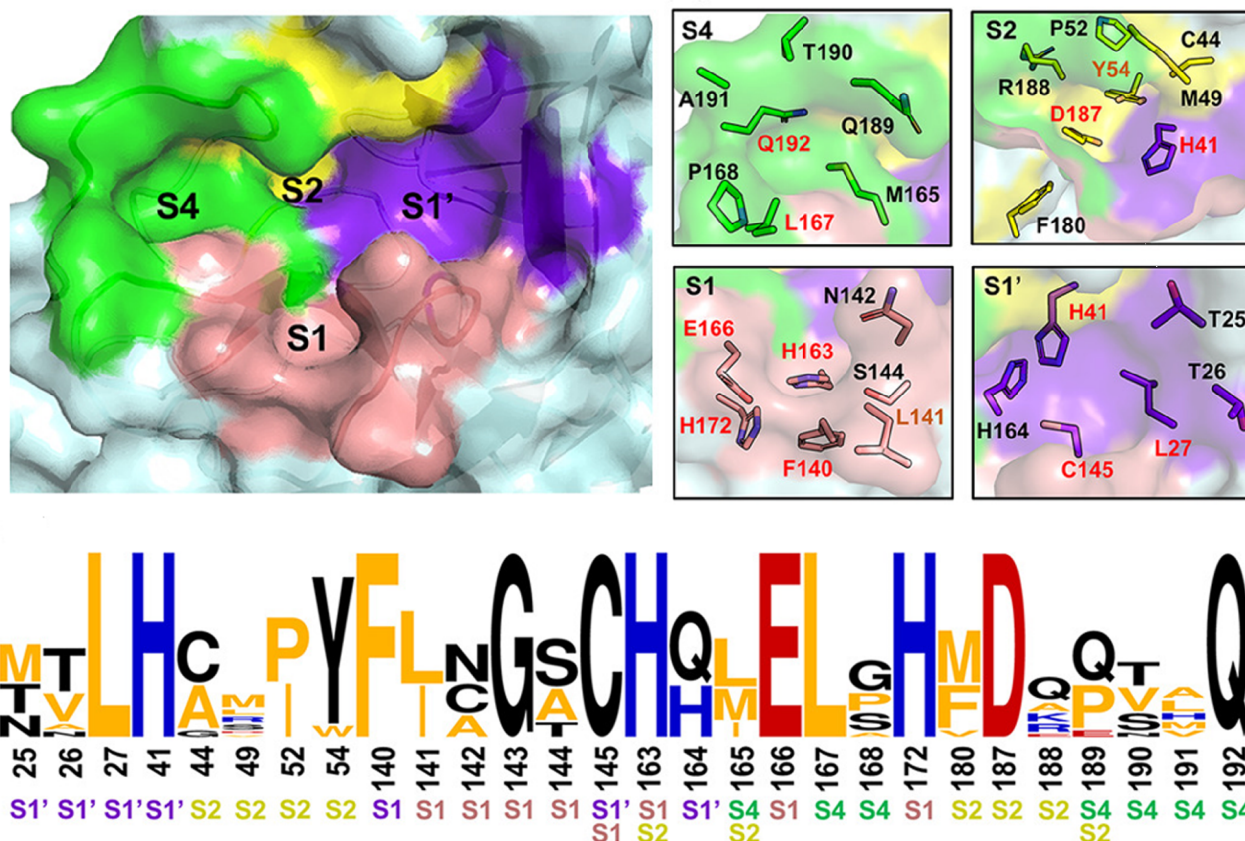


Рисунок 1. Схематичное изображение места связывания в главной протеазе SARS-CoV-2 с указанием консервативности аминокислотных остатков при сравнении различных коронавирусов [9] (лицензия открытого доступа).

они обладают гораздо более низкой токсичностью, поэтому представляют не меньший интерес для исследования.

Существует много работ, посвящённых предсказанию активности ингибиторов по отношению к главной протеазе SARS-CoV-2, как для ковалентных [14], так и нековалентных ингибиторов [15]. Аналогичные работы существуют и для ингибиторов главной протеазы SARS-CoV-1 [16]. Как правило, ингибиторы, использованные в таких работах, представляют серии химически “родственных” соединений. Например, авторами работы [17] был исследован набор из 110 соединений — производных N-оксида пиридина; работы [18] — набор из 28 соединений на основе N-фенил-2-(2-пиримидинилтио)ацетамида; в [19] были проанализированы соединения, представляющие собой полифенолы из морских водорослей. Максимально возможное число таких групп описано в работе [20]. Немало и работ, где используют разнородные выборки, включающие соединения разных химических групп [21]. Механистическое объединение разнородных выборок и обязательное наличие “выбросов” (объявление части выборки несоответствующей генеральной совокупности, если результат неудовлетворительный) в таких работах приводит к тому, что достоверность полученных результатов очень сомнительна, тем более что количество ошибок в таких выборках неизвестно.

Для случая, когда имеет место быть моделирование комплексов, результат тоже бывает неоднозначный, так как близкое к “реальному” положение лиганда не всегда находится в топе отобранных.

Задачей работы был анализ простых линейных моделей, основанных на использовании данных об известных или смоделированных комплексах. При этом собранные данные подвергали дополнительной проверке и фильтрации по несвязанным с предсказанием целевой величины критериям.

МЕТОДИКА

Первоначальный отбор соединений проводили с использованием исходных данных, либо результатов работ [22-50], в которых исследовались конкурентные ингибиторы M^{pro} SARS-CoV-2. Для части из таких соединений (1-39) существуют данные о 3D структурах комплексов, депонированные в Protein Data Bank [PDB]. Отбирали соединения, имеющие молекулярный вес от 250 а.е. Всего таких соединений было собрано 148 (рис. 2, табл. 1). Для определения качества выборки были проанализированы ссылки на первоисточники данных (нередко имела место перепубликация данных из более ранних работ), на методики измерения целевой величины (IC_{50}) и др. Например, ингибиторную активность соединений исследовали

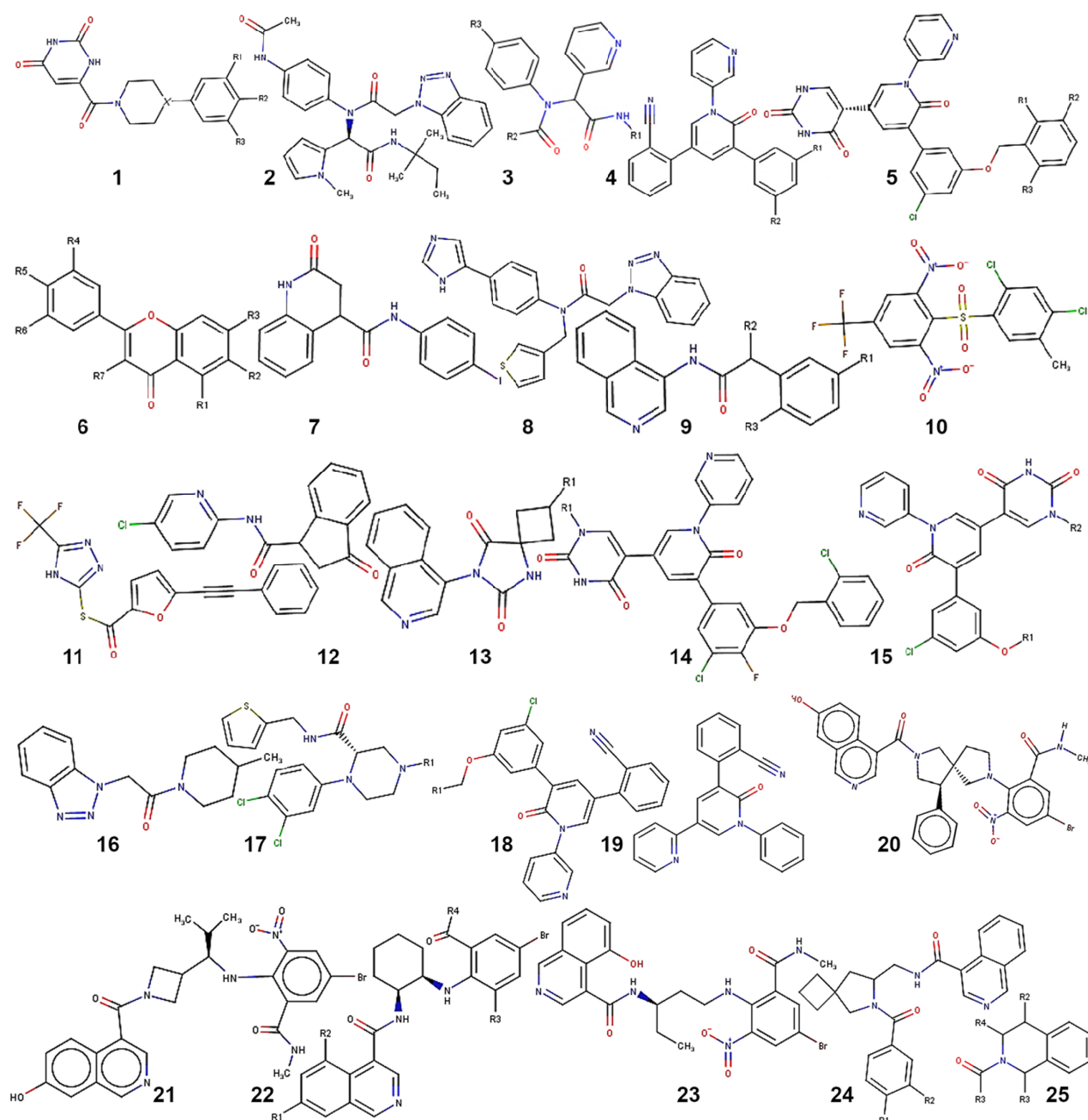


Рисунок 2. Структурные формулы соединений, использованных в работе (табл. 1).

с использованием 5 различных субстратов: (1) Dabcyl-KTSAVLQSGFRKME-Edans, всего 90 значений; (2) MCA-AVLQSGFR-K(Dnp)-K-NH₂ — 5; (3) HiyteFluor-488ESATLQSGRLKAK-(QXL)-NH₂ — 1; (4) Dabcyl-KNSTLQSGRLRKE-Edans — 40; (5) Ac-TSAVLQSGFRKK(biotin)-NH₂ — 9. В данной работе значения, полученные с использованием наиболее редких субстратов 2 и 3, не использовали. Кроме того, для трёх вариантов (4-6) данные о субстрате не получилось выявить. Для четырёх вариантов (4, 6, 16, 17) выяснилось, что значения IC_{50} взяты из работ по ингибированию главной протеазы SARS-CoV-1, и кристаллические структуры также принадлежали SARS-CoV-1. При этом для 4 и 6 было обнаружено, что данные их тоже не использовали. Ещё для одного соединения

выяснилось, что оно является неконкурентным ингибитором (83). Таким образом, выборка была сокращена до 136 соединений, из них для 30 структура комплексов с ферментом была известна из PDB.

Для лигандов с неизвестной структурой комплекса проводили моделирование комплексов с протеазой с использованием программы Schrodinger [51]. В качестве прототипа для докинга использовали вариант структуры M^{pro} из комплекса с соединением 39 (PDB ID 7N44). Область докинга определили как куб с размерами 20×20×20 Å вокруг лиганда 39. Процедура докинга была выполнена с использованием модуля Glide [52], при этом лиганды считали гибкими, а лучшую по оценочной функции Glide позицию отбирали после процедуры финальной минимизации структуры.

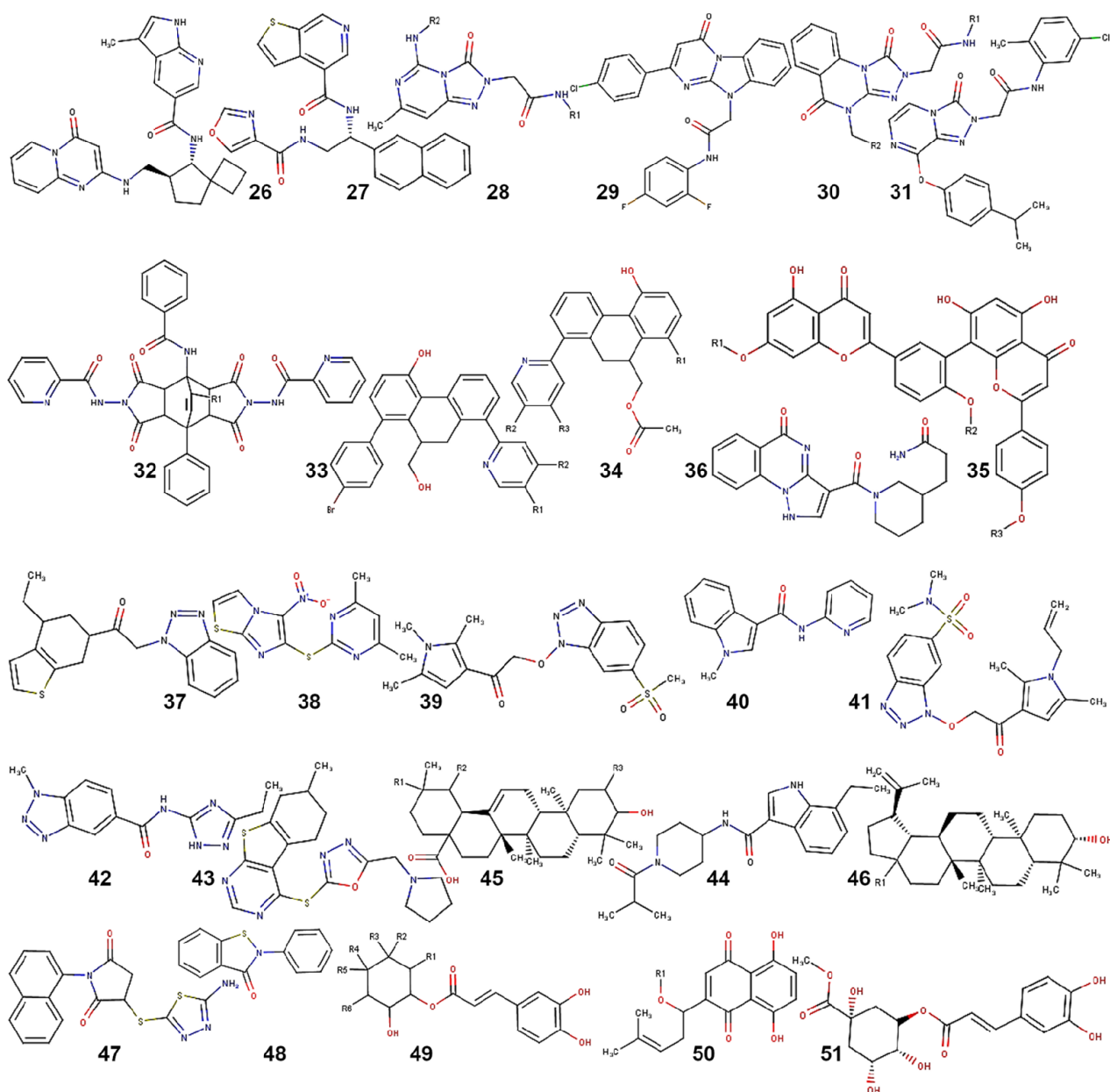


Рисунок 2. Структурные формулы соединений, использованных в работе (табл. 1) (продолжение).

Общую структуру комплексов оптимизировали с использованием поля сил OPLS3e [53]. Данную процедуру проводили для всех комплексов, включая структуры, полученные из PDB. Для каждого комплекса рассчитывали изменения свободной энергии с использованием метода MMGBSA. Использовали модель сольватации VSGB.

Изменение свободной энергии связывания рецептора и лиганда (ΔG_{bind}) методом MMGBSA рассчитывается по формуле:

$$\Delta G_{\text{bind}} = E_{\text{complex}} - E_{\text{ligand}} - E_{\text{receptor}}$$

где величина потенциальной энергия комплекса, рецептора (фермента) и лиганда вычисляется для оптимизированной структуры [51].

Для предсказания величины IC_{50} (в виде $\lg(IC_{50} \text{ нМ})$) использовали как конечную величину ΔG_{bind} (E1),

так и набор из 7 покомпонентных значений (E2), 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 обобщенной модели Борна, вклада водородного связывания и вклада, связанного с липофильностью.

Кроме того, для каждого из лигандов средствами программы Sybyl-X [54] был рассчитан набор из 6 параметров (P), характеризующих свойства лигандов (молекулярный вес, общий и полярный объем, площадь общей и полярной поверхности, число связей, по которым возможно вращение).

Таблица 1. Данные об ингибиторах главной протеазы SARS-CoV-2, использованные в работе

Ингибитор	PDB ID	IC ₅₀ SARS-CoV-2 (нМ)	lg(IC ₅₀)	IC ₅₀ SARS-CoV-1 (нМ)	Молекулярный вес	Суб-страт	Структура* ингибитора	Заместители	Выборка	Ссылка	Особые отметки
1	7RMB	610	2,785		413,65	1	1.1	R1 – Cl; R2 – Br; R3 – H	1	[22]	
2	7RME	1400	3,146		402,76	1	1.2	R1 – CF ₃ ; R2 – Cl; R3 – H	1	[22]	
3	7RMT	6200	3,792		362,77	1	1.3	R1 – Cl; R2 – CHO; R3 – H	1	[22]	
4	4MDS			6200	515,61	ND	2		0	[23]	1
5	6W63	2300	3,362		459,58	ND	3.1	R1 – cyclohexyl	0	[24]	
6	6WCO			1700	445,56	ND	3.2	R1 – cyclopentyl	0	[24]	1
7	7L0D	2500	3,398	4500	433,54	1	3.3		1	[25]	
8	7L11	140	2,146		441,91	1	4.1	R1 – O-Pr; R2 – Cl	1	[26]	
9	7L13	18	1,255		533,36	1	5.1	R1 – Cl; R2 – H; R3 – H	1	[26]	
10	7L14	170	2,230		453,92	1	4.2	R1 – CH ₂ -O-CH ₂ -cyclopropyl; R2 – Cl	1	[26]	
11	6M2N	940	2,973		270,24	2	6.1	R1 – OH; R2 – OH; R3 – OH; R4 – H; R5 – H; R6 – H; R7 – H	0	[27]	
12	7RN4	8800	3,944		368,21	1	1.4	R1 – Cl; R2 – Cl	1	[22]	
13	7P2G	1000	3,000		392,19	1	7		1	[28]	
14	7LMF	63	1,799	148	416,50	3	8		0	[29]	
15	7S3S	1600	3,204		296,75	1	9.1	R1 – Cl; R2 – H; R3 – H	1	[30]	
16	2GZ7			300	461,20	1	10		0	[31]	1
17	2GZ8			3000	363,31	1	11		0	[31]	1
18	7S4B	2100	3,322		294,32	1	9.2	R1 – F	1	[30]	
19	7S3K	1800	3,255		326,78	1	9.3	R1 – Cl; R2 – H; R3 – O-CH ₃	1	[30]	
20	7P51	5400	3,732		286,71	1	12		1	[32]	
21	7LTJ	680	2,833		369,20	1	1.5	R1 – Cl; R2 – Cl; R3 – H	1	[22]	
22	7L12	128	2,107		498,92	1	5.2	R1 – H; R2 – H; R3 – H	1	[26]	
23	7RMZ	6400	3,806		402,76	1	1.6	R1 – Cl; R2 – CF ₃ ; R3 – H	1	[22]	
24	7RM2	290	2,462		383,23	1	1.7	R1 – Cl; R2 – CH ₃ ; R3 – Cl	1	[22]	
25	7O46	330	2,519		321,37	1	13.1	R1 – cyclobutane	1	[33]	
26	7QBB	390	2,591		343,38	1	13.2	R1 – phenyl	1	[33]	
27	7M8P	20	1,301		551,35	1	14.1	R1 – H	1	[34]	
28	7M8M	120	2,079		450,87	1	15.1	R1 – C ₂ H ₅	1	[34]	
29	7M8N	100	2,000		512,94	1	15.2	R1 – o-chlorophenyl	1	[34]	
30	7KX5	660	2,820		501,57	1	3.4		1	[35]	
31	7RLS	290	2,462		403,65	1	1.8	R1 – Cl; R2 – Cl; R3 – Cl	1	[22]	

Таблица 1. Данные об ингибиторах главной протеазы SARS-CoV-2, использованные в работе (продолжение)

Ингибитор	PDB ID	IC ₅₀ SARS-CoV-2 (нМ)	lg(IC ₅₀)	IC ₅₀ SARS-CoV-1 (нМ)	Молекулярный вес	Суб-страт	Структура* ингибитора	Заместители	Выборка	Ссылка	Особые отметки
32	7B2J	20000	4,301		258,32	1	16		1	[33]	
33	7L10	4020	3,604		418,27	1	4.3	R1 – Cl; R2 – Cl	1	[26]	
34	7M90	250	2,398		534,95	1	15.3	R1 – 2-(3-oxo-piperazin-1-yl)	1	[26]	
35	7M91	25	1,398		504,85	1	15.4	R1 – CH ₂ CF ₃	1	[26]	
36	7M8Z	250	2,398		494,93	1	15.5	R1 – CH ₂ CF ₂ OH	1	[26]	
37	8ACD	400	2,602	1090	508,38	2	17.1	R1 – 2,6-dioxo-1,2,3,6-tetrahydropyrimidine-4-carbonyl	0	[36]	
38	8ACL	400	2,602	4610	475,39	2	17.2	R1 – 3,4-dichlorophenyl	0	[36]	
39	7N44	105	2,021		557,87	1	15.6	R1 – 4-(trifluoromethyl)-1,3-oxazol-5-yl group; R2 – H	1	[37]	
40	Модель	44	1,643		565,39	1	14.2	R1 – CH ₃	1	[37]	
41	Модель	38	1,580		573,94	1	15.7	R1 – 4-(trifluoromethyl)-1,3-thiazol-5-yl group; R2 – H	1	[37]	
42	Модель	59	1,771		518,87	1	15.8	R1 – CH ₂ CH ₂ CF ₃ ; R2 – CH ₃	1	[37]	
43	Модель	27	1,431		539,40	1	15.9	R1 – 2-chlorothiophen-3-yl group; R2 – H	1	[37]	
44	Модель	42	1,623		519,97	1	15.10	R1 – 5-methyl-1,3-thiazol-4-yl group; R2 – H	1	[37]	
45	Модель	131	2,117		503,90	1	15.11	R1 – 4-methyl-1,3-oxazol-5-yl group; R2 – H	1	[37]	
46	Модель	85	1,929		503,89	1	15.12	R1 – 5-methyl-1,3-oxazol-4-yl group; R2 – H	1	[37]	
47	Модель	61	1,785		476,92	1	15.13	R1 – CH ₂ -c-Pr; R2 – CH ₃	1	[37]	
48	Модель	83	1,919		494,94	1	21.1	R1 – 5-methyl-1,3-oxazol-4-yl group	1	[37]	
49	Модель	120	2,079		495,89	1	21.2	R1 – CH ₂ CF ₃	1	[37]	
50	Модель	328	2,516		481,98	1	21.3	R1 – cyclopentyl	1	[37]	
51	Модель	37	1,568		462,89	1	15.14	R1 – cyclopropyl	1	[37]	
52	Модель	10000	5,000		349,39	1	19		0	[37]	
53	Модель	21000	4,322	23800	286,24	1	6.2	R1 – OH; R2 – H; R3 – OH; R4 – OH; R5 – OH; R6 – H; R7 – OH	2	[38]	
54	Модель	71	1,851		630,50	1	20		0	[39]	
55	Модель	122	2,086		556,42	1	21.4		0	[39]	
56	Модель	467	2,669		506,40	1	22.1	R1 – H; R2 – H; R3 – CN; R4 – NHCH ₃	0	[39]	
57	Модель	6112	3,786		513,35	1	22.2	R1 – H; R2 – H; R3 – NO ₂ ; R4 – OH	0	[39]	
58	Модель	648	2,812		530,38	1	23		0	[39]	
59	Модель	221	2,344		542,39	1	22.3	R1 – OH; R2 – H; R3 – NO ₂ ; R4 – NHCH ₃	0	[39]	
60	Модель	72	1,857		526,39	1	22.4	R1 – H; R2 – H; R3 – NO ₂ ; R4 – NHCH ₃	0	[39]	
61	Модель	849	2,929		527,38	1	22.5	R1 – H; R2 – H; R3 – NO ₂ ; R4 – OCH ₃	0	[39]	
62	Модель	239	2,378		549,39	1	22.6	R1 – H; R2 – H; R3 – CF ₃ ; R4 – NHCH ₃	0	[39]	

Таблица 1. Данные об ингибиторах главной протеазы SARS-CoV-2, использованные в работе (продолжение)

Ингибитор	PDB ID	IC ₅₀ SARS-CoV-2 (нМ)	lg(IC ₅₀)	IC ₅₀ SARS-CoV-1 (нМ)	Молекулярный вес	Суб-страт	Структура ингибитора*	Заместители	Выборка	Ссылка	Особые отметки
63	Модель	135	2,130		542,39	1	22.7	R1 – H; R2 – OH; R3 – NO ₂ ; R4 – NHCH ₃	0	[39]	
64	Модель	11200	4,049		433,48	1	24.1	R1 – F; R2 – OH	0	[40]	
65	Модель	29570	4,471		430,51	1	24.2	R1 – NH ₂ ; R2 – OH	0	[40]	
66	Модель	18670	4,271		455,49	1	25.1	R1 – (1Z)-2-(1H-пуразол-4-yl)ethenyl group; R2 – H; R3 – [(6-fluoroisquinolin-4-yl)formamido]methyl group	0	[40]	
67	Модель	32730	4,515		456,55	1	26		0	[40]	
68	Модель	56780	4,754		444,47	1	25.2	R1 – 3-methyl-1,3-oxazol-4-yl group; R2 – [(6-fluoroisquinolin-4-yl)formamido]methyl group; R3 – H	0	[40]	
69	Модель	12130	4,084		442,49	1	27		0	[40]	
70	Модель	10840	4,035		427,47	1	25.3	R1 – 1,2,3,4[tetrazolo[1,5-a]pyridin-8-yl] group; R2 – [5-(methylpyridin-3-yl)formamido]methyl group; R3 – H	0	[40]	
71	Модель	56920	4,755		456,55	1	25.4	R1 – 2-(furan-2-yl)propyl group; R2 – H; R3 – H; R4 – [(5-methyl-1H-indazol-4-yl)formamido]methyl group	0	[40]	
72	Модель	10900	4,037		420,45	1	28.1	R1 – 2-fluorophenyl group; R2 – 3,5-dimethylphenyl group	0	[41]	
73	Модель	8700	3,940		406,42	1	28.2	R1 – 2-fluorophenyl group; R2 – 2-methylphenyl group	2	[41]	
74	Модель	5700	3,756		464,86	1	29		0	[41]	
75	Модель	8000	3,903		438,44	1	28.3	R1 – 2,5-difluorophenyl group; R2 – 3,5-dimethylphenyl group	2	[41]	
76	Модель	9500	3,978		475,46	1	30.1	R1 – 2,4-difluorophenyl group; R2 – 4-methylphenyl group	0	[41]	
77	Модель	12100	4,083		471,49	1	30.2	R1 – 4-ethylphenyl group; R2 – 4-fluorophenyl group	0	[41]	
78	Модель	8900	3,949		451,91	1	31		2	[41]	
79	Модель	102200	5,009		681,67	1	32.1	R1 – CH ₃	0	[42]	
80	Модель	200000	5,301		695,69	1	32.2	R1 – H	0	[42]	
81	Модель	5680	3,754	5800	286,24	1	6.3	R1 – OH; R2 – OH; R3 – OH; R4 – H; R5 – OH; R6 – H; R7 – H	0	[43]	
82	Модель	400	2,602		300,27	1	6.4	R1 – OH; R2 – OH; R3 – OH; R4 – H; R5 – OCH ₃ ; R6 – H; R7 – H	0	[43]	
83	Модель	16650	4,221		314,29	1	6.5	R1 – OCH ₃ ; R2 – OH; R3 – OH; R4 – H; R5 – OCH ₃ ; R6 – H; R7 – H	1	[43]	
84	Модель	10230	4,010		470,37	4	33.1	R1 – CH ₃ ; R2 – H	1	[44]	
85	Модель	3310	3,520		605,53	4	34.1	R1 – 4-Br-phenyl; R2 – N-benzyl; R3 – H	1	[44]	
86	Модель	2460	3,391		576,49	4	34.2	R1 – 4-Br-phenyl; R2 – H; R3 – phenyl	1	[44]	
87	Модель	7390	3,869		585,50	4	34.3	R1 – 4-Br-phenyl; R2 – H; R3 – N-morpholine	1	[44]	
88	Модель	1810	3,258		486,37	4	33.2	R1 – H; R2 – CHO	1	[44]	
89	Модель	8890	3,949		528,40	4	34.4	R1 – 4-Br-phenyl; R2 – CHO; R3 – H	0	[44]	
90	Модель	11230	4,050		569,50	4	34.5	R1 – 4-Br-phenyl; R2 – H; R3 – N-pyrrolidine	1	[44]	

Таблица 1. Данные об ингибиторах главной протеазы SARS-CoV-2, использованные в работе (продолжение)

Ингибитор	PDB ID	IC ₅₀ SARS-CoV-2 (нМ)	lg(IC ₅₀)	IC ₅₀ SARS-CoV-1 (нМ)	Молекулярный вес	Суб-страт	Структура ингибитора*	Заместители	Выборка	Ссылка	Особые отметки
91	Модель	5830	3,766		542,43	4	34,6	R1 – 4-Br-phenyl; R2 – acetyl; R3 – H	1	[44]	
92	Модель	1550	3,190		534,45	4	33,3	R1 – H; R2 – Ph	1	[44]	
93	Модель	5390	3,732		476,35	4	33,4	R1 – F; R2 – H	0	[44]	
94	Модель	4710	3,673		492,80	4	33,5	R1 – Cl; R2 – H	0	[44]	
95	Модель	6650	3,823		578,48	4	34,7	R1 – 4-Br-phenyl; R2 – methylsulfonyl; R3 – H	1	[44]	
96	Модель	10260	4,011		525,40	4	34,8	R1 – 4-Br-phenyl; R2 – H; R3 – CN	1	[44]	
97	Модель	8540	3,931		503,64	4	34,9	R1 – cyclohexyl; R2 – H; R3 – phenyl	1	[44]	
98	Модель	2720	3,435		534,83	4	34,10	R1 – 4-Br-phenyl; R2 – Cl; R3 – H	1	[44]	
99	Модель	2780	3,444		525,40	4	34,11	R1 – 4-Br-phenyl; R2 – CN; R3 – H	1	[44]	
100	Модель	10680	4,029		483,37	4	33,6	R1 – CN; R2 – H	1	[44]	
101	Модель	6730	3,828		579,29	4	34,12	R1 – 4-Br-phenyl; R2 – Br; R3 – H	1	[44]	
102	Модель	8280	3,918		576,49	4	34,13	R1 – 4-Br-phenyl; R2 – Ph; R3 – H	1	[44]	
103	Модель	5500	3,740		501,42	4	33,7	R1 – NCH ₃ CH ₃ ; R2 – H	0	[44]	
104	Модель	4790	3,680		534,83	4	34,14	R1 – 4-Br-phenyl; R2 – H; R3 – Cl	1	[44]	
105	Модель	3320	3,521		528,40	4	34,15	R1 – 4-Br-phenyl; R2 – H; R3 – CHO	1	[44]	
106	Модель	10000	4,000		543,46	4	34,16	R1 – 4-Br-phenyl; R2 – H; R3 – NCH ₃ CH ₃	0	[44]	
107	Модель	4290	3,632		518,38	4	34,17	R1 – 4-Br-phenyl; R2 – F; R3 – H	1	[44]	
108	Модель	5250	3,720		514,42	4	34,18	R1 – 4-Br-phenyl; R2 – CH ₃ ; R3 – H	1	[44]	
109	Модель	3660	3,563		543,46	4	34,19	R1 – 4-Br-phenyl; R2 – NCH ₃ CH ₃ ; R3 – H	0	[44]	
110	Модель	9820	3,992		514,42	4	34,20	R1 – 4-Br-phenyl; R2 – H; R3 – CH ₃	1	[44]	
111	Модель	1090	3,037	38400	580,54	4	35,1	R1 – CH ₃ ; R2 – CH ₃ ; R3 – CH ₃	1	[45]	
112	Модель	74860	4,874	20000	286,24	4	6,6	R1 – OH; R2 – H; R3 – OH; R4 – H; R5 – OH; R6 – OH; R7 – H	2	[45]	
113	Модель	10620	4,026		284,27	4	6,7	R1 – OH; R2 – H; R3 – OCH ₃ ; R4 – H; R5 – OH; R6 – H; R7 – H	0	[45]	
114	Модель	2330	3,367		566,51	4	35,2	R1 – H; R2 – CH ₃ ; R3 – CH ₃	1	[45]	
115	Модель	84940	4,929	280800	270,24	4	6,8	R1 – OH; R2 – H; R3 – OH; R4 – H; R5 – OH; R6 – H; R7 – H	2	[45]	
116	Модель	2980	3,474	32000	566,51	4	35,3	R1 – CH ₃ ; R2 – CH ₃ ; R3 – H	1	[45]	
117	Модель	11190	4,049	72300	552,48	4	35,4	R1 – H; R2 – CH ₃ ; R3 – H	1	[45]	
118	Модель	8,65	0,937	8300	538,46	4	35,5	R1 – H; R2 – H; R3 – H	0	[45]	2
119	Модель	31590	4,500	13130	316,26	4	6,9	R1 – OH; R2 – H; R3 – OH; R4 – H; R5 – OH; R6 – OCH ₃ ; R7 – OH	0	[45]	
120	Модель	21200	4,326		367,41	5	36		0	[46]	
121	Модель	5480	3,739		325,43	5	37		1	[46]	

Таблица 1. Данные об ингибиторах главной протеазы SARS-CoV-2, использованные в работе (продолжение)

Ингибитор	PDB ID	IC ₅₀ SARS-CoV-2 (нМ)	lg(IC ₅₀)	IC ₅₀ SARS-CoV-1 (нМ)	Молекулярный вес	Субстрат	Структура ингибитора*	Заместители	Выборка	Ссылка	Особые отметки
122	Модель	82200	4,915		307,35	5	38		0	[46]	
123	Модель	4670	3,669		362,40	5	39		1	[46]	
124	Модель	4530	3,656		251,29	5	40		0	[46]	
125	Модель	1670	3,223		417,48	5	41		1	[46]	
126	Модель	7370	3,867		271,28	5	42		1	[46]	
127	Модель	14760	4,169		387,52	5	43		0	[46]	
128	Модель	6450	3,810		341,46	5	44		0	[46]	
129	Модель	3220	3,508		472,71	4	45.1	R1 – CH ₃ ; R2 – H; R3 – OH	1	[47]	
130	Модель	10000	4,000	14550	455,71	4	46.1	R1 – COOH	1	[47]	
131	Модель	12570	4,099		455,71	4	45.2	R1 – H; R2 – CH ₃ ; R3 – H	1	[47]	
132	Модель	89670	4,953		442,73	4	46.2	R1 – CH ₂ OH	1	[47]	
133	Модель	1890	3,276		356,42	2	47		0	[48]	
134	Модель	200	2,301		227,28	2	48		0	[48]	
135	Модель	94100	4,974		478,41	1	6.10	R1 – OH; R2 – H; R3 – OH; R4 – H; R5 – OH; R6 – OCH ₃ ; R7 – O-glucoside	0	[49]	
136	Модель	34000	4,531		332,26	1	6.11	R1 – OH; R2 – OCH ₃ ; R3 – OH; R4 – H; R5 – OH; R6 – OH; R7 – OH	0	[49]	
137	Модель	42000	4,623		300,27	1	6.12	R1 – OH; R2 – OCH ₃ ; R3 – OH; R4 – H; R5 – OCH ₃ ; R6 – H; R7 – H	0	[49]	
138	Модель	31700	4,501		316,27	1	6.13	R1 – OH; R2 – OCH ₃ ; R3 – OH; R4 – H; R5 – OH; R6 – OH; R7 – H	0	[49]	
139	Модель	42540	4,629		368,34	1	49.1	R1 – OH; R2 – H; R3 – H; R4 – OH; R5 – methoxycarbonyl group; R6 – H	2	[50]	
140	Модель	37060	4,569		445,35	1	6.14	R1 – OH; R2 – OH; R3 – O-glucoronide; R4 – H; R5 – H; R6 – H; R7 – H	0	[50]	
141	Модель	27930	4,446		368,34	1	49.2	R1 – H; R2 – methoxycarbonyl; R3 – OH; R4 – H; R5 – H; R6 – OH	0	[50]	
142	Модель	27330	4,437		370,40	1	50.1	R1 – β,β-dimethylacetyl	0	[50]	
143	Модель	21830	4,339		372,42	1	50.2	R1 – isovaleryl	2	[50]	
144	Модель	3424	3,535		388,42	1	50.3	R1 – β-hydroxyisovaleryl	0	[50]	
145	Модель	3839	3,584		330,34	1	50.4	R1 – acetyl	0	[50]	
146	Модель	1570	3,196		288,30	1	50.5	R1 – H	0	[50]	
147	Модель	66390	4,822		368,34	1	51		2	[50]	
148	Модель	86530	4,937		515,46	1	49.3	R1 – H; R2 – carboxy group; R3 – OH; R4 – H; R5 – H; R6 – [(2E)-3-(3,4-dihydroxyphenyl)prop-2-enoyl]oxy	0	[50]	

Примечание. * – Структурные формулы лигандов представлены на рисунке 2 (первое число – номер базовой структуры, второе – порядковый номер варианта). Номер выборки: 0 – не прошёл отбора, 1 – обучающая выборка, 2 – тестовая. Особые отметки: 1 – данные для SARS-CoV-1; 2 – неконкурентный ингибитор.

Процедура докинга не даёт однозначного варианта положения лиганда в месте связывания, таким образом, нужно отобрать “правильные” позы. В качестве критерия “правильности” использовали следующие правила: (1) соединения с известной кристаллической структурой считались “правильными”; (2) для смоделированных комплексов “правильными” считали такие, в которых положение лиганда в пространстве соответствовало похожему соединению с известной структурой, либо положение группы близких по структуре соединений имели схожее расположение в месте связывания. Так как среди 148 лигандов были соединения различной химической природы, то провели кластеризацию по фрагментным отпечаткам (вектор из 1024 бинарных значений, где единица ставится при наличии в соединении определённого фрагмента химической структуры). Все соединения были разделены на 18 кластеров, 2 из 14 кластеров первого уровня были поделены на 2 и 4 подкластера.

При создании моделей предсказания величины IC_{50} использовали уравнения линейной регрессии. В качестве независимых переменных использовали как наборы E1, E2 и P, так и их комбинации E1+P и E2+P. Качество моделей оценивали по результатам процедуры скользящего контроля методом выкидывания по одному (“leave-one-out cross-validation”) и предсказанию для тестовой выборки. Для набора переменных E2+P дополнительно использовали тест на выкидывание 20% значений. Данные отсортировывали по значению IC_{50} , а затем отбирали каждый n-ый элемент, сформировав 5 выборок, имеющих сходные диапазоны значений IC_{50} .

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Результаты отбора по положению лигандов в месте связывания представлены в таблице 2. Из 18 структурных кластеров 5 имеют не более 3 структур, к тому же не совпадающих по положению с лигандами с известной структурой. Два кластера

Таблица 2. Формирование обучающей и тестовой выборок по положению ингибитора в месте связывания главной протеазы SARS-CoV-2

№ кластера	Известна структура комплекса		Результаты моделирования комплекса		Причина удаления соединения из выборки	Всего соединений	В обучающей выборке	В тестовой выборке
	Расположены в месте связывания единообразно	Положение не совпадает	Расположены в месте связывания единообразно	Положение не совпадает				
1		21	MC 1: 73, 75, 78	56, 72, 74, 76, 77, 79, 80, 120, 128		13	1	3
2	1-3, 12, 23, 24, 31, 37, 38	16			16 – SARS-CoV-1 37, 38 – субстрат 2	10	7	
3.1			83, 111, 114, 116-118 MC 2: 53, 112, 115	81, 82, 113, 119, 136-138, 142	118 – неконкурентный ингибитор	17	5	3
3.2			MC 2: 139, 143, 147	135, 141, 142-146		8		3
4				127		1		
5			84-88, 90-92, 95-102, 104, 105, 107, 108, 110	89, 93, 94, 103, 106, 109		27	21	
6		17		122	17 – SARS-CoV-1	2		
7	32		121, 123, 125, 126			5	5	
8		18		52, 124		3	1	
9		11		140, 148	11 – субстрат 2	3		
10.1		14		69	14 – субстрат 3	2		
10.2	5-7, 25, 26, 30	4		54, 55, 57-68, 70, 71	4, 6 – SARS-CoV-1 4-6 – субстрат не определён	23	4	
10.3	13, 20					2	2	
10.4	8-10, 15, 19, 22, 27-29, 33-36	15				13	13	
11			129-132			4	4	
12			39-51			13	13	
13				133		1		
14				134		1		

Примечание. MC – малая серия.

ПРЕДСКАЗАНИЕ ЗНАЧЕНИЙ IC_{50} ИНГИБИТОРОВ M^{pro} SARS-CoV-2

полностью состоят из соединений с известной структурой, ещё один — наполовину, причём среди близких структур в 2-х случаях имеются разные положения в кристаллах. Ещё один имеет 3 более-менее сходно лежащих лиганда из 12 (малая серия 1). В другом можно выделить 2 группы совпадающих по положению в пространстве лигандов из 6 и 3 (малая серия 2) соединений. К малой серии 2 можно отнести ещё 3 лиганда, отличающихся химически и приписанных другому подкластеру, но в целом сходным образом расположенных в месте связывания. В результате было отобрано 76 соединений, из которых была сформирована обучающая выборка (диапазон значений $lg(IC_{50})$ от 1,26 до 4,95). 9 соединений из малых серий составили тестовую выборку.

Было построено несколько моделей (уравнений) предсказания величины $lg(IC_{50})$ с использованием разных наборов независимых переменных (табл. 3). Следует отметить, что для данной выборки достаточно использовать один-единственный параметр $E1$ (ΔG_{bind}), чтобы получить удовлетворительный результат. Однако использование вариантов с большим числом независимых переменных даёт более устойчивый результат, что лучше всего видно, если анализировать R^2_T и величину максимальной ошибки тестовой выборки. Наиболее наглядно это демонстрируется на рисунке 3, где представлены 3 варианта предсказаний по уравнениям, полученным с использованием энергетических переменных. Следует отметить то, что если просто взять все 148 соединений и рассчитать для них

Таблица 3. Параметры моделей (уравнений) предсказания $lg(C_{50})$, полученные при обучении, и результаты тестирования

№	Тип модели (группы переменных)	Полное/оптимальное число переменных (компонент для 11 и 12)	R^2_L	ME_L	Q^2	ME_{LOO}	$MaxE_{LOO}$	R^2_T 80/20	ME_T 80/20	R^2_T	ME_T	$MaxE_T$
1	Линейная (E1)	1/1	0,73	0,38	0,71	0,38	1,55			0,67	0,10	1,27
2	Линейная (E2)	7/4	0,79	0,34	0,77	0,36	1,23			0,51	0,10	1,33
3	Линейная (P)	6/5	0,57	0,52	0,51	0,55	1,56			0,46	0,12	1,97
4	Линейная (E1+P)	7/3	0,75	0,35	0,73	0,36	1,4			0,63	0,10	1,41
5	Линейная (E2+P)	13/7	0,80	0,31	0,79	0,34	1,51			0,89	0,05	0,65
6	Линейная (E2+P)	13/8	0,83	0,28	0,78	0,36	1,24	0,79	0,42	0,88	0,05	0,64
7	Линейная (E2+P)	13/4	0,80	0,31	0,77	0,34	1,24	0,84	0,31	0,75	0,09	0,98
8	Линейная (E2+P)	13/5	0,85	0,28	0,82	0,31	1,19	0,67	0,47	0,87	0,08	0,92
9	Линейная (E2+P)	13/6	0,82	0,27	0,78	0,32	1,29	0,82	0,31	0,84	0,07	0,79
10	Линейная (E2+P)	13/8	0,85	0,26	0,81	0,30	1,03	0,63	0,39	0,91	0,07	0,73
11	CoMFA	2	0,82	0,39	0,76	0,45	1,12			0,50	0,14	1,70
12	CoMSIA	5	0,92	0,26	0,75	0,47	1,13			0,65	0,14	1,54

Примечание. R^2_L – R^2 обучения; MSE_L – средняя ошибка обучения; Q^2 – Q^2 модели; ME_{LOO} – средняя ошибка для метода скользящего контроля; $MaxE_{LOO}$ – максимальная ошибка для метода скользящего контроля; R^2_T 80/20 – R^2 тестирования в тесте 80/20; ME_T 80/20 – средняя ошибка тестирования в тесте 80/20; R^2_T – R^2 тестирования; ME_T – средняя ошибка тестирования; $MaxE_T$ – максимальная ошибка тестирования.

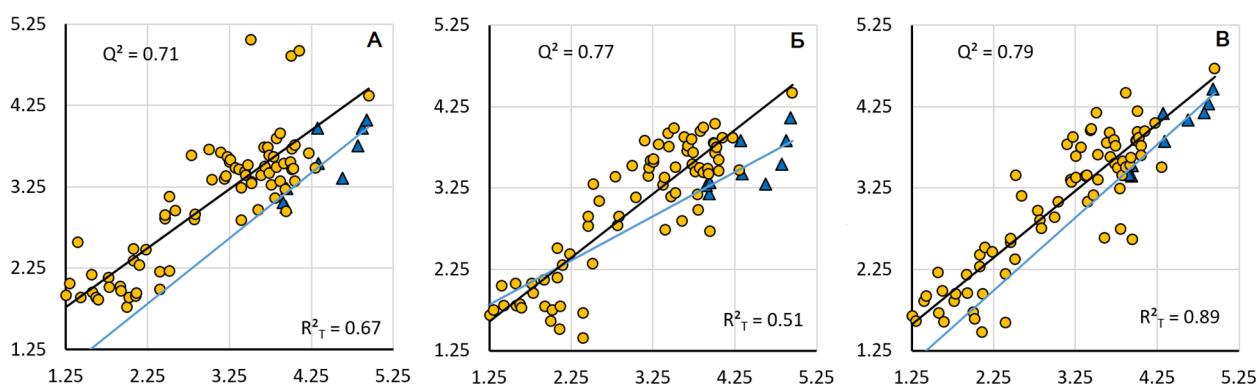


Рисунок 3. Сравнение экспериментальных величин $lg(IC_{50})$ и предсказанных в процедуре скользящего контроля (круг) и тестировании (треугольник) для моделей линейной регрессии E1 (А), E2 (Б) и E2+P (В). По оси ординат экспериментально определённые величины $lg(IC_{50})$. По оси абсцисс предсказанные.

13 параметров, характеризующих вклады в изменение энергии взаимодействия в комплексе и свойства лиганда, то, убрав, как это часто делают, 10 так называемых “выбросов”, можно получить формально удовлетворительную по параметрам описательную модель, показывающую в тесте выкидывания по одному величину Q^2 на уровне 0,65. Однако, при тестировании делением выборки на 80% (обучающая) и 20% (тестовая) результаты неудовлетворительные. При обучении же на 76 отобранных соединениях нет необходимости искать “выбросы”. Для набора E2+P (13 переменных) достаточно использовать только 7 переменных без потери качества ($Q^2=0,79$). При этом в тесте “80/20” усреднённый $R^2_{\text{предсказания}} = 0,7$, минимальный R^2_T для этих 5 уравнений — 0,75, максимальный — 0,91. Максимальная ошибка для этой группы уравнений (E2+P) не превышает 0,92, что позволяет уверенно относить ингибиторы к сильным, средним или слабым. Анализ переменных, входящих в число оптимальных в процедуре скользящего контроля показывает, что наиболее важными являются: молекулярный вес, электростатический вклад в энергию сольватации на основе обобщённой модели Борна, вклад водородного связывания, изменение энергии ван-дер-ваальсовых взаимодействий, площадь полярной поверхности и, иногда, полярный объём. Не имеют существенного значения изменения энергии кулоновского взаимодействия и ковалентных взаимодействий в лиганде и рецепторе, вклад, связанный с липофильностью, и число связей, по которым возможно вращение. Единственно возможного и минимально достаточного набора переменных выявить не получается, так как сами по себе некоторые из этих параметров коррелируют между собой, и в зависимости от набора соединений в выборке могут подменять друг друга.

Примечательно, что для тех же 76 соединений модели CoMFA и CoMSIA (использовали пространственное выравнивание, полученное в результате совмещения по структуре фермента кристаллических структур комплексов и данных докирования) показывают Q^2 равный 0,76 и 0,75 соответственно. Для тестовой выборки лучший вариант $R^2_{\text{предсказания}}$ для модели линейной регрессии составил 0,89, а для CoMFA и CoMSIA моделей — 0,5 и 0,65. Значение 0,5 формально неудовлетворительное, но, учитывая, что диапазон тестовых значений находится в пределах 1,03 логарифмической единицы (3,69 при обучении), то результат можно считать приемлемым.

Таким образом, при использовании несложных правил фильтрации обучающей выборки был создан набор уравнений, адекватно предсказывающий величину IC_{50} для ингибиторов главной протеазы SARS-CoV-2 с максимальной ошибкой не более одного порядка, то есть величиной, достаточной для эффективного скрининга новых соединений как потенциальных лекарственных соединений против SARS-CoV-2. Вновь тестируемые соединения должны удовлетворять правилам отбора в части идентификации положения лиганда в комплексе с главной протеазой.

ФИНАНСИРОВАНИЕ

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

СОБЛЮДЕНИЕ ЭТИЧЕСКИХ СТАНДАРТОВ

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

КОНФЛИКТ ИНТЕРЕСОВ

Авторы заявляют об отсутствии конфликта интересов.

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

ЛИТЕРАТУРА

1. Shereen M.A., Khan S., Kazmi A., Bashir N., Siddique R. (2020) COVID-19 infection: Emergence, transmission, and characteristics of human coronaviruses. *J. Adv. Res.*, **24**, 91-98. DOI: 10.1016/j.jare.2020.03.005
2. Hartenian E., Nandakumar D., Lari A., Ly M., Tucker J.M., Glaunsinger B.A. (2020) The molecular virology of coronaviruses. *J. Biol. Chem.*, **295**(37), 12910-12934. DOI: 10.1074/jbc.rev120.013930
3. Weiss S.R., Navas-Martin S. (2005) Coronavirus pathogenesis and the emerging pathogen severe acute respiratory syndrome coronavirus. *Microbiol. Mol. Biol. Rev.*, **69**(4), 635-664. DOI: 10.1128/mmbr.69.4.635-664.2005
4. Zhou P., Yang X.-L., Wang X.-G., Hu B., Zhang L., Zhang W., Si H.-R., Zhu Y., Li B., Huang C.-L., Chen H.-D., Chen J., Luo Y., Guo H., Jiang R.-D., Liu M.-Q., Chen Y., Shen X.-R., Wang X., Shi Z.-L. (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, **579**(7798), 270-273. DOI: 10.1038/s41586-020-2012-7
5. Muhammed Y., Yusuf Nadabo A., Pius M., Sani B., Usman J., Anka Garba N., Mohammed Sani J., Opeyemi Olayanju B., Zeal Bala S., Garba Abdullahi M., Sambo M. (2021) SARS-CoV-2 spike protein and RNA dependent RNA polymerase as targets for drug and vaccine development: A review. *Biosafety Health*, **3**(5), 249-263. DOI: 10.1016/j.bsheal.2021.07.003
6. Ugurel O.M., Mutlu O., Sariyer E., Kocer S., Ugurel E., Inci T.G., Ata O., Turgut-Balik D. (2020) Evaluation of the potency of FDA-approved drugs on wild type and mutant SARS-CoV-2 helicase (NSP13). *Int. J. Biol. Macromol.*, **163**, 1687-1696. DOI: 10.1016/j.ijbiomac.2020.09.138
7. Xu Y., Chen K., Pan J., Lei Y., Zhang D., Fang L., Tang J., Chen X., Ma Y., Zheng Y., Zhang B., Zhou Y., Zhan J., Xu W. (2021) Repurposing clinically approved drugs for COVID-19 treatment targeting SARS-CoV-2 papain-like protease. *Int. J. Biol. Macromol.*, **188**, 137-146. DOI: 10.1016/j.ijbiomac.2021.07.184
8. Ullrich S., Nitsche C. (2020) The SARS-CoV-2 main protease as drug target. *Bioorg. Med. Chem. Lett.*, **30**(17), 127377. DOI: 10.1016/j.bmcl.2020.127377

9. Xiong M., Su H., Zhao W., Xie H., Shao Q., Xu Y. (2021) What coronavirus 3C-like protease tells us: From structure, substrate selectivity, to inhibitor design. *Med. Res. Rev.*, **41**(4), 1965-1998. DOI: /10.1002/med.21783
10. la Monica G., Bono A., Lauria A., Martorana A. (2022) Targeting SARS-CoV-2 main protease for treatment of COVID-19: Covalent inhibitors structure-activity relationship insights and evolution perspectives. *J. Med. Chem.*, **65**(19), 12500-12534. DOI: 10.1021/acs.jmedchem.2c01005
11. Ramos-Guzmán C.A., Ruiz-Pernía J.J., Tuñón I. (2021) Computational simulations on the binding and reactivity of a nitrile inhibitor of the SARS-CoV-2 main protease. *Chemical Commun.*, **57**(72), 9096-9099. DOI: 10.1039/D1CC03953A
12. Mondal D., Warshel A. (2020) Exploring the mechanism of covalent inhibition: Simulating the binding free energy of α -ketoamide inhibitors of the main protease of SARS-CoV-2. *Biochemistry*, **59**(48), 4601-4608. DOI: 10.1021/acs.biochem.0c00782
13. Aljoundi A., Bjij I., El Rashedy A., Soliman M.E. (2020) Covalent versus non-covalent enzyme inhibition: Which route should we take? A justification of the good and bad from molecular modelling perspective. *Protein J.*, **39**(2), 97-105. DOI: 10.1007/s10930-020-09884-2
14. Sepehri B., Kohnepoushi M., Ghavami R. (2021) High predictive QSAR models for predicting the SARS coronavirus main protease inhibition activity of ketone-based covalent inhibitors. *J. Iranian Chem. Soc.*, **19**(5), 1865-1876. DOI: 10.1007/s13738-021-02426-2
15. Macchiagodena M., Pagliai M., Procacci P. (2020) Identification of potential binders of the main protease 3CLpro of the COVID-19 via structure-based ligand design and molecular modeling. *Chem. Phys. Lett.*, **750**, 137489. DOI: 10.1016/j.cplett.2020.137489
16. Toropov A.A., Toropova A.P., Veselinović A.M., Leszczynska D., Leszczynski J. (2020) SARS-CoV M^{pro} inhibitory activity of aromatic disulfide compounds: QSAR model. *J. Biomol. Struct. Dyn.*, **40**(2), 780-786. DOI: 10.1080/07391102.2020.1818627
17. Ghaleb A., Aouidate A., Ayouchia H.B., Aarjane M., Anane H., Stiriba S.-E. (2020) *In silico* molecular investigations of pyridine N-oxide compounds as potential inhibitors of SARS-CoV-2: 3D QSAR, molecular docking modeling, and ADMET screening. *J. Biomol. Struct. Dyn.*, **40**(1), 143-153. DOI: 10.1080/07391102.2020.1808530
18. Tsai K.-C., Chen S.-Y., Liang P.-H., Lu I.-L., Mahindroo N., Hsieh H.-P., Chao Y.-S., Liu L., Liu D., Lien W., Lin T.-H., Wu S.-Y. (2006) Discovery of a novel family of SARS-CoV protease inhibitors by virtual screening and 3D-QSAR studies. *J. Med. Chem.*, **49**(12), 3485-3495. DOI: 10.1021/jm050852f
19. Gentile D., Patamia V., Scala A., Sciortino M.T., Piperno A., Rescifina A. (2020) Putative inhibitors of SARS-CoV-2 main protease from a library of marine natural products: A virtual screening and molecular modeling study. *Marine Drugs*, **18**(4), 225. DOI: 10.3390/md18040225
20. Gao K., Wang R., Chen J., Tepe J. J., Huang F., Wei G.-W. (2021) Perspectives on SARS-CoV-2 main protease inhibitors. *J. Med. Chem.*, **64**(23), 16922-16955. DOI: 10.1021/acs.jmedchem.1c00409
21. Nandi S., Kumar M., Saxena A.K. (2022) QSAR of SARS-CoV-2 main protease inhibitors utilizing theoretical molecular descriptors. *Research Square* (online first), DOI: 10.21203/rs.3.rs-1526870/v1
22. Kneller D.W., Li H., Galanie S., Phillips G., Labbé A., Weiss K.L., Zhang Q., Arnould M.A., Clyde A., Ma H., Ramamathan A., Jonsson C.B., Head M.S., Coates L., Louis J.M., Bonnesen P.V., Kovalevsky A. (2021) Structural, electronic, and electrostatic determinants for inhibitor binding to subsites S1 and S2 in SARS-CoV-2 main protease. *J. Med. Chem.*, **64**(23), 17366-17383. DOI: 10.1021/acs.jmedchem.1c01475
23. Turlington M., Chun A., Tomar S., Eggler A., Grum-Tokars V., Jacobs J., Daniels J.S., Dawson E., Saldanha A., Chase P., Baez-Santos Y.M., Lindsley C.W., Hodder P., Mesecar A.D., Stauffer S.R. (2013) Discovery of n-(benzo[1,2,3]triazol-1-yl)-n-(benzyl)acetamido)phenyl carboxamides as severe acute respiratory syndrome coronavirus (SARS-CoV) 3CLpro inhibitors: Identification of ML300 and noncovalent nanomolar inhibitors with an induced-fit binding. *Bioorg. Med. Chem. Lett.*, **23**(22), 6172-6177. DOI: 10.1016/j.bmcl.2013.08.112
24. Mesecar A.D. (2020) A taxonomically-driven approach to development of potent, broad-spectrum inhibitors of coronavirus main protease including SARS-CoV-2 (COVID-19). Retrieved November 30, 2022, from: <https://www.rcsb.org/structure/6W63>
25. Lockbaum G.J., Reyes A.C., Lee J.M., Tilvawala R., Nalivaika E.A., Ali A., Kurt Yilmaz N., Thompson P.R., Schiffer C.A. (2021) Crystal structure of SARS-CoV-2 main protease in complex with the non-covalent inhibitor ML188. *Viruses*, **13**(2), 174. DOI: 10.3390/v13020174
26. Zhang C.-H., Stone E.A., Deshmukh M., Ippolito J.A., Ghahremanpour M.M., Tirado-Rives J., Spasov K.A., Zhang S., Takeo Y., Kudalkar S.N., Liang Z., Isaacs F., Lindenbach B., Miller S.J., Anderson K.S., Jorgensen W.L. (2021) Potent noncovalent inhibitors of the main protease of SARS-CoV-2 from molecular sculpting of the drug Perampanel guided by free energy perturbation calculations. *ACS Central Science*, **7**(3), 467-475. DOI: 10.1021/acscentsci.1c00039
27. Su H., Yao S., Zhao W., Li M., Liu J., Shang W., Xie H., Ke C., Hu H., Gao M., Yu K., Liu H., Shen J., Tang W., Zhang L., Xiao G., Ni L., Wang D., Zuo J., Xu Y. (2020) Anti-SARS-CoV-2 activities *in vitro* of Shuanghuanglian preparations and bioactive ingredients. *Acta Pharmacol. Sinica*, **41**(9), 1167-1177. DOI: 10.1038/s41401-020-0483-6
28. Rossetti G.G., Ossorio M.A., Rempel S., Kratzel A., Dionellis V.S., Barriot S., Tropia L., Gorgulla C., Arthanari H., Thiel V., Mohr P., Gamboni R., Halazonetis T.D. (2022) Non-covalent SARS-CoV-2 M^{pro} inhibitors developed from *in silico* screen hits. *Sci. Rep.*, **12**(1), 2505. DOI: 10.1038/s41598-022-06306-4
29. 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. (2021) Structure-based optimization of ML300-derived, noncovalent inhibitors targeting the severe acute respiratory syndrome coronavirus 3CL protease (SARS-CoV-2 3CLpro). *J. Med. Chem.*, **65**(4), 2880-2904. DOI: 10.1021/acs.jmedchem.1c00598
30. Glaser J., Sedova A., Galanie S., Kneller D.W., Davidson R.B., Maradzike E., del Galdo S., Labbé A., Hsu D.J., Agarwal R., Bykov D., Tharrington A., Parks J.M., Smith D.M., Daidone I., Coates L., Kovalevsky A., Smith J.C. (2022) Hit expansion of a noncovalent SARS-CoV-2 main protease inhibitor. *ACS Pharmacol. Transl. Sci.*, **5**(4), 255-265. DOI: 10.1021/acspstci.2c00026

31. Lu I.-L., Mahindroo N., Liang P.-H., Peng Y.-H., Kuo C.-J., Tsai K.-C., Hsieh H.-P., Chao Y.-S., Wu S.-Y. (2006) Structure-based drug design and structural biology study of novel nonpeptide inhibitors of severe acute respiratory syndrome coronavirus main protease. *J. Med. Chem.*, **49**(17), 5154-5161. DOI: 10.1021/jm060207o
32. Cantrelle F.X., Boll E., Brier L., Moschidi D., Belouzard S., Landry V., Leroux F., Dewitte F., Landrieu I., Dubuisson J., Deprez B., Charton J., Hanouille X. (2021) NMR spectroscopy of the main protease of SARS-CoV-2 and fragment-based screening identify three protein hotspots and an antiviral fragment. *Angewandte Chemie International Edition*, **60**(48), 25428-25435. DOI: 10.1002/anie.202109965
33. Luttens A., Gullberg H., Abdurakhmanov E., Vo D.D., Akaberi D., Talibov V.O., Nekhotiaeva N., Vangeel L., de Jonghe S., Jochmans D., Krambrich J., Tas A., Lundgren B., Gravenfors Y., Craig A.J., Atilaw Y., Sandström A., Moodie L.W., Lundkvist E., Carlsson J. (2022) Ultralarge virtual screening identifies SARS-CoV-2 main protease inhibitors with broad-spectrum activity against coronaviruses. *J. Am. Chem. Soc.*, **144**(7), 2905-2920. DOI: 10.1021/jacs.1c08402
34. Deshmukh M.G., Ippolito J.A., Zhang C.-H., Stone E.A., Reilly R.A., Miller S.J., Jorgensen W.L., Anderson K.S. (2021) Structure-guided design of a perampanel-derived pharmacophore targeting the SARS-CoV-2 main protease. *Structure*, **29**(8), DOI: 10.1016/j.str.2021.06.002
35. Ma C., Xia Z., Sacco M.D., Hu Y., Townsend J.A., Meng X., Choza J., Tan H., Jang J., Gongora M.V., Zhang X., Zhang F., Xiang Y., Marty M. T., Chen Y., Wang J. (2021) Discovery of di- and trihaloacetamides as covalent SARS-CoV-2 main protease inhibitors with high target specificity. *J. Am. Chem. Soc.*, **143**(49), 20697-20709. DOI: 10.1021/jacs.1c08060
36. Gao S., Sylvester K., Song L., Claff T., Jing L., Woodson M., Weiße R.H., Cheng Y., Schäkel L., Petry M., Gütschow M., Schiedel A.C., Sträter N., Kang D., Xu S., Toth K., Tavis J., Tollefson A.E., Müller C.E., Zhan P. (2022) Discovery and crystallographic studies of trisubstituted piperazine derivatives as non-covalent SARS-CoV-2 main protease inhibitors with high target specificity and low toxicity. *J. Med. Chem.*, **65**(19), 13343-13364. DOI: 10.1021/acs.jmedchem.2c01146
37. Zhang C.-H., Spasov K.A., Reilly R.A., Hollander K., Stone E.A., Ippolito J.A., Liosi M.-E., Deshmukh M.G., Tirado-Rives J., Zhang S., Liang Z., Miller S.J., Isaacs F., Lindenbach B.D., Anderson K.S., Jorgensen W.L. (2021) Optimization of triarylpyridinone inhibitors of the main protease of SARS-CoV-2 to low-nanomolar antiviral potency. *ACS Med. Chem. Lett.*, **12**(8), 1325-1332. DOI: 10.1021/acsmmedchemlett.1c00326
38. Mangiavacchi F., Botwina P., Menichetti E., Bagnoli L., Rosati O., Marini F., Fonseca S.F., Abenante L., Alves D., Dabrowska A., Kula-Pacurar A., Ortega-Alarcon D., Jimenez-Alesanco A., Ceballos-Laita L., Vega S., Rizzuti B., Abian O., Lenardão E.J., Velazquez-Campoy A., Santi C. (2021) Seleno-functionalization of quercetin improves the non-covalent inhibition of mpro and its antiviral activity in cells against SARS-CoV-2. *Int. J. Mol. Sci.*, **22**(13), 7048. DOI: 10.3390/ijms22137048
39. Hou N., Shuai L., Zhang L., Xie X., Tang K., Zhu Y., Yu Y., Zhang W., Tan Q., Zhong G., Wen Z., Wang C., He X., Huo H., Gao H., Xu Y., Xue J., Peng C., Zou J., Hu Q. (2022) Development of highly potent non-covalent inhibitors of SARS-CoV-2 3CLpro. *BioRxiv* (online first), DOI: 10.1101/2022.08.10.503531
40. 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
41. Lee J.-Y., Kuo C.-J., Shin J.S., Jung E., Liang P.-H., Jung Y.-S. (2021) Identification of non-covalent 3C-like protease inhibitors against severe acute respiratory syndrome coronavirus-2 via virtual screening of a Korean compound library. *Bioorg. Med. Chem. Lett.*, **42**, 128067. DOI: 10.1016/j.bmcl.2021.128067
42. Herlah B., Hoivik A., Jamšek L., Valjavec K., Yamamoto N., Hoshino T., Kranjc K., Perdih A. (2022) Design, synthesis and evaluation of fused bicyclo[2.2.2]octene as a potential core scaffold for the non-covalent inhibitors of SARS-CoV-2 3CLpro main protease. *Pharmaceuticals*, **15**(5), 539. DOI: 10.3390/ph15050539
43. Wu Q., Yan S., Wang Y., Li M., Xiao Y., Li Y. (2022) Discovery of 4'-O-methylscutellarein as a potent SARS-CoV-2 main protease inhibitor. *Biochem. Biophys. Res. Commun.*, **604**, 76-82. DOI: 10.1016/j.bbrc.2022.03.052
44. Zhang J.-W., Xiong Y., Wang F., Zhang F.-M., Yang X., Lin G.-Q., Tian P., Ge G., Gao D. (2022) Discovery of 9,10-dihydrophenanthrene derivatives as SARS-CoV-2 3CLpro inhibitors for treating COVID-19. *Eur. J. Med. Chem.*, **228**, 114030. DOI: 10.1016/j.ejmech.2021.114030
45. Xiong Y., Zhu G.-H., Wang H.-N., Hu Q., Chen L.-L., Guan X.-Q., Li H.-L., Chen H.-Z., Tang H., Ge G.-B. (2021) Discovery of naturally occurring inhibitors against SARS-CoV-2 3CLpro from *Ginkgo biloba* leaves via large-scale screening. *Fitoterapia*, **152**, 104909. DOI: 10.1016/j.fitote.2021.104909
46. Scholle M.D., O'Kane P.T., Dib S., Gurard-Levin Z.A. (2022) Label-free duplex samdi-MS screen reveals novel SARS-CoV-2 3CLpro inhibitors. *Antiviral Res.*, **200**, 105279. DOI: 10.1016/j.antiviral.2022.105279
47. Alhadrami H.A., Sayed A.M., Sharif A.M., Azhar E.I., Rateb M.E. (2021) Olive-derived triterpenes suppress SARS CoV-2 main protease: A promising scaffold for future therapeutics. *Molecules*, **26**(9), 2654. DOI: 10.3390/molecules26092654
48. Guo S., Xie H., Lei Y., Liu B., Zhang L., Xu Y., Zuo Z. (2021) Discovery of novel inhibitors against main protease (Mpro) of SARS-CoV-2 via virtual screening and biochemical evaluation. *Bioorg. Chem.*, **110**, 104767. DOI: 10.1016/j.bioorg.2021.104767
49. Kim J.H., Park Y.-I., Hur M., Park W.T., Moon Y.-H., Koo S.C., Yun-Chan H., Lee I. S., Park J. (2022) The inhibitory activity of methoxyl flavonoids derived from *Inula britannica* flowers on SARS-CoV-2 3CLpro. *Int. J. Biol. Macromol.*, **222**, 2098-2104. DOI: 10.1016/j.ijbiomac.2022.10.008
50. Zhang Y., Gao H., Hu X., Wang Q., Zhong F., Zhou X., Lin C., Yang Y., Wei J., Du W., Huang H., Zhou H., He W., Zhang H., Zhang Y., McCormick P. J., Fu J., Wang D., Fu Y., Li J. (2022) Structure-based discovery and structural basis of a novel broad-spectrum natural product against the main protease of coronavirus. *J. Virology*, **96**(1), e01253-21. DOI: 10.1128/jvi.01253-21
51. Schrodinger (Schrodinger, LLC, New York, NY). Retrieved September 02, 2022 from <https://www.schrodinger.com/>
52. Halgren T.A., Murphy R.B., Friesner R.A., Beard H.S., Frye L.L., Pollard W.T., Banks J.L. (2004) Glide: A new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.*, **47**(7), 1750-1759. DOI: 10.1021/jm030644s

53. Harder E., Damm W., Maple J., Wu C., Reboul M., Xiang J.Y., Wang L., Lupyán 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
54. SYBYL-X, Tripos, St. Louis, MO, USA.

Поступила в редакцию: 01. 11. 2022.
После доработки: 05. 12. 2022.
Принята к печати: 12. 12. 2022.

THE PREDICTION OF SARS-CoV-2 MAIN PROTEASE INHIBITION WITH FILTERING BY POSITION OF LIGAND

Ya.O. Ivanova, A.I. Voronina, V.S. Skvortsov*

Institute of Biomedical Chemistry,
10 Pogodinskaya str., Moscow, 119121 Russia; *e-mail: vladlen@ibmh.msk.su

A set of equations that adequately predict the IC₅₀ values for SARS-CoV-2 main protease inhibitors has been analyzed. The training set obtained using filtering by criteria independent of the target value prediction included 76 compounds, while the test set included nine compounds. We used the values of energy contributions obtained during calculation of the change of the free energy of enzyme-inhibitor complexes by MMGBSA method and a number of characteristics of the physical and chemical properties of the inhibitors as independent variables. According to results of this study, It is sufficient to use only seven independent variables without loss of the prediction quality ($Q^2 = 0.79$; $R^2_{\text{prediction}} = 0.89$). The maximum error in this case does not exceed 0.92 lg(IC₅₀) units with a full range of observed values from 1.26 to 4.95.

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: 01.11.2022; revised: 05.12.2022; accepted: 12.12.2022.