

КРАТКОЕ СООБЩЕНИЕ

©Коллектив авторов

ПРЕДСКАЗАНИЕ ИНГИБИРОВАНИЯ НЕЙРАМИНИДАЗЫ ВИРУСА ГРИППА РАЗЛИЧНЫХ ШТАММОВ С ИСПОЛЬЗОВАНИЕМ ОБОБЩЁННОЙ МОДЕЛИ, ПОСТРОЕННОЙ С УЧЁТОМ ДАННЫХ О ПОЛОЖЕНИИ ИЗВЕСТНЫХ ЛИГАНДОВ

А.В. Микурова, А.В. Рыбина, В.С. Скворцов*

Научно-исследовательский институт биомедицинской химии имени В.Н. Ореховича,
119121, Москва, ул. Погодинская, 10; *эл. почта: a.mikurova@ibmc.msk.ru

Представлены несколько вариантов моделей предсказания значений IC_{50} ингибиторов нейраминидазы вируса гриппа как для отдельных штаммов, так и для комбинаций данных для нейраминидаз нескольких штаммов. Они основаны на использовании расчётных энергетических вкладов в величину изменения свободной энергии комплексов фермент-ингибитор. В отличие от более ранних работ при моделировании комплекса была добавлена процедура сравнения варианта докирования с одним из ингибиторов нейраминидазы, для которых структура комплексов определена экспериментально. Выбор молекул для сравнения выполнен по подобию структур с использованием метрики Танимото, и RMSD между подобными парами атомов не должен бы быть больше 2 Å. При использовании данного ограничения, а также фильтрации серий данных для конкретного штамма по результатам процедуры скользящего контроля методом выкидывания по одному можно построить уравнения предсказания величины IC_{50} со значением Q^2 , близким к минимальному порогу достоверности (0,57 в данной работе). Учитывая, что в нашем варианте моделей используется минимальный набор энергетических вкладов, не предусматривающий дорогостоящих расчётов энтропийных вкладов, полученный результат свидетельствует в пользу корректности использования обобщенной модели, построенной по данным о положении известных лигандов, для предсказания ингибирования нейраминидазы вируса гриппа различных штаммов.

Ключевые слова: нейраминидаза вируса гриппа; ингибиторы; вычислительные методы; QSAR

DOI: 10.18097/PBMC20206606508

ВВЕДЕНИЕ

Настоящая работа развивает более ранние наши работы по созданию моделей для предсказания величины IC_{50} (концентрация, вызывающая 50% торможение) для ингибиторов нейраминидазы вируса гриппа [1-3]. Поиск новых препаратов против гриппа продолжается постоянно. Потребность в них обусловлена высокой изменчивостью вируса гриппа, которая приводит к быстрому возникновению резистентности к химиопрепаратам [4]. В настоящее время в качестве лекарственных средств второго поколения зарегистрированы ингибиторы нейраминидазы вирусов гриппа А и В осельтамивир, занамивир, перамивир и ланинамивир. Возникновение резистентности было неоднократно показано и по отношению к этим препаратам [4], так что переход к третьему и последующим поколениям неизбежен. И, несмотря на то, что коронавирус заметно снизил интерес общественности к вирусу гриппа, проблема противодействия лекарственной резистентности и создания универсальных лекарственных средств, эффективных в отношении нейраминидазы различных штаммов вируса гриппа, по-прежнему требует своего решения.

Основой для данной работы служит гипотеза о минимизации вероятности возникновения резистентности в случае, если ингибитор не обладает узкой специфичностью к какому-то отдельному серотипу, а связывается примерно в равной степени с нейраминидазами всех (или большинства) штаммов,

пусть и с меньшим сродством. При этом в качестве основы для создания предсказательных уравнений должны служить максимально общие расчётные параметры, не зависящие от выбора конкретного варианта нейраминидазы.

Ранее мы показали, что можно построить предсказательные уравнения на основании набора экспериментально определённых (методом рентгеноструктурного анализа) и смоделированных вариантов структуры нейраминидаз различных штаммов [2]. Кроме того, было показано [3], что использование покомпонентных энтальпийных вкладов в изменение свободной энергии комплекса, рассчитанных методом MMPBSA [5], достаточно для создания моделей, позволяющих разделить набор лигандов на слабые, средние и сильные ингибиторы. Данная работа рассматривает возможность использования априорной информации о положении уже известных лигандов для улучшения предсказательной силы модели. В процессе моделирования комплексов мы используем процедуру молекулярного докинга. Её основная проблема — отбор адекватного решения. Часто самое близкое к наблюдаемому решение не соответствует лучшему согласно оценочной функции. Как правило, если проанализировать лучшие 100 отобранных по оценочной функции вариантов, то близкий к экспериментально наблюдаемому варианту среди них есть. Нет другого способа выбрать из списка, кроме как по величине оценочной функции или по заранее известному положению отдельных

структурных групп. Например, для ингибиторов нейраминидазы это положения группы COO⁻ или локализация положительного заряда на атоме азота. За многие годы в Potein Data Bank (PDB) [6] было депонировано достаточное количество данных о положении отдельных лигандов в месте связывания различных нейраминидаз вируса гриппа. Проанализировав попарное сходство одной из этих молекул с моделируемым лигандом, можно выделить общие либо подобные части структуры и оценить насколько близко к известному они расположены в комплексе лиганд/фермент. Данную информацию можно использовать для отбора окончательного варианта комплекса. Так как в последующем используется оптимизация структуры комплекса с использованием молекулярной динамики, небольшие отклонения не играют существенной роли.

МЕТОДИКА

В работе использовали данные по структуре и ингибиторной активности (IC₅₀) химических соединений, которые включали: (1) набор различных низкомолекулярных соединений активных относительно нейраминидаз 5 штаммов A/Tokyo/3/67, A/tern/Australia/G70C/75, A/PR/8/34, A/Aichi/2/68 и B/Lee/40 (221 потенциальный комплекс) [1] и (2) набор трёх известных лекарственных соединений (осельтамивир, занамивир, перамивир), активных относительно 30 вариантов нейраминидаз штаммов: A/teal/Hong Kong/W312/97, A/duck/Alberta/35/76, A/duck/Singapore/3/97, A/duck/Germany/1215/73, A/turkey/Ontario/6118/68, A/shearwater/Australia/1/72, A/duck/Czechoslovakia/56, A/quail/Italy/1117/65, A/duck/Memphis/546/74, A/turkey/Minnesota/916/1980, A/duck/Memphis/546/74, A/turkey/Minnesota/916/1980, а также 8 вариантов мутаций, A/duck/Memphis/546/74 плюс 9 вариантов мутаций (90 комплексов) [2].

Общий план работ по моделированию комплексов проводили по схеме, описанной ранее [2, 3]. Создание предсказательных моделей проводили на основании параметров комплексов, рассчитанных методом ММ-PBSA (ММ-GBSA). Подготовку данных, докинг и молекулярную динамику выполняли средствами программ Dock 6.9 [7] и Amber 18.0 [8] (поля сил FF14SB и GAFF2). Время симуляции для молекулярной динамики было по 0,5 нс для каждого из подготовительных этапов. Конечную симуляцию выполняли для отрезка в 1 нс и записывали 25 вариантов комплексов через каждые 40 пс, по которым получали усреднённые значения энергетических вкладов в изменение энергии связывания, рассчитанных методом ММ-PBSA (ММ-GBSA). В уравнениях использовали следующий набор независимых переменных: молекулярная масса ингибитора; изменение величины электростатического взаимодействия (ELE); изменение величины ван-дер-ваальсовых взаимодействий (VDW), а также гидрофобный (PBSUR) и сольватационный (PBCAL) вклады в изменение свободной энергии, рассчитанной методом Пуассона-Больцмана (PB), аналогичные вклады, рассчитанные обобщённым

методом Борна (GB) (GBSUR и GB); Уравнения линейной регрессии оценивали по величине Q² в процедуре скользящего контроля методом исключения по одному.

Важным отличием данной работы от предыдущих было использование дополнительного критерия при отборе варианта комплекса после докирования. Процедура включала ряд последовательных шагов (в том числе и подготовительных):

1. Подбор комплексов нейраминидаз вируса гриппа из PDB с уникальными структурами лигандов. Всего в работе было использовано 12 структур (PDBID: 2QWB, 1F8E, 1F8D, 3K37, 4KS4, 4KS1, 2QWF, 2QWD, 3CL0, 2QWE, 2SIM, 1INF). Среди лигандов были: сиаловая кислота, ароматические ингибиторы, осельтамивир, занамивир и их производные (рис. 1).
2. Пространственное выравнивание всех комплексов с известными лигандами, а также вариантов нейраминидаз, использованных как место связывания при докировании. между собой (рис. 2).
3. Каждую из имеющихся молекулярных структур (всего 185 оригинальных соединений) попарно сравнивали с каждой из 12 структур, отобранных на шаге 1. С использованием пакета Instant JChem ("ChemAxon", Венгрия) [9] выделяли общую часть молекулярной структуры и определяли степень сходства структур (по метрике Танимото),
4. После процедуры докирования сохраняли до 100 вариантов положения лиганда. Рассчитав среднеквадратичное отклонение (RMSD) между атомами совпадающей части структуры молекулы кандидата и той из 12 молекул, с которой у кандидата было максимальное сходство (или идентичность), отбирали лучший вариант. В процессе отбора можно было варьировать следующие параметры: отсекающее значение оценочной функции (в нашей работе величина "Grid Score" не более -20), минимальную величину коэффициента Танимото (0,5), максимально допустимое значение RMSD (2 Å).

Для выполнения расчётов был использован гибридный высокопроизводительный вычислительный комплекс Федерального исследовательского центра "Информатика и управление" Российской Академии Наук (ФИЦ ИУ РАН) [10] IBM на базе CPU Power9 и графических ускорителей Nvidia Tesla V100.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Использованная в данной работе схема отбора комплексов изначально предполагает, что часть из них будет отброшена. В частности, на рисунке 3 приведено распределение максимально возможного коэффициента Танимото, полученного при сравнении 185 соединений из общей выборки с 12 ингибиторами, взятыми из кристаллических структур (для наглядности приведено также сравнение этих же соединений с одной единственной молекулой — осельтамивиром). Видно, что 20 соединений уже на этой стадии следует отбросить, так как они не удовлетворяют условию, что коэффициент Танимото должен быть больше 0,5.

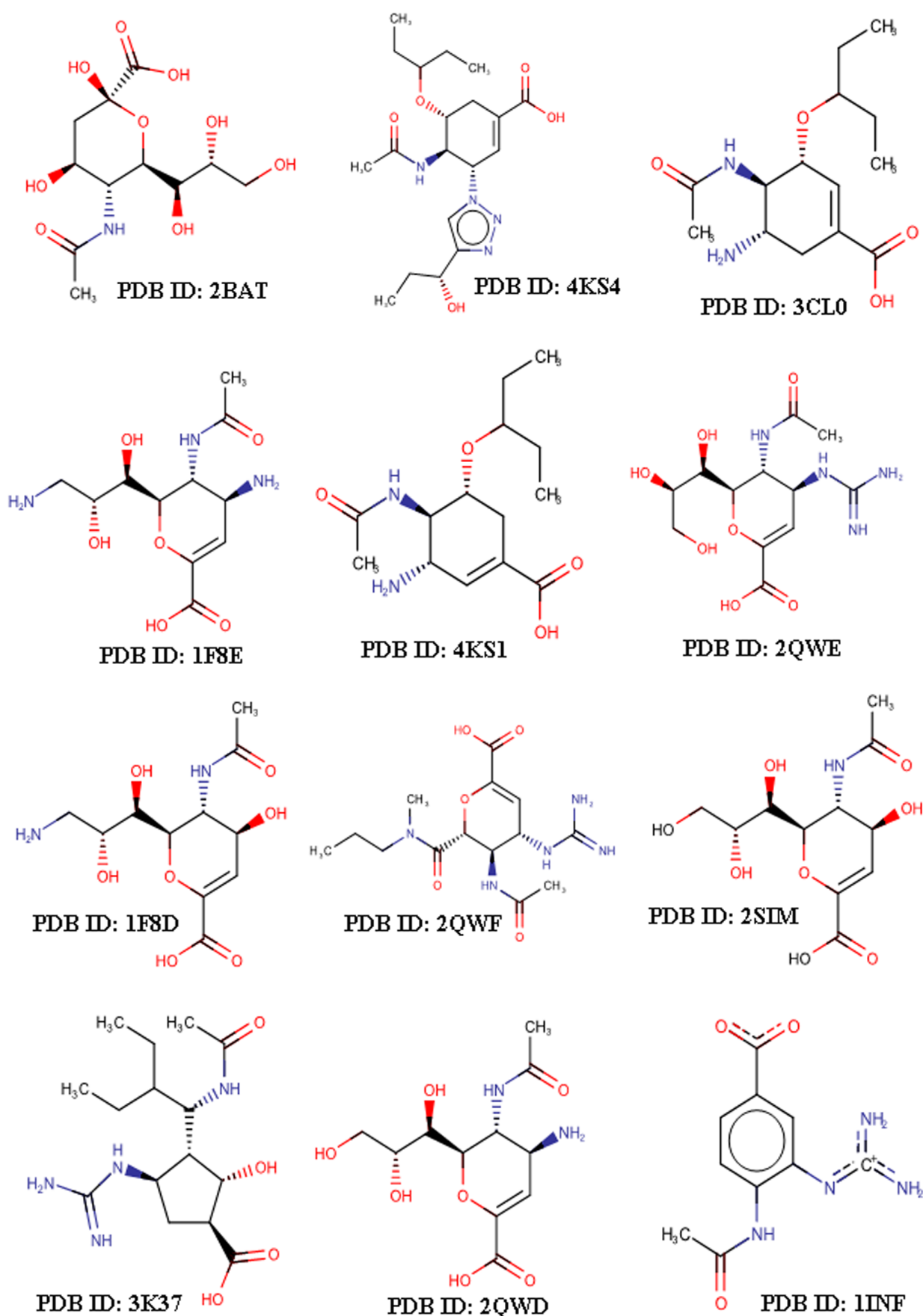


Рисунок 1. 12 структур ингибиторов нейраминидазы вируса гриппа, для которых известно положение в комплексе с нейраминидазой, депонированных в PDB.

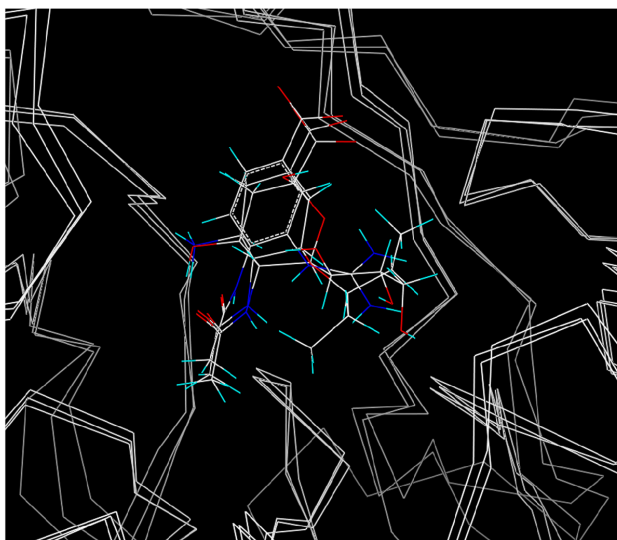


Рисунок 2. Пример пространственного выравнивания известных комплексов нейраминидазы и получившееся при этом наложение лигандов.

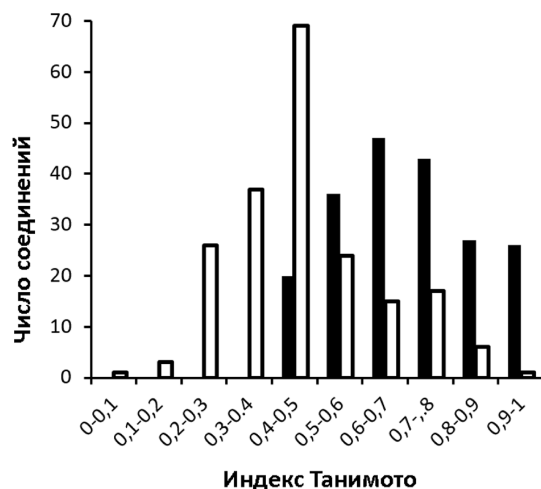


Рисунок 3. Распределение 185 соединений по величине коэффициента Танимото, полученного при попарном сравнении с лигандами с известным положением в кристаллах нейраминидазы вируса гриппа. Чёрная заливка — максимально возможные значения при сравнении со всеми 12 структурами. Без заливки — при сравнении с осельтамивиром.

Таблица. Параметры уравнений линейной регрессии предсказания pIC_{50} , полученные при обучении, и результаты тестирования

Обучающая выборка	R^2_L	MSE_L	n	$R^2_{СМЕШ}$	Q^2	MSE_{LOO}	N	Диапазон значений pIC_{50}
1	0,82	0,82	20	0,45	0,66	1,63	4	2,0:8,8
2	0,55	0,45	20	0,34	0,10	1,10	5	3,2:9,3
3	0,62	0,72	43	0,20	0,51	0,93	5	4,6:10,1
4	0,39	1,66	61	0,16	0,32	1,86	2	3,27:9,85
5	0,54	1,59	23	0,30	0,47	1,86	3	2,1:8,5
6	0,38	0,92	28	0,28	0,15	1,36	2	4,56:9,88
7	0,49	0,27	29	0,21	0,39	0,35	3	6,58:9,5
8	0,55	0,31	28	0,28	0,44	0,39	3	6,94:10,22
С 1 по 5	0,35	2,30	167	0,155	0,24	2,70	4	2,0:10,1
С 6 по 8	0,42	0,45	85	0,20	0,04	0,70	1	4,56:10,22
С 1 по 8	0,45	2,40	252	0,10	0,22	1,70	6	2,0:10,22
1+3+5	0,47	2,20	86	0,07	0,38	2,50	3	2,0:10,1
7+8	0,37	0,50	58	0,10	0,17	0,60	6	6,58:10,22
1+3+5+7+8	0,59	1,75	143	0,01	0,57	1,84	4	2,0:10,22

Примечание. R^2_L — R^2 обучения; $MSEL$ — среднеквадратичная ошибка обучения; n — число наблюдений при обучении (отобранных комплексов); $R^2_{СМЕШ}$ — усреднённый R^2 в тесте случайного перемешивания зависимого значения; Q^2 — Q^2 модели; MSE_{LOO} — среднеквадратичная ошибка для метода скользящего контроля; N — оптимальное число переменных в процедуре скользящего контроля (максимально 7 + константа). Состав выборок: 1 — данные по ингибированию штамма A/Tokyo/3/67; 2 — штамма A/tern/Australia/G70C/75; 3 — A/PR/8/34; 4 — штамма A/Aichi/2/68; 5 — штамма B/Lee/40; 6 — данные по ингибированию осельтамивиром 30 вариантов нейраминидазы вируса гриппа А; 7 — аналогично предыдущей, ингибирование занамивиром; 8 — ингибирование перамивиром.

Кроме того, докинг может не найти решения совсем, либо при отсечении по заданным параметрам не найдётся ни одного близкого решения (RMSD не более 2 Å). В результате, из потенциально возможных 311 вариантов после процедуры отбора осталось 252 (167 для первой выборки и 85 для второй).

Основные параметры уравнений для предсказания величины IC_{50} , или точнее величины $pIC_{50} = -\lg(IC_{50})$, построенных для различных комбинаций подвыборок данных, представлены в таблице. Высокие значения R^2

обучения для выборки 1 могут быть результатом переобучения, несмотря на то, что в тесте скользящего контроля методом выбрасывания по одному она даёт очень хороший результат (в отличие от выборок 2 и 4). В качестве стартового всегда использовали уравнение с 7 независимыми переменными и константой. Другими словами, число наблюдений должно быть не меньше 40, что хорошо демонстрирует такой параметр, как $R^2_{СМЕШ}$ — усреднённый по 10 обучающим процедурам, когда значения pIC_{50}

смешиваются случайным образом. Простое объединение выборок не даёт значимого улучшения. Однако если объединять те из них, для которых Q^2 больше 0,4, то результат становится значимым (выборка 1+3+5+7+8, $Q^2=0,57$). К сожалению, качество данных, собранных из литературных источников, не может быть достоверно подтверждено, тем более, что часто наборы данных сами по себе скомпилированы из различных источников, опубликованных разными группами исследователей и в разное время. Другой вариант — это выборки 6, 7 и 8. Данные получены двумя группами исследователей, но проблема этих выборок — узкий диапазон значений IC₅₀.

Таким образом, можно заключить, что использование дополнительного ограничения при отборе вариантов положения лигандов после процедуры докинга даёт возможность получить адекватный набор расчётных параметров, на основании которых можно построить уравнение, объединяющее данные по ингибированию различных штаммов нейраминидазы вируса гриппа. При включении в общую выборку данных по ингибированию отдельных штаммов имеет смысл использовать только те наборы данных, которые внутри себя в тесте скользящего контроля показывают минимально допустимый результат (например, имеют Q^2 больше чем 0,4).

ФИНАНСИРОВАНИЕ

Работа выполнена в рамках Программы фундаментальных научных исследований государственных академий наук на 2013-2020 годы. Адаптация программного обеспечения под гибридную платформу Power9 в ЦОД ФИЦ ИУ РАН выполнена при поддержке гранта РФФИ № 18-29-03100.

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.

ЛИТЕРАТУРА

1. Микурова А.В., Рыбина А.В., Скворцов В.С. (2016) Биомедицинская химия, **62**(6), 691-703. [Mikurova A.V., Rybina A.V., Skvortsov V.S. (2016) Biomeditsinskaya Khimiya, **62**(6), 691-703.]
2. Микурова А.В., Скворцов В.С. (2018) Биомедицинская химия, **64**(3), 247-252. [Mikurova A.V., Skvortsov V.S. (2018) Biomeditsinskaya Khimiya, **64**(3), 247-252.]
3. Микурова А.В., Скворцов В.С. (2019) Биомедицинская химия, **65**(6), 520-525. [Mikurova A.V., Skvortsov V.S. (2019) Biomeditsinskaya Khimiya, **65**(6), 520-525.]
4. Бреслав Н.В., Шевченко Е.С., Абрамов Д.Д., Прилипов А.Г., Журавлева М.М., Оскерко Т.А., Колобухина Л.В., Меркулова Л.Н., Щелканов М.Ю., Бурцева Е.И., Львов Д.К. (2013) Вопросы вирусологии, **58**(1), 28-32. [Breslav N.V., Shevchenko E.S., Abramov D.D., Prilipov A.G., Zhuravleva M.M., Oskerko T.A., Kolobukhina L.V., Merkulova L.N., Shchelkanov M.Yu., Burtzeva E.I., Lvov D.K. (2013) Voprosy Virusologii, **58**(1), 28-32.]
5. Kollman P.A., Massova I., Reyes C., Kuhn B., Huo S., Chong L., Lee M., Lee T., Duan Y., Wang W., Donini O., Cieplak P., Srinivasan J., Case D.A., Cheatham T.E., 3rd (2000) Acc. Chem. Res, **33**, 889-897.
6. <https://www.rcsb.org> (application date: 01/10/2020).
7. Allen W.J., Balius T.E., Mukherjee S., Brozell S.R., Moustakas D.T., Lang P.T., Case D.A., Kuntz I.D., Rizzo R.C. (2015) J. Comput. Chem., **36**(15), 1132-1156.
8. Salomon-Ferrer R., Case D.A., Walker R.C. (2013) WIREs Comput. Mol. Sci., **3**, 198-210.
9. <https://www.chemaxon.com> (application date: 01/10/2020).
10. Federal Research Center Computer Science and Control of Russian Academy of Sciences [Electronic resource]: site. — Moscow: FRC CS RAS. - URL: <http://hhpcc.frcsc.ru> (application date: 09/11/2020)

Поступила в редакцию: 30. 10. 2020.
После доработки: 20. 11. 2020.
Принята к печати: 27. 11. 2020.

**PREDICTION OF THE INHIBITION OF INFLUENZA VIRUS NEURAMINIDASE
OF VARIOUS STRAINS BY USING A GENERALIZED MODEL CONSTRUCTED
USING THE DATA ON THE POSITION OF KNOWN LIGANDS**

A.V. Mikurova, A.V. Rybina, V.S. Skvortsov*

Institute of Biomedical Chemistry,
10 Pogodinskaya str., Moscow, 119121 Russia; *e-mail: a.mikurova@ibmc.msk.ru

Several variants of models for predicting the IC_{50} values of inhibitors of influenza virus neuraminidase are presented for both individual strains and for combinations of data for neuraminidases of several strains. They are based on the use of calculated energy contributions to the amount of change in the free energy of enzyme-inhibitor complexes. In contrast to previous works, aimed at the complex modeling, we added a procedure of comparison of the docking variants with one of the neuraminidase inhibitors, for which the structure of the complexes was determined experimentally. The choice of the comparison structure was made according to the similarity of structures evaluated using the Tanimoto metrics and the limit of the RMSD value for a similar part of the structure was no more than 2 Å. Using this limitation and filtering datasets for a particular strain by the Q^2 value obtained in the leave-one-out control procedure it is possible to construct equations for predicting the IC_{50} value with a Q^2 value close to the minimum confidence threshold (0.57 in this work). Taking into consideration that in this version of the prediction, a minimum set of energy contributions is used, which does not provide for expensive calculations of entropy contributions, the result obtained supports the correctness of using a generalized model based on the data on the position of known ligands to predict the inhibition of neuraminidase of the influenza virus of various strains.

Key words: influenza virus neuraminidase; inhibitors; computational methods; QSAR

Funding. This work was performed within the framework of the Program for Basic Research of State Academies of Sciences for 2013-2020. The software porting on hybrid Power 9 cluster was supported by the Russian Foundation for Basic Research (project 18-29-03100).

Received: 30.10.2020, revised: 20.11.2020, accepted: 27.11.2020.