

УДК 544.165:615.917

## ВИРТУАЛЬНЫЙ СКРИНИНГ И МОЛЕКУЛЯРНЫЙ ДИЗАЙН ПОТЕНЦИАЛЬНЫХ ИНГИБИТОРОВ SARS-COV-2

О.В. Тиньков<sup>1,2\*</sup>, В.Ю. Григорьев<sup>3</sup>, Л.Д. Григорьева<sup>4</sup>

(<sup>1</sup> Кафедра фармакологии и фармацевтической химии медицинского факультета Приднестровского государственного университета имени Т.Г. Шевченко, Тирасполь, Приднестровье, Молдова; <sup>2</sup> Военный институт Министерства обороны, Тирасполь, Приднестровье, Молдова; <sup>3</sup> Институт физиологически активных веществ РАН, Черноголовка, Россия; <sup>4</sup> Факультет фундаментальной физико-химической инженерии МГУ имени М.В. Ломоносова, Москва, Россия; \* e-mail: oleg.tinkov.chem@mail.ru)

Согласно недавно проведенным исследованиям, главная протеаза M<sup>pro</sup> вируса SARS-CoV-2, которая является важнейшей мишенью при разработке перспективных лекарственных средств для лечения COVID-19, эволюционно консервативна и не претерпела значительных изменений по сравнению с главной протеазой M<sup>pro</sup> вируса SARS-CoV. Многие исследователи отмечают высокую схожесть сайтов связывания главной протеазы M<sup>pro</sup> вирусов SARS-CoV и SARS-CoV-2, поэтому в условиях нарастающей эпидемии выглядят логичными дальнейшие исследования ингибиторов главной протеазы M<sup>pro</sup> вируса SARS-CoV для борьбы с COVID-19. В ходе исследования с применением симплексных, фрактальных, HYBOT-дескрипторов, методов частичных наименьших квадратов, случайного леса, опорных векторов, градиентного бустинга, а также интернет-платформы OCHEM (<https://ochem.eu>), в которой реализованы различные типы молекулярных дескрипторов и методы машинного обучения, построены удовлетворительные QSAR-модели. Для полученных моделей проведена структурная интерпретация, позволившая выделить молекулярные фрагменты, повышающие и понижающие активность ингибиторов SARS-CoV. Результаты структурной интерпретации использованы для рационального молекулярного дизайна потенциальных ингибиторов SARS-CoV-2. Полученные QSAR-модели использованы для виртуального скрининга 2087 одобренных FDA лекарственных средств.

**Ключевые слова:** протеаза M<sup>pro</sup>, QSAR, молекулярные дескрипторы, машинное обучение, структурная интерпретация.

В 2002 г. мировое сообщество столкнулось с коронавирусом SARS-CoV, который вызвал эпидемию атипичной пневмонии (тяжелый острый респираторный синдром). Первые случаи заражения коронавирусом SARS-CoV были отмечены в южном Китае, далее эпидемия распространилась на 29 стран, в результате чего было заражено более 8 тыс. человек, при этом 916 человек погибли [1]. В 2012 г. в Саудовской Аравии была зафиксирована вторая эпидемия, вызванная распространением коронавируса, в данном случае вируса MERS-CoV, для которого характерна более высокая летальность. К концу августа 2015 г. из общего числа зараженных (1511 пациентов) погибли 574 человека [2], т.е. летальность от коронавируса MERS-CoV составила более 37%, в отличие от SARS-CoV, для которого летальность оценивается примерно в 10%.

Ряд экспертов спрогнозировали возникновение угрозы для человечества, которая будет вызвана новым типом коронавируса [3]. Авторы настоящего

исследования предположили, что новый коронавирус, циркулирующий в популяциях китайских подковоносых летучих мышей, будет связываться с человеческим ангиотензинпревращающим ферментом 2 (angiotensin converting enzyme II, ACE2) с последующей эффективной репликацией в клетках дыхательной системы.

К сожалению, прогнозы вышеуказанных экспертов оказались верными и 8 декабря 2019 г. китайские органы здравоохранения зафиксировали первый случай заражения новым коронавирусом SARS-CoV-2 [4]. Уровень летальности от нового коронавируса SARS-CoV-2 по оценкам меньше, чем у SARS-CoV и MERS-CoV (примерно 7%), однако было показано, что для заражения первых 1000 пациентов коронавирусом SARS-CoV-2 необходимо всего 48 дней, в то время как для заражения SARS-CoV требуется 130 дней, для MERS-CoV – два с половиной года [5].

В настоящее время пандемия, вызванная новым коронавирусом SARS-CoV-2, представляет серьез-

ную медицинскую и социально-экономическую проблему для всего человечества.

Рекомендованный в Российской Федерации препарат «Фавипиравир» для лечения COVID-19 [6], по мнению ряда исследователей, имеет тератогенный эффект [7]. Таким образом, поиск и разработка высокоэффективных, безопасных лекарственных средств, способных остановить распространение пандемии COVID-19, является актуальной задачей.

Существенную помощь в снижении временных и финансовых затрат при репозиционировании и разработке новых лекарственных средств могут оказать методы хемоинформатики [8–10]. В этом направлении с момента появления нового коронавируса SARS-CoV-2 проведено достаточно много исследований с использованием методов молекулярного докинга, молекулярной динамики и фармакофорного анализа [11–22].

В настоящее время молекулярную структуру потенциальных ингибиторов SARS-CoV-2 рассматривают с разных точек зрения. Получаемая при этом разноплановая информация представляет значительный интерес для мирового научного и медицинского сообщества. В проведенных исследованиях наиболее значимой мишенью для разрабатываемых лекарственных средств служит главная протеаза вируса  $M^{pro}$ , также известная как 3-химотрипсиноподобная протеаза ( $3CL^{pro}$ ), которая играет ключевую роль в репликации коронавирусов. Установлено, что данный фермент, будучи эволюционно консервативным, не претерпел значительных изменений, в отличие от главной протеазы  $M^{pro}$  вируса SARS-CoV, который вызвал вспышку острого респираторного синдрома в 2002–2003 гг. [23–27]. Исследования, проведенные ранее, также указывают на консервативность последовательностей и пространственных структур  $M^{pro}$  разных видов коронавирусов [28]. При этом в организме человека близкие гомологи данного фермента не выявлены, что положительно влияет на специфичность и снижение числа потенциальных побочных эффектов ингибиторов главной протеазы  $M^{pro}$  [29].

Так, в работе [25] на основании схожести сайтов связывания главной протеазы  $M^{pro}$  вирусов SARS-CoV и SARS-CoV-2 с применением метода молекулярного докинга проведен виртуальный скрининг библиотеки химических соединений DrugBank [30]. В результате предложен список из десяти потенциальных ингибиторов главной протеазы  $M^{pro}$ , которые, по мнению авторов [25], наиболее перспективны для борьбы с SARS-CoV-2.

В исследовании [31] первоначально были отобраны соединения, входящие в список лекарственных средств традиционной китайской медицины. Для этих соединений дана оценка таких фармакокинетических характеристик, как адсорбция, распределение, метаболизм и выведение. Наиболее перспективные соединения изучены с помощью молекулярного докинга. Следующим этапом стал отбор лекарственных трав, которые содержат, как минимум, два соединения, предложенных в ходе молекулярного докинга. В результате проведенного исследования авторами [31] определены 26 лекарственных трав китайской медицины, потенциально перспективных для лечения заболевания COVID-19, вызванного коронавирусом SARS-CoV-2.

Заслуживает особого внимания публикация [23], в которой приведены результаты консенсусного *in vitro* и *in silico* скрининга. Авторы исследовали базу, представленную более чем 10 тыс. соединений, для которых экспериментально измеряли связывание с  $M^{pro}$  коронавируса SARS-CoV-2 методом флуоресцентного резонансного переноса энергии. Наиболее перспективными оказались семь соединений, которые дополнительно исследовали на способность предотвращать инфицирование клеток вирусом SARS-CoV-2. Скрининг *In silico* для этих 10 тыс. соединений был выполнен с помощью программ Glide v8.2 и Maestro (Schrödinger). Согласно консенсусным результатам, наиболее перспективными оказались эбселен (2-фенил-1,2-бензоселеназол-3-он, CAS номер 60940-34-3), кармофур (1-гексилкарбамоил-5-фторурацил, CAS номер 61422-45-5), соединение TDZD-8 (2-метил-4-(фенилметил)-1,2-адиазолидин-3,5-дион, CAS номер 327036-89-5), а также предложенный ранее ингибитор SARS-CoV – пепдитомиметик N3, который по реакции Михаэля ковалентно связывается с  $M^{pro}$ .

Авторы [32] провели виртуальный скрининг 1,3 млрд молекул в целях выявления наиболее активных ингибиторов главной протеазы  $M^{pro}$  вируса SARS-CoV-2. В данном исследовании был использован недавно разработанный алгоритм глубокого докинга (Deep Docking), который интегрировал классический докинг и методические основы построения моделей QSAR (количественное соотношение структура – активность; Quantitative Structure – Activity Relationship), что позволило увеличить производительность скрининга по сравнению с традиционными методами докинга. Авторы отмечают, что оценочные функции определяются используемыми методами докинга, а модели QSAR применяются для оптимизации

виртуального скрининга. По результатам виртуального скрининга предложен хит-лист из 1000 соединений, который доступен к свободному скачиванию по ссылке <https://drive.google.com/drive/folders/1xgA8ScPrqIunxEAXFrUEkavS7y3tLIMN>.

С использованием глубокого обучения проведено исследование [33], в котором разработаны модели, описывающие структуру соединений с помощью строк символов, составленных по правилам SMILES [34]. Принцип использованного метода прогнозирования основан на технологии под названием «Обработка естественного языка» [35], применяемой при анализе человеческой речи компьютером, только в данном случае в качестве языка выступает строка символов, записанная по правилам SMILES, и последовательность целевого протеина. Для выявления и анализа закономерностей использовали сверточные нейронные сети [36]. В ходе исследования была спрогнозирована активность ингибиторов главной протеазы  $M^{pro}$ , РНК репликазы (RNA-dependent RNA-polymerase, RdRP), хеликазы и ряда других ферментов вируса SARS-CoV-2. Для сравнительного исследования применяли молекулярный докинг (в частности, программу AutoDock Vina v.1.1.2). В результате авторы выделили три препарата против ВИЧ (ритонавир, атазанавир, эфавиренц), а также противовирусное средство ганцикловир.

Другой мишенью для борьбы с коронавирусом служит TMPRSS2-протеаза (Transmembrane protease, serine 2, мембрано-связанная сериновая протеаза), ингибиторы которой могут предотвращать проникновение вириона в клетку [37]. Однако исследований, посвященных компьютерному моделированию ингибиторов TMPRSS2-протеазы, значительно меньше, чем главной протеазы  $M^{pro}$ . Так, можно отметить публикацию [38], в которой с помощью молекулярного докинга и фармакофорного анализа был проведен виртуальный скрининг базы данных, содержащей более 30 тыс. природных соединений. Для отобранных 12 соединений авторы оценили адсорбцию, распределение, метаболизм, выведение и токсичность. Наиболее перспективным оказалось низкомолекулярное соединение генипозид (CAS номер 24512-63-8).

Таким образом, большинство работ выполнены с использованием метода молекулярного докинга, который, как любой метод исследования, имеет ограничения. В частности, методологические трудности докинга связаны с учетом конформаций лиганда, выбором методов построения оценочной функции и гибкостью рецепторов. Основные методы молекулярного докинга, а также присущие

им ограничения подробно освещены в обзорных публикациях [39–41].

Альтернативным методом компьютерной разработки лекарственных средств является QSAR, который успешно себя зарекомендовал при решении различных задач [42]. В литературе имеются сведения о схожести сайтов связывания главной протеазы  $M^{pro}$  вирусов SARS-CoV и SARS-CoV-2, подтвержденные в ходе независимых исследований [23–27]. В связи с этим мы предположили, что оценка потенциальных ингибиторов SARS-CoV-2 при разработке лекарственных средств может быть реализована с помощью QSAR-моделей ингибиторов SARS-CoV.

Авторы [43] разработали QSAR-модель ингибиторов SARS-CoV с использованием методов 3D-QSAR (CoMFA, CoMSIA), ограничительной особенностью которых является неоднозначность трехмерного выравнивания структур исследуемых соединений [44]. В другом исследовании [45] были разработаны 33 QSAR-модели ингибиторов SARS-CoV, однако авторами не приведены показатели прогнозирующей способности, оцененные с помощью соединений тестовой выборки.

В соответствии с пятым принципом QSAR-моделирования, разработанным экспертной группой OECD [46], желательна интерпретация полученных моделей. В рассмотренных публикациях [43, 45] отсутствует структурная интерпретация QSAR-моделей, что не позволяет проводить молекулярный дизайн и ограничивает применение результатов моделирования для исследования механизмов биологических реакций [47].

Недавно опубликовано исследование [48], в котором в рамках поиска эффективных лекарственных средств против COVID-19 предложены приемлемые QSAR-модели ингибиторов главной протеазы  $M^{pro}$  вируса SARS-CoV, которые разработаны с помощью PaDEL- и Dragon-дескрипторов, а также метода множественной линейной регрессии (Multiple Linear Regression, MLR). Используя разработанные QSAR-модели, авторы провели виртуальный скрининг более 50 тыс. различных соединений в целях выявления наиболее активных ингибиторов главной протеазы  $M^{pro}$  вируса. Исходя из предложенных регрессионных уравнений, а именно вкладов некоторых значимых дескрипторов, авторы [48] проанализировали влияние структурных особенностей изучаемых соединений на изменение ингибирующей активности.

В работе [49] для 54 пептидомиметиков разработаны адекватные QSAR-модели ингибито-

ров SARS-CoV, в ходе структурной интерпретации которых определены молекулярные фрагменты, понижающие и повышающие данный вид активности. Для построения моделей авторы [49] также использовали метод MLR.

Примененный в [48, 49] регрессионный метод анализа данных MLR может давать адекватные результаты только при наличии линейной связи между структурой и активностью [50]. Одним из способов преодоления вышеуказанного недостатка может быть использование непараметрических методов, в частности различных методов машинного обучения (machine learning, ML).

Особого внимания заслуживает исследование [27], в ходе которого разработаны высококачественные классификационные QSAR-модели ингибиторов SARS-CoV. Достоверность сконструированных QSAR-моделей была подтверждена последующими экспериментальными исследованиями, в результате которых некоторые соединения продемонстрировали высокую активность и были рекомендованы для дальнейшего изучения. Параллельно QSAR-анализу авторы [27] проводили исследование методами молекулярного докинга, при этом выявленный неприемлемо низкий уровень предсказательной способности не позволил принимать во внимание результаты изучения ингибиторов SARS-CoV методами молекулярного докинга. Несомненно, работа [27] очень успешна, однако в ней отсутствует структурная интерпретация QSAR-моделей.

Настоящее исследование состояло из следующих этапов:

- 1) построение QSAR-моделей ингибиторов главной протеазы M<sup>pro</sup> SARS-CoV;
- 2) проведение виртуального скрининга наиболее перспективных соединений – потенциальных лекарственных средств для лечения COVID-19;

3) структурная интерпретация QSAR-моделей и рациональный молекулярный дизайн ингибиторов главной протеазы M<sup>pro</sup>.

### Методы исследований

Источником формирования выборки для QSAR-моделирования послужила общеизвестная база данных ChEMBL (ID: ChEMBL3927) [51]. Из полученной выборки были удалены неорганические соединения, полимеры, смеси, соединения в солевой форме. Окончательный набор ингибиторов главной протеазы M<sup>pro</sup> SARS-CoV содержал 65 соединений.

Приведенные в первоисточниках экспериментальные значения активности ингибиторов главной протеазы M<sup>pro</sup> SARS-CoV, выраженные с помощью концентрации полумаксимального ингибирования (IC<sub>50</sub>, nM или μM), были сконвертированы (1) в отрицательный десятичный логарифм величины pIC<sub>50</sub>, который общепринят в QSAR-исследованиях и используется в тех случаях, когда линейный рост концентрации вызывает экспоненциальный рост эффекта:

$$pIC_{50} = -\log_{10}(IC_{50}). \quad (1)$$

Экспортированные экспериментальные значения и структуры соединений приведены в Приложении (табл. 1 П).

Для описания молекулярной структуры использовали 2D-симплексные дескрипторы, которые рассчитываются в рамках симплексного представления молекулярной структуры (СПМС, Simplex representation of molecular structure, SIRMS) [52]. В рамках SIRMS молекула рассматривается как система различных симплексов –

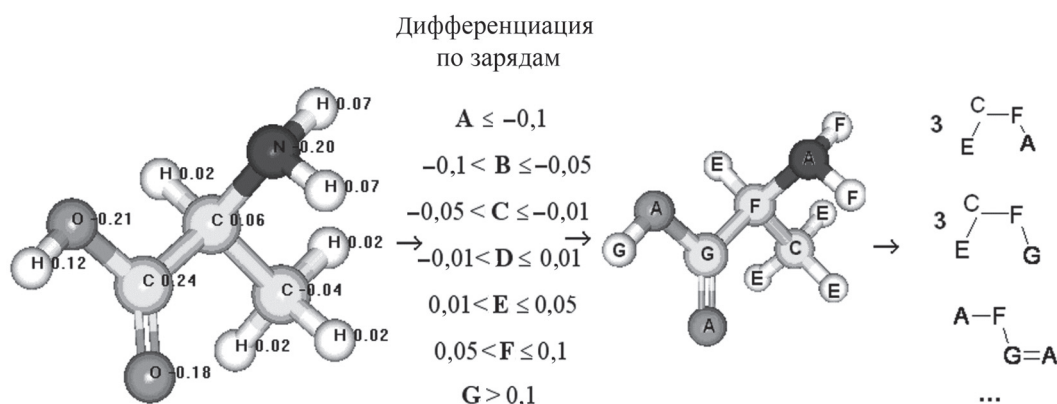


Рис. 1. Пример 2D-генерации симплексных дескрипторов для аланина на 2D-уровне при использовании дифференциации атомов по их парциальным зарядам

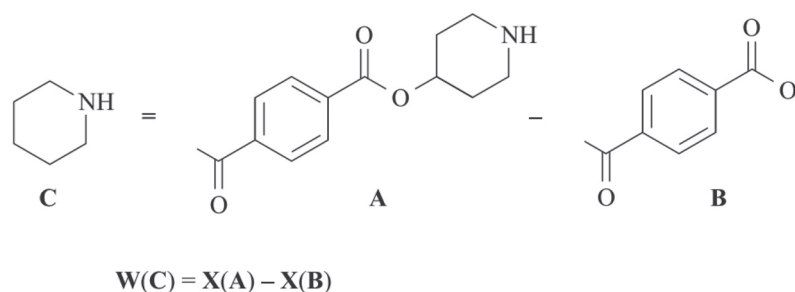


Рис. 2. Использованный принцип структурной интерпретации.  $W(C)$  – вклад фрагмента (С),  $X(A)$  – предсказанное значение активности родительской структуры (А);  $X(B)$  – предсказанное значение активности для гипотетической структуры (В)

четырёхатомных молекулярных фрагментов фиксированной структуры (рис. 1).

Дескриптором в данном случае служит число симплексов определенного вида. На 2D-уровне атомы (вершины симплекса) дифференцируют не только по природе «метке» атома, но и с учетом различных физико-химических свойств (частичного заряда на атоме, липофильности, рефракции, способности выступать донором или акцептором водорода при образовании водородной связи).

Структурную интерпретацию проводили в соответствии с подходом [53], при котором вклад исследуемого фрагмента (С) рассчитывали путем нахождения разницы между рассчитанными значениями активности для родительской структуры (А) и гипотетической структуры (В), полученной удалением исследуемого фрагмента (С) из родительской структуры (А) (рис. 2).

При моделировании с использованием симплексных дескрипторов применяли пакет «Scikit-learn» [54] для языка программирования Python, в котором реализованы методы частичных наименьших квадратов (Partial Least Squares, PLS), случайного леса (Random Forest, RF), опорных векторов (Support Vector Machine, SVM), градиентного бустинга (Gradient Boosting Method, GBM).

Ввиду небольшого числа изучаемых соединений и их структурного разнообразия была проведена пятикратная внутренняя кросс-валидация (CV, 5-fold cross validation). Для этого все соединения обучающей выборки случайным образом разделяются на пять частей. Затем на четырех частях данных, объединенных в обучающую выборку, производят построение (обучение) QSAR-моделей, а оставшуюся часть данных используют как внешнюю тестовую выборку, т.е. на соединениях этой группы проверяют предсказательную способность модели. Эта процедура повторяется 5 раз, в результате чего каждая из пяти частей данных последо-

вательно используется для тестирования. Важно подчеркнуть, что исследуемые соединения никогда не являются одновременно частью как обучающей, так и внешней тестовой выборки.

При QSAR-моделировании для тестовых выборок учитывали вхождение соединений в область применимости (applicability domain, AD) [55], при этом если значение хотя бы одного дескриптора выходило за пределы его минимального или максимального значения для обучающей выборки, то соединение тестовой выборки, содержащее этот дескриптор, не входило в область применимости (bounding box). Вышеописанный подход для QSAR-моделирования с использованием симплексных дескрипторов реализован в виде программного обеспечения «SPCI», которое находится в свободном доступе по адресу [http://qsar4u.com/pages/sirms\\_qsar.php](http://qsar4u.com/pages/sirms_qsar.php).

Помимо этого, для QSAR-анализа мы использовали интернет-платформу «OCHEM» (<https://ochem.eu>). Наилучшие результаты при моделировании были достигнуты с использованием ряда дескрипторов (ALogPS, OEstate, Dragon, CDK, ISIDA Fragments, StructuralAlerts, alvaDesc), а также методов случайного леса и ассоциативных нейронных сетей (Associative Neural Networks, ASNN). Построение консенсусной модели проводили путем усреднения прогнозов лучших индивидуальных моделей. При этом область применимости оценивалась с помощью концепции «расстояние до модели» (в частности, подхода «CLASS-LAG»). Краткое описание использованных методов, дескрипторов, а также ссылки на оригинальные работы приведены в руководстве пользователя «OCHEM» [56].

В интернет-ресурсе «OCHEM» реализован метод молекулярных пар [57], который также позволяет проводить интерпретацию моделей, построенных на любых дескрипторах.

Оценку точности и прогностической способности предложенных в работе моделей и их сравнение с другими QSAR-моделями выполняли на основе следующих критериев.

1. Коэффициента детерминации ( $R^2$ ):

$$R^2 = 1 - \frac{\sum_{i=1}^m (y_i - \hat{y}_i)^2}{\sum_{i=1}^m (y_i - y_{i\text{mean}})^2}, \quad (2)$$

где  $\hat{y}_i$  – вычисленное значение свойства для  $i$ -й молекулы,  $y_i$  – наблюдаемое (экспериментальное) значение свойства для  $i$ -й молекулы,  $m$  – число молекул в выборке,  $y_{i\text{mean}}$  – среднее значение наблюдаемого свойства.

2. Среднеквадратичной ошибки (RMSE):

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^m (y_i - \hat{y}_i)^2}{m}}. \quad (3)$$

С учетом того, что рассмотренное выше QSAR-моделирование требует привлечения большого числа дескрипторов (сотни и тысячи), для сравнения были сконструированы с применением метода RF [58] регрессионные модели, содержащие небольшое число HYBOT-переменных [59] с добавлением ряда топологических и фрактальных

характеристик [60]. В этом случае оценку AD осуществляли с помощью интервального метода.

Для виртуального скрининга мы использовали базу данных DrugBank [30], представленную одобренными FDA лекарственными средствами. Из экспортированной базы данных DrugBank были удалены неорганические соединения, полимеры, смеси и соединения в солевой форме. Окончательная выборка для виртуального скрининга содержала 2087 одобренных FDA лекарственных средств.

## Результаты и обсуждение

Результаты QSAR-моделирования приведены в табл. 1–3. Консенсусная модель находится в свободном доступе по ссылке, приведенной в табл. 1. Все сконструированные модели имеют удовлетворительные статистические характеристики и обладают сопоставимой предсказательной способностью.

Для полученных консенсусных QSAR-моделей (табл. 1, 2) была проведена структурная интерпретация. При интерпретации консенсусной модели, построенной с помощью симплексных дескрипторов (табл. 2), были определены вклады молекулярных фрагментов в величину активности ингибиторов главной протеазы M<sup>pro</sup> SARS-CoV (рис. 3).

Т а б л и ц а 1

Статистические характеристики QSAR-моделей, разработанных с помощью интернет-ресурса «OCHEM»

Метод	Дескриптор	$R^2_{cv}$	$\text{RMSE}_{cv}$
ASNN	ISIDA Fragments	0,67	0,50
	ALogPS, OEstate	0,68	0,49
	Dragon	0,66	0,50
	CDK	0,60	0,53
	alvaDesc	0,65	0,51
RF	StructuralAlerts	0,63	0,52
	ISIDA Fragments	0,67	0,49
Консенсусная модель ( <a href="https://ochem.eu/model/43078789">https://ochem.eu/model/43078789</a> )		0,70	0,47

Т а б л и ц а 2

Статистические характеристики QSAR-моделей, разработанных с использованием 2D-симплексных дескрипторов

Метод	Дескрипторы	$R^2_{cv}$	RMSE <sub>cv</sub>
GBM	SIRMS	0,57	0,57
RF		0,65	0,51
SVM		0,52	0,60
PLS		0,64	0,52
Консенсусная модель		0,64	0,51

Т а б л и ц а 3

Статистические характеристики QSAR-модели, разработанной с использованием HUBOT, топологических и фрактальных дескрипторов

Метод	Дескриптор	$R^2_{cv}$	RMSE <sub>cv</sub>
RF	MaxE <sub>a</sub> ; MaxC <sub>a</sub> ; Nv2; IC0; D <sub>unb</sub> *	0,62	0,53

Примечание. MaxE<sub>a</sub> – максимальный энтальпийный H-акцепторный дескриптор; MaxC<sub>a</sub> – максимальный свободноэнергетический H-акцепторный дескриптор; Nv2 – число вершин со степенью 2; IC0 – среднее информационное содержание 0-го порядка; D<sub>unb</sub>\* – фрактальная плотность несвязанных атомов. Использованные дескрипторы рассмотрены в работах [59, 60].

Данный набор молекулярных фрагментов сформирован из стандартных функциональных групп (51 фрагмент) и шести молекулярных фрагментов, полученных в ходе автоматической фрагментации соединений обучающей выборки с помощью программы «SPCI» при использовании SMART-шаблона [#6+0;!\$(\*=#,#[!#6])!@!=#[\*], кодирующего разрываемые связи [53]. Интерпретации подверглись только те молекулярные фрагменты, которые были обнаружены в трех и более соединениях, что, с нашей точки зрения, позволило сконцентрировать внимание на фрагментах, устойчиво влияющих на ингибирующую активность и избежать в некоторой степени влияния случайных факторов, например ошибок в экспериментальных данных или предсказанных значениях активности и вкладах фрагментов. Полный список идентифицированных молекулярных фрагментов, записанных в виде SMARTS, с рассчитанными средними вкладами в активность приведен (Приложение, табл. 2 П).

Проведенная интерпретация позволила количественно описать и ранжировать влияние моле-

кулярных фрагментов на изменение активности ингибиторов M<sup>pro</sup> SARS-CoV, детализировать молекулярное окружение известных функциональных групп, выделив производные фрагменты, повышающие и понижающие вклады в указанный вид активности. Например, при детализации пиримидина выделен молекулярный фрагмент 2-сульфанилпиримидин-4-ол (**f9** на рис. 3), который в значительной степени уменьшает активность ингибиторов M<sup>pro</sup> SARS-CoV. При этом карбоксильные производные фурана и пиридина (фрагменты **f1** и **f2** на рис. 3), наоборот, увеличивают активность ингибиторов M<sup>pro</sup> SARS-CoV.

Интерпретация была также проведена для консенсусной модели (табл. 1), построенной с помощью интернет-ресурса «OCHEM». В табл. 4 приведены результаты интерпретации, согласно которым при замене атомов водорода на хлор или метильную группу возрастает ингибирующая активность. Рост активности ингибиторов M<sup>pro</sup> SARS-CoV наблюдается также при замене фенольных, *n*-пропильных радикалов на нафтиловые. Результаты интерпретаций рассмотренных

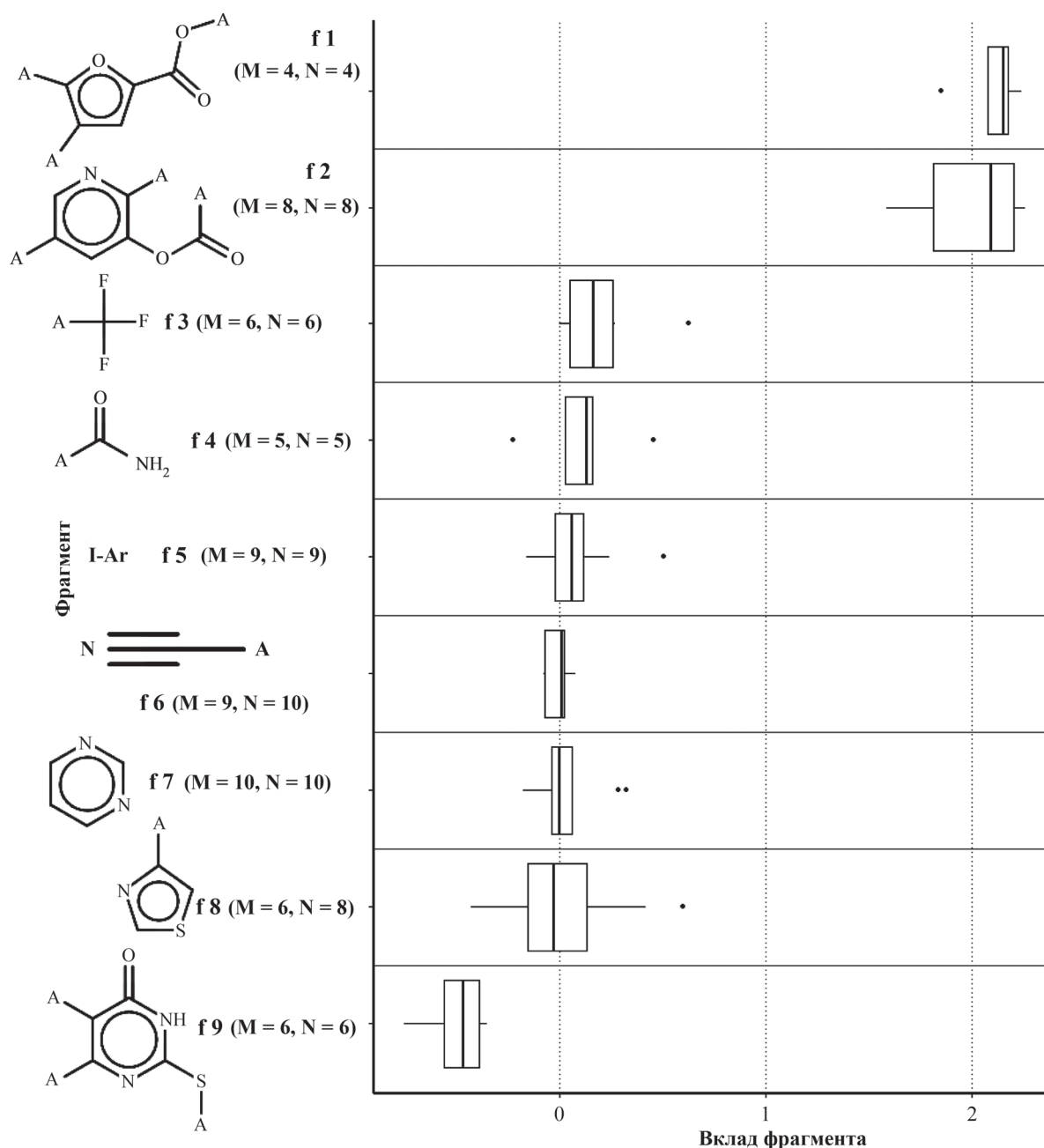


Рис. 3. Вклады молекулярных фрагментов в способность соединений ингибировать главную протеазу  $M^{\text{pro}}$  SARS-CoV. А – место присоединения фрагмента к остальной части молекулы, f – порядковый номер фрагмента, М – количество соединений, содержащих данный фрагмент, N – количество обнаружений соответствующего фрагмента в выборке

выше консенсусных моделей согласованно указывают на повышение активности ингибиторов  $M^{\text{pro}}$  SARS-CoV при замене фрагментов, содержащих иод (f5), на карбамоил группу (f4).

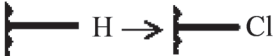
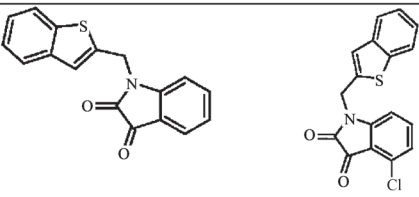

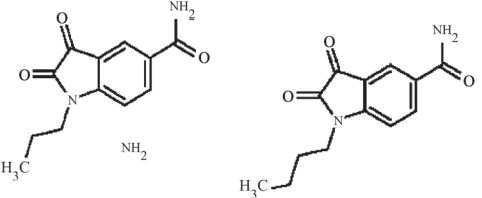
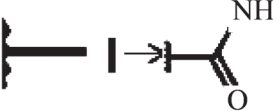
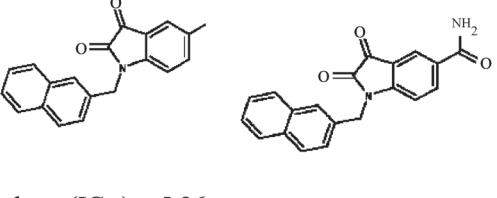
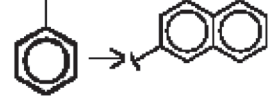
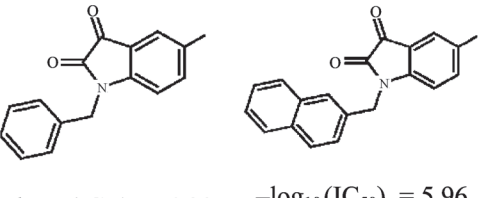
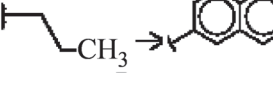
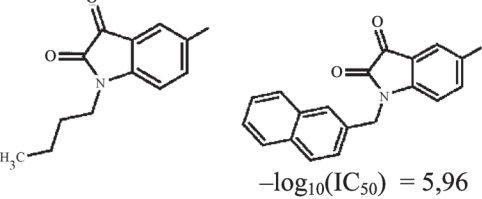
С учетом тенденций влияния на изменение активности структуры соединений, выявленных в ходе интерпретации, был проведен рациональный молекулярный дизайн и предложен ряд перспективных агентов против COVID-19. При этом была выполнена замена молекулярных фрагментов, уменьшающих активность, на фрагменты, которые, согласно результатам интерпретации, повышают активность ингибито-

ров  $M^{\text{pro}}$ . В результате были предложены гипотетические соединения (табл. 5, вещества 2, 4, 6), обладающие значительной рассчитанной ингибирующей активностью и входящие в область применимости консенсусной QSAR-модели, разработанной с помощью «OCHEM». Например, при замене остатка 6-метил-2-сульфанилпиридин-4-ола (соединение 1, табл. 5) на карбоксильное производное пиридина (соединение 2, табл. 5) отмечается значительный рост активности ингибиторов главной протеазы  $M^{\text{pro}}$  SARS-CoV. Также характерно увеличение активности при замене фрагмента, содержащего нитрил (соединение 3)



Т а б л и ц а 4

## Результаты интерпретации консенсусной модели, разработанной с помощью интернет-ресурса «ОСЕМ»

Молекулярная трансформация	Пары соединений и их экспериментальные характеристики	N	$\Delta_{\text{mean}}$	SMIRKS
 H → Cl	 $-\log_{10}(\text{IC}_{50}) = 4,88$ $-\log_{10}(\text{IC}_{50}) = 4,95$	2	$0,035 \pm 0,049$	*[H] → *Cl
 H → CH <sub>3</sub>	 $-\log_{10}(\text{IC}_{50}) = 4,60$ $\log_{10}(\text{IC}_{50}) = 4,72$	2	$0,06 \pm 0,085$	*[H] → C*
 H → NH <sub>2</sub>	 $-\log_{10}(\text{IC}_{50}) = 5,96$ $-\log_{10}(\text{IC}_{50}) = 6,43$	3	$0,54 \pm 0,065$	*I → NC(*)=O
	 $-\log_{10}(\text{IC}_{50}) = 4,30$ $-\log_{10}(\text{IC}_{50}) = 5,96$	2	$1,6 \pm 0,092$	*c1cccc1 → →*c1ccc2ccccc2 c1
	 $-\log_{10}(\text{IC}_{50}) = 4,18$ $-\log_{10}(\text{IC}_{50}) = 5,96$	2	$1,7 \pm 0,049$	CCC* → →*c1ccc2ccccc2 c1

Пр и м е ч а н и я. N – Количество молекулярных пар, для которых выполняется молекулярная трансформация;  $\Delta_{\text{mean}}$  – средняя разница значений  $-\log_{10}(\text{IC}_{50})$  при проведении молекулярной трансформации; SMIRKS – формат записи молекулярной трансформации (<https://www.daylight.com/dayhtml/doc/theory/theory.smirks.html>).

на трифторметил (соединение 4) или же остатка 4-(1,3-тиазол-4-ил)пиримидин-2-тиола (соединение 5, таблица 5) на упомянутое выше карбоксильное производное пиридина (соединение 6). Необходимо отметить, что при молекулярном дизайне в данных примерах учитывались результаты интерпретации QSAR-модели, построенной с помощью симплексных дескрипторов, в то время как прогнозирование активности ингибиторов проводили с помощью QSAR-модели, построенной интернет-ресурсом «OCHEM».

При определении стратегии синтеза и испытаний помимо целевого свойства (активности) важно оценивать различные виды токсичности и липофильность, которые являются существенными факторами при принятии решения о рекомендации к использованию соединения в качестве действующего вещества лекарственного средства. В этих целях для соединений 1–9 была проведена оценка острой токсичности ( $LD_{50}$ ) при пероральном введении крысам и вероятности проявления мутагенности (тест Эймса) с помощью программы «T.E.S.T. v.4.2.», разработанной специалистами Агентства по охране окружающей среды США [61]. Также для данных соединений с помощью интернет-платформы швейцарского Института биоинформатики «swissADME» (<http://www.swissadme.ch/>) [62] оценивали липофильность ( $\text{Log Po/w}$ ), соответствие правилам Липински [63], присутствие PAINS фрагментов [64], синтетическую доступность по десятибалльной шкале (0 – максимальная степень синтетической доступности, 10 – минимальная степень синтетической доступности) [65], что крайне важно для предложенных, но еще не синтезированных соединений. Результаты прогнозирования приведены в табл. 6, из которой видно, что предложенные в ходе молекулярного дизайна соединения 2, 4, 6, 8, 9 обладают сопоставимой синтетической доступностью по сравнению с уже синтезированными веществами 1, 3, 5, 7. За исключением соединения 4 все вещества удовлетворяют правилам пяти Липински и не содержат PAINS-фрагментов. Для изучаемых соединений характерен широкий диапазон значений липофильности, что должно учитываться при исследовании фармакокинетики и выборе лекарственных форм. Согласно проведенным расчетам, соединение 7 обладает ненулевой вероятностью проявления мутагенности, что может снизить его привлекательность как соединения-лидера, даже несмотря на то, что оно обладает максимальной экспериментально измеренной ингибирующей активностью (табл. 5) среди соединений экспорти-

рованной выборки из базы данных ChEMBL (ID: ChEMBL3927). При модификации соединения 7 предложено гипотетическое соединение 9, которое согласно расчетным данным не обладает вероятностью проявления мутагенности, при этом показатели  $LD_{50}$  и ингибирующей активности сопоставимы с исходным соединением 7. Кроме того, модификация соединения 1 в соединение 2 позволила не только увеличить ингибирующую активность почти на два порядка, достигнув сопоставимых значений с наиболее активными веществами в экспортируемой выборке, но и уменьшить токсичность ( $LD_{50}$ ) почти в 2,7 раза. Таким образом, к синтезу и дальнейшим испытаниям может быть рекомендовано соединение 2.

Поскольку синтез новых соединений и их клинические испытания занимают достаточно много времени, важнейшим средством борьбы с новой, быстро распространяющейся пандемией является репозиционирование уже одобренных лекарственных средств, которые прошли все необходимые клинические исследования. В целях выявления перспективных ингибиторов главной протеазы  $M^{pro}$  SARS-CoV-2 был проведен скрининг 2087 одобренных FDA лекарственных средств. Для скрининга использовали консенсусную модель, построенную экспертной системой «OCHEM», поскольку она обладает более высокими статистическими характеристиками и может быть использована всеми заинтересованными лицами для виртуального скрининга собственных наборов соединений. QSAR-модель, полученную с помощью симплексных дескрипторов не использовали, ввиду особенностей вышеописанного метода определения области применимости, сильно ограничивающего структурное пространство модели.

В ходе проведения виртуального скрининга предложено 10 соединений (табл. 7), входящих в область применимости консенсусной QSAR-модели, разработанной с помощью «OCHEM» и обладающих наибольшей рассчитанной ингибирующей активностью. Среди данных соединений встречаются противоопухолевые, антипротозойные, противогрибковые, антибактериальные, антигипертензивные и противовоспалительные лекарственные средства. Особое внимание привлекают Тилброхинол, Хлороксин, Хлорхинальдол, которые являются галогенпроизводными хинолина, как и широко известные хлорохин и гидроксихлорохин, которые уже применяются при лечении COVID-19 [66].

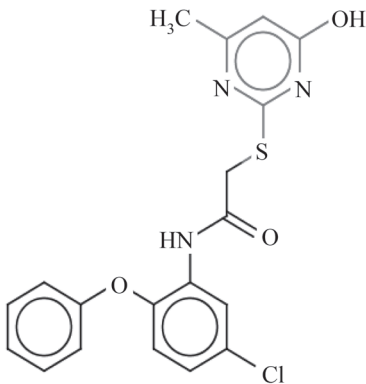
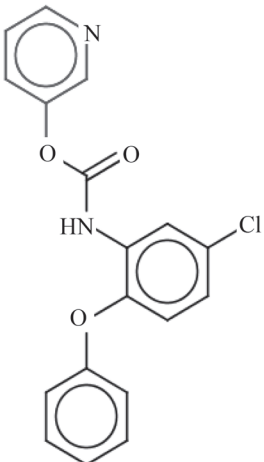
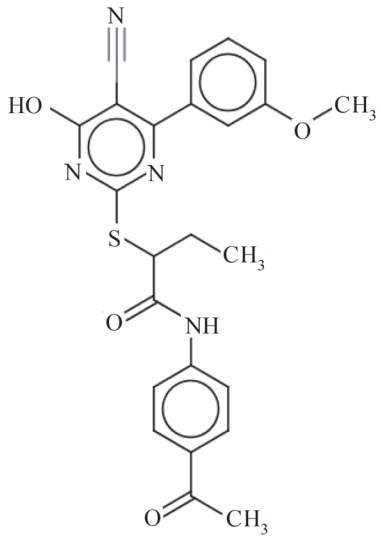
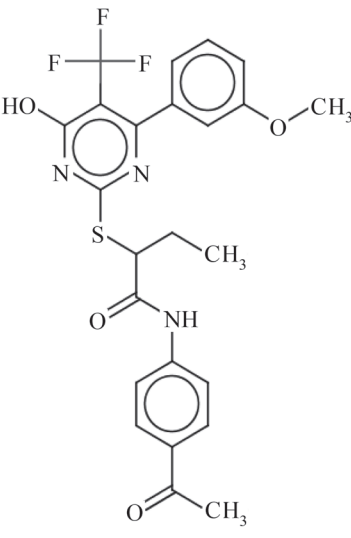
Согласно данным виртуального скрининга, наибольшей ингибирующей активностью сре-

ди одобренных FDA лекарственных средств обладает Пексидартиниб, который является противоопухолевым средством, ингибитором тирозинкиназы. В исследовании [12], проведенном с использованием молекулярного докинга и молекулярной динамики, противоопухолевое средство Нератиниб (Neratinib), блокирующее

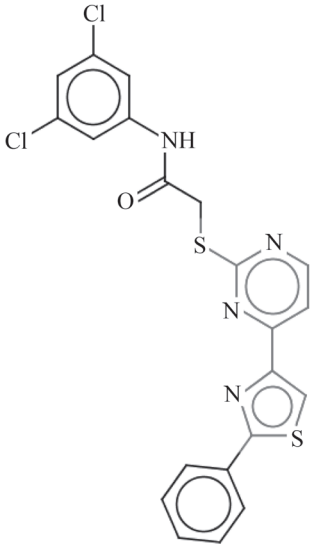
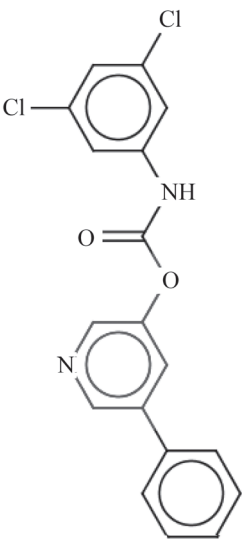
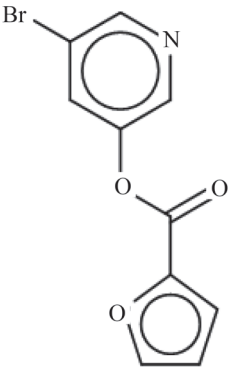
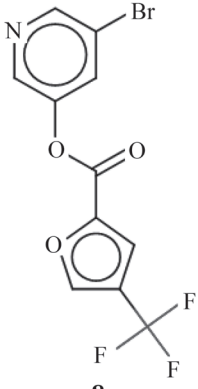
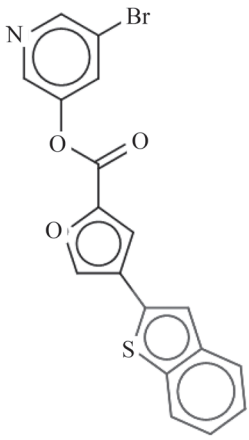
работу рецепторных тирозинкиназ, также предложено в качестве перспективного ингибитора главной протеазы M<sup>pro</sup> SARS-CoV-2. Выводы авторов [12] основаны на предположении об аналогичном связывании данного противоопухолевого средства с остатком цистеина в активных центрах киназных доменов рецеп-

Таблица 5

## Результаты молекулярного дизайна

Соединения выборки	Экспериментальные значения, $-\log_{10}(IC_{50})$	Гипотетические соединения	Предсказанные значения, $-\log_{10}(IC_{50})$
 <p style="text-align: center;"><b>1</b></p>	4,00	 <p style="text-align: center;"><b>2</b></p>	5,93
 <p style="text-align: center;"><b>3</b></p>	4,22	 <p style="text-align: center;"><b>4</b></p>	4,34

Окончание табл. 5

Соединения выборки	Экспериментальные значения, $-\log_{10}(IC_{50})$	Гипотетические соединения	Предсказанные значения, $-\log_{10}(IC_{50})$
 <p><b>5</b></p>	<p>5,52</p>	 <p><b>6</b></p>	<p>6,19</p>
 <p><b>7</b></p>	<p>7,3</p>	 <p><b>8</b></p>	<p>6,90</p>
		 <p><b>9</b></p>	<p>6,94</p>

Т а б л и ц а 6

Оценка некоторых физико-химических свойств, видов токсичности и синтетической доступности соединений, исследуемых в рамках молекулярного дизайна (структуры соединений представлены в табл. 5)

Соединение	Log Po/w	Число несоответствий правилам Липински	Число фрагментов PAINS	Синтетическая доступность	LD50 крысы при пероральном введении, мг/кг	Вероятность проявления мутагенности**
1	3,75	0	0	2,86	1382,28	0
2*	3,71	0	0	2,94	3788,68	0
3	3,35	0	0	3,76	956,51	0
4*	4,45	1 (Молекулярная масса свыше 500)	0	3,89	–	0
5	4,98	0	0	3,33	663,22	0
6*	4,33	0	0	2,78	772,67	0
7	2,18	0	0	2,5	581,59	1
8*	3,28	0	0	2,81	193,04	0
9*	4,58	0	0	3,26	561,22	0

\*Гипотетические соединения, предложенные в результате молекулярного дизайна; \*\* 0 – отрицательная, 1 – положительная; «←» соединение не входит в область применимости QSAR модели.

торных тирозинкиназ и главной протеазы M<sup>pro</sup> SARS-CoV-2. Другое противоопухолевое средство кармофур по результатам высокопроизводительного скрининга в упомянутом выше исследовании [23] также было выделено в качестве перспективного ингибитора главной протеазы M<sup>pro</sup> SARS-CoV-2. Проведенные недавно дополнительные исследования с использованием метода рентгеноструктурного анализа [67] описывают механизм ингибирования главной протеазы M<sup>pro</sup> SARS-CoV-2 кармофуrom через ковалентное связывание с остатком цистеина Cys145 в активном центре. На основании вышесказанного логичным выглядит предложение о перепрофилировании выявленного в ходе виртуального скрининга пексидартиниба для лечения COVID-19. Необходимо отметить, что подтверждение эффективности в борьбе с COVID-19 предложенных для перепрофилирования лекарственных средств требует значительных дополнительных экспериментальных исследований. Прием лекарственных средств должен проводиться только по назначению лечащего врача.

Таким образом, в ходе проведения вычислительных экспериментов с применением концептуально различающихся дескрипторов и методов машинного обучения были разработаны приемлемые QSAR-модели ингибиторов главной протеазы M<sup>pro</sup>.

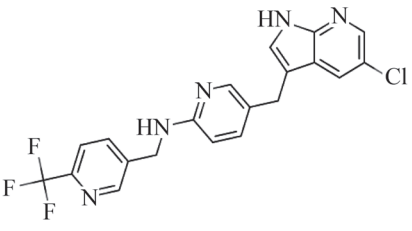
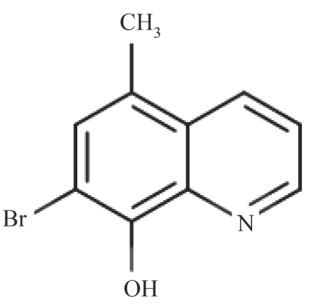
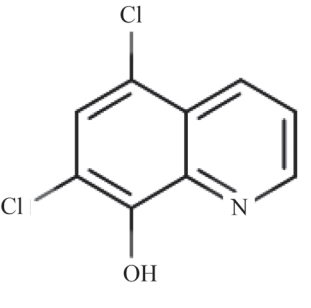
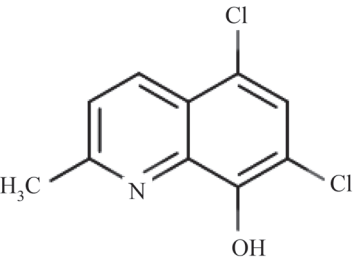
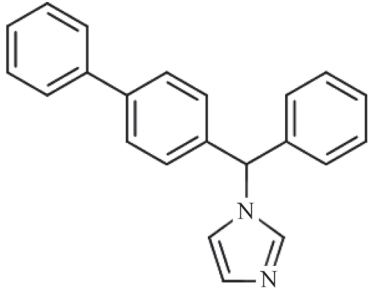
Проведенная структурная интерпретация QSAR-моделей позволила выявить общие закономерности влияния структуры химических соединений на их ингибирующую активность путем выделения молекулярных фрагментов и трансформаций, повышающих и понижающих активность ингибиторов SARS-COV. Результаты структурной интерпретации использованы при проведении рационального молекулярного дизайна, в ходе которого предложен ряд перспективных соединений для борьбы с COVID-19.

В процессе виртуального скрининга одобренных FDA лекарственных средств выделено 10 веществ, которые могут быть рекомендованы для перепрофилирования в качестве лекарств против новой коронавирусной инфекции.

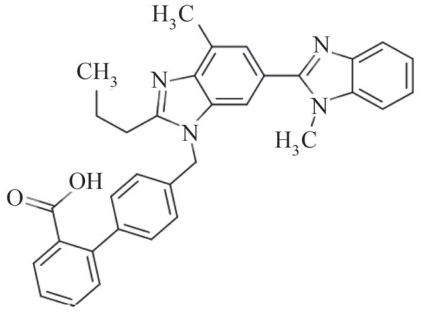
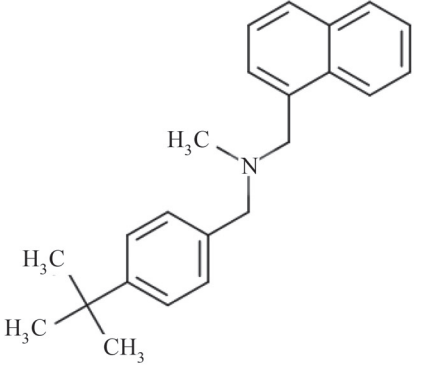
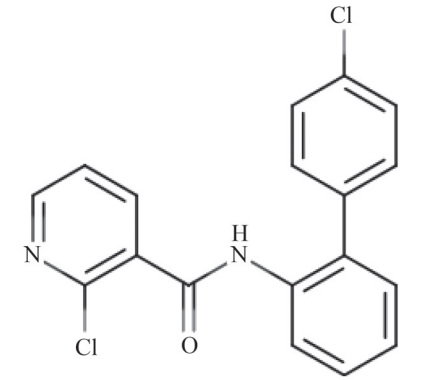
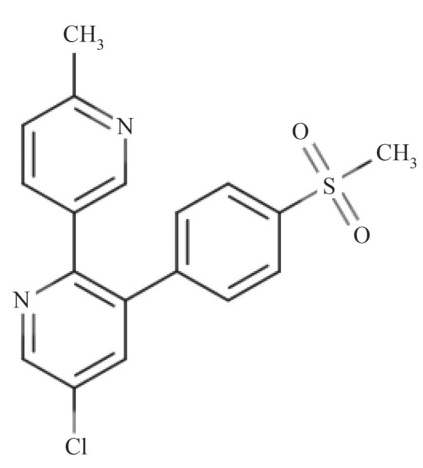
Результаты проведенного исследования могут способствовать снижению финансовых,

Т а б л и ц а 7

Одобрены FDA лекарственные средства, которые согласно результатам виртуального скрининга наиболее перспективны для ингибирования репликации SARS-CoV-2

$-\log_{10}(IC_{50})$	Название	Химическая структура	Описание
6,09	Пексидартиниб (Pexidartinib)		противоопухолевое средство, ингибитор тирозинкиназы
5,66	Тилброхинол (Tilbroquinol)		антипротозойное средство, эффективное против амебиаза, препарат также был использован против холерного вибриона
5,61	Хлороксин (Chloroxine)		лекарственные средства, обладающие бактериостатическими, фунгистатическими и антипротозойными свойствами
5,6	Хлорхинальдол (Chlorquinaldol)		
5,6	Бифоназол (Bifonazole)		противогрибковый препарат

Продолжение табл. 7

$-\log_{10}(IC_{50})$	Название	Химическая структура	Описание
5,59	Телмисартан (Telmisartan)		антигипертензивное средство, антагонист рецепторов ангиотензина II
5,57	Бутенафин (Butenafine)		синтетический противогрибковый бензиламин
5,54	Боскалид (Boscalid)		обладает фунгицидными свойствами
5,53	Эторикоксиб (Etoricoxib)		противовоспалительное, анальгезирующее средство, являющиеся селективным ингибитор циклооксигеназы-2

Окончание табл. 7

$-\log_{10}(\text{IC}_{50})$	Название	Химическая структура	Описание
5,51	Рофлумиласт (Roflumilast)		противовоспалительное средство, представляющее ингибитор фосфодиэстеразы-4 (ФДЭ4)

временных и трудовых затрат при определении стратегии разработки новых лекарственных средств, а также при репозиционировании существующих лекарственных средств – ингибиторов SARS-CoV-2.

Часть работы выполнена в рамках Государственного задания ИФАВ РАН 2020 года (тема № 0090-2020-0004).

Конфликта интересов нет.

Имеются дополнительные материалы.

## СПИСОК ЛИТЕРАТУРЫ

1. Enserink M. // *Science*. 2013. Vol. 339. N 6125. P. 1266 (doi:10.1126/science.339.6125.1266).
2. WHO. Middle East respiratory syndrome coronavirus (MERS-CoV) — Republic of Korea. Global Alert and Response (GAR). Available from: <https://www.who.int/csr/don/01-june-2015-mers-korea/en/> (accessed: 28.09.2020).
3. Menachery V.D., Yount B.L., Jr, Debbink K., Agnihothram S., Gralinski L.E., Plante J.A., Graham R.L., Scobey T., Ge X.Y., Donaldson E.F., Randell S.H., Lanzavecchia A., Marasco W.A., Shi Z.L., Baric R.S. // *Nature medicine*. 2015. Vol. 21. N 12. P. 1508 (doi:10.1038/nm.3985)
4. Chen N., Zhou M., Dong X., Qu J., Gong F., Han Y., Qiu Y., Wang J., Liu Y., Wei Y., Xia J., Yu T., Zhang X., Zhang L. // *Lancet*. 2020. Vol. 395. N 10223. P. 507 (doi: 10.1016/S0140-6736(20)30211-7).
5. Comparing the Wuhan coronavirus outbreak with SARS and MERS. Available from: <https://graphics.reuters.com/CHINA-HEALTH-VIRUS-COMPARISON/0100B5BY3CY/index.html> (accessed: 28.09.2020).
6. Минздрав России зарегистрировал первый препарат от коронавируса. Сообщение на официальном сайте Министерства здравоохранения РФ от 01 июня 2020 г. Available from: <https://www.rosminzdrav.ru/news/2020/06/01/14086-minzdrav-rossii-zaregistririval-pervyy-preparat-ot-koronavirusa> (accessed: 28.09.2020).
7. Shiraki K., Daikoku T. // *Pharmacology & therapeutics*. 2020. Vol. 209. P. 107512 (doi:10.1016/j.pharmthera.2020.107512).
8. Xu J., Hagler A. // *Molecules*. 2002. Vol. 7. N 8. P. 566 (doi:10.3390/70800566).
9. Lo Y.C., Rensi S.E., Torng W., Altman R.B. // *Drug Discov Today*. 2018. Vol. 23. N 8. P. 1538 (doi:10.1016/j.drudis.2018.05.010).
10. Yu W., MacKerell A.D. Jr. // *Methods Mol Biol*. 2017. Vol. 1520. P. 85 (doi: 10.1007/978-1-4939-6634-9\_5).
11. Pant S., Singh M., Ravichandiran V., Murty U., Srivastava H.K. // *J Biomol Struct Dyn*. 2020. Advance online publication. P. 1 (doi: 10.1080/07391102.2020.1757510).
12. Skvortsov V.S., Druzhilovskiy D.S., Veselovsky A.V. // *Biomedical Chemistry: Research and Methods*. 2020. Vol. 3. N 1. P. e00124 (doi: 10.18097/BMCRM00124).
13. Wang J. // *J Chem Inf Model*. 2020. Vol. 60. N 6. P. 3277 (doi:10.1021/acs.jcim.0c00179).
14. Mittal L, Kumari A, Srivastava M, Singh M, Asthana S. // *J Biomol Struct Dyn*. 2020. Advance online publication. 1–19 (doi:10.1080/07391102.2020.1768151).
15. Gyebi G.A., Ogunro O.B., Adegunloye A.P., Ogundemi O.M., Afolabi S.O. // *J Biomol Struct Dyn*. 2020. Advance online publication. 1–13 (doi:10.1080/07391102.2020.1764868).
16. Enmozhi S.K., Raja K., Sebastine I., Joseph J. // *J Biomol Struct Dyn*. 2020. Advance online publication. 1–7 (doi:10.1080/07391102.2020.1760136).
17. Elmezayen A.D., Al-Obaidi A., Şahin A.T., Yeleği K. // *J Biomol Struct Dyn*. 2020. Advance online publication. 1–13 (doi:10.1080/07391102.2020.1758791).
18. Mahanta S, Chowdhury P, Gogoi N, Goswami N, Borah D, Kumar R, Chetia D, Borah P, Buragohain A. K., Gogoi B. // *J Biomol Struct Dyn*. 2020. Advance online publication. 1–10 (doi:10.1080/07391102.2020.1768902).
19. Kandeel M., Al-Nazawi M. // *Life Sci*. 2020. Vol. 251. P. 117627 (doi:10.1016/j.lfs.2020.117627).
20. Gentile D., Patamia V., Scala A., Sciortino M.T., Piperno A., Rescifina A. // *Mar Drugs*. 2020. Vol. 18. N 4. P. 225 (doi:10.3390/md18040225).
21. Sepay N., Sepay N., Al Hoque A., Mondal R., Halder U. C., Muddassir M. // *Struct Chem*. 2020. Vol. 31. P. 1831 (doi:10.1007/s11224-020-01537-5).



22. Shamsi A., Mohammad T., Anwar S., AlAjmi M.F., Hussain A., Rehman M.T., Islam A., Hassan M.I. // *Bio-science reports*. 2020. Vol. 40. N 6. P. BSR20201256 (doi:10.1042/BSR20201256).
23. Jin Z., Du X., Xu Y., Deng Y., Liu M., Zhao Y., Zhang B., Li X., Zhang L., Peng C., Duan Y., Yu J., Wang L., Yang K., Liu F., Jiang R., Yang X., You T., Liu X., Yang X., Yang H. // *Nature*. 2020. Vol. 582. N 7811. P. 289 (doi: 10.1038/s41586-020-2223-y).
24. Chen Y.W., Yiu, C.B., Wong, K.Y. // *F1000Research*. 2020, Vol. 9, P.129 (doi:10.12688/f1000research.22457.2).
25. Liu X, Wang X.J. // *J Genet Genomics*. 2020. Vol. 47. N 2. P. 119 (doi:10.1016/j.jgg.2020.02.001).
26. Ul Qamar M.T., Alqahitani S.M., Alamri M.A., Chen L.L. // *J Pharm Anal*. 2020. Vol. 10. N 4. P. 313 (doi:10.1016/j.jpha.2020.03.009).
27. Alves V.M., Bobrowski T., Melo-Filho C.C., Korn D.I., Auerbach S., Schmitt C., Muratov E.N., Tropsha A. // *Mol. Inform.* 2020. Advance online publication (doi:10.1002/minf.202000113).
28. Xue X., Yu H., Yang H., Xue F., Wu Z., Shen W., Li J., Zhou Z., Ding Y., Zhao Q., Zhang X.C., Liao M., Bartlam M., and Rao Z. // *J. Virol*. 2008. Vol. 82. N 5. P. 2515 (doi: 10.1128/JVI.02114-07).
29. Pillaiyar T., Manickam M., Namasivayam V., Hayashi Y., Jung S.H. // *J Med Chem*. 2016. Vol. 59. N 14. P. 6595 (doi:10.1021/acs.jmedchem.5b01461).
30. The DrugBank database, Available from: <https://www.drugbank.ca/> (accessed: 28.09.2020).
31. Zhang D.H., Wu K.L., Zhang X., Deng S.Q., Peng B. // *J Integr Med*. 2020. Vol. 18. N 2. P. 152 (doi:10.1016/j.joim.2020.02.005).
32. Ton A.T., Gentile F., Hsing M., Ban F., Cherkasov A. // *Mol Inform*. 2020. Vol. 39. N 8. P. e2000028 (doi:10.1002/minf.202000028).
33. Beck B.R., Shin B., Choi Y., Park S., Kang K. // *Comput Struct Biotechnol J*. 2020. Vol.18, P.784 (doi:10.1016/j.csbj.2020.03.025).
34. Weininger D. // *J. Chem. Inf. Comput. Sci*. 1988. Vol. 28. N 1. P. 31 (doi: 10.1021/ci00057a005).
35. Kreimeyer K., Foster M., Pandey A., Arya N., Halford G., Jones S.F., Forshee R., Walderhaug M., Botsis T. // *J Biomed Inform*. 2017. Vol. 73. P. 14 (doi: 10.1016/j.jbi.2017.07.012).
36. Valueva M.V., Nagornov N.N., Lyakhov P.A., Valuev G.V., Chervyakov N.I. // *Mathematics and Computers in Simulation*. 2020. Vol. 177. P. 232 (doi: 10.1016/j.matcom.2020.04.031).
37. Hoffmann M., Kleine-Weber H., Schroeder S., Krüger N., Herrler T., Erichsen S., Schiergens T.S., Herrler G., Wu N.H., Nitsche A., Müller M.A., Drosten C., Pöhlmann S. // *Cell*. 2020. Vol. 181. N 2. P. 271. (doi: 10.1016/j.cell.2020.02.052).
38. Rahman N., Basharat Z., Yousuf M., Castaldo G., Rasrelli L., Khan H. // *Molecules*. 2020. Vol. 25. N 10. P. E2271 (doi:10.3390/molecules25102271).
39. Zsoldos Z., Reid D., Simon A., Sadjad S.B., Johnson A.P. // *J Mol Graph Model*. 2007. Vol. 26. N 1. P. 198 (doi:10.1016/j.jmkgm.2006.06.002).
40. Pyrkov T.V., Ozerov I.V., Balitskaya E.D., Efremov R.G. // *Russ J Bioorg Chem*. 2010. Vol. 36. P. 446–455 (doi: 10.1134/S1068162010040023).
41. Šledž P, Caflisch A. // *Curr Opin Struct Biol*. 2018. Vol. 48. P. 93 (doi:10.1016/j.sbi.2017.10.010).
42. Muratov E.N., Bajorath J., Sheridan R.P., Tetko I.V., Filimonov D., Poroikov V., Oprea T.I., Baskin I.I., Varnek A., Roitberg A., Isayev O., Curtalolo S., Fourches D., Cohen Y., Aspuru-Guzik A., Winkler D.A., Agrafiotis D., Cherkasov A., Tropsha A. // *Chem Soc Rev*. 2020. QSAR without borders. Vol. 49. N 11. P. 3525 (doi:10.1039/d0cs00098a).
43. 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. // *J Med Chem* 2006. Vol. 49. N 12. P. 3485 (doi:10.1021/jm050852f).
44. Todeschini R, Consonni V, Ballabio D., Grisoni F. (2020) Chemometrics for QSAR Modeling // *Comprehensive Chemometrics / Eds Brown S., Tauler R., Walczak B.* Elsevier, 2944p (doi:10.1016/B978-0-12-409547-2.14703-1).
45. Adhikari N., Baidya S.K., Saha A., Jha T. // *Viral Proteases and Their Inhibitors* 2017. Structural Insight Into the Viral 3C-Like Protease Inhibitors: Comparative SAR/QSAR Approaches., Published online, P. 317 (doi: 10.1016/B978-0-12-809712-0.00011-3).
46. OECD. 2014. Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models, OECD Series on Testing and Assessment, No. 69, OECD Publishing, Paris (doi: 10.1787/9789264085442-en).
47. Polishchuk P. // *J. Chem. Inf. Model*. 2017. Vol. 57. N 11. P. 2618 (doi:10.1021/acs.jcim.7b00274).
48. Kumar V., Roy K. // *SAR QSAR Environ. Res*. 2020. Vol. 31. N 7. P. 511 (doi:10.1080/1062936X.2020.1776388).
49. Masand V.H., Rastija V., Patil M.K., Gandhi A., Chapolikar A. // *SAR QSAR Environ*. 2020. Vol. 31. N 9. P. 643 (doi: 10.1080/1062936X.2020.1784271).
50. <https://scienceng.com/disadvantages-linear-regression-8562780.html> (accessed: 28.09.2020).
51. ChEMBL Database. Available from: [https://www.ebi.ac.uk/chembl/g/#browse/activities/filter/target\\_chembl\\_id%3ACHEMBL3927%20AND%20standard\\_type%3A\(%22IC50%22\)](https://www.ebi.ac.uk/chembl/g/#browse/activities/filter/target_chembl_id%3ACHEMBL3927%20AND%20standard_type%3A(%22IC50%22)) (accessed: 28.09.2020).
52. Simplex representation of molecular structure – a chemoinformatic tool for calculation of simplex (fragment) descriptors. Available from: <https://github.com/DrrDom/sirms> (accessed: 28.09.2020).
53. Polishchuk P., Tinkov O., Khristova T., Ognichenko L., Kosinskaya A., Varnek A., Kuz'min V. // *J. Chem. Inf. Model*. 2016. Vol. 56. N 8. P. 1455 (doi: 10.1021/acs.jcim.6b00371).
54. Scikit-learn. Free software machine learning library for the Python programming language. Available from: <https://scikit-learn.org/stable/> (accessed: 28.09.2020).
55. Jaworska J., Nikolova-Jeliazkova N., Aldenberg T. // *Altern Lab Anim*. 2005. Vol. 33. N 5. P. 445 (doi:10.1177/026119290503300508).
56. Web-based platform OCHEM. OCHEM user's manual. Available from: <http://docs.ochem.eu/display/MAN/OCHEM+Introduction> (accessed: 28.09.2020).
57. Sushko Y., Novotarskyi S., Korner R., Vogt J., Abdelaziz, A., Tetko I. // *J. Cheminf*. 2014. Vol. 6. N 1. P. 48 (doi: 10.1186/s13321-014-0048-0).
58. Breiman L. RRforest software. Available from: [http://www.stat.berkeley.edu/~breiman/RandomForests/reg\\_examples/RFR](http://www.stat.berkeley.edu/~breiman/RandomForests/reg_examples/RFR) (accessed: 28.09.2020).

59. Raevsky O.A., Grigorev V.Yu., Kireev D.B., Zefirov N.S. // *Quant. Struct.-Act. Relat.* 1992. Vol. 11. N 1. P. 49 (doi: 10.1002/qsar.19920110109).
60. Grigorev V.Yu., Grigoreva L.D. // *Moscow Univ. Chem. Bull.* 2016. Vol. 71. N 3. P. 199 (doi: 10.3103/S0027131416030056).
61. Martin T., Harten P., Young D., TEST (Toxicity Estimation Software Tool) V 4.1, US Environmental Protection Agency, Washington DC, USA, 2012; Available from: <https://www.epa.gov/chemical-research/toxicity-estimation-software-tool-test> (accessed: 28.09.2020).
62. Daina A., Michielin O., Zoete V. // *Sci Rep.* 2017. Vol. 7. P. 42717. (doi:10.1038/srep42717).
63. Lipinski C.A., Lombardo F., Dominy B.W., Feeney P.J. // *Adv. Drug Deliv. Rev.* 2001. Vol. 46. P. 3 (doi:10.1016/S0169-409X(00)00129-0).
64. Dahlin J.L., Nissink J.W., Strasser J.M., Francis S., Higgins L., Zhou H., Zhang Z., Walters M.A. // *J. Med. Chem.* 2015. Vol. 58. N 5. P. 2091 (doi:10.1021/jm5019093).
65. Ertl P., Schuffenhauer A. // *J. Cheminform.* 2009. Vol. 1. N 1. P. 8 (doi:10.1186/1758-2946-1-8).
66. Pastick K.A., Okafor E.C., Wang F., Lofgren S.M., Skipper C.P., Nicol M.R., Pullen M.F., Rajasingham R., McDonald E.G., Lee T.C., Schwartz I.S., Kelly L.E., Lothar S.A., Mitjà O., Letang E., Abassi M., Boulware D.R. // *Open Forum Infect Dis.* 2020. Vol. 7. N. 4. P. ofaa130. (doi:10.1093/ofid/ofaa130).
67. Jin Z., Zhao Y., Sun Y., Zhang B., Wang H., Wu Y., Zhu Y., Zhu C., Hu T., Du X., Duan Y., Yu J., Yang X., Yang K., Liu X., Guddat L.W., Xiao G., Zhang L., Yang H., Rao Z. // *Nat Struct Mol Biol.* 2020. Vol. 27. N. 6. P. 529 (doi:10.1038/s41594-020-0440-6).

Поступила в редакцию 11.01.2021  
Получена после доработки 14.01.2021  
Принята к публикации 20.01.2021

## VIRTUAL SCREENING AND MOLECULAR DESIGN OF POTENTIAL SARS-COV-2 INHIBITORS

O.V. Tinkov<sup>1,2\*</sup>, V.Yu. Grigorev<sup>3</sup>, L.D. Grigoreva<sup>4</sup>

(<sup>1</sup>*Department of Pharmacology and Pharmaceutical Chemistry, Medical Faculty, Shevchenko Transnistria State University, Tiraspol, Moldova;* <sup>2</sup>*Military Institute of the Ministry of Defense, Tiraspol, Moldova;* <sup>3</sup>*Institute of Physiologically Active Compounds of the Russian Academy of Science, Chernogolovka, Russia;* <sup>4</sup>*Department of Fundamental Physical-Chemical Engineering, Lomonosov Moscow State University;* \*e-mail: [oleg.tinkov.chem@mail.ru](mailto:oleg.tinkov.chem@mail.ru))

According to recent studies, the main M<sup>pro</sup> protease of the SARS-CoV-2 virus, which is the most important target in the development of promising drugs for the treatment of COVID-19, is evolutionarily conservative and has not undergone significant changes compared to the main M<sup>pro</sup> protease of the SARS-CoV virus. Many researchers have noted a high similarity between the binding sites of the main M<sup>pro</sup> protease of the SARS-CoV and SARS-CoV-2 viruses, so in the context of an increasing epidemic, further research of inhibitors of the main mpro protease of the SARS-CoV virus to fight COVID-19 seems logical. QSAR methods (Quantitative Structure – Activity Relationship), which were applied in this paper in the analysis of inhibitors of SARS-CoV main M<sup>pro</sup> protease, can significantly help in reducing time and financial costs for repositioning and developing new drugs. In the study using simplex, fractal, HYBOT descriptors, methods Partial Least Squares (PLS), Random Forest (RF), Support Vectors, Gradient Boosting (GBM), as well as the OCHEM Internet platform (<https://ochem.eu>), where different types of molecular descriptors, machine learning methods are implemented, QSAR models were built. For the obtained models, a structural interpretation was performed, which allowed us to identify molecular fragments that increase and decrease the activity of SARS-CoV inhibitors. The structural interpretation results were used for the rational molecular design of potential SARS-CoV-2 inhibitors. The QSAR models were also used for a virtual screening of 2087 FDA-approved drugs.

**Key words:** M<sup>pro</sup> protease, QSAR, molecular descriptors, machine learning, structural interpretation.

**Сведения об авторах:** Тиньков Олег Викторович – доцент кафедры фармакологии и фармацевтической химии медицинского факультета Приднестровского государственного университета; заместитель начальника учебной части Военного института, канд. хим. наук ([oleg.tinkov.chem@mail.ru](mailto:oleg.tinkov.chem@mail.ru)); Григорьев Вениамин Юрьевич – зав. отделом компьютерного молекулярного дизайна ИФАВ РАН, докт. хим. наук ([beng@ipac.ac.ru](mailto:beng@ipac.ac.ru)); Григорьева Людмила Дмитриевна – зам. декана по учебной работе, доцент факультета фундаментальной физико-химической инженерии МГУ имени М.В. Ломоносова, канд. физ.-матем. наук ([ldg@physchem.msu.ru](mailto:ldg@physchem.msu.ru)).

## Приложение

Таблица 1 П

## Структуры и активности исследуемых соединений

Номер	SMILES	Molecule ChEMBL ID	IC50, nM	pChEMBL Value	Document ChEMBL ID
1	<chem>Brc1cnc(OC(=O)c2ccco2)c1</chem>	CHEMBL427404	50	7,3	CHEMBL1144475
2	<chem>Cle1cnc(OC(=O)c2ccco2)c1</chem>	CHEMBL426898	60	7,22	CHEMBL1144475
3	<chem>Cle1ccc(cc1)-c1ccc(o1)C(=O)Oc1cnc(Cl)c1</chem>	CHEMBL426082	63	7,2	CHEMBL1144475
4	<chem>Cle1cnc(OC(=O)c2cc3ccccc3[nH]2)c1</chem>	CHEMBL384739	65	7,19	CHEMBL1144475
5	<chem>Cle1cnc(OC(=O)c2cc3ccccc3s2)c1</chem>	CHEMBL383725	95	7,02	CHEMBL1144475
6	<chem>Cle1cnc(OC(=O)c2cc3ccccc3o2)c1</chem>	CHEMBL380470	170	6,77	CHEMBL1144475
7	<chem>Cle1cnc(OC(=O)c2escn2)c1</chem>	CHEMBL380403	270	6,57	CHEMBL1144475
8	<chem>Cc1cc(c(Cl)cc1Cl)S(=O)(=O)c1c(cc(cc1[N+]([O-])=O)C(F)(F)F)[N+]([O-])=O</chem>	CHEMBL379727	300	6,52	CHEMBL1145342
9	<chem>COc1cccc(c1)C(=O)Oc1cnc(Cl)c1</chem>	CHEMBL379642	340	6,47	CHEMBL1144475
10	<chem>NC(=O)c1ccc2N(Cc3ccc4ccccc4c3)C(=O)C(=O)c2c1</chem>	CHEMBL378700	370	6,43	CHEMBL1148529
11	<chem>ClC(Cl)=C(Cl)C(=O)Oc1ccc(cc1)S(=O)(=O)c1ccc(OC(=O)C(Cl)=C(Cl)Cl)cc1</chem>	CHEMBL378674	900	6,05	CHEMBL1145342
12	<chem>Ic1ccc2N(Cc3cc4ccccc4s3)C(=O)C(=O)c2c1</chem>	CHEMBL378342	950	6,02	CHEMBL1139624
13	<chem>Brc1cccc2C(=O)C(=O)N(Cc3cc4ccccc4s3)c12</chem>	CHEMBL377253	980	6,01	CHEMBL1139624
14	<chem>Ic1ccc2N(Cc3ccc4ccccc4c3)C(=O)C(=O)c2c1</chem>	CHEMBL377150	1100	5,96	CHEMBL1148529
15	<chem>[O-][N+](=O)c1ccc2C(=O)C(=O)N(Cc3ccc4ccccc4s3)c12</chem>	CHEMBL375130	2000	5,7	CHEMBL1139624
16	<chem>FC(F)(F)c1nnc(SC(=O)c2ccc(o2)C#Cc2ccccc2)[nH]1</chem>	CHEMBL370923	3000	5,52	CHEMBL1145342
17	<chem>Cle1cc(Cl)cc(NC(=O)CSc2nccc(n2)-c2csc(n2)-c2ccccc2)c1</chem>	CHEMBL365469	3000	5,52	CHEMBL1148632
18	<chem>Fe1ccc2N(Cc3cc4ccccc4s3)C(=O)C(=O)c2c1</chem>	CHEMBL365134	4820	5,32	CHEMBL1139624
19	<chem>Cc1noc(NC(=O)c2ccc(s2)-c2cc(nn2C)C(F)(F)F)c1[N+](O-)=O</chem>	CHEMBL358279	5000	5,3	CHEMBL1145342
20	<chem>Nc1ccc(c(N)n1)S(=O)(=O)c1ccc(Cl)cc1</chem>	CHEMBL348660	6000	5,22	CHEMBL1145342
21	<chem>Cc1noc(C)c1CN1C(=O)C(=O)c2cc(c(c12)C#N</chem>	CHEMBL225515	7200	5,14	CHEMBL1139624
22	<chem>Fe1ccc(CN2C(=O)C(=O)c3cc(I)ccc23)c(Cl)c1</chem>	CHEMBL222893	9400	5,03	CHEMBL1139624
23	<chem>Cn1nc(cc1C(F)(F)F)-c1ccc(s1)-c1cnc(SCC(=O)Nc2ccc(Cl)cc2)n1</chem>	CHEMBL222840	10000	5	CHEMBL1148632
24	<chem>Cc1oc(cc1-c1cc(NS(=O)(=O)c2cccs2)[nH]n1)C(C)C</chem>	CHEMBL222769	10000	5	CHEMBL1145342
25	<chem>CSc1sc(c(C)c1-c1cnc(SCC(=O)Nc2ccc(Cl)cc2)n1)-c1nc(C)cs1</chem>	CHEMBL222735	11000	4,96	CHEMBL1148632
26	<chem>Cle1cccc2N(Cc3cc4ccccc4s3)C(=O)C(=O)c12</chem>	CHEMBL222628	11200	4,95	CHEMBL1139624
27	<chem>[O-][N+](=O)c1cc(ccc1S(=O)(=O)c1ccc(Cl)cc1)C(F)(F)F</chem>	CHEMBL222234	12000	4,92	CHEMBL1145342
28	<chem>CSc1sc(c(C)c1-c1cnc(SCC(=O)Nc2ccccc2Cl)n1)-c1nc(C)cs1</chem>	CHEMBL215732	12000	4,92	CHEMBL1148632
29	<chem>NC(=O)c1ccc2N(Cc3ccccc3)C(=O)C(=O)c2c1</chem>	CHEMBL215397	12500	4,9	CHEMBL1148529
30	<chem>Cle1ccc(NC(=O)c2ccc(CN3C(=O)C(=O)c4cc(I)ccc34)s2)cc1</chem>	CHEMBL214372	12570	4,9	CHEMBL1139624
31	<chem>Cc1nc(c(C#N)c(C)c1[N+](O-)=O)S(=O)(=O)c1ccccc1</chem>	CHEMBL213581	13000	4,89	CHEMBL1145342
32	<chem>Cc1ccc(cc1)S(=O)(=O)c1nc(C)c(c(C)c1C#N)[N+](O-)=O</chem>	CHEMBL212504	13000	4,89	CHEMBL1145342

Окончание табл. 1 П

Номер	SMILES	Molecule ChEMBL ID	IC50, nM	pChEMBL Value	Document ChEMBL ID
33	<chem>O=C1N(Cc2cc3ccccc3s2)c2ccccc2C1=O</chem>	CHEMBL212454	13110	4,88	CHEMBL1139624
34	<chem>Ic1ccc2N(CC3COc4ccccc4O3)C(=O)C(=O)c2c1</chem>	CHEMBL212399	13500	4,87	CHEMBL1139624
35	<chem>Cc1nc(es1)-c1nc(es1)-c1cnc(SCC(=O)Nc2ccc(Cl)cc2)n1</chem>	CHEMBL212240	14000	4,85	CHEMBL1148632
36	<chem>[O-][N+](=O)c1ccc([n+](c1)[O-])S(=O)(=O)c1ccc(Cl)cc1</chem>	CHEMBL212218	15000	4,82	CHEMBL1145342
37	<chem>CSc1[nH]nc(NC(=O)c2cccs2)c1S(=O)(=O)c1ccccc1</chem>	CHEMBL212190	15000	4,82	CHEMBL1145342
38	<chem>FC(F)(F)c1ccc(NC(=O)CSc2nccc(n2)-c2cc(no2)-c2ccc(Cl)cc2Cl)cc1</chem>	CHEMBL212019	15000	4,82	CHEMBL1148632
39	<chem>Clc1ccc(NC(=O)CSc2nccc(n2)-c2cc(no2)-c2ccccc2Cl)cc1</chem>	CHEMBL211969	15000	4,82	CHEMBL1148632
40	<chem>Clc1ccc(NC(=O)CSc2nccc(n2)-c2cc(no2)-c2ccccc2)cc1</chem>	CHEMBL210632	15000	4,82	CHEMBL1148632
41	<chem>CC(=O)c1ccccc1S(=O)(=O)c1ccccc1C(O)=O</chem>	CHEMBL210612	16000	4,8	CHEMBL1145342
42	<chem>CCO\C(O)=C(\C=N\c1ccc(cc1)S(=O)(=O)c1ccc(N\C=C(/C#N)C(=O)OCC)cc1)/C#N</chem>	CHEMBL210525	16000	4,8	CHEMBL1145342
43	<chem>OC(=O)c1ccc(cc1)S(=O)(=O)c1cc(Br)c(O)c(Br)c1</chem>	CHEMBL210497	16000	4,8	CHEMBL1145342
44	<chem>CC1(C)Cc2c(se(NCc3ccco3)c2C(=O)C1)C#N</chem>	CHEMBL210487	16000	4,8	CHEMBL1145342
45	<chem>Ic1ccc2N(Cc3ccc(s3)C(=O)N3CCCC3)C(=O)C(=O)c2c1</chem>	CHEMBL210097	17500	4,76	CHEMBL1139624
46	<chem>CSc1nn(c(-c2cccs2)c1C#N)-c1c(c(C)nn1C)[N+](c1)[O-]=O</chem>	CHEMBL210092	18000	4,75	CHEMBL1145342
47	<chem>CCCCN1C(=O)C(=O)c2cc(ccc12)C(N)=O</chem>	CHEMBL209667	19000	4,72	CHEMBL1148529
48	<chem>Cc1nn(C)c(NC2ccc(s2)-c2cccs2)c1[N+](c1)[O-]=O</chem>	CHEMBL209287	20000	4,7	CHEMBL1145342
49	<chem>Ic1ccc2N(C\C=C\c3cc4ccccc4s3)C(=O)C(=O)c2c1</chem>	CHEMBL209227	23500	4,63	CHEMBL1139624
50	<chem>[O-][N+](=O)c1ccc(cc1)S(=O)(=O)c1ccc(cc1)[N+](c1)[O-]=O</chem>	CHEMBL208763	25000	4,6	CHEMBL1145342
51	<chem>CCCN1C(=O)C(=O)c2cc(ccc12)C(N)=O</chem>	CHEMBL208732	25000	4,6	CHEMBL1148529
52	<chem>CCCc1cc(O)nc(SCC(=O)Nc2ccc(Cl)cc2)n1</chem>	CHEMBL208584	30000	4,52	CHEMBL1148632
53	<chem>CCO\C(O)=C(\C=N\c1ccc(cc1)S(=O)(=O)c1ccc(NC=C(C(=O)OCC)C(=O)OCC)cc1)/C(=O)OCC</chem>	CHEMBL207207	32000	4,5	CHEMBL1145342
54	<chem>O=C(Cc1nccs1)c1nccs1</chem>	CHEMBL196635	40000	4,4	CHEMBL1145342
55	<chem>CC(C)c1ccc(NC(=O)CSc2nccc(n2)-c2cccs2)cc1</chem>	CHEMBL194398	40000	4,4	CHEMBL1148632
56	<chem>COc1ccccc1)-c1nc(SCC(=O)Nc2ccc(cc2)S(N)(=O)=O)nc(O)c1C#N</chem>	CHEMBL191575	40000	4,4	CHEMBL1148632
57	<chem>COc1ccc(NC(=O)CSc2nc(O)cc(n2)-c2ccccc2)cc1OC</chem>	CHEMBL190743	45000	4,35	CHEMBL1148632
58	<chem>CCOC(=O)C=C\C[C@H](C[C@@H]1CCNC1=O)NC(=O)[C@@H](CC(=O)[C@@H](NC(=O)c1cc(C)on1)C(C)C)Cc1ccccc1</chem>	CHEMBL188983	45000	4,35	CHEMBL1141032
59	<chem>Ic1ccc2N(Cc3ccccc3)C(=O)C(=O)c2c1</chem>	CHEMBL188487	50000	4,3	CHEMBL1148529
60	<chem>COc1ccccc1)-c1nc(SCC(=O)Nc2ccc(cc2)C(C)=O)nc(O)c1C#N</chem>	CHEMBL187717	60000	4,22	CHEMBL1148632
61	<chem>CCC(Sc1nc(O)c(C#N)c(n1)-c1ccccc1OC)c1C(=O)Nc1ccc(cc1)C(C)=O</chem>	CHEMBL187598	60000	4,22	CHEMBL1148632
62	<chem>CCCCN1C(=O)C(=O)c2cc(I)ccc12</chem>	CHEMBL187579	66000	4,18	CHEMBL1148529
63	<chem>CCOC(=O)C=C\C[C@H](C[C@@H]1CCNC1=O)NC(=O)[C@H](CC(=O)C(C)C)CC(=O)[C@@H](NC(=O)c1cc(C)on1)C(C)C</chem>	CHEMBL185698	70000	4,16	CHEMBL1141032
64	<chem>CNIC(=O)C(=O)c2cc(ccc12)C(N)=O</chem>	CHEMBL148483	71000	4,15	CHEMBL1148529
65	<chem>Cc1cc(O)nc(SCC(=O)Nc2cc(Cl)ccc2O)c2ccccc2)n1</chem>	CHEMBL118596	100000	4	CHEMBL1148632

Т а б л и ц а 2 П

## Полный список идентифицированных молекулярных фрагментов, записанных в виде SMARTS

SMARTS	M	N	Средний вклад фрагмента
<chem>O=C(O[*])c1cc[*]c[*]o1</chem>	3	3	2,08424025
<chem>O=C(Oc1cc[*]cnc1[*])[*]</chem>	8	8	2,0206205
<chem>Clc1c[*]ncc[*]c1[*]</chem>	7	7	1,8665205
<chem>O=C(O[*])[*]</chem>	9	9	1,59494225
<chem>O=C1C(=O)N(C[*])[*]c2c1cccc2[*]</chem>	3	3	0,6641145
<chem>c1c[*]cc2cc[*]sc2c1[*]</chem>	8	8	0,652451375
<chem>Clc1cc(Cl)c[*]c[*]c1[*]</chem>	3	3	0,30955275
<chem>FC(F)(F)[*]</chem>	6	6	0,146696125
<chem>NC(=O)[*]</chem>	5	5	0,13180325
<chem>O=S(=O)(c1cc[*]c(Cl)cc1[*])[*]</chem>	4	4	0,1251
<chem>c1cc(C[*])c[*]cc1[*]</chem>	4	4	0,12120975
<chem>c1c[*]sc(C[*])c1[*]</chem>	3	3	0,09563475
<chem>Cl[*]</chem>	25	30	0,04179175
<chem>Cn1nc[*]c[*]c1[*]</chem>	4	4	0,030965625
<chem>F[*]</chem>	8	8	0,027254125
<chem>N#C[*]</chem>	7	7	0,0052065
<chem>O=C1C(=O)N[*]c2c[*]cc1c[*]c21</chem>	9	9	0,00389875
<chem>c1cc(-c2cc[*]on2)c[*]cc1[*]</chem>	3	3	0,002501
SMARTS	M	N	Средний вклад фрагмента
<chem>c1nc[*]nc[*]c1[*]</chem>	10	10	0,001548625
<chem>O=[N+]{([O-])[*]}</chem>	10	10	0
<chem>I[*]</chem>	9	9	-0,000999
<chem>n1c[*]sc[*]c1[*]</chem>	6	8	-0,025514375
<chem>O=C(Nc1ccc[*]cc1[*])[*]</chem>	12	12	-0,0292055
<chem>O=C(N[*])[*]</chem>	17	17	-0,1267215
<chem>O=C(CS[*])N[*]</chem>	14	14	-0,17362825
<chem>NC(=O)c1c[*]cc2c(c1[*])C(=O)C(=O)N2[*]</chem>	4	4	-0,178003875
<chem>O=c1[nH]c(S[*])nc[*]c1[*]</chem>	5	5	-0,3823725

О б о з н а ч е н и я: M – количество соединений, содержащих данный фрагмент; N – количество обнаружений соответствующего фрагмента в выборке.