

# Ингибиторы новой коронавирусной протеазы

## Идентифицировано виртуальным скринингом 687 миллионов соединений

Андре Фишер, Мануэль Селнер, Сантош Неранджан, Маркус А. Лилль и Мартин Смиеско

Компьютерная аптека, Департамент Фармацевтических Наук, Университет Базеля

E-mail: markus.lill@unibas.ch; martin.smiesko@unibas.ch

### Аннотация

Быстрая вспышка нового тяжелого острого респираторного синдрома коронавируса 2 (SARS-CoV-2) в Китае, сопровождаемая его распространением по всему миру, представляет серьезную глобальную проблему для общественного здравоохранения: почти 90000 человек заразились и тысячи смертельных случаев. Чтобы победить вирусные инфекции, ингибирование протеаз, необходимых для протеолитического процессинга вирусных полипротеинов, является традиционной терапевтической стратегией. На сегодняшний день нет конкретных лекарств или вакцин для лечения SARS-CoV-2, несмотря на его тесную связь с вирусом SARS-CoV-1, который вызвал подобную эпидемию в 2003 году. Таким образом, остается острая необходимость в разработке специальных противовирусных терапевтических средств для борьбы с SARS-CoV-2. Для того чтобы найти новые ингибиторы, мы провели компьютерный скрининг библиотеки соединений из более 687 миллионов соединений для связывания в недавно решенной кристаллической структуре основной протеазы SARS-CoV-2. Скрининг такого обширного химического пространства для ингибиторов протеазы SARS-CoV-2 ранее не проводился. После скрининга формы применяли два протокола стыковки с последующим определением фармакокинетически релевантных молекулярных дескрипторов для сужения начальных совпадений. Далее молекулярное динамическое моделирование проводилось для подтверждения стабильности состыкованных режимов связывания и всестороннего количественного определения энергий связывания лигандов. После оценки связывания о-мишени мы публикуем список из 11 лекарствоподобных соединений с улучшенной свободной энергией связывания с целевой протеазой в отличие от сокристаллизованного пептидомиметического соединения свинца, которое обладает плохими фармакокинетическими свойствами. Кроме того, мы определили одно сильное связующее с сопоставимыми свойствами из библиотеки природных соединений.

### Введение

В конце 2019 года новый коронавирус, названный тяжелым острым респираторным синдромом коронавируса 2 (SARS-CoV-2), был определен в качестве причины для нескольких случаев респираторного заболевания в Китае. Несмотря на то, что большинство инфицированных пациентов страдают только от легких симптомов, таких как повышение температуры и кашель, ассоциирующихся с хорошим прогнозом, заболевание может прогрессировать в смертельные случаи

пневмонии и острой дыхательной недостаточности, особенно у пожилых мужчин с сопутствующими заболеваниями.<sup>1,2</sup> Вирус быстро распространился из Китая в более чем 55 стран мира и на сегодняшний день заразил почти 90000 человек, унесших более 2800 смертей<sup>3</sup> (по состоянию на 1 марта 2020 года), что уже превышает эпидемию SARS-CoV-1 2003 года.<sup>4</sup> Скорее всего, вирус возник в результате зоонозной передачи между животными, такими как летучие мыши и люди, но прогрессировал для передачи от человека человеку через обычную воздушно-капельную инфекцию.<sup>2,5,6</sup> Всемирная организация здравоохранения объявила вспышку SARS-CoV-2 чрезвычайной ситуацией в области общественного здравоохранения, имеющей международное значение<sup>7,8</sup>, и основываясь на эпидемии SARS-CoV-1 2003 года, это приведет к гибели многих людей, огромное социальное воздействие и экономические потери в размере миллиардов долларов.<sup>9</sup> В настоящее время нет конкретных терапевтических вариантов лечения этой инфекции, что означает, что клиническое лечение заболевания в основном симптоматическое в сочетании с повторным использованием уже имеющихся на рынке противовирусных препаратов, таких как ритонавир и антибиотики, для лечения вторичных инфекций.<sup>1</sup> Таким образом, сохраняется острая необходимость в разработке специальных противовирусных препаратов и вакцин против SARS-CoV-2.<sup>4,10-12</sup> Ингибирование вирусных протеаз, необходимых для протеолитического процессинга полипротеинов, было успешной стратегией в фармакологическом лечении вируса иммунодефицита человека (ВИЧ) и гепатита С, соответственно<sup>13,14</sup>, доказывая потенциал ингибиторов протеазы для лечения вирусных инфекций. Точно так же считается, что основная протеаза SARS-CoV-2 необходима для репликации вируса и, следовательно, рассматривается как перспективная мишень для противовирусной фармакотерапии.<sup>10,15</sup> Кристаллическая структура основной протеазы SARS-CoV-2 была недавно определена<sup>16</sup>, что позволило рационально создать специфические ингибирующие соединения. Тесная связь SARS-CoV-2 с SARS-CoV-1 подтверждается высокой идентичностью последовательностей 96,1% среди их протеаз.<sup>10,17</sup> В этой связи было высказано предположение, что лекарственные средства, разработанные против SARS-CoV-1, могут быть эффективными для лечения SARS-CoV-2.<sup>12</sup> Однако эти соединения оставались на доклинической или ранней клинической стадии, не приводя к проверенному лекарственному средству<sup>18</sup> и их эффективность против нового вируса может отличаться из-за различий в отдельных аминокислотах.<sup>19</sup> Соответственно, разработка специфических ингибиторов для основной протеазы SARS-CoV-2 остается насущной необходимостью в научном сообществе<sup>10-12</sup>, что отражено в многочисленных проектах, посвященных этому белку. Например, в трех исследованиях, посвященных повторному использованию имеющихся на рынке лекарств, было предложено несколько кандидатов на лечение SARS-CoV-2.<sup>10,15,20</sup> В другом недавнем исследовании был применен подход глубокого поиска в сочетании с гомологической моделью протеазы для скрининга библиотеки приблизительно одного миллиона соединений, включая уже одобренные лекарства, трипептиды и натуральные продукты.<sup>12</sup> Однако недостатком применения терапевтических средств, изначально предназначенных для другой цели, является риск нежелательных фармакологических эффектов и побочных реакций.<sup>21</sup>

В этом исследовании мы провели скрининг большой библиотеки из более чем 687 миллионов соединений с целью найти новые ингибиторы для основной протеазы SARS-CoV-2. На первом этапе мы провели скрининг на основе форм с известными связующими для основной протеазы SARS-CoV-1 и соответствующих субструктур в качестве шаблонных молекул. После начального раунда скрининга формы были применены два различных протокола стыковки с последующей оценкой молекулярных дескрипторов, относящихся к фармакокинетике, чтобы сузить количество связующих. Кластеризация на основе молекулярных отпечатков была применена для обеспечения структурного разнообразия соединений, которые на следующем этапе были подвергнуты молекулярно-динамическому моделированию (MD). На основании полученных траекторий, свободную энергию связывания лигандов определяли количественно, используя постобработку Molecular Mechanics / Generalized Born Surface Area (MM / GBSA). На последнем этапе мы оценили потенциальные токсические эффекты соединений из-за взаимодействия с 16 известными лекарственными целями для окончательного отбора 11 соединений. Насколько нам известно, о таком всестороннем исследовании химического пространства с целью обнаружения ингибиторов протеазы SARS-CoV-2 ранее не сообщалось.

## Результаты и их обсуждения

### Сравнение протеаз между SARS-CoV-2 и SARS-CoV-1

Протеаза SARS-CoV-2 состоит из трех доменов (рис. 1A)<sup>18</sup> и обрабатывает полипротеины с использованием каталитической диады, состоящей из гистидина и цистеина в качестве каталитических остатков. Его активный сайт расположен между доменами I и II. Поскольку SARS-CoV-2 тесно связан с SARS-CoV-1, их протеазы демонстрируют высокую степень сходства последовательностей (96,1%; Рис. 1D).<sup>10,17</sup> В непосредственной близости от активного сайта только одна аминокислота (Ser46) отличается между двумя протеазами. Тем не менее, топология поверхности активного участка среди двух белков отличается, особенно в окрестности петли с центром вокруг Asn142 (Рисунки 1B и 1C). Кроме того, размер и глубина кармана S1' показывает заметные различия, и в центре S1, S1' и S3 протеаза SARS-CoV-2 представляет собой более отчетливую субполость, в отличие от SARS-CoV-1 фермента. Следовательно, ингибиторы протеазы SARS могут представлять измененные связывающие свойства для протеазы SARS-CoV-2. Сходства с другими вирусными протеазами, такими как ВИЧ-инфекция или короновироз респираторного синдрома Ближнего Востока (MERS-CoV), сравнительно низки.<sup>22</sup> Хотя основная протеаза, по-видимому, является наиболее перспективной терапевтической мишенью для ослабления репликации вируса, также рассматривалось ингибирование других функциональных белков, таких как папаин-подобная протеаза или взаимодействие между белком вирусного спайка и его рецептором проникновения в клетки человека.<sup>8,22</sup>

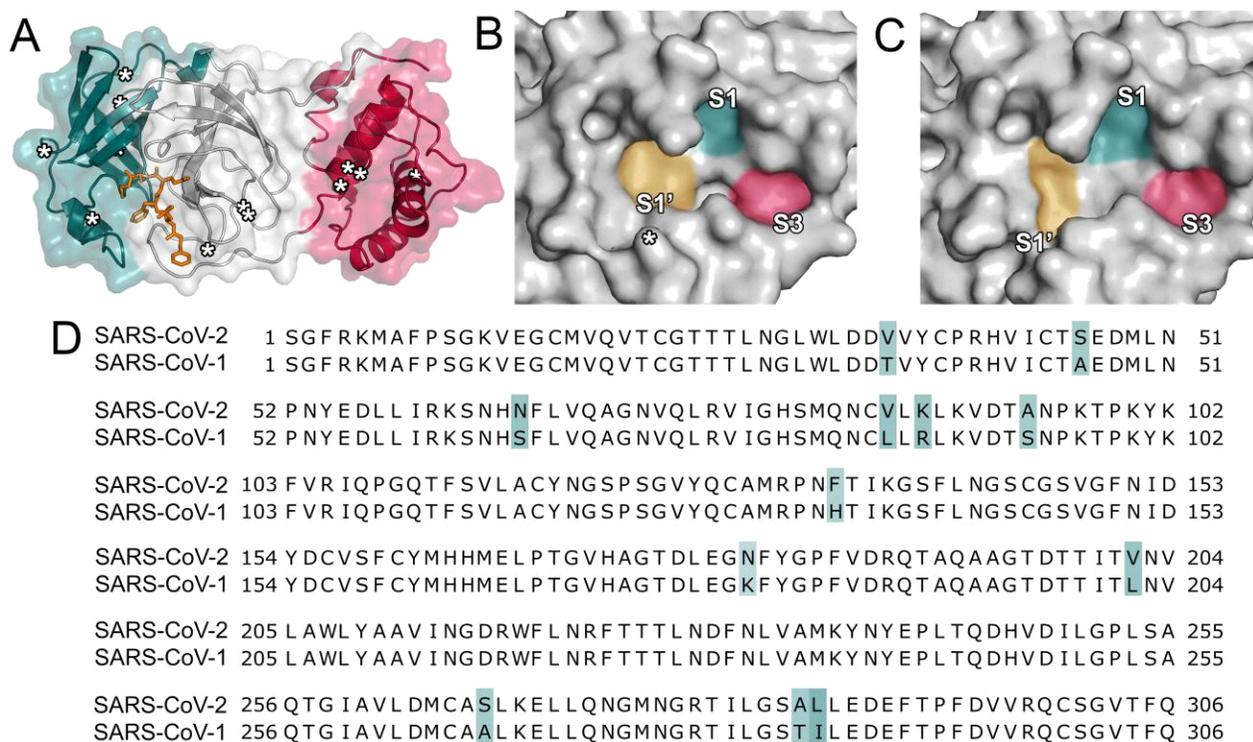


Рисунок 1. Структурный обзор и выравнивание последовательности. (A) Показаны три домена (домен I в синем, домен II в сером, домен III в красном) основной протеазы SARS-CoV-2. Изменения аминокислот между SARS-CoV-1 и SARS-CoV-2 отмечены звездочками. Сокристаллизованный лиганд (PDB ID 6LU7) представлен в виде органа. (B) Топология поверхности связывающего кармана основной протеазы SARS-CoV-2 (PDB ID 6LU7). Местоположение Ser46 обозначено звездочкой. (C) Топология поверхности связывающего кармана основной протеазы SARS-CoV-1 (PDB ID 2A5I). (D) Выравнивание последовательностей протеаз SARS-CoV-1 и SARS-CoV-2. Несоответствия отмечены синим цветом.

## Процедуры виртуального скрининга

Виртуальный скрининг - это широко используемый метод на ранней стадии обнаружения лекарств, который позволяет идентифицировать потенциально биологически активные соединения с высокой пропускной способностью.<sup>23</sup> Из-за своей скорости и экономической эффективности, его использование для выявления кандидатов в лекарственные средства против распространяющегося во всем мире SARS-CoV-2 предлагает многообещающий подход, особенно когда время имеет существенное значение. Было доказано, что ингибирование протеаз в качестве лечения против вирусных инфекций для ВИЧ делает основную протеазу SARS-CoV-2 привлекательной мишенью для лекарственного средства, в частности, поскольку ее кристаллическая структура была недавно определена. В нашей работе по виртуальному скринингу (рис. 2) мы начали с 687 миллионов соединений из трехмерной базы данных ZINC в процедуре скрининга на основе сходства форм. Такой грубый протокол с GPU-ускорением позволяет быстро просматривать большие базы данных на основе известных связующих в качестве шаблона. Предыдущие попытки виртуального скрининга были ограничены значительно меньшими библиотеками составов.<sup>10,12,15,17,20</sup>

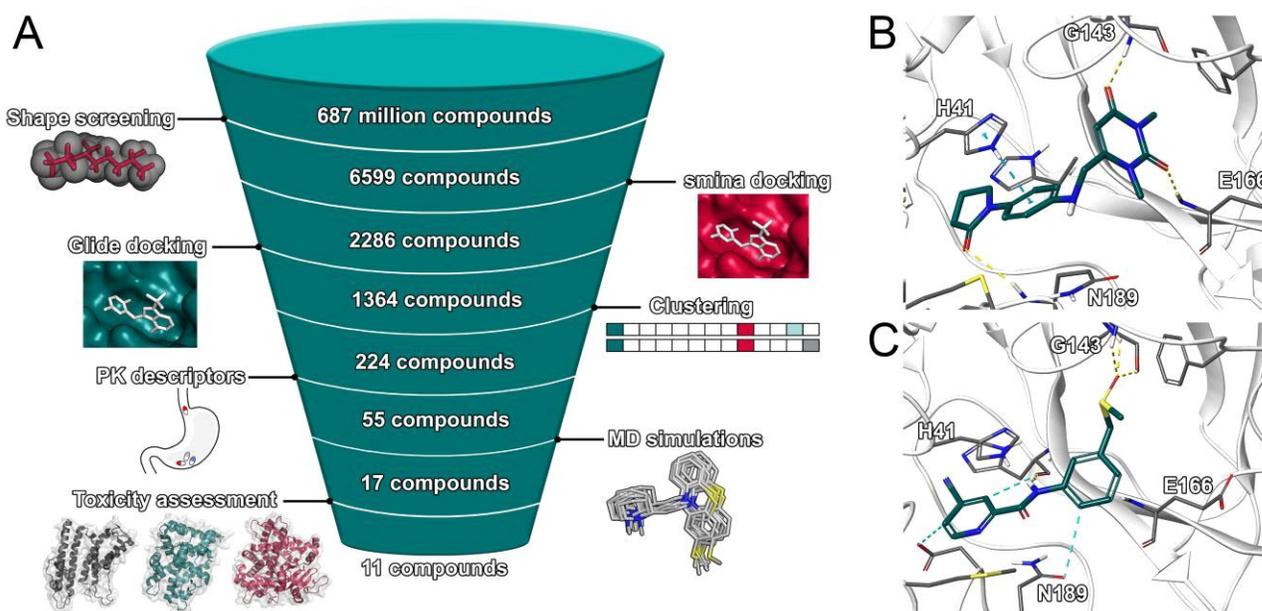


Рисунок 2: Работа по виртуальному скринингу и связывание конфигураций двух верхних лигандов (А) Виртуальный скрининг процесса (В) Связывающая конфигурация CP-1. (С) Связывающая конфигурация CP-2.

Мы определили большое количество начальных совпадений (395132 соединения) для 6599 соединений, выбрав лучшие совпадения в отношении перекрытия формы. Оставшиеся соединения были состыкованы в активный участок репрезентативных структур протеазы с использованием протокола стыковки smina, что привело к 2286 попаданиям с оценкой ниже порога  $-7,0$  ккал / моль, при этом большинство соединений оценивали ниже  $-5,0$  ккал / моль. В то время как большинство предыдущих исследований, касающихся SARS-CoV-2, ограничивались гомологической моделью протеазы<sup>10,12,15,17</sup>, мы могли положиться на недавно решенную кристаллическую структуру для наших процедур стыковки. Фармакокинетические дескрипторы этих соединений показаны на рисунке 3. Из-за происхождения выбранных соединений в базе данных ZINC, можно было ожидать сходство лекарственного средства в отношении молекулярной массы (MW). Однако некоторые структуры нарушают общепринятые критерии распределения коэффициента  $\log D$  и полярной площади поверхности (PSA), что позволяет проводить их разграничение. Благодаря каталитической функции

целевого фермента, пептиды и пептидомиметики широко применяются для нацеливания протеаз.<sup>18</sup> Тем не менее, недостатки пептидов или пептидомиметиков включают ограниченную пероральную биодоступность из-за их большой молекулярной массы, PSA и большого количества вращающихся связей, а также плохой метаболической стабильности и более высоких производственных затрат.<sup>24</sup> Таким образом, разработка малых молекул со сбалансированными и благоприятными фармакокинетическими свойствами, способствующими оральному всасыванию, предлагает многообещающую альтернативу.

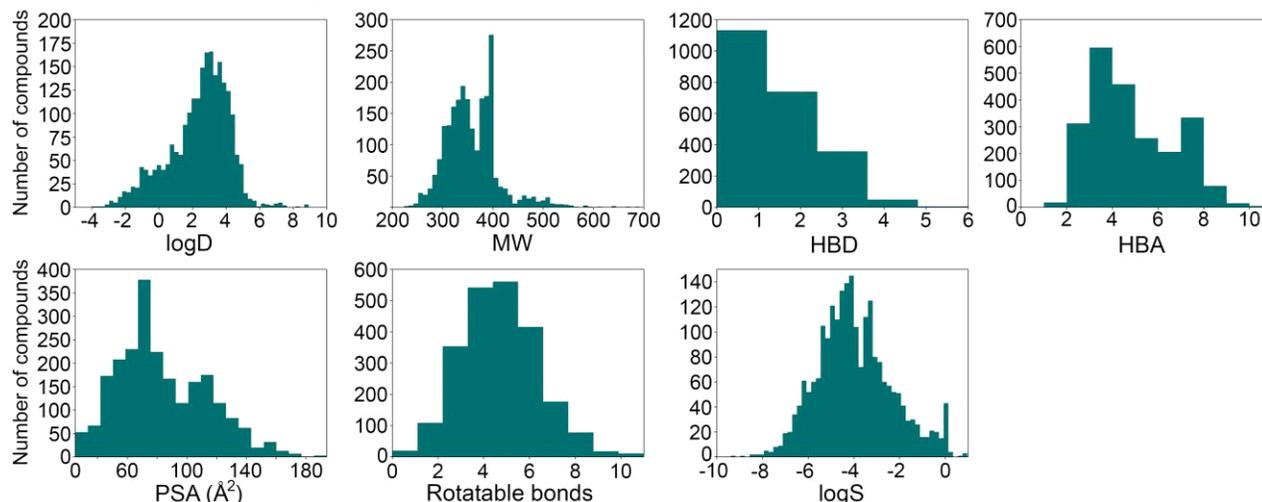


Рисунок 3: Фармакокинетически значимые дескрипторы соединений, которые были подвергнуты протоколу стыковки Glide SP.

В попытке достичь консенсуса в сходстве в связывании мы использовали протокол Glide SP чтобы оценить взаимодействие оставшихся 2286 соединений с активным центром белка. За исключением двух соединений, правильная поза связывания была обнаружена во всех случаях с большинством соединений ниже  $-5,0$  ккал / моль. Набор соединений с показателем Glide ниже  $-6,5$  ккал / моль представлял большое количество родственных лигандов, включая многие структурно сходные природные соединения. Поэтому, чтобы стимулировать структурное разнообразие набора соединений, мы вычислили расширенные отпечатки связности, и сравнили их в соответствии с коэффициентом Танимото для последующей кластеризации. Для каждого кластера были отобраны два соединения с лучшим показателем Glide и оценены в отношении их фармакокинетических дескрипторов. Мы не допустили нарушения критериев Липинского и Вебера, в результате чего было оставлено 55 соединений для моделирования МД. На основе полученного ансамбля лиганд-белковых комплексов, эволюционировавших во времени, свободные энергии связывания лигандов оценивали с использованием протокола MM / GBSA. Этот протокол был недавно применен для прогнозирования энергии взаимодействия между *nel navig* и основной протеазой SARS-CoV-2.<sup>10</sup> Соединения с улучшением свободной энергии связывания в отличие от сокристаллизованного лиганда были рассмотрены в отношении их потенциальной токсичности. VirtualToxLab оценивает токсический потенциал небольшой молекулы на основе индивидуального связывания с 16 утвержденными мишенями, включая ядерные рецепторы, метаболические ферменты и человеческий генный канал, связанный с Ether-a-go-go-go Related Gene channel (hERG). Такая оценка на ранней стадии обнаружения лекарств может снизить степень истощения лекарств из-за токсичности и безопасности, которые представляют собой большую долю доклинических и клинических неудач программ разработки лекарств.<sup>25</sup> Остальные 17 соединений включали несколько соединений с токсическим потенциалом выше  $0,5$ , которые можно было бы исключить из окончательного набора. В лигандах, подвергнутых моделированию MD, мы дополнительно выбрали природное соединение с самой низкой свободной энергией связывания, которое было (-) - **таксифолин**.

## Выбор окончательного состава

Мы идентифицировали 11 соединений с более низкой свободной энергией связывания в сочетании с более высоким теоретическим потенциалом абсорбции после перорального введения по сравнению с сокристаллизованным

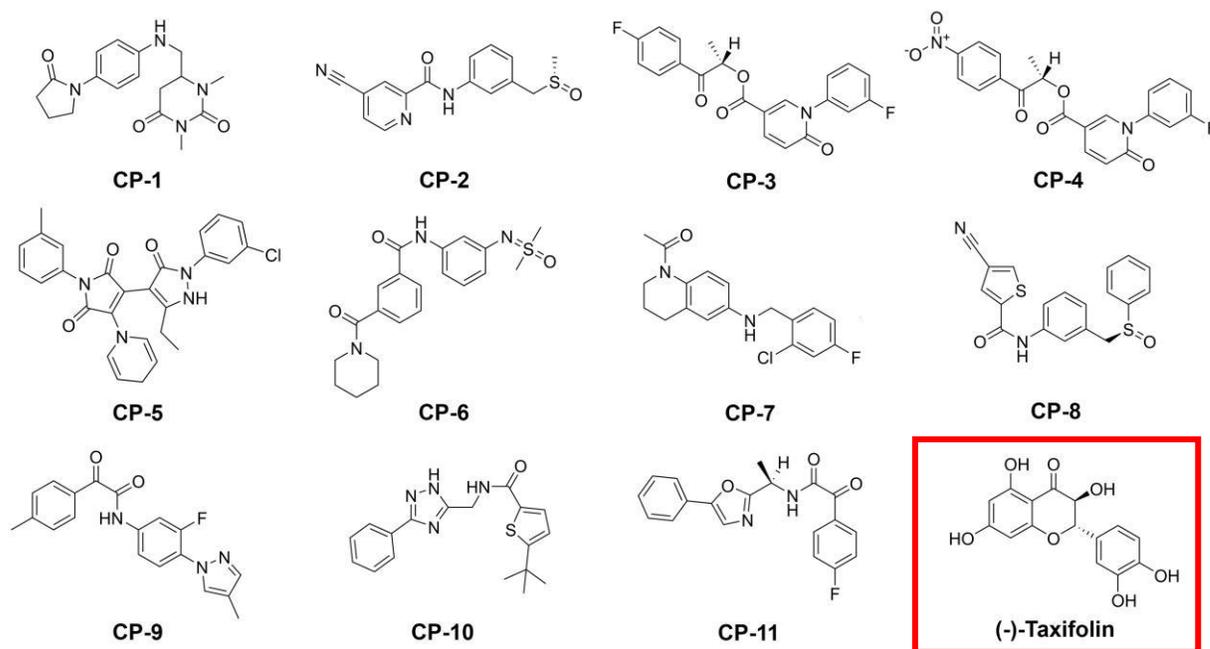


Рисунок 4: Структуры нашего окончательного выбора соединений и самого высокого оцененного природного соединения.

лиганд N3 (таблица 1 и рисунок 4). Сообщенный набор лигандов следует рассматривать как ранние соединения свинца, так как не проводилась экспериментально подтвержденная оптимизация. В то время как некоторые соединения имели лишь незначительное снижение в отношении энергии связи, CP-1-CP-4 продемонстрировали существенное улучшение по сравнению с сокристаллизованным лигандом почти до 30%. Исследование, в котором в основном использовался искусственный интеллект в форме нейронной сети для обнаружения ингибиторов SARS-CoV-2, дополнительно тестировало их соединение с использованием AutoDock Vina. Их соединения имеют показатели стыковки в диапазоне от -7,3 до -8,5 ккал / моль,<sup>12</sup> что сопоставимо с нашим выбором лигандов (Таблица S1, Дополнительная информация). Предлагаемые нами противовирусные препараты взаимодействовали с мишенью, по крайней мере, с одной водородной связью в среднем более двух (рис. 2A и 2B, таблица S2). Кроме того, они показали отличные фармакокинетические дескрипторы со значениями logD в диапазоне от 0 до 4,6, MWt ниже 487 г / моль, PSA ниже 105,5 Å и акцепторами водородных связей, а также донорами в диапазоне, описанном Lipinski.<sup>26</sup> Следовательно, соединения обладают преимуществом перед обычно описанными ингибиторами пептидомиметиков.<sup>18</sup> В отличие от других предлагаемых ингибиторов SARS-CoV-2,<sup>18</sup> доступность соединений помечена как для продажи на складе или по требованию в базе данных ZINC (таблица S1). Что касается потенциальной токсичности при связывании о-цели, соединения регулярно демонстрировали токсические потенциалы, близкие к 0,5. Два верхних соединения CP-1 и CP-2 имели сравнительно низкие значения и поэтому являются отличными кандидатами для дальнейшего исследования *in vitro*.

В то же время природное соединение (-) - **таксифолин** проявляет только низкий потенциал связывания с 16 внецелевыми веществами. Ранее считалось, что природные соединения, такие как авоноиды, ингибируют основную протеазу SARS-CoV-1 и их эффективность была продемонстрирована с помощью резонансного переноса энергии флуоресценции (FRET).<sup>18,27</sup> Кроме того, *in vivo* эксперименты с (-) - **таксифолином** проявляют активность против других вирусов, таких как вирус Коксаки В4.<sup>28</sup> Известно, что **лиственница сибирская (Larix sibirica)** вырабатывает (-) - **таксифолин**, являющийся природным ресурсом для его экстракции. Кроме того, пищевые препараты, содержащие (-) - **таксифолин**, легко доступны в аптеках и других компаниях, обеспечивая прямой и быстрый доступ к потенциальному противовирусному препарату. Примечательно, что (-) - **таксифолин** переносит семь водородных связей, что является самым высоким показателем в нашем выборе соединений.

**Поскольку водородные связи являются ключевым фактором, определяющим особенность лекарственного средства<sup>29</sup> (-) - таксифолин может предложить природную альтернативу предлагаемым ингибиторам CP-1-CP-11.**

## Выводы

Вспышка SARS-CoV-2 по всему миру и связанные с ней последствия являются впечатляющим напоминанием об угрозе зоонозных заболеваний для общественного здравоохранения и экономики многих стран.<sup>30</sup> Отсутствие специальной терапии против нового вируса требует открытия новых лекарственных соединений, для которых вычислительные методы предлагают быстрый и экономически эффективный подход. Здесь мы использовали виртуальный скрининг, состоящий из семи отдельных этапов, чтобы в конечном итоге определить 11 потенциальных связующих. Для всех описанных соединений мы оценили токсичность от низкой до умеренной, вызванную связыванием о-мишени, улучшенным сходством с лекарственным средством, а также повышенным связыванием с терапевтической мишенью по сравнению с сокристаллизованным лигандом. **Кроме того, мы сообщаем о природном соединении таксифолин, встречающемся в сибирской лиственнице, которое предлагает многообещающую альтернативу с коммерчески доступными растительными экстрактами и производными добавками.**

Таблица 1: Окончательный выбор соединений по сравнению с сокристаллизованным лигандом и природным соединением с наибольшим количеством баллов.

Соединение	$\Delta G^a$ (Ккал / моль)	Оценка <sup>b</sup> (Ккал / моль)	М.в. <sup>c</sup> (г/моль)	logD	logS	PSA <sup>d</sup> (Å)	ToxPot <sup>e</sup>
CP-1	-70,6 3,9	-14,5	328,4	0	-2,3	73,0	0,352
CP-2	-69,9 4,0	-14,6	299,4	0,6	-2,5	82,9	0,245
CP-3	-66,7 4,7	-15,8	383,4	3,8	-5,4	63,7	0,492
CP-4	-65,3 3,6	-15,2	391,4	3,5	-6,0	105,5	0,494
CP-5	-60,3 4,8	-15,5	487,0	3,5	-5,9	73,0	0,291
CP-6	-58,2 6,0	-15,6	399,5	2,3	-5,2	78,8	0,257
CP-7	-56,3 6,2	-14,6	332,8	3,5	-4,9	32,3	0,482
CP-8	-56,3 5,7	-15,3	366,5	3,4	-5,2	70,0	0,154
CP-9	-56,1 4,6	-14,7	337,4	4,2	-5,1	64,0	0,418
CP-10	-56,0 4,5	-14,9	340,5	4,6	-5,2	70,7	0,499
CP-11	-55,2 3,7	-14,9	338,3	2,9	-4,9	72,2	0,482
N3	-55,0 6,9	-16,0	587,8	-3,1	-0,6	205,3	нет данных <sup>f</sup>
(-) - <b>taxifolin</b>	-53,3 5,1	-16,8	304,3	1,7	-2,1	127,5	0,289

Энергия связывания без лигандов, предсказанная методом MM / GBSA (исключая энтропийные вклады) со стандартным отклонением; <sup>b</sup> Консенсусная стыковочная оценка, основанная на сумме наивысших баллов *smina* и *Glide*; <sup>c</sup> Молекулярный вес; <sup>d</sup> площадь полярной поверхности; <sup>e</sup> Токсический потенциал вычислено *VirtualToxLab*. <sup>f</sup> Значение недоступно.

К этой дате наше исследование является самым обширным аналитическим скринингом для обнаружения основных ингибиторов протеазы SARS-CoV-2. Экспериментальная проверка и последующая оптимизация предложенных нами ранних соединений свинца могут предложить ценную стратегию по борьбе с SARS-CoV-2.

## Материалы и Методы

### Подготовка системы и генерация согласованности

Структура белка основной протеазы SARS-CoV-2 была получена из Банка данных белка (PDB ID 6LU7) и обработана с помощью мастера подготовки белка, который входит в комплект Maestro Small-Molecule Drug Discovery Suite (v2019-4)<sup>31</sup> путем добавления атомов водорода, прогнозирование протонного состояния при физиологическом pH (pH = 7,4) и переориентация связывающей водорода сети. Выравнивание последовательностей по протеазе SARS-CoV-1 проводили с использованием набором инструментов Ugene в сочетании с алгоритмом ClustalW.<sup>32,33</sup> Затем систему подвергли сдержанной минимизации с преобразованием среднеквадратичного отклонения (RMSD) с конвергенцией порогового значения 0,3Å с использованием силового поля OPLS 2005. Чтобы получить согласованность структур для следующих процедур было выполнено моделирование MD длительностью 20 нс с использованием механизма моделирования Desmond (v2019-1).<sup>34</sup> Чтобы сольватировать орторомбическую периодическую границу в системах, выходящих на 10Å из комплекса во всех измерениях, мы выбрали модель воды TIP3P. Сначала был выполнен протокол расслабления по умолчанию Десмонда, после чего последовала фаза производства при 310 K в комплекте NPT. Для всех производственных симуляций в этом исследовании мы применяли баростат Martyna {Tobias {Klein, со временем релаксации 2,0 пс при 300 K и термостат Носа {Гувера, со временем релаксации 1,0 пс. Мы использовали U-серию алгоритма для обработки далекодействующих взаимодействий с отсечением 9Å для ближних взаимодействий<sup>35</sup> и алгоритм M-SHAKE для ограничения связей с атомами водорода. Временной шаг интегратора RESPA оставался равным 2,0 фс, а снимки с атомными координатами сохранялись с интервалом 10 пс. Используя полученную траекторию моделирования, мы использовали скрипт trj cluster.py, предоставленный в Maestro, который основан на алгоритме распространения объектов, для определения наиболее представительных структур протеазы.

### Фармакокинетические дескрипторы

Химические дескрипторы по Липински и Веберу<sup>26</sup>, а также растворимость в воде (logS), известная как влияющая на фармакокинетику лекарств-кандидатов, определяли с использованием модуля sxcalc из набора Marvin (v20.4.0).<sup>36</sup> Вместо обычно используемого значения logP мы решили использовать более репрезентативное значение logD для прогнозирования разделения лигандов по состояниям ионизации.<sup>37</sup> Свойства, зависящие от ионизации соединения, рассчитывали при физиологическом значении pH 7,4. Дескрипторы были вычислены для всех лигандов, которые были предоставлены протоколу стыковки smina.

### Стыковка и проверка формы

На первом этапе был использован протокол скрининга формы с ускорением GPU от Maestro для быстрого сравнения лекарствовподобного средства части трехмерной библиотеки ZINC<sup>38</sup>. Кроме того, из-за сравнительно большого размера сокристаллизованного пептидомиметического ингибитора были рассмотрены соединения с молекулярной массой выше 500 г / моль. В качестве матриц - субструктура сокристаллизованного лиганда протеазы SARS-CoV-2 (соединение 4), сокристаллизованного лиганда (соединение 9) и три субструктуры эквивалентных ингибиторов протеазы SARS-CoV-1 (соединения 1, 3 и 7), а также были выбраны четыре известных ингибитора SARS-CoV-1 со связыванием ниже 10 M (соединения 2, 5, 6 и 8), полученные из базы данных PubChem<sup>39</sup> (ср. Рисунок 2). Полученные

экспериментально ингибиторы были отобраны на основе их структурного разнообразия, оцененного по отпечаткам молекулярной расширенной связности, с последующим сравнением Tanimoto и последующей кластеризацией на панели Discovery Informatics в Maestro. Для сравнения формы был выбран порог 45% перекрытия формы. Лучшие 5000 совпадений по схожести формы, а также 500 лучших составов каждого входного шаблона были отобраны для дальнейших шагов.

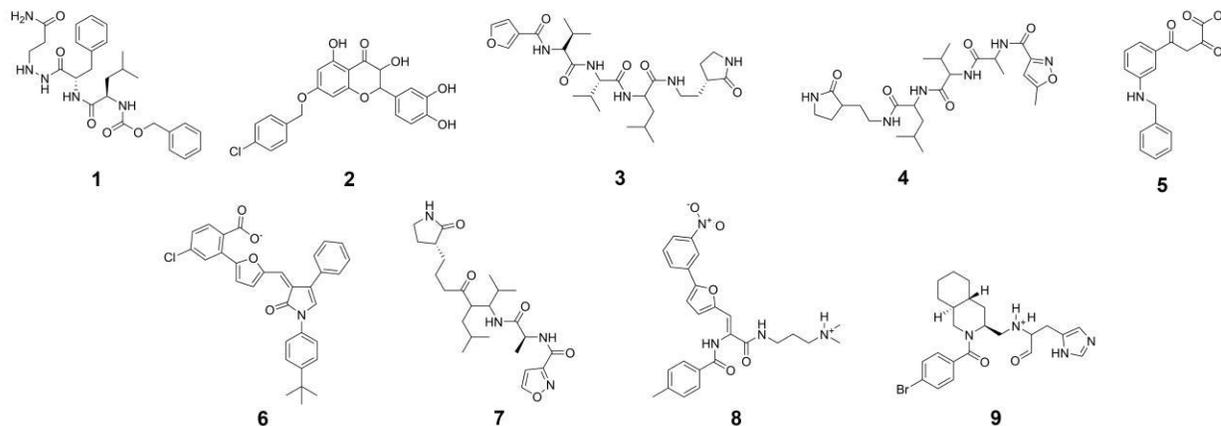


Рисунок 5: Входные структуры для скрининга формы, полученные из кристаллических структур и базы данных Pub-Chem.

Затем мы использовали протокол стыковки *smina*<sup>40</sup>, ответвление AutoDock Vina (v1.1.2),<sup>41</sup> для выбранных соединений, чтобы количественно оценить их взаимодействие с каждой из структур-представителей из сгенерированной согласованности. Файлы лигандов PDBQT для стыковки были созданы с использованием *prepare ligand4.py* script, поставляемого с AutoDockTools (v1.5.4).<sup>42</sup> Стыковка была выполнена при исчерпывающей 16 пространство поиска, было определено с использованием автоматического метода с *smina*, и было выбрано случайное начальное число 42. Попадания со свободной энергией связывания ниже  $-7,0$  ккал / моль были затем подвергнуты связывающему протоколу стандартной точности (SP) Glide по умолчанию в Maestro.<sup>43</sup> Аналогично, взаимодействие со всеми ансамблевыми структурами было количественным. Соединения со свободной энергией связи ниже  $-6,5$  ккал / моль затем подвергали кластеризации на основе сравнения Tanimoto молекулярных отпечатков на атом. Для каждого из 138 полученных кластеров два лучших соединения (одно соединение в случае одной структуры на кластер) в соответствии с показателем Glide были выбраны для дальнейшей оценки. Коммерческая доступность соединений была получена с веб-сайта ZINC. Конечный набор соединений, которые были подвергнуты моделированию MD, был выбран в соответствии со следующими критериями: (i) доступность не уменьшилась, (ii) нет нарушений критериев Липинского и Вебера, (ii.) Средний балл по ансамблю ниже  $-6,5$  ккал / моль в стыковке *smina*, (iv.) средний балл по ансамблю ниже  $-6,0$  ккал / моль при стыковке Glide SP, и (v.) сумма лучших показателей стыковки в любом протоколе ниже  $-14,5$  ккал / моль. Такую же процедуру стыковки и ранжирования проводили с сокристаллизованным лигандом N3.

## Моделирование MD и постобработка

Чтобы понять энергетику стабильности и взаимодействия ранее отфильтрованного набора лигандов, мы провели моделирование MD каждого соединения. В качестве исходных данных для моделирования мы выбрали комплекс лиганд-белок с самым высоким показателем стыковки Glide SP среди ансамблевых структур. Продолжительность моделирования была выбрана равной 25 нс с записью атомных координат с интервалом 25 пс, а остальные настройки были оставлены, как описано выше.

Стабильность режимов привязки определялась с помощью RMSD на панели Simulation Inter-action Diagram в Maestro. Свободную энергию связи количественно определяли, используя процедуру mal mmgbsa.py, которая поставляется с Maestro для последних 100 кадров траектории с размером шага два. Эта процедура основана на вычислениях MM / GBSA, которые по существу сравнивают апо и голо состояния фермента с различными энталпическими терминами. Полученные свободные энергии связи были усреднены для выбранных систем траекторий.

## **Оценка токсичности и выбор конечного набора соединений**

Потенциальные токсические эффекты соединений, которые были подвергнуты моделированию MD, оценивали с использованием VirtualToxLab21,<sup>44</sup> который оценивает связывание соединения-зонда с 16 известными о-мишенями. Структуры-мишени включают рецептор андрогена (AR), арильный углеводородный рецептор (AhR), цитохром P450 1A2 (CYP1A2), CYP2C9, CYP2D6, CYP3A4, рецептор эстрогена (ER), ER, глюкокортикоидный рецептор (GR), минералокортикоидный рецептор (MR), X рецептор печени (LXR), канал hERG, рецептор, активируемый пролифератором пероксисима (PPAR), рецептор прогестерона (PR), рецептор щитовидной железы (TR) и TR. Процедура количественно определяет степень связывания с каждой мишенью и, основываясь на них, оценивает числовое значение токсичности, называемое токсическим потенциалом, в диапазоне от 0,0 до 1,0.

Набор лигандов, подвергнутых моделированию MD, был основан на их токсическом потенциале и средней свободной энергии связывания для получения конечного набора соединений. Были рассмотрены соединения с энергией связи MM / GBSA ниже -55,0 ккал / моль и токсическим потенциалом ниже 0,5.

## **Финансирование**

Внешнее финансирование не было получено для этого исследования.

## **Выражение признательности**

Авторы выражают благодарность за поддержку корпорации NVIDIA пожертвованием графического процессора Titan Xp, использованного для этого исследования.

## **Доступна вспомогательная информация**

Подробная информация обо всех 55 соединениях, подвергнутых моделированию MD, а также информация о связывающих позах, фармакокинетических свойствах и водородных связях для окончательного выбора соединений доступна в отдельном документе PDF с дополнительной информацией.

## Справочные документы

1. 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. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *The Lancet* 2020, 395, 507{513.
2. Zhou, P. et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 2020,
3. World Health Organisation, Novel Coronavirus (2019-nCoV) situation reports. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/>.
4. Wang, F.; Chen, C.; Tan, W.; Yang, K.; Yang, H. Structure of Main Protease from Human Coronavirus NL63: Insights for Wide Spectrum Anti-Coronavirus Drug Design. *Scientific Reports* 2016, 6, 22677.
5. Chen, Z.-M. et al. Diagnosis and treatment recommendations for pediatric respiratory infection caused by the 2019 novel coronavirus. *World journal of pediatrics : WJP* 2020,
6. Wu, F. et al. A new coronavirus associated with human respiratory disease in China. *Nature* 2020,
7. World Health Organisation, International Health Regulations Emergency Committee on novel coronavirus in China. <https://www.who.int/news-room/events/detail/2020/01/30/default-calendar/international-health-regulations-emergency-committee-on-novel-coronavirus-in-china>.
8. Stoermer, M. J. Homology Models of the Papain-Like Protease PLpro from Coronavirus 2019-nCoV. 2020,
9. Keogh-Brown, M. R.; Smith, R. D. The economic impact of SARS: How does the reality match the predictions? *Health Policy* 2008, 88, 110{120.
10. Zhijian Xu, Y. S. Z. Z. K. M. X. W. W. Z., Cheng Peng Nel navir was predicted to be a potential inhibitor of 2019-nCoV main protease by an integrative approach combining homology modelling, molecular docking and binding free energy calculation. 2020, 1201, 1{20.
11. Zumla, A.; Hui, D. S.; Azhar, E. I.; Memish, Z. A.; Maeurer, M. Reducing mortality from 2019-nCoV: host-directed therapies should be an option. *The Lancet* 2020, 395, e35{e36.
12. Zhang, H.; Saravanan, K. M.; Yang, Y.; Hossain, T. Deep learning based drug screening for novel coronavirus 2019-nCoV. 2020, 19, 1{17.
13. Yang, S. et al. Synthesis, Crystal Structure, Structure-Activity Relationships, and Antiviral Activity of a Potent SARS Coronavirus 3CL Protease Inhibitor. *Journal of Medicinal Chemistry* 2006, 49, 4971{4980.
14. Ghosh, A. K.; Osswald, H. L.; Prato, G. Recent Progress in the Development of HIV-1 Protease Inhibitors for the Treatment of HIV/AIDS. *Journal of Medicinal Chemistry* 2016, 59, 5172{5208.
15. Liu, X.; Wang, X.-J. Potential inhibitors against 2019-nCoV coronavirus M protease from clinically approved medicines. *Journal of Genetics and Genomics* 2020,
16. Liu, Z.-B. J. Z. Y. H. R. Z., X. The crystal structure of COVID-19 main protease in complex with an inhibitor N3. <http://www.rcsb.org/structure/6LU7>.
17. Li, Y.; Zhang, J.; Wang, N.; Li, H.; Shi, Y.; Guo, G.; Liu, K.; Zeng, H.; Zou, Q. Therapeutic Drugs Targeting 2019-nCoV Main Protease by High-Throughput Screening. *bioRxiv* 2020, 2020.01.28.922922.
18. Pillaiyar, T.; Manickam, M.; Namasivayam, V.; Hayashi, Y.; Jung, S. H. An overview of severe acute respiratory syndrome-coronavirus (SARS-CoV) 3CL protease inhibitors: Peptidomimetics and small molecule chemotherapy. *Journal of Medicinal Chemistry* 2016, 59, 6595{6628.
19. Raugi, D. N.; Smith, R. A.; Gottlieb, G. S. Four Amino Acid Changes in HIV-2 Protease Confer Class-Wide Sensitivity to Protease Inhibitors. *Journal of Virology* 2016, 90, 1062{1069.
20. Chen, Y. W.; Yiu, C.-p.; Wong, K.-y. Prediction of the 2019-nCoV 3C-like protease (3CLpro) structure virtual screening reveals velpatasvir, ledipasvir, and other drug repurposing candidates.pdf.crdownload. 2020,
21. Vedani, A.; Dobler, M.; Hu, Z.; Smiesko, M. OpenVirtualToxLab-A platform for generating and exchanging in silico toxicity data. *Toxicology Letters* 2015, 232, 519{532.
22. Gao, K.; Nguyen, D. D.; Wang, R.; Wei, G.-W. Machine intelligence design of 2019-nCoV drugs. *bioRxiv* 2020, 2020.01.30.927889.
23. Lionta, E.; Spyrou, G.; Vassilatis, D.; Cournia, Z. Structure-Based Virtual Screening for Drug Discovery: Principles, Applications and Recent Advances. *Current Topics in Medicinal Chemistry* 2014, 14, 1923{1938.

24. Recio, C.; Maione, F.; Iqbal, A. J.; Mascolo, N.; De Feo, V. The potential therapeutic application of peptides and peptidomimetics in cardiovascular disease. *Frontiers in Pharmacology* 2017, 7, 1{11}.
25. Waring, M. J.; Arrowsmith, J.; Leach, A. R.; Leeson, P. D.; Mandrell, S.; Owen, R. M.; Pairaudeau, G.; Pennie, W. D.; Pickett, S. D.; Wang, J.; Wallace, O.; Weir, A. An analysis of the attrition of drug candidates from four major pharmaceutical companies. *Nature Reviews Drug Discovery* 2015, 14, 475{486}.
26. Walters, W. P. Going further than Lipinski's rule in drug design. *Expert Opinion on Drug Discovery* 2012, 7, 99{107}.
27. Nguyen, T. T. H.; Woo, H. J.; Kang, H. K.; Nguyen, V. D.; Kim, Y. M.; Kim, D. W.; Ahn, S. A.; Xia, Y.; Kim, D. Flavonoid-mediated inhibition of SARS coronavirus 3C-like protease expressed in *Pichia pastoris*. *Biotechnology Letters* 2012, 34, 831{838}.
28. Galochkina, A. V.; Anikin, V. B.; Babkin, V. A.; Ostrouhova, L. A.; Zarubaev, V. V. Virus-inhibiting activity of dihydroquercetin, a flavonoid from *Larix sibirica*, against coxsackievirus B4 in a model of viral pancreatitis. *Archives of Virology* 2016, 161, 929{938}.
29. Wade, R. C.; Goodford, P. J. The role of hydrogen-bonds in drug binding. *Progress in clinical and biological research* 1989, 289, 433{444}.
30. Hui, D. S.; I Azhar, E.; Madani, T. A.; Ntoumi, F.; Kock, R.; Dar, O.; Ippolito, G.; Mchugh, T. D.; Memish, Z. A.; Drosten, C.; Zumla, A.; Petersen, E. The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health | The latest 2019 novel coronavirus outbreak in Wuhan, China. *International Journal of Infectious Diseases* 2020, 91, 264{266}.
31. Schrodinger LCC, Maestro Small-Molecular Drug Discovery Suite 2019-4. 2019,
32. Okonechnikov, K.; Golosova, O.; Fursov, M.; Varlamov, A.; Vaskin, Y.; Efremov, I.; German Grehov, O. G.; Kandrov, D.; Rasputin, K.; Syabro, M.; Tleukenov, T. Unipro UGENE: A unified bioinformatics toolkit. *Bioinformatics* 2012, 28, 1166{1167}.
33. Thompson, J. D.; Higgins, D. G.; Gibson, T. J. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* 1994, 22, 4673{4680}.
34. Bowers, K.; Chow, E.; Xu, H.; Dror, R.; Eastwood, M.; Gregersen, B.; Klepeis, J.; Kolossvary, I.; Moraes, M.; Sacerdoti, F.; Salmon, J.; Shan, Y.; Shaw, D. Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. *ACM/IEEE SC 2006 Conference (SC'06)* 2006, 43.
35. Shaw, D. E. et al. Anton 2: Raising the Bar for Performance and Programmability in a Special-Purpose Molecular Dynamics Supercomputer. *International Conference for High Performance Computing, Networking, Storage and Analysis, SC 2014, 2015-Janua*, 41{53}.
36. ChemAxon, Marvin (v.20.4.0). 2020.
37. Bhal, S. K.; Kassam, K.; Peirson, I. G.; Pearl, G. M. The rule of five revisited: Applying log D in place of log P in drug-likeness filters. *Molecular Pharmaceutics* 2007, 4, 556{ 560}.
38. Sterling, T.; Irwin, J. J. ZINC 15 - Ligand Discovery for Everyone. *Journal of Chemical Information and Modeling* 2015, 55, 2324{2337}.
39. Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B. A.; Thiessen, P. A.; Yu, B.; Zaslavsky, L.; Zhang, J.; Bolton, E. E. PubChem 2019 update: improved access to chemical data. *Nucleic acids research* 2019, 47, D1102{D1109}.
40. Koes, D. R.; Baumgartner, M. P.; Camacho, C. J. Lessons learned in empirical scoring with smina from the CSAR 2011 benchmarking exercise. *Journal of Chemical Information and Modeling* 2013, 53, 1893{1904}.
41. Brooks, B. R. et al. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *Journal of computational chemistry* 2009, 30, 1545{1614}.
42. Morris, G.; Huey, R. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem.* 2009, 30, 2785{2791}.
43. Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *Journal of Medicinal Chemistry* 2004, 47, 1750{1759}.
44. Vedani, A.; Dobler, M.; Smiesko, M. VirtualToxLab - A platform for estimating the toxic potential of drugs, chemicals and natural products. *Toxicology and Applied Pharmacology* 2012, 261, 142{153}.