

ISSN 1561-8323 (Print)  
ISSN 2524-2431 (Online)

**ХИМИЯ**  
**CHEMISTRY**

УДК 004.94+547.7+616-006+616-085  
<https://doi.org/10.29235/1561-8323-2023-67-3-197-206>

Поступило в редакцию 15.02.2023  
Received 15.02.2023

**А. М. Андрианов<sup>1</sup>, К. В. Фурс<sup>2</sup>, Н. А. Шульдow<sup>2</sup>, член-корреспондент А. В. Тузиков<sup>2</sup>**

<sup>1</sup>Институт биоорганической химии Национальной академии наук Беларуси,  
Минск, Республика Беларусь

<sup>2</sup>Объединенный институт проблем информатики Национальной академии наук Беларуси,  
Минск, Республика Беларусь

**DE NOVO ДИЗАЙН ПОТЕНЦИАЛЬНЫХ ИНГИБИТОРОВ ОСНОВНОЙ ПРОТЕАЗЫ  
КОРОНАВИРУСА SARS-CoV-2 С ПОМОЩЬЮ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО  
ИНТЕЛЛЕКТА И МОЛЕКУЛЯРНОГО МОДЕЛИРОВАНИЯ**

**Аннотация.** С помощью генеративной нейронной сети глубокого обучения, разработанной ранее на основе технологий искусственного интеллекта, осуществлен *de novo* дизайн 95 775 потенциальных лигандов основной протеазы (Mpro) SARS-CoV-2, играющей важную роль в процессе репликации вируса. Методами молекулярного докинга и молекулярной динамики выполнена оценка аффинности связывания этих молекул с каталитическим сайтом фермента. В результате проведенных исследований отобраны 7 соединений-лидеров, которые характеризуются низкими значениями свободной энергии Гиббса, сопоставимыми с величинами, полученными с помощью идентичного вычислительного протокола для двух мощных нековалентных ингибиторов Mpro SARS-CoV-2, использованных в расчетах в качестве позитивного контроля. Полученные результаты свидетельствуют о перспективности использования идентифицированных соединений в работах по созданию новых противовирусных препаратов, терапевтическое действие которых основано на ингибировании каталитической активности Mpro SARS-CoV-2.

**Ключевые слова:** генеративные нейронные сети глубокого обучения, SARS-CoV-2, основная протеаза, молекулярный докинг, молекулярная динамика, противовирусные препараты

**Для цитирования.** *De novo* дизайн потенциальных ингибиторов основной протеазы коронавируса SARS-CoV-2 с помощью технологий искусственного интеллекта и молекулярного моделирования / А. М. Андрианов [и др.] // Докл. Нац. акад. наук Беларуси. – 2023. – Т. 67, № 3. – С. 197–206. <https://doi.org/10.29235/1561-8323-2023-67-3-197-206>

**Alexander M. Andrianov<sup>1</sup>, Konstantin V. Furs<sup>2</sup>, Mikita A. Shuldau<sup>2</sup>,  
Corresponding Member Alexander V. Tuzikov<sup>2</sup>**

<sup>1</sup>Institute of Bioorganic Chemistry of the National Academy of Sciences of Belarus, Minsk, Republic of Belarus

<sup>2</sup>United Institute of Informatics Problems of the National Academy of Sciences of Belarus, Minsk, Republic of Belarus

**DE NOVO DESIGN OF POTENTIAL SARS-CoV-2 MAIN PROTEASE INHIBITORS USING ARTIFICIAL  
INTELLIGENCE AND MOLECULAR MODELING TECHNOLOGIES**

**Abstract.** *De novo* design of 95 775 potential ligands of SARS-CoV-2 main protease (Mpro), playing an important role in the process of virus replication, was carried out using a deep learning generative neural network that was developed previously based on artificial intelligence technologies. Molecular docking and molecular dynamics methods were used to evaluate the binding affinity of these molecules to the catalytic site of the enzyme. As a result, 7 leading compounds exhibiting Gibbs free energy low values comparable with the values obtained using an identical computational protocol for two potent non-covalent SARS-CoV-2 Mpro inhibitors used in calculations as a positive control were selected. The results obtained indicate the promise of applying identified compounds for development of new antiviral drugs able to inhibit the catalytic activity of SARS-CoV-2 Mpro.

**Keywords:** deep learning generative neural networks, SARS-CoV-2, main protease, molecular docking, molecular dynamics, antiviral drugs

**For citation.** Andrianov A. M., Furs K. V., Shuldau M. A., Tuzikov A. V. *De novo* design of potential SARS-CoV-2 main protease inhibitors using artificial intelligence and molecular modeling technologies. *Doklady Natsional'noi akademii nauk Belarusi* = *Doklady of the National Academy of Sciences of Belarus*, 2023, vol. 67, no. 3, pp. 197–206 (in Russian). <https://doi.org/10.29235/1561-8323-2023-67-3-197-206>

**Введение.** В последние годы генеративные модели глубокого обучения нашли широкое применение в исследованиях по разработке лекарств *de novo* [1]. Благодаря огромному прогрессу методов глубокого обучения в настоящее время разработаны генеративные модели с различной архитектурой и разными методами обучения, использующие разные типы и структуры данных. Применение этих моделей уже показало их способность генерировать соединения, которые могут быть синтезированы, активны *in vitro*, стабильны и проявляют активность *in vivo* в моделях, связанных с различными заболеваниями [2]. Однако несмотря на то что генеративные модели глубокого обучения становятся все более распространенными в химической и биологической информатике, их потенциал в этой области еще не раскрыт в полном объеме. В связи с этим разработка и применение генеративных методов глубокого обучения для компьютерного конструирования потенциальных лекарственных препаратов имеют большое научное и практическое значение. При этом использование технологий глубокого обучения в сочетании с методами виртуального скрининга существенно расширяет их возможности и позволяет значительно сократить время и затраты, необходимые для создания новых терапевтических средств [3]. Среди методов виртуального скрининга важная роль принадлежит молекулярному докингу, который широко используется для предсказания ориентации лигандов в сайте связывания белковых мишеней, предсказания свободной энергии образования комплексов и исследования профиля межмолекулярных взаимодействий [4]. Молекулярная динамика (МД) также является мощным вычислительным инструментом для решения задач, ориентированных на идентификацию малых молекул, формирующих перспективные базовые структуры для разработки новых лекарственных препаратов [5]. В отличие от молекулярного докинга, МД моделирует движения каждого атома в поле остальных атомов, более эффективно, чем другие алгоритмы, отражает гибкость как лиганда, так и белка и позволяет с более высокой точностью предсказывать аффинность связывания [5]. Примеры успешных приложений этих вычислительных подходов к виртуальному скринингу химических соединений с заданными свойствами наглядно демонстрируют их возможности для создания новых низкомолекулярных лекарств-кандидатов [5].

Настоящая работа продолжает исследования, начатые в [6], в которой были разработаны две генеративные модели глубокого обучения для *de novo* дизайна потенциальных ингибиторов основной протеазы (Mpro; Main Protease) SARS-CoV-2 – фермента, критически важного для репликации и транскрипции вируса, и, поэтому, представляющего перспективную мишень для конструирования эффективных противовирусных препаратов [7]. В связи с этим в настоящее время ведутся интенсивные исследования по разработке новых эффективных ингибиторов Mpro SARS-CoV-2, подробная информация о которых представлена в [8]. Эти исследования позволили обнаружить ряд перспективных природных ингибиторов фермента из растений, морских организмов и микроорганизмов, синтезировать ковалентные пептидомиметики и малые молекулы, проявляющие анти-SARS-CoV-2 активность в инфицированных клетках человека [8]. Однако мутации генома коронавируса, ведущие к появлению новых вариантов COVID-19, потенциально устойчивых к используемым в клинике препаратам, обуславливают необходимость разработки новых противовирусных соединений, способных ингибировать каталитическую активность фермента.

Цель данного исследования заключалась в применении разработанных ранее [6] генеративных моделей глубокого обучения для *de novo* дизайна потенциальных ингибиторов Mpro SARS-CoV-2 и последующей идентификации соединений-лидеров методами молекулярного моделирования. Для решения поставленной задачи были выполнены исследования, включавшие:

- 1) *de novo* дизайн потенциальных ингибиторов SARS-CoV-2 с заданной энергией связывания с каталитическим сайтом фермента;
- 2) молекулярный докинг и оценку аффинности связывания сгенерированных соединений с терапевтической мишенью;
- 3) молекулярную динамику комплексов потенциальных лигандов с Mpro SARS-CoV-2 и расчет свободной энергии их образования;
- 4) анализ полученных результатов и отбор соединений, перспективных для тестирования на моделях *in vitro*.

**Материалы и методы исследования.** *De novo* дизайн потенциальных лигандов Mpro SARS-CoV-2. Разработанную ранее [6] генеративную нейронную сеть глубокого обучения использовали

для *de novo* дизайна высокоаффинных лигандов каталитического сайта Mpro SARS-CoV-2 с целью последующей идентификации потенциальных ингибиторов этого фермента методами молекулярного моделирования. В результате были получены линейные представления SMILES (Simplified Molecular Input Line Entry System) ([https://dbpedia.org/page/Simplified\\_molecular-input\\_line-entry\\_system](https://dbpedia.org/page/Simplified_molecular-input_line-entry_system)) для 128 955 молекул. Трехмерные (3D) структуры химических соединений восстанавливали из данного формата с помощью алгоритма, включавшего следующие этапы:

1) чтение структуры и генерация 2D проекций координат атомов молекул из описаний SMILES;

2) генерацию 3D координат атомов молекул;

3) добавление атомов водорода;

4) оптимизацию сгенерированной 3D структуры соединения;

5) проверку сохранения исходной стереохимии молекулы в полученных структурах.

Суммарное количество молекул, отобранных для построения комплексов этих соединений с Mpro SARS-CoV-2 методами молекулярного докинга, составило 95 775.

**Молекулярный докинг.** Молекулярный докинг проводили с помощью программы AutoDock Vina (<https://vina.scripps.edu>) в приближении жесткого рецептора и гибких лигандов. В качестве рецептора использовали структуру Mpro SARS-CoV-2 в кристалле (идентификационный код в Банке данных белков 6Y84; <https://www.rcsb.org>). Ячейка для докинга включала каталитический сайт Mpro SARS-CoV-2 со следующими параметрами:  $\Delta X = 19 \text{ \AA}$ ,  $\Delta Y = 21 \text{ \AA}$ ,  $\Delta Z = 23 \text{ \AA}$  с центром в точках  $X = -20 \text{ \AA}$ ,  $Y = 19 \text{ \AA}$ ,  $Z = -26 \text{ \AA}$ . Значение параметра ширины охвата конформационного пространства лигандов, определяющего количество независимых запусков процедуры поиска, задавали равным 100. Значения свободной энергии связывания лигандов с Mpro SARS-CoV-2 рассчитывали с помощью классической оценочной функции AutoDock Vina (<https://vina.scripps.edu>) и двух функций машинного обучения – RF-Score-4 (<https://pjballester.wordpress.com/software/>) и NNScore 2.0 (<https://git.durrantlab.pitt.edu/jdurrant/nnscore2>). Затем для всех соединений определяли их ранги согласно каждой оценочной функции и на основе этих данных вычисляли величину экспоненциального консенсусного ранга (ECR) [9]. В результате из 95 775 исходных молекул были отобраны лиганды, которые принадлежали 11 группам с высшими значениями ECR, что позволило идентифицировать 39 соединений, комплексы которых с Mpro SARS-CoV-2 исследовали методами молекулярной динамики.

**Молекулярная динамика.** Молекулярную динамику комплексов лиганд/Mpro в воде выполняли в программном пакете Amber18 с использованием силовых полей Amber ff14SB (Mpro) и GAFF (лиганды) (<https://ambermd.org/doc12/Amber18.pdf>). Для задания парциальных зарядов атомов (модель AM1-BCC) использовали модуль Antechamber программного пакета AmberTools18 (<https://ambermd.org/doc12/Amber18.pdf>). Атомы водорода добавляли с помощью программы tleap пакета AmberTools18. Комплексы помещали в кубическую коробку, заполняли растворителем (модель воды TIP3P; <https://ambermd.org/doc12/Amber18.pdf>) и добавляли ионы  $\text{Na}^+$  и  $\text{Cl}^-$  до значения ионной силы, равного 0,15 М. Систему минимизировали методами наискорейшего спуска (500 шагов) и сопряженных градиентов (500 шагов), нагревали от 0 до 300 К в течение 50 пс в рамках статистического ансамбля NVT и термостата Ланжевена, а затем уравнивали в течение 50 пс при давлении 1,0 атм (ансамбль NPT, баростат Берендсена). На заключительном шаге систему уравнивали в течение 0,5 нс при постоянном объеме и проводили молекулярную динамику длительностью 200 нс в изобарно-изотермических условиях при температуре 300 К и давлении 1 атм. Интегрирование уравнений движения Ньютона осуществляли с помощью алгоритма “leap-frog” (<https://ambermd.org/doc12/Amber18.pdf>) с шагом интегрирования 2,0 фс. Для фиксации длин связей, в образовании которых участвуют атомы водорода, применяли алгоритм SHAKE (<https://ambermd.org/doc12/Amber18.pdf>). Максимальное расстояние, на котором учитывали электростатические взаимодействия, задавали равным 8,0 Å. Для расчета энергии электростатических взаимодействий использовали метод Эвальда (<https://ambermd.org/doc12/Amber18.pdf>).

Средние значения энергии связывания для динамических моделей комплексов лиганд/Mpro рассчитывали с помощью метода MM/GBSA [10] в программном пакете Amber18 (<https://ambermd.org/doc12/Amber18.pdf>). При расчете свободной энергии первые 50 нс МД моделирования отводили на релаксацию системы и не учитывали в расчетах. Энергию связывания вычисляли для 150 комплексов МД траектории, разделенных интервалом 1 нс. Для расчета полярной

составляющей энергии сольватации использовали континуальную модель растворителя Пуассона–Больцмана с ионной силой 0,15 М. Неполарные компоненты свободной энергии гидратации вычисляли на основе расчетов площади поверхности, доступной растворителю (<https://ambermd.org/doc12/Amber18.pdf>). Энтропийную компоненту свободной энергии Гиббса рассчитывали с использованием программного модуля Nmode (<https://ambermd.org/doc12/Amber18.pdf>). Анализ МД траекторий выполняли с помощью программного модуля CPPTRAJ пакета AmberTools18 (<https://ambermd.org/doc12/Amber18.pdf>). В качестве контрольных соединений в расчетах использовали два мощных нековалентных ингибитора Mpro SARS-CoV-2 со значениями констант ингибирования  $IC_{50}$ , равными  $18 \pm 2$  нМ (ингибитор I) и  $20 \pm 5$  нМ (ингибитор II). Эти соединения представлены в [11] под номерами 21 и 23 соответственно.

**Результаты и их обсуждение.** Анализ данных молекулярного моделирования позволил выявить семь соединений-лидеров, которые продемонстрировали высокое химическое сродство к каталитическому сайту Mpro SARS-CoV-2 согласно низким значениям свободной энергии связывания, предсказанным для статических и динамических моделей комплексов лиганд/Mpro. Химические структуры этих соединений показаны на рис. 1, а в табл. 1 и 2 приведены их физико-химические параметры, традиционно используемые в качестве основных фильтров для скрининга лигандов на их способность быть эффективными при пероральном применении. Анализ данных табл. 1 показывает, что лиганд I полностью удовлетворяет требованиям, предъявляемым к потенциальному лекарству «правилом пяти» Липинского, которые обеспечивают такие важные характеристики, как всасывание, распределение, метаболизм и экскреция [12]. В то же время лиганды II–VII обнаруживают лишь одно нарушение этого правила, связанное с небольшим превышением их молекулярной массы (табл. 1), что позволяет предположить, что эти соединения также обладают лекарственными свойствами [12]. Кроме того, данные качественной оценки молярной растворимости анализируемых соединений, являющейся одним из основных свойств, влияющих на абсорбцию, указывают на то, что эти молекулы растворимы в воде, о чем свидетельствуют значения  $\log S$ , рассчитанные с помощью программного обеспечения веб-сервера SwissADME (<http://www.swissadme.ch>) (табл. 2). Наконец, расчеты показывают, что анализируемые соединения (рис. 1) могут быть синтезированы, что является основным фактором для выбора наиболее перспективных молекул, идентифицированных методами виртуального скрининга. Это предположение подтверждается оценкой синтетической доступности (СД) этих молекул (табл. 3), которая классифицирует значения СД в диапазоне от 1 (очень легко) до 10 (очень сложно) (<http://www.swissadme.ch>).

На рис. 2 приведены аминокислотные остатки Mpro SARS-CoV-2, которые участвуют в межмолекулярных взаимодействиях, реализующихся в структурных комплексах идентифицированных соединений с каталитическим центром фермента. Анализ рис. 2 показывает, что эти лиганды образуют широкую сеть ван-дер-ваальсовых контактов с такими функционально важными остатками фермента, как His-41, Met-49 (кроме соединения VI), Met-165, Glu-166 и Gln-189. Наряду с ван-дер-ваальсовыми взаимодействиями анализируемые соединения образуют водородные связи с Gly-143 (соединения I, III, IV), His-41 (соединение II), Ser-46 (соединение II), Thr-24 (соединение VI), Thr-26 (соединения II и VI), Glu-166 (соединения II, III, IV, V и VII), Cys-145 (соединение IV), His-163 (соединение V) и Gln-192 (соединение VII) (рис. 2). Кроме того, соединения III, V и VI участвуют в специфических катион- $\pi$  взаимодействиях с His-41, который является частью каталитической диады Mpro, образованной этим остатком и Cys-145 [14]. Наконец,  $\pi$ -сопряженные системы соединений VI и VII образуют  $\pi$ -стэкинг с боковой цепью His-41, а соединение III формирует солевой мостик с Glu-166 (рис. 2). При этом химическое сродство соединений к Mpro обеспечивается главным образом межмолекулярными водородными связями и ван-дер-ваальсовыми взаимодействиями.

Эффективность межмолекулярных взаимодействий, реализующихся в статических моделях комплексов лиганд/Mpro, подтверждается низкими значениями свободной энергии связывания и констант диссоциации, свидетельствующими о высоком сродстве соединений I–VII к каталитическому сайту Mpro SARS-CoV-2 (табл. 3). Данные табл. 3 позволяют предположить, что эти значения, рассчитанные с использованием трех различных оценочных функций, по меньшей мере, сравнимы со значениями, полученными с помощью идентичного вычислительного протокола для контрольных ингибиторов I и II. Результаты МД расчетов приводят к тому же выводу, что



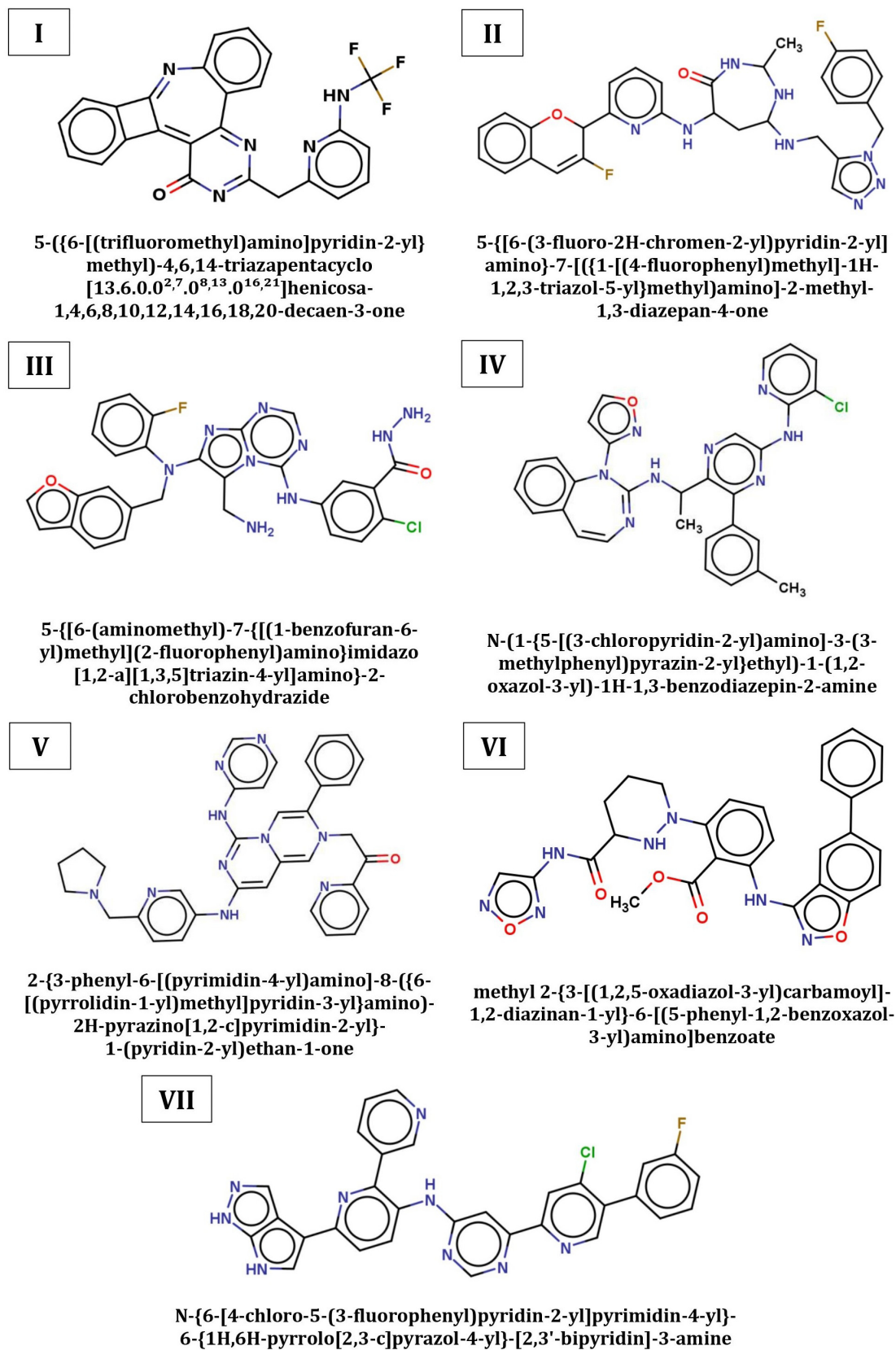


Рис. 1. Химические структуры идентифицированных соединений. Приведены названия соединений согласно систематической номенклатуре ИЮПАК

Fig. 1. Chemical structures of the identified compounds. The names of the compounds are given according to the IUPAC systematic nomenclature

Т а б л и ц а 1. Физико-химические параметры идентифицированных соединений

T a b l e 1. Physicochemical parameters of the identified compounds

Лиганд Ligand	Химическая формула <sup>1</sup> Chemical formula	Молекулярная масса <sup>1</sup> (Да) Molecular mass (Da)	LogP <sup>2</sup>	Число доноров водородной связи <sup>1</sup> Number of H-bond donors	Число акцепторов водородной связи <sup>1</sup> Number of H-bond acceptors
I	C <sub>25</sub> H <sub>14</sub> F <sub>3</sub> N <sub>5</sub> O	457,4	4,43	1	8
II	C <sub>30</sub> H <sub>30</sub> F <sub>2</sub> N <sub>8</sub> O <sub>2</sub>	572,6	2,93	4	9
III	C <sub>28</sub> H <sub>23</sub> C <sub>1</sub> FN <sub>9</sub> O <sub>2</sub>	572,0	3,48	5	8
IV	C <sub>30</sub> H <sub>25</sub> C <sub>1</sub> N <sub>8</sub> O	549,0	4,73	2	6
V	C <sub>34</sub> H <sub>32</sub> N <sub>10</sub> O	596,7	3,62	2	7
VI	C <sub>28</sub> H <sub>25</sub> N <sub>7</sub> O <sub>5</sub>	539,5	3,43	3	9
VII	C <sub>30</sub> H <sub>19</sub> C <sub>1</sub> FN <sub>9</sub>	560,0	4,93	3	7

П р и м е ч а н и я: <sup>1</sup> – физико-химические параметры рассчитывали с помощью веб-сервера открытого доступа SwissADME (<http://www.swissadme.ch>); <sup>2</sup> – LogP (липофильность соединения) вычисляли с использованием веб-платформы SciFinder (<https://scifinder.cas.org/>).

N o t e s: <sup>1</sup> – physical and chemical parameters were calculated using the SwissADME open access web server (<http://www.swissadme.ch>); <sup>2</sup> – LogP (lipophilicity of the compound) was calculated using the SciFinder web platform (<https://scifinder.cas.org/>).

Т а б л и ц а 2. Данные о молярной растворимости в воде и синтетической доступности, предсказанные для сконструированных соединений с помощью веб-сервера SwissADME

T a b l e 2. Data on the molar solubility in water and synthetic accessibility predicted for the designed compounds by the SwissADME web server

Лиганд Ligand	Десятичный логарифм молярной растворимости в воде Decimal logarithm of the molar solubility in water, LogS	Синтетическая доступность Synthetic accessibility
I	–5,68	3,55
II	–5,40	5,70
III	–6,37	4,07
IV	–6,62	5,23
V	–6,79	4,52
VI	–6,28	4,83
VII	–6,75	3,89

Т а б л и ц а 3. Значения свободной энергии связывания ( $\Delta G$ ) и констант диссоциации ( $K_d$ ) для статических моделей комплексов лиганд/Mpro согласно оценочным функциям AutoDock Vina, RFScore4 и NNScore2T a b l e 3. Binding free energy ( $\Delta G$ ) and dissociation constants ( $K_d$ ) for static models of the ligand/Mpro complexes according to the AutoDock Vina, RFScore4, and NNScore2 scoring functions

Лиганд Ligand	$\Delta G_{\text{Vina}}^1$ , ккал/моль $\Delta G_{\text{Vina}}^1$ , kcal/mol	$K_{d\text{Vina}}^1$ , мкМ $K_{d\text{Vina}}^1$ , $\mu\text{M}$	$\Delta G_{\text{RFScore4}}^2$ , ккал/моль $\Delta G_{\text{RFScore4}}^2$ , kcal/mol	$K_{d\text{RFScore4}}^2$ , мкМ $K_{d\text{RFScore4}}^2$ , $\mu\text{M}$	$\Delta G_{\text{NNScore2.0}}^2$ , ккал/моль $\Delta G_{\text{NNScore2.0}}^2$ , kcal/mol	$K_{d\text{NNScore2.0}}^2$ , мкМ $K_{d\text{NNScore2.0}}^2$ , $\mu\text{M}$
I	–9,1	0,384	–10,9	0,022	–11,9	0,0041
II	–10,3	0,055	–11,0	0,016	–11,6	0,0069
III	–8,7	0,735	–11,1	0,015	–12,7	0,0012
IV	–10,0	0,089	–11,1	0,014	–13,0	0,0007
V	–9,2	0,326	–10,9	0,022	–13,4	0,0004
VI	–9,6	0,171	–11,2	0,012	–11,9	0,0043
VII	–9,9	0,105	–11,2	0,012	–12,9	0,0007
Ингибитор I	–8,3	1,407	–11,0	0,018	–8,1	1,9
Ингибитор II	–8,5	1,017	–11,1	0,015	–7,9	2,9

П р и м е ч а н и я: <sup>1</sup> – значения  $\Delta G$ , предсказанные программой AutoDock Vina, были переведены в  $K_d$  по формуле  $\Delta G = R T \ln(K_d)$ , где  $\Delta G$  – свободная энергия связывания;  $R$  – универсальная газовая постоянная;  $T$  – абсолютная температура, равная 310 К [13]. <sup>2</sup> – Эту формулу также использовали для преобразования значений  $K_d$ , оцененных с помощью RFScore 4 и NNScore 2.0, в значения  $\Delta G$ .

N o t e s: <sup>1</sup> – the  $\Delta G$  values predicted by AutoDock Vina were converted to  $K_d$  using the formula  $\Delta G = R T \ln(K_d)$ , where  $\Delta G$  is the binding free energy,  $R$  is the universal gas constant,  $T$  is the absolute temperature of 310 K [13]. <sup>2</sup> – This formula was also used to convert  $K_d$  values estimated using RFScore 4 and NNScore 2.0 to the  $\Delta G$  values.

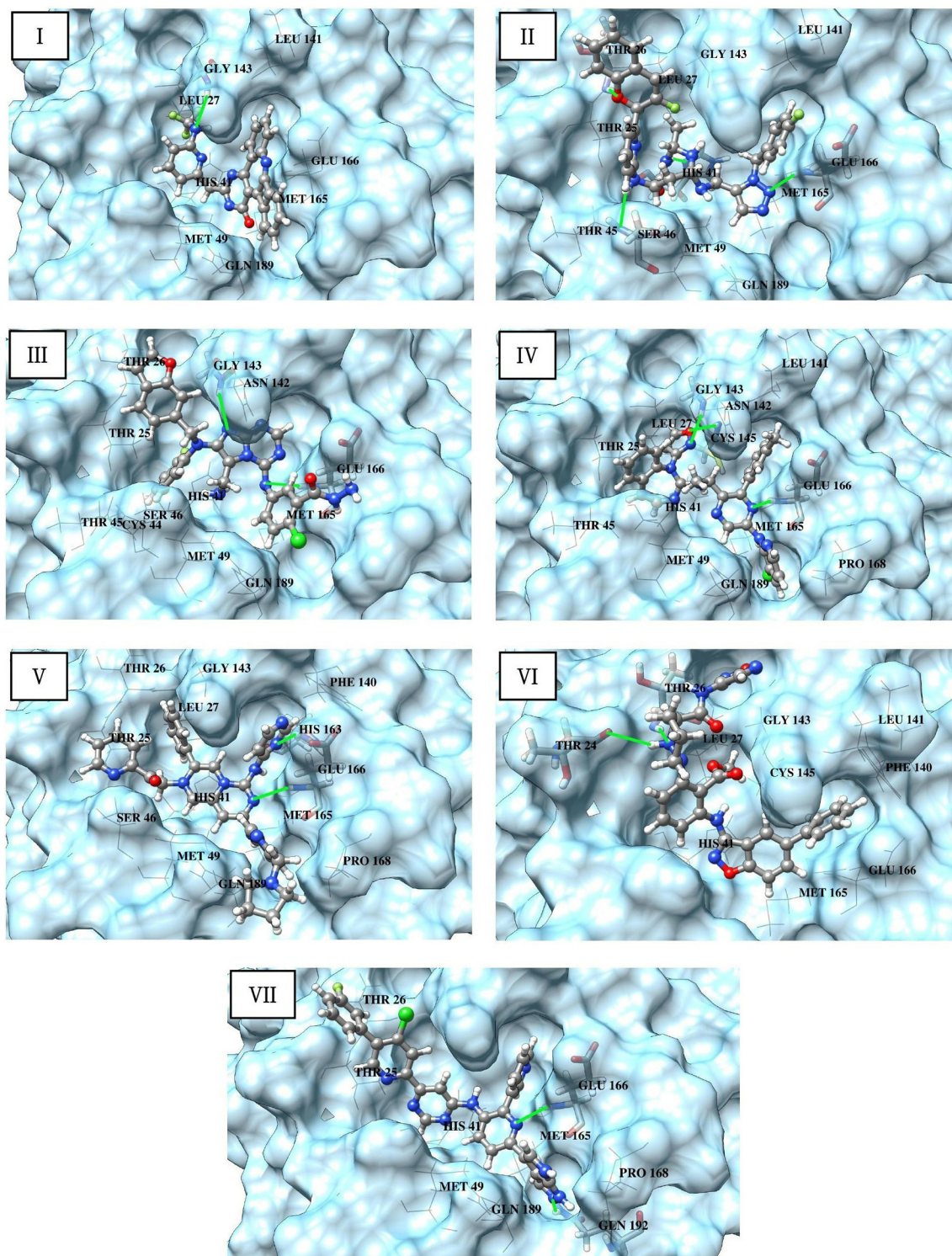


Рис. 2. Структурные комплексы идентифицированных соединений с Mpro SARS-CoV-2, построенные методом молекулярного докинга. Соединения представлены моделью «шарик–палочка–шарик». Указаны остатки Mpro, образующие межатомные контакты с лигандами. Остатки Mpro, участвующие в водородных связях, обозначены палочковой моделью. Водородные связи показаны сплошными зелеными линиями. Проволочная модель использована для обозначения остатков Mpro, образующих ван-дер-ваальсовы контакты, катион- $\pi$  взаимодействия и  $\pi$ -стэкинг

Fig. 2. Structural complexes of the identified compounds with SARS-CoV-2 Mpro generated by molecular docking. The compounds are represented by a ball–stick–ball model. The Mpro residues forming interatomic contacts with the ligands are indicated. Residues of Mpro involved in hydrogen bonding are indicated by a stick model. Hydrogen bonds are shown by green solid lines. The wire model is used to designate the Mpro residues forming van der Waals contacts, cation- $\pi$  interactions and  $\pi$ -stacking



и данные молекулярного докинга: комплексы идентифицированных соединений с Mpro SARS-CoV-2 энергетически стабильны, на что указывают низкие значения энергии связывания, сопоставимые с величинами, полученными для контрольных ингибиторов I и II (табл. 4). Этот вывод подтверждают данные о временных зависимостях среднеквадратичных отклонений (RMSD, Root-Mean-Square Deviations) координат атомов динамических структур комплексов от их стартовых моделей. Для предсказанных соединений средние значения RMSD, равные  $2,14 \pm 0,31$  Å (соединение I),  $2,39 \pm 0,24$  Å (соединение II),  $2,05 \pm 0,32$  Å (соединение III),  $2,26 \pm 0,28$  Å (соединение IV),  $1,98 \pm 0,23$  Å (соединение V),  $2,39 \pm 0,33$  Å (соединение VI),  $1,80 \pm 0,24$  Å (соединение VII), близки к величинам  $1,94 \pm 0,30$  и  $1,99 \pm 0,27$  Å, рассчитанным для ингибиторов I и II соответственно. Эти результаты согласуются с данными о временных зависимостях свободной энергии образования комплексов лиганд/Mpro, показывающими, что их величины не обнаруживают тенденции к увеличению во времени.

Т а б л и ц а 4. Средние значения свободной энергии связывания  $\langle \Delta G \rangle$  и соответствующие им стандартные отклонения  $\Delta G_{\text{STD}}$ , рассчитанные для динамических моделей комплексов лиганд/Mpro

Table 4. Mean binding free energies  $\langle \Delta G \rangle$  and corresponding standard deviations  $\Delta G_{\text{STD}}$  calculated for dynamic models of the ligand/Mpro complexes

Лиганд Ligand	$\langle \Delta H \rangle$ , ккал/моль $\langle \Delta H \rangle$ , kcal/mol	$\Delta H_{\text{STD}}$ , ккал/моль $\Delta H_{\text{STD}}$ , kcal/mol	$\langle T\Delta S \rangle$ , ккал/моль $\langle T\Delta S \rangle$ , kcal/mol	$(T\Delta S)_{\text{STD}}$ , ккал/моль $(T\Delta S)_{\text{STD}}$ , kcal/mol	$\langle \Delta G \rangle$ , ккал/моль $\langle \Delta G \rangle$ , kcal/mol	$\Delta G_{\text{STD}}$ , ккал/моль $\Delta G_{\text{STD}}$ , kcal/mol
I	-45,3	5,2	-22,3	4,5	-23,0	6,9
II	-46,0	6,2	-24,4	4,7	-21,5	7,3
III	-46,6	6,0	-27,6	6,0	-19,0	8,1
IV	-44,6	3,9	-25,9	5,3	-18,7	6,7
V	-45,5	4,3	-28,2	4,3	-17,3	6,0
VI	-40,9	6,2	-24,8	4,4	-16,1	7,0
VII	-37,4	4,8	-23,4	3,9	-13,9	6,2
A	-42,8	4,1	-24,5	4,9	-18,3	6,1
B	-39,0	4,1	-24,5	4,5	-14,4	6,4

П р и м е ч а н и я:  $\langle \Delta H \rangle$  и  $\langle T\Delta S \rangle$  – соответственно средние значения энтальпийной и энтропийной составляющих свободной энергии;  $(\Delta H)_{\text{STD}}$  и  $(T\Delta S)_{\text{STD}}$  – соответствующие этим значениям стандартные отклонения.

N o t e s:  $\langle \Delta H \rangle$  and  $\langle T\Delta S \rangle$  are the average values of the enthalpy and entropy components of free energy, respectively;  $(\Delta H)_{\text{STD}}$  and  $(T\Delta S)_{\text{STD}}$  are the standard deviations corresponding to these values.

Изучение данных о вкладах отдельных аминокислот Mpro в энтальпию связывания позволило выявить остатки фермента, доминирующие в интерфейсе комплексов лиганд/Mpro. Согласно расчетным данным, это остатки His-41 (соединения I–V, VII), Met-49 (соединения I–VII), Asn-142 (соединения III–VI), Gly-143 (соединения III–VI), Cys-145 (соединения III–VII), Met-165 (соединения I–V, VII), Glu-166 (соединения I, III–VII), Asp-187 (соединения I, II, VII) и Gln-189 (соединения I–V, VII). Для контрольных ингибиторов I и II остатками, критически важными для связывания с ферментом, являются Met-49, Asn-142, Gly-143, Cys-145, Met-165, Glu-166, Asp-187 и Gln-189. Кроме того, ингибитор I формирует межмолекулярные контакты с остатком His-41, вносящим значительный вклад в энтальпийную компоненту свободной энергии Гиббса. Важно отметить, что большинство этих остатков используются предсказанными соединениями для эффективных взаимодействий с Mpro SARS-CoV-2. Полученные данные свидетельствуют о том, что эти основные вкладчики в интерфейс исследуемых комплексов играют роль «якорных» остатков, обеспечивающих их энергетическую стабилизацию. Среди этих «горячих точек» связывания в первую очередь следует отметить His-41 и Cys-145, образующие каталитическую диаду Mpro SARS-CoV-2 [14].

Расчет величин среднеквадратичных флуктуаций (RMSF, Root-Mean-Square Fluctuations) отдельных остатков Mpro, позволяющих судить об их гибкости на МД траектории, показывает, что большинство остатков фермента демонстрируют небольшие колебания. Причем этот вывод относится как к остаткам Mpro в комплексах с предсказанными соединениями и контрольными ингибиторами, так и в свободном состоянии. Средние значения RMSF в обоих случаях примерно одинаковы и находятся в интервале от 0,88 до 1,15 Å. Важно отметить, что колебательные



движения остатков Mpro, доминирующих в интерфейсе комплексов лиганд/Mpro, также ограничены в пространстве. Для этих остатков значения RMSF не превышают 2,2 Å, а в большинстве случаев они меньше 1,5 Å. В то же время значения RMSF, рассчитанные для этих «горячих точек» фермента, связанного с идентифицированными и контрольными соединениями, близки друг к другу, что подтверждает их ключевую роль во взаимодействии с каталитическим сайтом Mpro SARS-CoV-2.

**Заключение.** Данные о величинах энергии связывания идентифицированных соединений с Mpro, полученные методами молекулярного докинга и молекулярной динамики, согласуются друг с другом и указывают на их высокое сродство к каталитическому центру фермента (табл. 3 и 4). Это обстоятельство позволяет предполагать, что использованный нами вычислительный подход, сочетающий *de novo* дизайн на основе технологий искусственного интеллекта с молекулярным моделированием, позволил избежать ложноположительных результатов и правильно оценить силу межмолекулярных взаимодействий. Это предположение косвенно подтверждается результатами недавнего исследования [15], в котором было показано, что использование оценочной функции машинного обучения NNScore 2.0 совместно с 1–4 классическими оценочными функциями обеспечивает наилучшую точность предсказания аффинности связывания. В совокупности полученные данные дают веские основания предполагать, что предсказанные соединения могут иметь низкие значения свободной энергии связывания с Mpro, близкие к величинам, рассчитанным для контрольных ингибиторов I и II. На основании этих данных можно ожидать, что низкомолекулярные соединения I–VII (рис. 1) обладают хорошим терапевтическим потенциалом для ингибирования каталитической активности фермента и, следовательно, могут служить базовыми структурами для разработки противовирусных препаратов, терапевтическое действие которых основано на блокаде Mpro SARS-CoV-2.

**Благодарности.** Работа поддержана Белорусским республиканским фондом фундаментальных исследований (проект Ф21АРМГ-001) и Альянсом международных научных организаций ANSO (ANSO-CR-PP-2021-04). Авторы выражают благодарность А. М. Юшкевичу за разработку алгоритма восстановления трехмерной структуры соединений по SMILES представлению.

**Acknowledgments.** The work was supported by the Belarusian Republican Foundation for Basic Research (project Ф21АРМГ-001) and the Alliance of International Scientific Organizations ANSO (ANSO-CR-PP-2021-04). The authors are grateful to Yushkevich A. M. for developing an algorithm for restoring the 3D structure of compounds from the SMILES representation.

#### Список использованных источников

1. Advances and perspectives in applying deep learning for drug design and discovery / C. F. Lipinski [et al.] // *Front. Robotics and AI*. – 2019. – Vol. 6, N 108. <https://doi.org/10.3389/frobt.2019.00108>
2. Deep learning enables rapid identification of potent DDR1 kinase inhibitors / A. Zhavoronkov [et al.] // *Nat. Biotechnol.* – 2019. – Vol. 37, N 9. – P. 1038–1040. <https://doi.org/10.1038/s41587-019-0224-x>
3. Application of deep learning and molecular modeling to identify small drug-like compounds as potential HIV-1 entry inhibitors / A. M. Andrianov [et al.] // *J. Biomol. Struct. Dyn.* – 2022. – Vol. 40, N 16. – P. 7555–7573. <https://doi.org/10.1080/07391102.2021.1905559>
4. Review on the use of Molecular Docking as the First Line Tool in Drug Discovery and Development / R. N. Sahoo [et al.] // *Indian J. Pharm. Sci.* – 2022. – Vol. 84, N 5. – P. 1334–1337. <https://doi.org/10.36468/pharmaceutical-sciences.1031>
5. Hollingsworth, S. A. Molecular dynamics simulation for all / S. A. Hollingsworth, R. O. Dror // *Neuron*. – 2018. – Vol. 99, N 6. – P. 1129–1143. <https://doi.org/10.1016/j.neuron.2018.08.011>
6. Разработка генеративной нейронной сети глубокого обучения для компьютерного дизайна потенциальных ингибиторов коронавируса SARS-CoV-2 / Н. А. Шульдов [и др.] // *Математическая биология и биоинформатика*. – 2022. – Т. 17, № 2. – С. 188–207. <https://doi.org/10.17537/2022.17.188>
7. Ullrich, S. The SARS-CoV-2 main protease as drug target / S. Ullrich, C. Nitsche // *Bioorg. Med. Chem. Lett.* – 2020. – Vol. 30, N 17. – Art. 127377. <https://doi.org/10.1016/j.bmcl.2020.127377>
8. Review on development of potential inhibitors of SARS-CoV-2 main protease (M<sup>Pro</sup>) / S. G. Katre [et al.] // *Futur. J. Pharm. Sci.* – 2022. – Vol. 8, N 1. – Art. 36. <https://doi.org/10.1186/s43094-022-00423-7>
9. Exponential consensus ranking improves the outcome in docking and receptor ensemble docking / K. Palacio-Rodriguez [et al.] // *Sci. Rep.* – 2019. – Vol. 9, N 1. – Art. 5142. <https://doi.org/10.1038/s41598-019-41594-3>
10. Genheden, S. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities / S. Genheden, U. Ryde // *Expert Opin. Drug Discov.* – 2015. – Vol. 10, N 5. – P. 449–461. <https://doi.org/10.1517/17460441.2015.1032936>
11. Potent noncovalent inhibitors of the main Protease of SARS-CoV-2 from molecular sculpting of the drug perampanel guided by free energy perturbation calculations / C. H. Zhang [et al.] // *ACS Cent. Sci.* – 2021. – Vol. 7, N 3. – P. 467–475. <https://doi.org/10.1021/acscentsci.1c00039>
12. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings / C. A. Lipinski [et al.] // *Adv. Drug Deliv. Rev.* – 1997. – Vol. 23, N 1–3. – P. 3–25. [https://doi.org/10.1016/s0169-409x\(96\)00423-1](https://doi.org/10.1016/s0169-409x(96)00423-1)

13. Structural basis of SARS-CoV-2 3CL<sup>pro</sup> and anti-COVID-19 drug discovery from medicinal plants / M. T. Qamar [et al.] // J. Pharm. Anal. – 2020. – Vol. 10, N 4. – P. 313–319. <https://doi.org/10.1016/j.jpha.2020.03.009>
14. Sharma, G. Thermodynamic analysis reveals a temperature-dependent change in the catalytic mechanism of *Bacillus stearothermophilus* tyrosyl-tRNA synthetase / G. Sharma, E. A. First // J. Biol. Chem. – 2009. – Vol. 284, N 7. – P. 4179–4190. <https://doi.org/10.1074/jbc.m808500200>
15. Can machine learning consistently improve the scoring power of classical scoring functions? Insights into the role of machine learning in scoring functions / C. Shen [et al.] // Brief. Bioinf. – 2021. – Vol. 22, N 1. – P. 497–514. <https://doi.org/10.1093/bib/bbz173>

## References

1. Lipinski C. F., Maltarollo V. G., Oliveira P. R., da Silva A. B. F., Honorio K. M. Advances and perspectives in applying deep learning for drug design and discovery. *Frontiers in Robotics and AI*, 2019, vol. 6, no. 108. <https://doi.org/10.3389/frobt.2019.00108>
2. Zhavoronkov A., Ivanenkov Y. A., Aliper A., Veselov M. S., Aladinskiy V. A., Aladinskaya A. V., Terentiev V. A. [et al.]. Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nature Biotechnology*, 2019, vol. 37, no. 9, pp. 1038–1040. <https://doi.org/10.1038/s41587-019-0224-x>
3. Andrianov A. M., Nikolaev G. I., Shuldov N. A., Bosko I. P., Anischenko A. I., Tuzikov A. V. Application of deep learning and molecular modeling to identify small drug-like compounds as potential HIV-1 entry inhibitors. *Journal of Biomolecular Structure and Dynamics*, 2022, vol. 40, no. 16, pp. 7555–7573. <https://doi.org/10.1080/07391102.2021.1905559>
4. Sahoo R. N., Pattanaik S., Pattnaik G., Mallick S., Mohapatra R. Review on the use of Molecular Docking as the First Line Tool in Drug Discovery and Development. *Indian Journal of Pharmaceutical Sciences*, 2022, vol. 84, no. 5, pp. 1334–1337. <https://doi.org/10.36468/pharmaceutical-sciences.1031>
5. Hollingsworth S. A., Dror R. O. Molecular dynamics simulation for all. *Neuron*, 2018, vol. 99, no. 6, pp. 1129–1143. <https://doi.org/10.1016/j.neuron.2018.08.011>
6. Shuldau N. A., Yushkevich A. M., Furs K. V., Tuzikov A. V., Andrianov A. M. Development of a deep learning generative neural network for computer-aided design of potential SARS-CoV-2 inhibitors. *Mathematical Biology and Bioinformatics*, 2022, vol. 17, no. 2, pp. 188–207. <https://doi.org/10.17537/2022.17.188>
7. Ullrich S., Nitsche C. The SARS-CoV-2 main protease as drug target. *Bioorganic & Medicinal Chemistry Letters*, 2020, vol. 30, no. 17, art. 127377. <https://doi.org/10.1016/j.bmcl.2020.127377>
8. Katre S. G., Asnani A. J., Pratyush K., Sakharkar N. G., Bhoje A. G., Sawarkar K. T., Nimbekar V. S. Review on development of potential inhibitors of SARS-CoV-2 main protease (M<sup>pro</sup>). *Future Journal of Pharmaceutical Sciences*, 2022, vol. 8, no. 1, art. 36. <https://doi.org/10.1186/s43094-022-00423-7>
9. Palacio-Rodriguez K., Lans I., Cavasotto C. N., Cossio P. Exponential consensus ranking improves the outcome in docking and receptor ensemble docking. *Scientific Reports*, 2019, vol. 9, no. 1, art. 5142. <https://doi.org/10.1038/s41598-019-41594-3>
10. Genheden S., Ryde U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opinion on Drug Discovery*, 2015, vol. 10, no. 5, pp. 449–461. <https://doi.org/10.1517/17460441.2015.1032936>
11. Zhang C. H., Stone E. A., Deshmukh M., Ippolito J. A., Ghahremanpour M. M., Tirado-Rives J., Spasov K. A. [et al.]. Potent noncovalent inhibitors of the main protease of SARS-CoV-2 from molecular sculpting of the drug perampanel guided by free energy perturbation calculations. *ACS Central Sciences*, 2021, vol. 7, no. 3, pp. 467–475. <https://doi.org/10.1021/acscentsci.1c00039>
12. Lipinski C. A., Lombardo F., Dominy B. W., Feeney P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews*, 1997, vol. 23, no. 1–3, pp. 3–25. [https://doi.org/10.1016/s0169-409x\(96\)00423-1](https://doi.org/10.1016/s0169-409x(96)00423-1)
13. Qamar M. T., Alqahtani S. M., Alamri M. A., Chen L.-L. Structural basis of SARS-CoV-2 3CL<sup>pro</sup> and anti-COVID-19 drug discovery from medicinal plants. *Journal of Pharmaceutical Analysis*, 2020, vol. 10, no. 4, pp. 313–319. <https://doi.org/10.1016/j.jpha.2020.03.009>
14. Sharma G., First E. A. Thermodynamic analysis reveals a temperature-dependent change in the catalytic mechanism of *Bacillus stearothermophilus* tyrosyl-tRNA synthetase. *Journal of Biological Chemistry*, 2009, vol. 284, no. 7, pp. 4179–4190. <https://doi.org/10.1074/jbc.m808500200>
15. Shen C., Hu Y., Wang Z., Zhang X., Zhong H., Wang G., Yao X., Xu L., Cao D., Hou T. Can machine learning consistently improve the scoring power of classical scoring functions? Insights into the role of machine learning in scoring functions. *Briefings in Bioinformatics*, 2021, vol. 22, no. 1, pp. 497–514. <https://doi.org/10.1093/bib/bbz173>

## Информация об авторах

Андреанов Александр Михайлович – д-р хим. наук, профессор, гл. науч. сотрудник. Институт биоорганической химии НАН Беларуси (ул. Купrevича, 5/2, 220084, Минск, Республика Беларусь). E-mail: alexande.andriano@yandex.ru.

Фурс Константин Викторович – инженер-программист. Объединенный институт проблем информатики НАН Беларуси (ул. Сурганова, 6, 220012, Минск, Республика Беларусь). E-mail: ky6ujlo@gmail.com.

Шульдов Никита Андреевич – инженер-программист. Объединенный институт проблем информатики НАН Беларуси (ул. Сурганова, 6, 220012, Минск, Республика Беларусь). E-mail: nickshuldov29@gmail.com.

Тузиков Александр Васильевич – член-корреспондент, д-р физ.-мат. наук, профессор, заведующий лабораторией. Объединенный институт проблем информатики НАН Беларуси (ул. Сурганова, 6, 220012, Минск, Республика Беларусь). E-mail: tuzikov@newman.bas-net.by.

## Information about the authors

Andrianov Alexander M. – D. Sc. (Chemistry), Professor, Chief Researcher. Institute of Bioorganic Chemistry of the National Academy of Sciences of Belarus (5/2, Kuprevich Str., 220084, Minsk, Republic of Belarus). E-mail: alexande.andriano@yandex.ru.

Furs Konstantin V. – Software Engineer. United Institute of Informatics Problems of the National Academy of Sciences of Belarus (6, Surganov Str., 220012, Minsk, Republic of Belarus). E-mail: ky6ujlo@gmail.com.

Shuldau Mikita A. – Software Engineer. United Institute of Informatics Problems of the National Academy of Sciences of Belarus (6, Surganov Str., 220012, Minsk, Republic of Belarus). E-mail: nickshuldov29@gmail.com.

Tuzikov Alexander V. – Corresponding Member, D. Sc. (Physics and Mathematics), Professor, Head of the Laboratory. United Institute of Informatics Problems of the National Academy of Sciences of Belarus (6, Surganov Str., 220012, Minsk, Republic of Belarus). E-mail: tuzikov@newman.bas-net.by.