

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

## ИНФОРМАТИКА INFORMATICS

УДК 004.032.26+616-002.5+51-76  
<https://doi.org/10.29235/1561-8323-2025-69-1-13-22>

Поступило в редакцию 28.11.2024  
Received 28.11.2024

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

<sup>1</sup>Объединенный институт проблем информатики Национальной академии наук Беларуси,  
Минск, Республика Беларусь

<sup>2</sup>Институт биоорганической химии Национальной академии наук Беларуси,  
Минск, Республика Беларусь

### ГЕНЕРАТИВНАЯ СОСТЯЗАТЕЛЬНАЯ НЕЙРОННАЯ СЕТЬ НА ОСНОВЕ ГРАФОВЫХ ЭМБЕДИНГОВ ДЛЯ *DE NOVO* ДИЗАЙНА НИЗКОМОЛЕКУЛЯРНЫХ ИНГИБИТОРОВ ФЕРМЕНТА KasA МИКОБАКТЕРИИ ТУБЕРКУЛЕЗА

**Аннотация.** Разработана генеративная состязательная нейронная сеть с частичным привлечением учителя, обученная на графовых эмбедингах и предназначенная для *de novo* дизайна потенциальных ингибиторов бета-кетоацил-[ацил-белок-носитель] синтазы I (KasA) – фермента, критически важного для биосинтеза миколовых кислот клеточной стенки микобактерии туберкулеза. Проведено обучение и тестирование созданной нейронной сети на наборе соединений из виртуальной библиотеки малых молекул, содержащих элементы структуры, способные к селективным взаимодействиям с терапевтической мишенью. С помощью разработанной нейронной сети осуществлен *de novo* дизайн 3637 соединений с последующей оценкой потенциала их ингибиторной активности против белка KasA методами молекулярного докинга. На основе анализа полученных данных отобраны шесть соединений, проявляющих высокое сродство к малонил-связывающему сайту фермента. Предполагается, что идентифицированные соединения формируют перспективные базовые структуры для проведения дальнейших теоретических и экспериментальных исследований по разработке новых эффективных ингибиторов лекарственно-устойчивых форм туберкулеза.

**Ключевые слова:** микобактерия туберкулеза, фермент KasA, генеративная состязательная нейронная сеть, обучение с частичным привлечением учителя, графовые эмбединги, виртуальный скрининг, молекулярный докинг

**Для цитирования.** Генеративная состязательная нейронная сеть на основе графовых эмбедингов для *de novo* дизайна низкомолекулярных ингибиторов фермента KasA микобактерии туберкулеза / А. В. Гончар, К. В. Фурс, А. В. Тузиков, А. М. Андрианов // Доклады Национальной академии наук Беларуси. – 2025. – Т. 69, № 1. – С. 13–22. <https://doi.org/10.29235/1561-8323-2025-69-1-13-22>

**Anna V. Gonchar<sup>1</sup>, Konstantin V. Furs<sup>1</sup>, Corresponding Member Alexander V. Tuzikov<sup>1</sup>, Alexander M. Andrianov<sup>2</sup>**

<sup>1</sup>United Institute of Informatics Problems of the National Academy of Sciences of Belarus, Minsk, Republic of Belarus

<sup>2</sup>Institute of Bioorganic Chemistry of the National Academy of Sciences of Belarus, Minsk, Republic of Belarus

### GENERATIVE ADVERSARIAL NEURAL NETWORK WITH GRAPH EMBEDDINGS FOR *DE NOVO* DESIGNING SMALL-MOLECULE INHIBITORS AGAINST *MYCOBACTERIUM TUBERCULOSIS* KasA ENZYME

**Abstract.** A generative semi-supervised adversarial neural network trained on graph embeddings was developed for *de novo* design of potential inhibitors against beta-ketoacyl-[acyl-carrier protein] synthase I (KasA), an enzyme critically important for biosynthesis of mycolic acids of the *Mycobacterium tuberculosis* cell wall. The designed model was trained and tested on a set of compounds from a virtual library of small molecules containing structural elements capable of selective interactions with the therapeutic target. Using the developed neural network, 3,637 compounds were *de novo* designed, followed by assessment of their inhibitory activity against the KasA protein using molecular docking methods. Based on the analysis of the obtained data, six compounds exhibiting high affinity to the malonyl-binding site of the enzyme were selected. The identified compounds are assumed to form promising basic structures for further theoretical and experimental studies on the development of new effective inhibitors of drug-resistant tuberculosis.

**Keywords:** *Mycobacterium tuberculosis*, KasA enzyme, generative adversarial neural network, semi-supervised learning, graph embeddings, virtual screening, molecular docking

**For citation.** Gonchar A. V., Furs K. V., Tuzikov A. V., Andrianov A. M. Generative adversarial neural network with graph embeddings for *de novo* designing small-molecule inhibitors against *Mycobacterium tuberculosis* KasA enzyme. *Doklady Natsional'noi akademii nauk Belarusi = Doklady of the National Academy of Sciences of Belarus*, 2025, vol. 69, no. 1, pp. 13–22 (in Russian). <https://doi.org/10.29235/1561-8323-2025-69-1-13-22>

**Введение.** Туберкулез (ТБ) по-прежнему входит в десятку основных причин смертей во всем мире (<https://deathmeters.info>) и является основной причиной гибели пациентов с ВИЧ и сахарным диабетом, что в первую очередь связано с устойчивостью новых штаммов микобактерии туберкулеза (МБТ) – этиологического агента этого опасного инфекционного заболевания – к используемым в клинике противомикробным препаратам. Увеличение распространенности лекарственной устойчивости представляет серьезную проблему для эффективной борьбы с ТБ. В связи с этим одним из ключевых вызовов, стоящих перед научным сообществом, является лечение ТБ с множественной и широкой лекарственной устойчивостью. Очевидно, что это обстоятельство обуславливает актуальность, важность и практическую значимость исследований по разработке новых эффективных ингибиторов МБТ.

В настоящее время технологии виртуального скрининга представляют собой одно из важнейших и востребованных направлений исследований в области создания терапевтических агентов и являются важным вычислительным инструментом на первом этапе сложного и многостадийного процесса разработки новых лекарственных препаратов, позволяя существенно сократить время и затраты, необходимые для их создания [1; 2]. Использование этих технологий в исследованиях ТБ уже привело к идентификации ряда антибактериальных препаратов, которые находятся на различных стадиях биомедицинских испытаний [3–5]. Однако необходимо проделать большую дополнительную работу, чтобы ускорить создание эффективных противотуберкулезных терапевтических средств. Несмотря на то что эффективность компьютерных методов в создании лекарственных препаратов в настоящее время является общепризнанной, разработка новых математических подходов в сочетании с доступностью мощных и дешевых вычислительных ресурсов способствует их постоянному совершенствованию. Среди этих подходов важное место занимают методы машинного обучения и, в частности, методы глубокого обучения, которые предлагают большой потенциал для дальнейшего прогресса в данной области исследований [2]. Одним из преимуществ машинного обучения для конструирования лекарств является помощь исследователям в понимании и использовании взаимосвязей между химическими структурами и их биологической активностью. Современные методы машинного обучения могут использоваться для моделирования взаимосвязи структура–активность или количественных отношений структура–свойство, разработки интеллектуальных инструментов, способных достаточно точно предсказывать влияние химических модификаций соединения на его биологическую активность и фармакологические характеристики, а также для генерации новых химических соединений с заданными свойствами [2]. В настоящее время исследования ТБ, проводимые с привлечением технологий машинного обучения, в основном охватывают диагностику и результаты лечения этой болезни. Это является одним из пробелов, который необходимо устранить, чтобы иметь возможность ускорить процесс создания новых и эффективных препаратов для терапии лекарственно-устойчивых форм ТБ и облегчить тяжелое бремя этого инфекционного заболевания.

Анализ доклинических исследований в области поиска противотуберкулезных препаратов показал, что среди белковых мишеней, играющих важную роль в биосинтезе клеточной стенки МБТ, следует особое внимание уделить бета-кетоацил-[ацил-белок-носитель]синтазе I (KasA) – одному из ключевых ферментов пути FAS-II синтеза жирных кислот [6]. Известно, что потеря активности белка KasA приводит к лизису клеток бактерии [7], свидетельствуя о том, что этот фермент имеет ключевое значение для жизненного цикла МБТ и, следовательно, является важной терапевтической мишенью для разработки новых эффективных ингибиторов лекарственно-устойчивого ТБ [6].

Цель настоящего исследования заключалась в создании генеративной модели нейронной сети для *de novo* дизайна малых молекул, потенциально активных против фермента KasA МБТ. Для достижения этой цели были проведены исследования, которые включали:

1. Разработку архитектуры нейронной сети для генерации низкомолекулярных химических соединений, обладающих высоким сродством к ферменту KasA.
2. Формирование обучающей библиотеки малых молекул, содержащих элементы структуры, способные к селективным взаимодействиям с активным центром фермента.
3. Обучение модели и ее тестирование на наборе соединений из созданной молекулярной библиотеки.
4. Генерацию новых молекул, потенциально активных против белка KasA.
5. Молекулярный докинг сгенерированных молекул с целевым белком, оценку аффинности связывания и отбор соединений, перспективных для дальнейших исследований по разработке эффективных антибактериальных агентов.

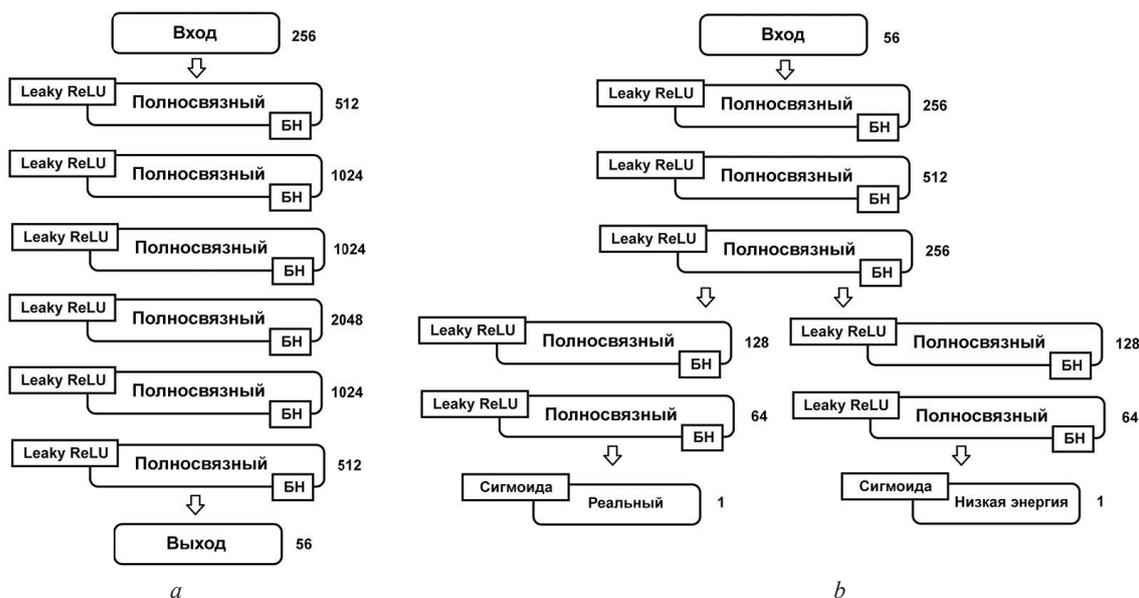
**Материалы и методы исследования.** *Архитектура модели нейронной сети.* Для решения поставленных задач нами была разработана модель генеративной состязательной нейронной сети с частичным привлечением учителя (Semi-Supervised Generative Adversarial Network, SGAN) [8], которая использует графовые эмбединги, полученные из латентного пространства вариационного автоэнкодера JTVAE (Junction Tree Variational Autoencoder) [9]. Обучение SGAN проводили на наборе молекул из обучающей выборки с использованием в неявном виде значений энергии связывания с белком KasA. Это должно было повысить вероятность генерации молекул с более высоким сродством к молекулярной мишени. Молекулы в обучающем наборе данных разделяли на две группы, включавшие соответственно соединения с низкими и высокими значениями энергии связывания с целевым белком, рассчитанными с помощью методов молекулярного докинга. Это позволило SGAN генерировать новые молекулы, похожие на соединения с высоким сродством к мишени.

Модель JTVAE представляет собой вариационный автоэнкодер (VAE), который работает с молекулярными графами с помощью кодировщика и декодировщика [9], в то время как большинство других методов используют для представления молекул формат SMILES [10]. Кодировщик является нейронной сетью, используемой для вычисления скрытого представления молекулы в многомерном пространстве молекулярных эмбедингов, а декодировщик – это нейронная сеть, предназначенная для декодирования химического соединения из вектора в этом пространстве обратно в молекулярный граф. В JTVAE каждая молекула рассматривается как набор допустимых химических подструктур, которые выбираются из словаря компонентов, сформированного из обучающего набора данных [9]. Эти компоненты используются в качестве строительных блоков для молекулы во время процессов кодирования и декодирования. На основе компонентов для каждой молекулы строится дерево соединений с помощью специального алгоритма разложения. Исходный молекулярный граф и связанное с ним дерево соединений являются двумя взаимодополняющими представлениями молекулы [9]. Результирующий скрытый вектор представляет собой латентный вектор молекулярного графа, объединенный с латентным вектором дерева соединений. Известно, что JTVAE генерирует только молекулы с корректной стереохимией благодаря подходу покомпонентного кодирования-декодирования, тогда как генеративные нейронные сети на основе SMILES также генерируют и некорректные соединения.

Модель SGAN состоит из двух нейронных сетей: генератора и дискриминатора, которые обучаются одновременно. Генератор пытается генерировать данные, похожие на данные из обучающего набора, в то время как дискриминатор пытается различать реальные и ложные данные. GAN относится к классу алгоритмов без учителя, поскольку явные метки любого класса данных не используются, за исключением неявных «поддельных» и «реальных» меток. SGAN является GAN с дополнительным классом для дискриминатора, позволяющим различать «низкие» и «высокие» значения свободной энергии связывания с молекулярной мишенью, рассчитанные методами молекулярного докинга. В настоящем исследовании молекулы со значениями свободной энергии связывания ниже  $-8,2$  ккал/моль были отнесены к классу «низкой» энергии, а остальные соединения – к классу «высокой» энергии.

Архитектура SGAN, расширяющая эту модель до графовых эмбедингов латентного пространства JTVAE, показана на рис. 1, где БН означает пакетную нормализацию (батч-нормализацию).

*Структура белковой мишени и контрольные ингибиторы.* Поскольку настоящее исследование было направлено на *de novo* дизайн потенциальных ингибиторов белка KasA, то были необ-

Рис. 1. Архитектура SGAN: *a* – генератор, *b* – дискриминаторFig. 1. The SGAN architecture: *a* – generator, *b* – discriminator

ходимы структуры этого фермента и его известных ингибиторов, которые могли бы выступить в роли положительных контролей при оценке сродства новых сгенерированных соединений к молекулярной мишени. В качестве такого контроля в расчетах использовали два ингибитора каталитической активности фермента KasA – тиолактомицин (TLM) и его аналог TLM5 [11]. Молекулы TLM и TLM5 являются медленно действующими (slow onset) ингибиторами, преимущественно взаимодействующими с ацилированной формой фермента [11], структура которой в настоящее время не установлена методом рентгеноструктурного анализа. Однако было показано [12], что ацилированный фермент KasA может быть заменен мутантной формой C171Q, поскольку мутация Cys-Gln в положении 171 полипептидной цепи приводит к структурным изменениям в активном центре фермента, имитирующим ацилирование остатка Cys-171. Поэтому в данной работе нами была использована структура комплекса TLM5 с C171Q KasA в кристалле (ID PDB: 4C72, <https://www.rcsb.org/structure/4c72>).

**Формирование обучающего набора данных. Виртуальный скрининг.** Для формирования обучающего набора данных был проведен виртуальный скрининг трех молекулярных библиотек веб-сервера Pharmit (<https://pharmit.cs.pitt.edu>): Zinc15, ChemSpace и ChemDiv. С этой целью на основе анализа комплекса белка C171Q KasA с ингибитором МБТ TLM5 (PDB ID: 4C72), выполненного с помощью программного обеспечения веб-сервера Pharmit, были построены две фармакофорные модели, отличающиеся друг от друга отсутствием в одной из этих моделей гидрофобной функциональной группы. Также для того, чтобы расширить тренировочный набор данных, была сгенерирована третья фармакофорная модель на основе комплекса белка C171Q KasA с молекулой TLM (PDB ID: 4C6X). Главным отличием первых двух моделей от третьей является присутствие у них дополнительной фармакофорной группы – акцептора водородной связи. Наличие этой группы является основной причиной более сильного, по сравнению с TLM, ингибирования фермента KasA молекулой TLM5, так как она лучше имитирует субстрат, связывающийся с KasA в клетке [13].

В процессе скрининга использовали ряд фильтров, накладывающих ограничения на физико-химические параметры молекул, которые принято считать основными критериями их эффективности при пероральном приеме (табл. 1). Последние два фильтра ( $\Delta G$ , RMSD) применяли для оптимизации конформаций, полученных после основного поиска.

Используя Python 3 и программный пакет RDKit (<https://www.rdkit.org>) из отобранных молекул были удалены дубликаты и для каждого соединения получены канонические представления SMILES. В результате размер обучающей молекулярной библиотеки, сформированной с помощью фармакофорного анализа баз данных веб-сервера Pharmit, составил 58815.

Таблица 1. Фильтры, использованные для фармакофорного поиска в молекулярных библиотеках веб-сервера Pharmit

Table 1. Filters used for pharmacophore search in molecular libraries of the Pharmit web server

М, Да M, Da	logP	HBD	HBA	$\Delta G$ , ккал/моль $\Delta G$ , kcal/mol	RMSD, Å
$\leq 500$	$\leq 5$	$\leq 5$	$\leq 10$	$< 0$	$< 2,0$

Примечания: М – молекулярная масса; logP – липофильность соединения, вычисленная с помощью программы OpenBabel (<http://openbabel.org/index.html>); HBD и HBA – количество доноров и акцепторов водородной связи соответственно;  $\Delta G$  – величина энергии связывания, рассчитанная с помощью оценочной функции AutoDock Vina (<https://vina.scripps.edu>) и ее ускоренной версии smina (<https://sourceforge.net/projects/smina/>); RMSD – среднеквадратическое отклонение полученной структуры от стартовой модели после ее оптимизации.

Notes: M – molecular weight; logP – compound lipophilicity calculated using the OpenBabel program (<http://openbabel.org/index.html>); HBD and HBA – numbers of hydrogen bond donors and acceptors, respectively;  $\Delta G$  – value of binding energy calculated using the AutoDock Vina scoring function (<https://vina.scripps.edu>) and its accelerated version smina (<https://sourceforge.net/projects/smina/>); RMSD – the root-mean-square deviation of the obtained structure after its optimization from the starting model.

**Молекулярный докинг.** Для расчета значений свободной энергии связывания молекул из обучающего набора данных с целевым белком с помощью программы QuickVina2 (<https://qvina.github.io>) был проведен молекулярный докинг этих лигандов с ферментом C171Q KasA в свободном состоянии. Структуры белка и лигандов были подготовлены к расчетам с помощью программного пакета MGLTools (<https://ccsb.scripps.edu/mgltools/>). Ячейка для докинга включала малонил-связывающий сайт KasA [6] и имела следующие размеры:  $\Delta X = 20,67$  Å,  $\Delta Y = 24,8$  Å,  $\Delta Z = 16,46$  Å с центром  $X = -7,24$  Å,  $Y = -19,9$  Å,  $Z = 6,75$  Å (ID PDB: 4C72, <https://www.rcsb.org/structure/4c72>). Значение параметра охвата конформационного пространства, определяющего широту поиска, было установлено равным 100.

Гистограмма распределения значений свободной энергии связывания с белком KasA, рассчитанных методом молекулярного докинга для молекул из обучающего набора данных, приведена на рис. 2.

**Получение графовых эмбедингов.** Для создания графовых эмбедингов из формата SMILES, полученного на этапе фармакофорного поиска, использовали предобученную модель нейронной сети JTVAE [9], которую обучали на выборке из 250000 соединений, отобранных случайным образом в библиотеке ZINC15. Используя кодировщик JTVAE для обучающей выборки были сгенерированы графовые эмбединги, а именно векторы латентного пространства длиной 56.

**Обучение, тестирование и оценка работы нейронной сети на наборе соединений из обучающей молекулярной библиотеки.** Для 58815 малых молекул из сформированного обучающего набора данных с помощью сети JTVAE были получены графовые эмбединги, а затем эта виртуальная библиотека была разделена на две выборки – тренировочную и тестовую, состоящие из 47052 и 11763 векторов соответственно.

Перед процессом финального обучения гиперпараметры SGAN были скорректированы. Первый параметр, задающий количество эпох обучения, был экспериментально выбран равным 150. Второй параметр представлял собой процент того, как часто нужно обучать генератор и дискриминатор. Для того чтобы определить эти параметры, для различных значений наблюдались графики функций потерь генера-

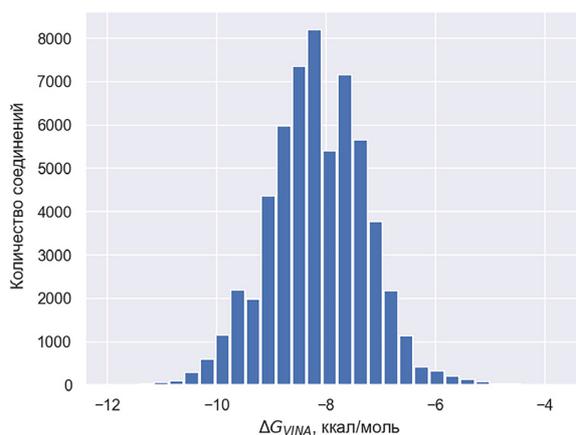


Рис. 2. Гистограмма распределения значений свободной энергии связывания с белком KasA, рассчитанных методом молекулярного докинга для молекул из обучающего набора данных

Fig. 2. Histogram of the distribution of binding free energy values calculated by molecular docking for molecules from the training dataset

тора и дискриминатора и графики предсказаний дискриминатора на наборах данных тестовой выборки и сгенерированных молекул. В результате эти параметры были выбраны равными 0,3 для генератора и 0,7 для дискриминатора.

Функция потерь дискриминатора представлена в формуле

$$D_{\text{loss}} = \text{BCE}(D_{\text{out1}}(G(\text{noise})), 0) + \text{BCE}(D_{\text{out2}}(\text{real\_data}), \text{energy\_class}),$$

где BCE – функция потерь «бинарная кросс-энтропия»;  $D$  – дискриминатор;  $D_{\text{out1}}$  – выход дискриминатора, отвечающий за предсказание реальности молекулы;  $D_{\text{out2}}$  – выход дискриминатора, отвечающий за предсказание класса энергии молекулы;  $G$  – генератор; noise – вектор гауссовского шума размерности 256; real\_data – вектор, соответствующий молекуле из тренировочной выборки; energy\_class – 1 или 0 в зависимости от того, принадлежит ли молекула классу с низкой энергией связывания или нет.

В процессе экспериментов были опробованы две функции потерь генератора. Сначала была использована  $G_{\text{loss}}$ , а затем она была модифицирована в  $G_{\text{loss\_new}}$ :

$$G_{\text{loss}} = \text{BCE}(D_{\text{out1}}(G(\text{noise})), 1) + \text{BCE}(D_{\text{out2}}(G(\text{noise})), 1), \quad (1)$$

$$G_{\text{loss\_new}} = \text{BCE}(D_{\text{out1}}(G(\text{noise})), 1). \quad (2)$$

На рис. 3 представлены результаты предсказаний дискриминатора на векторах молекул, сгенерированных генератором на основе случайного гауссовского шума (синий, 23000 векторов), и векторах из тестовой выборки (оранжевый, 11763 векторов) для SGAN с функцией потерь (1) и с функцией потерь (2) соответственно.

Как следует из рис. 3, *a*, при использовании генератором функции потерь (1) дискриминатор усвоил закономерность того, что молекулы с низкой вероятностью низкого энергетического класса реальны. Известно, что при наличии комплементарности лиганда активному центру белка значения энергии связывания часто зависят от размера соединения (чем больше молекулярная масса, тем ниже значение энергии), что очевидно отражается на результатах, полученных с помощью оценочных функций молекулярного докинга, и, в частности, оценочной функции программы QuickVina 2. Это означает, что сеть также усвоила закономерность того, что соединения

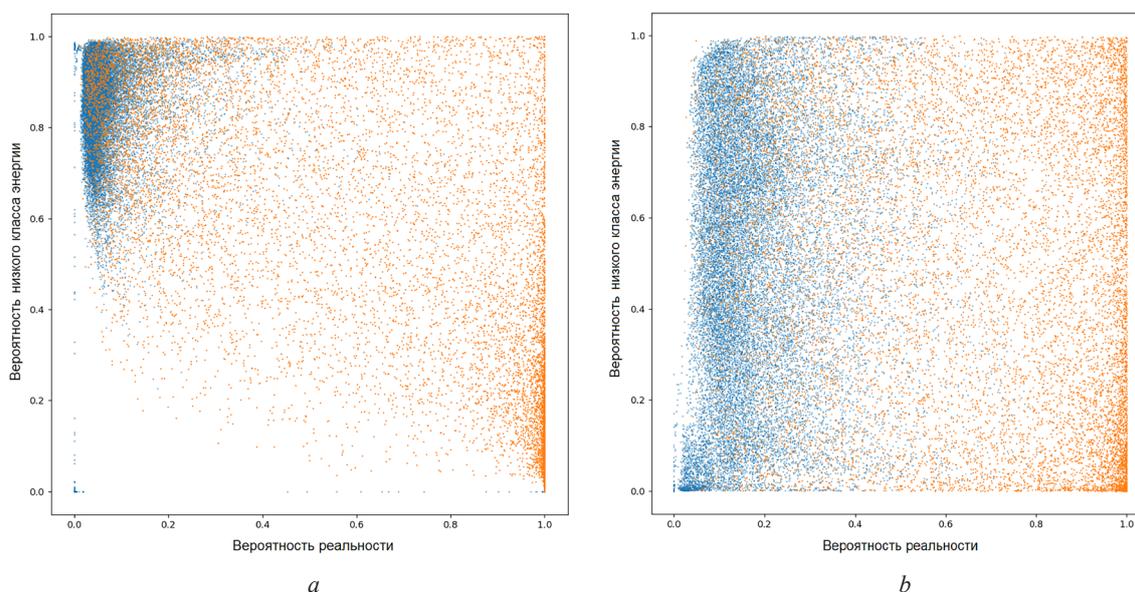


Рис. 3. Результаты предсказаний дискриминатора SGAN на векторах случайного гауссовского шума (синий, 23000 векторов) и векторах из валидационной выборки (оранжевый, 11763 векторов); *a*) SGAN обученный с функцией потерь (1), *b*) SGAN обученный с функцией потерь (2)

Fig. 3. Prediction results of the SGAN discriminator on random Gaussian noise vectors (blue, 23,000 vectors) and vectors from the validation set (orange, 11,763 vectors); *a*) SGAN trained with loss function (1), *b*) SGAN trained with loss function (2)

с низкой вероятностью низкого класса энергии связывания скорее всего являются не только реальными, но и имеют низкую молекулярную массу. Именно поэтому результатом генерации соединений с высокой вероятностью низкой энергии связывания не могли быть достаточно малые молекулы. Чтобы преодолеть эту проблему, функция потерь (1) генератора была изменена. В (1) второй член представляет штраф, который дискриминатор назначает генератору за создание соединения класса с высокой энергией связывания. Однако дискриминатору изначально не хватает данных для правильного определения энергетического класса, из-за чего функция потерь генератора быстро растет. В связи с этим второй член был удален из (1) и в дальнейших исследованиях была использована новая функция потерь (2).

Графики функций потерь генератора и дискриминатора SGAN с функцией потерь генератора (2) и соотношением обучения генератора к дискриминатору 0,3 к 0,7 показаны на рис. 4. Сплошная линия (1) соответствует потерям генератора, а пунктирная линия (2) – потерям дискриминатора. Анализ графиков функций потерь и предсказаний дискриминатора дает основания полагать, что модель была обучена корректно, так как функции потерь сходятся, но не пересекаются, и дискриминатор с высокой точностью отличает вектора молекул тестовой выборки (реальных) от случайных.

*Генерация новых молекул.* На основе случайных векторов длины 256, имеющих гауссовское распределение, с помощью генератора SGAN было сгенерировано 200000 векторов длиной 56 и получены соответствующие предсказания дискриминатора. Для оценки генеративных возможностей SGAN были выбраны два региона векторов. Первый (перспективный) регион включал векторы с вероятностью реальности выше 50 % и вероятностью низкого энергетического класса выше 50 %, а второй регион представлял область с вероятностью реальности молекул выше 50 % и вероятностью низкого энергетического класса ниже 50 %. Количество графовых эмбедингов, удовлетворявших условиям первого и второго регионов, составило 2565 и 2745 соответственно. Эти векторы были переданы в декодер JTVAE и далее преобразованы в формат SMILES. Затем дубликаты SMILES были удалены и сгенерированные молекулы проверены на корректность. После этого шага были отобраны 1755 и 1882 молекулы для первого и второго регионов соответственно.

*Генерация трехмерных структур новых молекул.* Трехмерные структуры сгенерированных молекул были получены на основе их представлений в формате SMILES. Для этого использовали стохастический алгоритм генерации конформеров ETKDG (Experimental-Torsion “Basic Knowledge” Distance Geometry) из пакета RDKit (<https://www.rdkit.org>). Затем полученные струк-

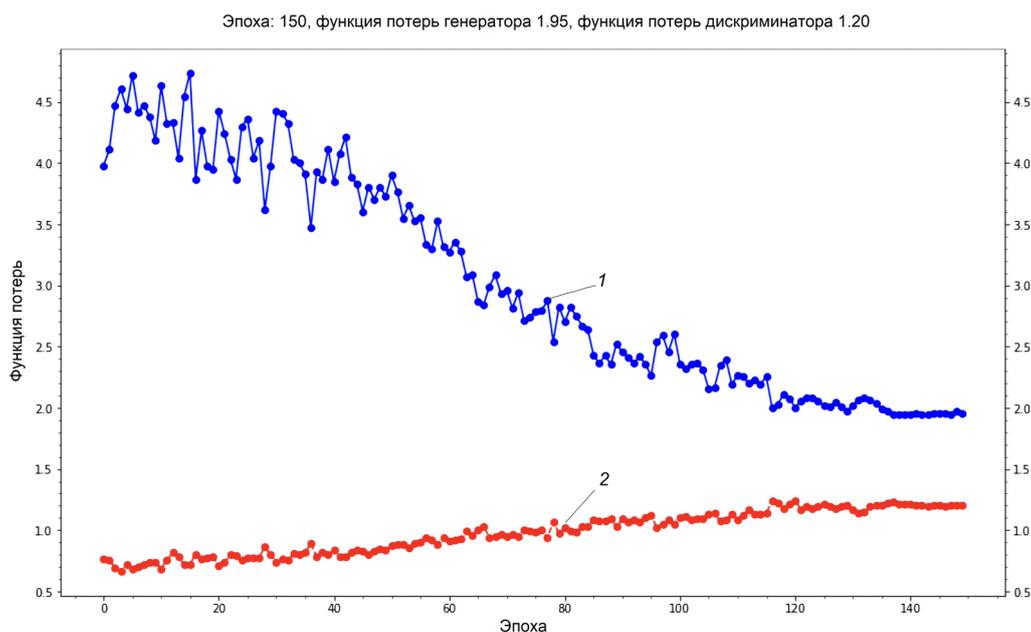


Рис. 4. Функции потерь генератора (1) и дискриминатора (2) SGAN

Fig. 4. Loss functions of the SGAN generator (1) and discriminator (2)

туры оптимизировали в силовом поле Merck (MMFF) или универсальном силовом поле UFF в тех случаях, когда оптимизация в поле MMFF оказалась неудачной.

**Результаты и их обсуждение.** Как было отмечено выше, сгенерированные SGAN новые молекулы были разделены на два набора, которые включали 1755 соединений из перспективного региона и 1882 молекулы из второго региона. Кроме этого, для оценки потенциала разработанной модели нейронной сети нами были получены прогнозные показатели дискриминатора для 11763 векторов из тестового набора данных и отобраны векторы, удовлетворяющие условиям перспективной и второй областей. В выборку вошли соответственно 3438 и 4448 молекулярных эмбедингов. Для сгенерированных молекул и соединений из тестового набора был проведен молекулярный докинг с белком KasA с использованием того же вычислительного протокола, что и для соединений из обучающего набора данных. Результаты молекулярного докинга представлены в табл. 2.

Т а б л и ц а 2. Результаты молекулярного докинга для сгенерированных соединений и соединений из тестовой выборки

Table 2. Molecular docking results for generated compounds and compounds from the test set

Соединение Compound	Область Region	Количество соединений Compound number	Процент соединений с низкой энергией связывания ( $<-8,2$ ккал/моль) Percentage of compounds with low binding energy ( $<-8,2$ kcal/mol)	Средняя энергия связывания, (ккал/моль) Average binding energy (kcal/mol)
Сгенерированные соединения Generated compounds	1	1755	67	-8,5
	2	1882	40	-8,0
Соединения из тестовой выборки Compounds from the test sample	1	3438	76	-8,7
	2	4448	21	-7,6

Примечание: 1 – перспективная область; 2 – вторая область.

Note: 1 – promising region; 2 – second region.

Данные табл. 2 указывают на то, что дискриминатор SGAN может с высокой точностью предсказывать класс энергии на реальных (не сгенерированных) данных, что подтверждает корректность обучения модели. Например, для тестовых соединений из перспективного региона 76 % молекул имели энергию связывания ниже  $-8,2$  ккал/моль, а среднее значение энергии составило  $-8,7$  ккал/моль. Для тестовых соединений второго региона только 21 % молекул имел энергию ниже указанного порога, при этом средняя величина энергии составила  $-7,6$  ккал/моль. Сгенерированные соединения продемонстрировали схожие закономерности. В перспективном регионе 67 % молекул имели энергию ниже  $-8,2$  ккал/моль, со средней энергией  $-8,5$  ккал/моль, а во втором регионе только 40 % соединений соответствовали этому порогу, и средняя величина энергии составила  $-8,0$  ккал/моль. Полученные результаты свидетельствуют о способности сети SGAN генерировать различные химические соединения с высоким сродством к белку KasA, особенно в перспективном регионе. Эти данные также указывают на то, что разделение молекул на классы на основе энергии связывания при обучении модели является эффективной стратегией для создания соединений с высоким сродством к целевому белку.

Т а б л и ц а 3. Значения оценочных функций для шести наиболее перспективных соединений

Table 3. Values of the scoring functions for the six most promising compounds

Лиганд Ligand	$\Delta G_{VINA}$ , ккал/моль $\Delta G_{VINA}$ , kcal/mol	$\Delta G_{RFScore-4}$ , ккал/моль $\Delta G_{RFScore-4}$ , kcal/mol	$\Delta G_{NNScore2,0}$ , ккал/моль $\Delta G_{NNScore2}$ , kcal/mol	ECR
I	-9,491	-11,49	-11,51	0,2786
II	-9,449	-11,52	-12,21	0,2803
III	-9,5	-11,66	-11,12	0,2783
IV	-9,572	-11,45	-12,32	0,2824
V	-10,126	-11,06	-12,97	0,2706
VI	-9,599	-11,16	-12,08	0,2687
TLM5	-8,0	-8,27	-6,97	-

Для выбора наиболее перспективных лигандов среди всех сгенерированных соединений (из обеих областей), значения свободной энергии связывания были переоценены с использованием оценочных функций RFScore-4 (<https://pjballester.wordpress.com/software/>) и NNScore 2.0 (<https://git.durrantlab.pitt.edu/jdurrant/nnscore2>). На основе предсказанных с помощью оценочных функций AutoDock Vina, RFScore-4, NNScore 2.0 значений энергии связывания для каждого соединения был рассчитан экспоненциальный консенсусный ранг (ECR) [14] и отобраны шесть соединений-кандидатов в лекарственные средства. В табл. 3 представлены результаты, полученные для этих молекул и ингибитора KasA TLM5, использованного в расчетах в качестве положительного контроля.

Анализ данных табл. 3 показывает, что отобранные соединения имеют величины энергии связывания ниже, чем значение, предсказанное для контрольного ингибитора TLM5, что может свидетельствовать об их высоком сродстве к ферменту KasA.

**Заключение.** Разработана генеративная состязательная нейронная сеть с частичным привлечением учителя, обученная на графовых эмбедингах, полученных из латентного пространства вариационного автоэнкодера JTVAE. Для обучения модели собрана виртуальная библиотека низкомолекулярных соединений, имеющих схожие функциональные группы с известными ингибиторами белка KasA – фермента, играющего важную роль в биосинтезе клеточной стенки МБТ. Выполнен молекулярный докинг сгенерированных нейронной сетью новых молекул с этой молекулярной мишенью, и с помощью оценочных функций AutoDock Vina, RFScore-4 и NNScore 2.0 осуществлена оценка энергии связывания построенных комплексов с последующим расчетом для каждого соединения экспоненциального консенсусного ранга. На основе полученных данных отобраны шесть лучших по величине ECR соединений, формирующих перспективные базовые структуры для проведения дальнейших теоретических и экспериментальных исследований по разработке новых эффективных ингибиторов лекарственно-устойчивых форм ТБ.

**Благодарности.** Работа поддержана Белорусским республиканским фондом фундаментальных исследований (проект Ф23-007).

**Acknowledgments.** The work was supported by the grant from the Belarusian Republican Foundation for Fundamental Research (project Ф23-007).

### Список использованных источников

1. A comprehensive survey of prospective structure-based virtual screening for early drug discovery in the past fifteen years / H. Zhu, Y. Zhang, W. Li, N. Huang // *International Journal of Molecular Sciences*. – 2022. – Vol. 23, N 24. – Art. 15961. <https://doi.org/10.3390/ijms232415961>
2. Virtual screening algorithms in drug discovery: A review focused on machine and deep learning methods / T. A. D. Oliveira, M. P. D. Silva, E. H. B. Maia [et al.] // *Drugs and Drug Candidates*. – 2023. – Vol. 2, N 2. – P. 311–334. <https://doi.org/10.3390/ddc2020017>
3. Identification of new *Mycobacterium tuberculosis* proteasome inhibitors using a knowledge-based computational screening approach / T. M. Almelebia, M. A. Shahrani, M. Y. Alshahrani [et al.] // *Molecules*. – 2021. – Vol. 26, N 8. – Art. 2326. <https://doi.org/10.3390/molecules26082326>
4. A deep learning approach to antibiotic discovery / J. M. Stokes, K. Yang, K. Swanson [et al.] // *Cell*. – 2020. – Vol. 180, N 4. – P. 688–702. <https://doi.org/10.1016/j.cell.2020.01.021>
5. Drug discovery for *Mycobacterium tuberculosis* using structure-based computer-aided drug design approach / M. A. Ejalonibu, S. A. Ogundare, A. A. Elrashedy [et al.] // *International Journal of Molecular Sciences*. – 2021. – Vol. 22, N 24. – Art. 13259. <https://doi.org/10.3390/ijms222413259>
6. Identification of KasA as the cellular target of an anti-tubercular scaffold / K. A. Abrahams, C. W. Chung, S. Ghidelli-Disse [et al.] // *Nature Communications*. – 2016. – Vol. 7. – Art. 12581. <https://doi.org/10.1038/ncomms12581>
7. Conditional depletion of KasA, a key enzyme of mycolic acid biosynthesis, leads to mycobacterial cell lysis / A. Bhatt, L. Kremer, A. Z. Dai [et al.] // *Journal of Bacteriology*. – 2005. – Vol. 187, N 22. – P. 7596–7606. <https://doi.org/10.1128/jb.187.22.7596-7606.2005>
8. Odena, A. Semi-supervised learning with generative adversarial networks / A. Odena. – 2016. – URL: <https://arxiv.org/pdf/1606.01583> (date of access: 13.01.2025).
9. Jin, W. Junction tree variational autoencoder for molecular graph generation / W. Jin, R. Barzilay, T. Jaakkola // *International conference on machine learning*. – 2018. – P. 2323–2332.
10. Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules / D. Weininger // *Journal of Chemical Information and Computer Sciences*. – 1988. – Vol. 28, N 1. – P. 31–36. <https://doi.org/10.1021/ci00057a005>
11. Slow onset inhibition of bacterial beta-ketoacyl-acyl carrier protein synthases by thiolactomycin / C. A. Machutta, G. R. Bommineni, S. R. Luckner [et al.] // *Journal of Biological Chemistry*. – 2010. – Vol. 285, N 9. – P. 6161–6169. <https://doi.org/10.1074/jbc.m109.077909>
12. Conversion of a beta-ketoacyl synthase to a malonyl decarboxylase by replacement of the active-site cysteine with glutamine / A. Witkowski, A. K. Joshi, Y. Lindqvist, S. Smith // *Biochemistry*. – 1999. – Vol. 38, N 36. – P. 11643–11650. <https://doi.org/10.1021/bi990993h>

13. Structural basis for the recognition of mycolic acid precursors by KasA, a condensing enzyme and drug target from *Mycobacterium tuberculosis* / J. Schiebel, K. Kapilashrami, A. Fekete [et al.] // *Journal of Biological Chemistry*. – 2013. – Vol. 288, N 47. – P. 34190–34204. <https://doi.org/10.1074/jbc.m113.511436>

14. Exponential consensus ranking improves the outcome in docking and receptor ensemble docking / K. Palacio-Rodriguez, I. Lans, C. N. Cavasotto, P. Cossio // *Scientific Reports*. – 2019. – Vol. 9. – Art. 5142. <https://doi.org/10.1038/s41598-019-41594-3>

## References

1. Zhu H., Zhang Y., Li W., Huang N. A comprehensive survey of prospective structure-based virtual screening for early drug discovery in the past fifteen years. *International Journal of Molecular Sciences*, 2022, vol. 23, no. 24, art. 15961. <https://doi.org/10.3390/ijms232415961>

2. Oliveira T. A. D., Silva M. P. D., Maia E. H. B., Silva A. M. D., Taranto A. G. Virtual screening algorithms in drug discovery: A review focused on machine and deep learning methods. *Drugs and Drug Candidates*, 2023, vol. 2, no. 2, pp. 311–334. <https://doi.org/10.3390/ddc2020017>

3. Almeleebia T. M., Shahrani M. A., Alshahrani M. Y., Ahmad I., Alkahtani A. M., Alam M. J., Kausar M. A., Saeed A., Saeed M., Iram S. Identification of new *Mycobacterium tuberculosis* proteasome inhibitors using a knowledge-based computational screening approach. *Molecules*, 2021, vol. 26, no. 8, art. 2326. <https://doi.org/10.3390/molecules26082326>

4. Stokes J. M., Yang K., Swanson K., Jin W., Cubillos-Ruiz A., Donghia N. M., MacNair C. R. [et al.]. A deep learning approach to antibiotic discovery. *Cell*, 2020, vol. 180, no. 4, pp. 688–702. <https://doi.org/10.1016/j.cell.2020.01.021>

5. Ejalonibu M. A., Ogunbare S. A., Elrashedy A. A., Ejalonibu M. A., Lawal M. M., Mhlongo N. N., Kumalo H. M. Drug discovery for *Mycobacterium tuberculosis* using structure-based computer-aided drug design approach. *International Journal of Molecular Sciences*, 2021, vol. 22, no. 24, art. 13259. <https://doi.org/10.3390/ijms222413259>

6. Abrahams K. A., Chung C., Ghidelli-Disse S., Rullas J., Rebollo-Lopez M. J., Gurcha S. S., Cox J. A. G. [et al.]. Identification of KasA as the cellular target of an anti-tubercular scaffold. *Nature Communications*, 2016, vol. 7, art. 12581. <https://doi.org/10.1038/ncomms12581>

7. Bhatt A., Kremer L., Dai A. Z., Sacchettini J. C., Jacobs W. R. Jr. Conditional depletion of KasA, a key enzyme of mycolic acid biosynthesis, leads to mycobacterial cell lysis. *Journal of Bacteriology*, 2005, vol. 187, no. 22, pp. 7596–7606. <https://doi.org/10.1128/jb.187.22.7596-7606.2005>

8. Odena A. Semi-supervised learning with generative adversarial networks, 2016. Available at: <https://arxiv.org/pdf/1606.01583> (accessed 13 January 2025).

9. Jin W., Barzilay R., Jaakkola T. Junction tree variational autoencoder for molecular graph generation. *International conference on machine learning*, 2018, pp. 2323–2332.

10. Weininger D. SMILES, a chemical language and information system. I. Introduction to methodology and encoding rules. *Journal of Chemical Information and Computer Sciences*, 1988, vol. 28, no. 1, pp. 31–36. <https://doi.org/10.1021/ci00057a005>

11. Machutta C. A., Bommineni G. R., Luckner S. R., Kapilashrami K., Ruzsicska B., Simmerling C., Kisker C., Tonge P. J. Slow onset inhibition of bacterial beta-ketoacyl-acyl carrier protein synthases by thiolactomycin. *Journal of Biological Chemistry*, 2010, vol. 285, no. 9, pp. 6161–6169. <https://doi.org/10.1074/jbc.m109.077909>

12. Witkowski A., Joshi K., Lindqvist Y., Smith S. Conversion of a beta-ketoacyl synthase to a malonyl decarboxylase by replacement of the active-site cysteine with glutamine. *Biochemistry*, 1999, vol. 38, no. 36, pp. 11643–11650. <https://doi.org/10.1021/bi990993h>

13. Schiebel J., Kapilashrami K., Fekete A., Bommineni G. R., Schaefer C. M., Mueller M. J., Tonge P. J., Kisker C. Structural basis for the recognition of mycolic acid precursors by KasA, a condensing enzyme and drug target from *Mycobacterium tuberculosis*. *Journal of Biological Chemistry*, 2013, vol. 288, no. 47, pp. 34190–34204. <https://doi.org/10.1074/jbc.m113.511436>

14. 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, art. 5142. <https://doi.org/10.1038/s41598-019-41594-3>

## Информация об авторах

Гончар Анна Викторовна – мл. науч. сотрудник. Объединенный институт проблем информатики НАН Беларуси (ул. Сурганова, 6, 220012, Минск, Республика Беларусь). E-mail: [hanna.hanchar@gmail.com](mailto:hanna.hanchar@gmail.com).

Фурс Константин Викторович – мл. науч. сотрудник. Объединенный институт проблем информатики НАН Беларуси (ул. Сурганова, 6, 220012, Минск, Республика Беларусь). E-mail: [kvfurs@gmail.com](mailto:kvfurs@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](mailto:tuzikov@newman.bas-net.by).

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

## Information about the authors

Gonchar Anna V. – Junior Researcher. United Institute of Informatics Problems of the National Academy of Sciences of Belarus (6, Surganov Str., 220012, Minsk, Republic of Belarus). E-mail: [hanna.hanchar@gmail.com](mailto:hanna.hanchar@gmail.com).

Furs Konstantin V. – Junior Researcher. United Institute of Informatics Problems of the National Academy of Sciences of Belarus (6, Surganov Str., 220012, Minsk, Republic of Belarus). E-mail: [kvfurs@gmail.com](mailto:kvfurs@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](mailto:tuzikov@newman.bas-net.by).

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](mailto:alexande.andriano@yandex.ru).