

doi: 10.17586/2226-1494-2026-26-4-683-694

УДК 575.112:547.96+577.1:004.94+615.012:004.89

Оценочные функции моделей молекулярного докинга белков в автоматизированных системах обработки и анализа биоинформационных данных

Матвей Владимирович Спринджук^{1✉}, Владислав Викторович Кукарцев²,
Анна Валерьевна Ноздрачева³, Андрей Петрович Кончиц⁴,

Александр Станиславович Владыко⁵, Василий Иванович Берник⁶, Елена Ивановна Сукач⁷

¹ Объединенный институт проблем информатики Национальной академии наук Беларуси, Минск, 220072, Беларусь

² Российский государственный аграрный университет – МСХА имени К.А. Тимирязева, Москва, 127434, Российская Федерация

² Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет), Москва, 105005, Российская Федерация

³ Национальный исследовательский центр эпидемиологии и микробиологии имени Н.Ф. Гамалеи, Москва, 123098, Российская Федерация

⁴ Институт леса Национальной академии наук Беларуси, Гомель, 246050, Беларусь

⁵ Научно-исследовательский институт гигиены, токсикологии, эпидемиологии, вирусологии и микробиологии, государственного учреждения «Республиканский центр гигиены, эпидемиологии и общественного здоровья», Минск, 220114, Беларусь

⁶ Институт математики Национальной академии наук Беларуси, Минск, 220072, Беларусь

⁷ Гомельский государственный университет, Гомель, 246050, Беларусь

¹ stepanenkomatvei@yandex.ru ✉, <https://orcid.org/0000-0001-9500-2954>

² vlad_saa_2000@mail.ru, <https://orcid.org/0000-0001-6382-1736>

³ nozdracheva0506@gmail.com, <https://orcid.org/0000-0002-8521-1741>

⁴ konchits@yandex.ru, <https://orcid.org/0000-0002-8823-0630>

⁵ vladalex1703@gmail.com, <https://orcid.org/0009-0001-2188-1893>

⁶ bernik@im.bas-net.by, <https://orcid.org/0000-0001-8096-1079>

⁷ konchits@yandex.ru, <https://orcid.org/0000-0002-3693-1494>

Аннотация

Представлена обобщенная интегрированная классификация оценочных (скоринговых функций) молекулярного докинга моделей взаимодействия белков. Приведено описание различных видов скоринговых функций. Рассмотрены отдельные релевантные вопросы биофармацевтики, медицинской кибернетики, вычислительной биологии и биофизики, прикладной математики. Работа может быть интересна ученым различных специальностей, инженерам-программистам, разрабатывающим автоматизированные системы обработки биоинформации; преподавателям-лекторам по данной тематике; магистрантам и аспирантам по направлениям подготовки: биоинформатика, системный анализ, метрология и приборостроение, системы медицинского назначения и теоретические основы информатики.

Ключевые слова

молекулярный докинг, вычислительная биология, геномика, искусственный интеллект, биоинформатика, вычислительный анализ, компьютерные системы медицинского назначения, приборостроение и метрология

Благодарности

Исследование выполнено при поддержке грантов ISTC (International Science and Technology Center) PR150, Advancing Data Integration, AI-Based Image Analysis, and Genomic Research in Belarus, а также государственных научных программ ОИПИ НАН Беларуси.

Ссылка для цитирования: Спринджук М.В., Кукарцев В.В., Ноздрачева А.В., Кончиц А.П., Владыко А.С., Берник В.И., Сукач Е.И. Оценочные функции моделей молекулярного докинга белков в автоматизированных системах обработки и анализа биоинформационных данных // Научно-технический вестник информационных технологий, механики и оптики. 2026. Т. 26, № 4. С. 683–694. doi: 10.17586/2226-1494-2026-26-4-683-694

Scoring functions of protein molecular docking models in automated systems for processing and analyzing bioinformatics data

Matvey V. Sprindzuk¹✉, Vladislav V. Kukartsev², Anna V. Nozdracheva³,
Andrey P. Konchits⁴, Alexander S. Vladiko⁵, Vassily I. Bernik⁶, Elena I. Sukach⁷

¹ State Scientific Institution “United Institute of Informatics Problems” of the National Academy of Sciences of Belarus, Minsk 220072, Belarus

² Russian State Agrarian University – Moscow Timiryazev Agricultural Academy, Moscow, 127434, Russian Federation

³ Bauman Moscow State Technical University, Moscow, 127434, Russian Federation

⁴ The National Research Center for Epidemiology and Microbiology named after Honorary Academician N.F. Gamaleya, Moscow, 123098, Russian Federation

⁵ Forest Institute of the National Academy of Sciences of the Republic of Belarus, Gomel, 246050, Belarus

⁶ Scientific Research Institute of Hygiene, Toxicology, Epidemiology, Virology and Microbiology, State Institution “Republican Center of Hygiene, Epidemiology and Public Health”, Minsk, 220114, Belarus

⁷ Institute of Mathematics of the National Academy of Sciences of Belarus, Minsk, 220072, Belarus

⁷ Gomel State University of Belarus, Gomel, 246050, Belarus

¹ stepanenkomatvei@yandex.ru✉, <https://orcid.org/0000-0001-9500-2954>

² vlad_saa_2000@mail.ru, <https://orcid.org/0000-0001-6382-1736>

³ nozdracheva0506@gmail.com, <https://orcid.org/0000-0002-8521-1741>

⁴ konchits@yandex.ru, <https://orcid.org/0000-0002-8823-0630>

⁵ vladalex1703@gmail.com, <https://orcid.org/0009-0001-2188-1893>

⁶ bernik@im.bas-net.by, <https://orcid.org/0000-0001-8096-1079>

⁷ konchits@yandex.ru, <https://orcid.org/0000-0002-3693-1494>

Abstract

This review article presents a generalized, integrated classification of scoring functions for molecular docking of protein interaction models. It describes various types of scoring functions and examines relevant issues in biopharmaceuticals, medical cybernetics, computational biology and biophysics, applied mathematics. This article may be of interest to a wide range of readers, including scientists from various fields, software engineers developing automated bioinformatics processing systems, lecturers in relevant topics, graduate and postgraduate students in bioinformatics, systems analysis, medical software systems research.

Keywords

molecular docking, medical cybernetics, computational biology, genomics, pattern recognition, artificial intelligence, bioinformatics, computational analysis, applied mathematics, medical computer systems, measurement software tools, hardware and software engineering

Acknowledgements

The study was supported by grants from ISTC (International Science and Technology Center) PR150, Advancing Data Integration, AI-Based Image Analysis, and Genomic Research in Belarus, as well as State scientific programs of the Joint Institute for Informatics and Problems of the National Academy of Sciences of Belarus.

For citation: Sprindzuk M.V., Kukartsev V.V., Nozdracheva A.V., Konchits A.P., Vladiko A.S., Bernik V.I., Sukach E.I. Scoring functions of protein molecular docking models in automated systems for processing and analyzing bioinformatics data. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2026, vol. 26, no. 4, pp. 683–694 (in Russian). doi: 10.17586/2226-1494-2026-26-4-683-694

Введение

Компьютерные методы, применяемые для разработки, перепрофилирования [1–4], исследования новых лекарственных средств, подразделяются на методы изучения лигандов и на так называемые структурные (рис. 1–3). Методы, основанные на исследовании лиганд, требуют объекта активных лигандов и используют модели количественной зависимости структура-активность (Quantitative Structure–Activity Relationship, QSAR), фармакофор (в широком смысле слова это ансамбль характеристик целевых структур, взаимодействие которых подвергается изучению и анализу)

и химическое сходство для прогнозирования новых соединений. Структурные методы требуют трехмерной структуры рецептора и лигандов, а открытие новых активных соединений основано на определении физических взаимодействий между рецептором и небольшими молекулами с образованием биологически активного комплекса [5–7].

В настоящей работе предпринята попытка интеграции знаний по данной тематике, описание и обобщение релевантной информации.

Программное обеспечение, реализующее алгоритмы и модели молекулярного докинга (МД), сегодня широко используется в автоматизированных системах обра-

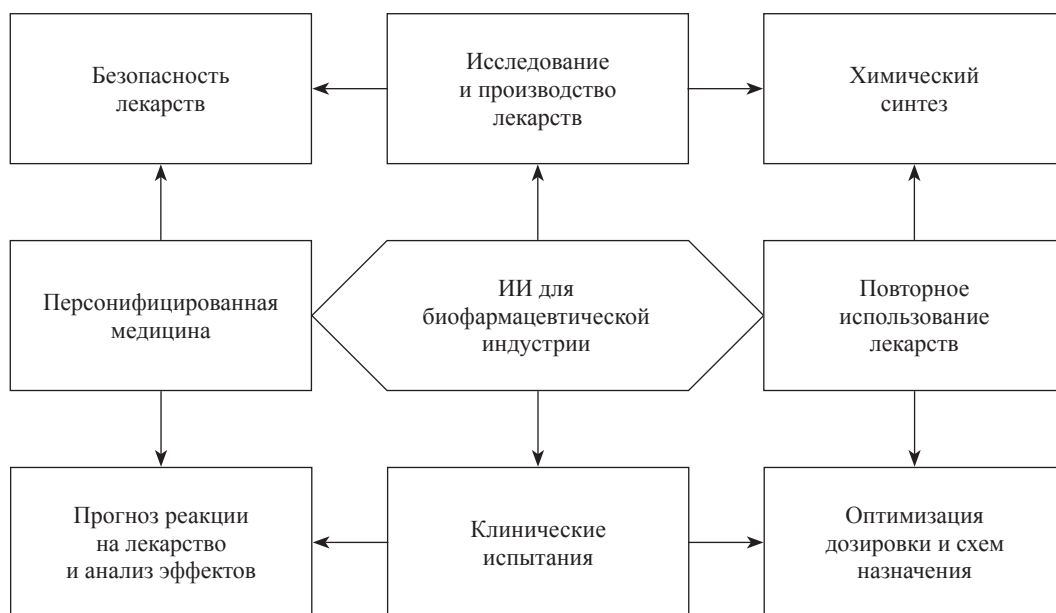


Рис. 1. Сферы применения технологий искусственного интеллекта (ИИ) в биофармацевтике

Fig. 1. Applications of artificial intelligence technologies in biopharmaceuticals

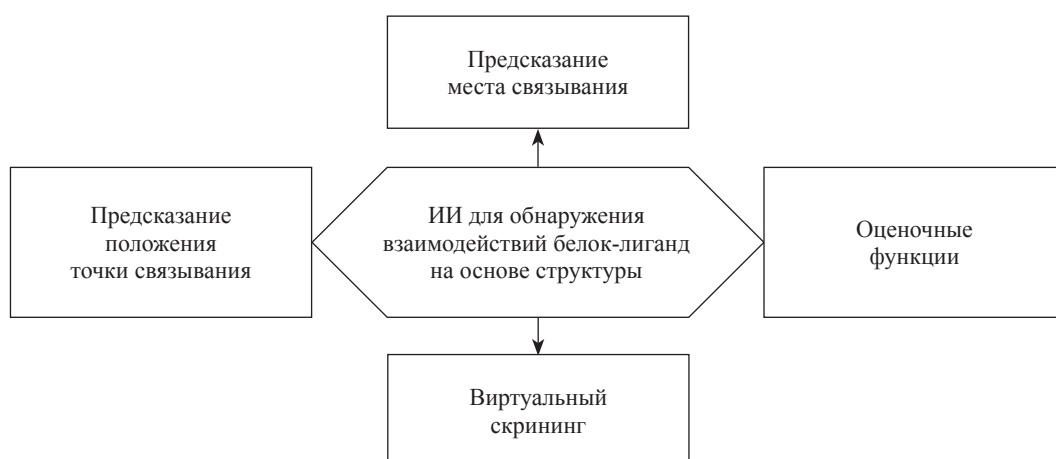


Рис. 2. Области применения технологий искусственного интеллекта для изучения структурной биоинформатики протеинов

Fig. 2. Application areas of artificial intelligence technologies for studying protein structural bioinformatics

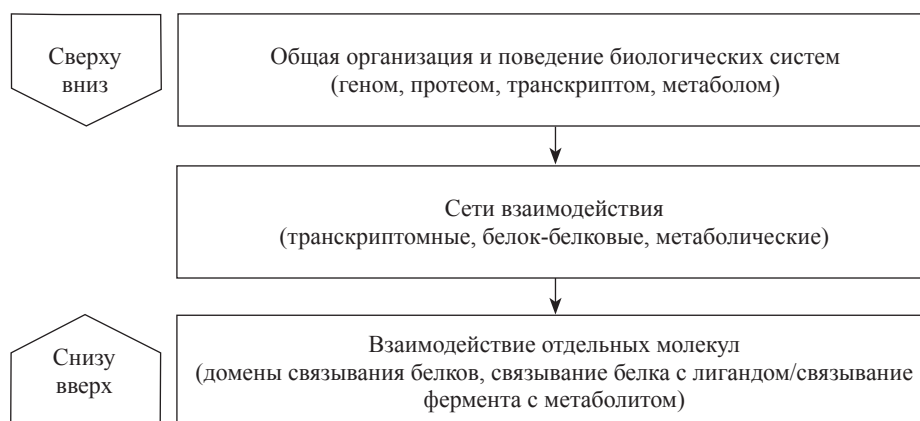


Рис. 3. Организация сфер исследования биоинформатики белковых комплексов

Fig. 3. Organization of research spheres in protein complex bioinformatics

ботки биоинформации для задач и целей хемоинформатики, прикладного биомедицинского приборостроения, вычислительной биологии, ботаники, геномики и транскриптомики, технологий и компьютерных систем медицинского назначения.

Одной из областей применения моделей молекулярного МД взаимодействия белков является изучение устойчивости микробов к антибиотикам. Такая резистентность микроорганизмов представляет собой серьезную угрозу для общественного здравоохранения. Ассоциируемые с этим феноменом высокие показатели заболеваемости и смертности от инфекций, вызванных антибиотикоустойчивыми микроорганизмами, являются следствием неправильного назначения антибиотиков и широко распространенного бесконтрольного их использования. Поиск стратегий борьбы с лекарственной устойчивостью имеет важное значение и высокую актуальность сегодня. Информационные технологии активно применяются для указанной медицинской проблемы.

Программное обеспечение МД основано на двух основных компонентах: алгоритме конформационного поиска и оценочной функции (ОФ) (рис. 4 и 5) [8].

Сравнивать программное обеспечение для докинга достаточно проблемно, так как сайты связывания и про-

цессы распознавания имеют уникальные особенности, которые в итоге делают взаимодействия белок-лиганд специфичными, что может потребовать индивидуального анализа (для конкретных мишеней или связанных рецепторов).

МД [9] и виртуальный скрининг [10–12] — современные эффективные инструменты для выявления новых кандидатов в лекарственные препараты, обладающих антагонистической активностью в отношении ферментов или белковых рецепторов, имеющих медицинское и биологическое значения. Докинг прогнозирует и ранжирует конформации лиганда в конкретной мишени [8].

ОФ (рис. 4 и 5) представляют собой вычислительные методы, широко применяемые при разработке структурно-ориентированных лекарств для оценки взаимодействий белок-лиганд. Большой набор ОФ был опубликован с начала 1990-х годов [13]. Возможность подсчета баллов является основным элементом виртуального скрининга [14]. Докинг белков с белками является ключом к пониманию того, как белки работают вместе. Алгоритмы вычислительного докинга предсказывают трехмерную структуру комплекса, используя данные о трехмерных структурах отдельных белков.

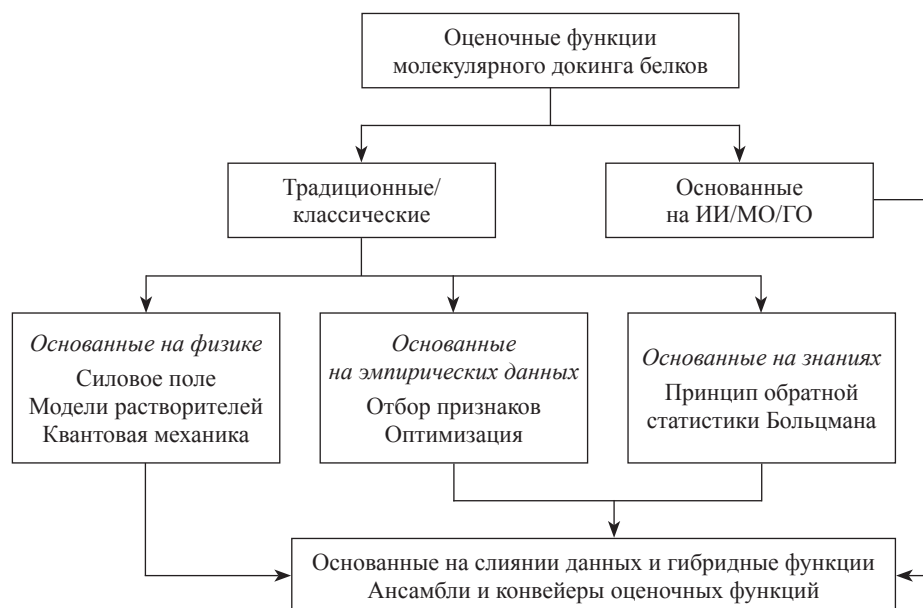


Рис. 4. Современная интегрированная классификация оценочных функций моделей молекулярного докинга.

ИИ — искусственный интеллект; МО — машинное обучение; ГО — глубокое обучение

Fig. 4. Developed integrated modern classification of scoring functions of molecular docking models.

ИИ is artificial intelligence; MO is machine learning; GO is deep learning

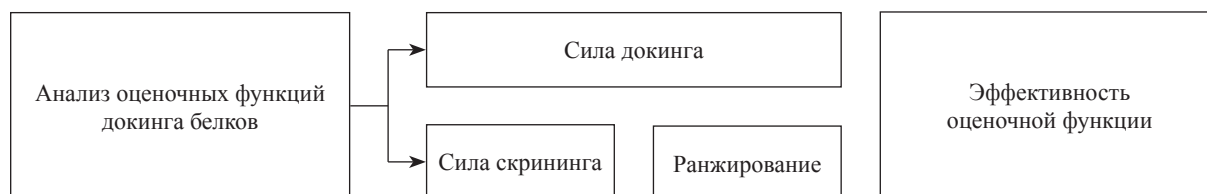


Рис. 5. Компоненты характеристики эффективности оценочных функций

Fig. 5. Components of the performance characteristics of scoring functions

Подсчет очков проводится для выявления конформаций, сходных с нативной структурой. Хорошая ОФ должна правильно и эффективно оценивать модели, близкие к нативным, и присваивать низкие оценки к ненативным конформациям. Задача подсчета очков определяется в конкурсе под названием CAPRI (Critical Assessment of Predicted Interactions, критическая оценка прогнозирования взаимодействий) [15]. Надежность МД зависит от точности применяемых ОФ, которые могут направлять и определять положения лигандов, когда их генерируются тысячи. ОФ может использоваться для определения режима связывания и точного местоположения лиганда, прогнозирования аффинности (меры прочности) связывания и идентификации потенциальных лекарственных средств для заданной белковой мишени. Таким образом, как наиболее важный компонент МД, ОФ имеют три доминирующие функции: определение способа связывания и места связывания лиганда с белком; предсказание абсолютного сродства связывания между белком и лигандом при оптимизации отведения; виртуальный скрининг, который может идентифицировать потенциальные лекарственные средства для данного комплекса белка-мишени путем поиска в большой базе данных лигандов [16].

Метрики сравнительного анализа и характеристик ОФ

Стандартные бенчмарки (проверенные стандартные эталоны) имеют решающее значение для объективного изучения и характеристики ОФ (рис. 5), обеспечивая воспроизводимый и надежный способ сравнения различных методов. PDBbind [17], DUD-E [18], DEKOIS 2.0 [17] являются примерами широко используемых бенчмарков для анализа ОФ.

Для количественной оценки эффективности ОФ учитываются эффективность прогнозирования позиций конформации, классификации «активный/неактивный» и прогнозирования сродства. ОФ обычно сами оцениваются по четырем аспектам, связанным с тремя доминирующими функциями, о которых в тексте статьи уже сообщено.

Сила докинга — способность ОФ определять нативный режим связывания как наиболее правильное ранжированное решение. Среднеквадратичное отклонение является наиболее часто используемой метрикой для характеристики эффективности ОФ. В свою очередь скрининговая сила — способность ОФ правильно отличать активные соединения от неактивных молекул. Тест скрининговой силы не требует, чтобы скоринговая функция правильно предсказывала абсолютную аффинность связывания. Скрининговая сила обычно количественно оценивается с помощью *больцмановского* усиления дискриминации рабочих характеристик приемника и фактора обогащения.

Ранжирующая мощность: способность ОФ правильно ранжировать соединения в соответствии с их аффинностями связывания с одним и тем же целевым белком. Коэффициент корреляции Спирмена и тау Кендалла — метрики, широко используемые для оценки ранжирующей мощности ОФ.

Прогностическая эффективность ОФ может различаться в разных экспериментах по бенчмаркингу из-за таких факторов, как: состав набора данных, структурное качество комплексов, уровень опыта исследователей, проводящих эксперименты, и протокол приготовления комплексов [19].

Поскольку в большинстве исследований бенчмаркинга ОФ определяются на нескольких сотнях комплексов, небольшие различия в коэффициенте корреляции Спирмена, например, между 0,05 и 0,15, не имеют статистической значимости. Таким образом, для надежного сравнения методологий стыковки и ОФ требуются более крупные бенчмаркинговые наборы, состоящие из высококачественных структур комплексов белок-лиганд [19].

Оценочные функции

Физические оценочные функции. Физические ОФ имеют в основе алгоритмы на основе вычислений моделей силовых полей и квантовой механики. Классическая ОФ на основе силового поля определяет энергию связи путем накопления ван-дер-ваальсовского и электростатического взаимодействий между парами атомов белок-лиганд, который учитывает вклад энтальпии (сумма внутренней энергии термодинамической системы и произведения ее давления и объема) в энергию. Поскольку такой метод пренебрегает энтропией (это научная концепция, которая чаще всего ассоциируется с состоянием беспорядка, случайности или неопределенности) и эффектом растворителя, производительность ОФ на основе силового поля недостаточно хороша. Таким образом, ОФ на основе силового поля улучшается за счет включения торсионной энтропии лигандов и эффекта сольватации/десольватации (в водном растворе это гидратация/дегидратация), описываемого явными и неявными моделями растворителя [16].

Protein Interfaces, Surfaces and Assemblies (PISA) — крупный биоинформатический ресурс, разработанный в Европейском институте биоинформатики (Великобритания). Интерактивный сервер вычисляет и анализирует геометрические и физико-химические свойства поверхностей макромолекул, белковых комплексов, определяет их вероятную четвертичную структуру и интерфейсы взаимодействий. Исследователи используют Protein Data Bank in Europe PISA для оценки площади контакта молекул, энергии диссоциации, водородных связей и гидрофобных взаимодействий после проведения белок-белкового докинга. Atomic Potential Protein Interfaces, Surfaces and Assemblies (AP-PISA) и Coarse-grained Potentials — Protein Interaction Energy (CP-PIE) относятся к числу известных классических функций для распознавания правильных моделей белковых комплексов. AP-PISA — оценка на основе атомных потенциалов, учитывающая расположение и свойства отдельных атомов белка. CP-PIE — крупнозернистая (упрощенная) оценка, основанная на вычислении энергии белкового взаимодействия и учете площади перекрытия взаимодействующих молекул. В ряде научных публикаций CP-PIE часто выделяется как отличный, быстрый фильтр для отсеивания не-

правильных структур до того, как будут применены более ресурсоемкие алгоритмы. Обе функции являются классическими инструментами, которые до сих пор используются в качестве надежного эталона для сравнения с новыми нейросетевыми моделями и алгоритмами [20, 21].

Оценочные функции, основанные на знаниях. ОФ, основанные на знаниях, получают желаемые парные потенциалы из трехмерных структур большого набора комплексов белок-лиганд, используя обратный статистический принцип Больцмана. Предполагается, что частота различных пар атомов на разных расстояниях связана с взаимодействием двух атомов и преобразует частоту в зависящий от расстояния потенциал средней силы. Самым большим преимуществом применения ОФ, основанных на знаниях, является снижение стоимости вычислений и повышение точности прогнозирования по сравнению с физическими и эмпирическими ОФ [16].

Отметим, что таким функциям проблематично найти эталонное состояние. В настоящее время существуют две классические стратегии определения эталонного состояния. Первая стратегия — аппроксимация эталонного состояния случайным распределением пар атомов в наборе обучающих данных; вторая представляет элемент коррекций, основанный на первой стратегии для повышения точности и достоверности ОФ. При этом достоверность ОФ второй стратегии опирается на результаты вычисления алгоритма коррекции объемного коэффициента, основанного на физических процессах и явлениях итеративного вычислительного метода.

Эмпирические оценочные функции. Эмпирические ОФ [22–28] отражают объединение комплекса путем суммирования основных энергетических факторов для связывания белок-лиганд, таких как водородные связи, гидрофобные эффекты, стерические столкновения и т. д. Обычно используется обучающий набор с известными сродствами связывания для оптимизации весов энергетических факторов для эмпирических ОФ с помощью линейного регрессионного анализа. Эмпирические ОФ изучаются по двум основным направлениям исследований: применение большого и высококачественного набора обучающих данных для оптимизации белок-лигандных структур; выбор подходящих энергетических условий с помощью пошаговых переменных и систематического отбора в отношении целевого белка. В настоящее время эмпирические ОФ обычно реализуются в программных системах докинга белков и лигандов. Хотя эмпирические ОФ разлагают сродство связывания белок-лиганд на несколько отдельных энергетических характеристик, подобно ОФ, основанных на физических процессах, как правило, они применяют гибкую и интуитивную функциональную форму, отличную от использования устоявшихся моделей, которые имеют ОФ, основанные на таких физических вычислениях. Эмпирические ОФ эффективны для прогнозирования сродства связывания, положения лиганда и виртуального скрининга с низкими вычислительными затратами, однако плохо подходят для описания взаимосвязей между сродством связывания и кристаллоидными элементами и сталки-

ваются с проблемой двойного счета, которая возникает при оценке энергии взаимодействия.

Оценочные функции на основе машинного обучения. В отличие от классических ОФ с предполагаемой математической функциональной формой, ОФ на основе машинного обучения [29–32] используют различные вычислительные методы машинного обучения, такие как машины опорных векторов, случайные леса, регрессии, дискриминантные модели, бустинг и пакетирование, кластеризация и различные нейронные сети, включая сложные архитектуры ГО. Заметим, что согласно данным современных научных публикаций, ОФ на основе машинного обучения превзошли классические ОФ, они редко напрямую включаются в программное обеспечение для докинга, но обычно используются для повторной его оценки. Причина такого решения заключается в том, что ОФ на основе машинного обучения зависят от наборов обучающих данных. Если белок и лиганд в моделировании стыкуются с помощью классического программного обеспечения для докинга, а затем стыкованная структура анализируется с помощью ОФ машинного обучения, возможно, что точность докинга будет значительно улучшена.

Глубокое обучение (ГО). По сравнению с другими ОФ, таковые с применением ГО способны извлекать признаки вместе с процессом подбора параметров модели к доступным данным. Модель сверточной нейронной сети — одна из популярных моделей ГО для ОФ. Данную модель можно применять для классификации связывания лекарственных средств и прогнозирования аффинности прямого связывания. В настоящее время подтверждено, что модель сверточной нейронной сети более эффективна, чем классические методы машинного обучения, хотя она становится чрезвычайно трудоемкой по мере увеличения узлов и глубины сети. При этом, так как глубокие нейронные сети подходят для разработки многозадачных классификаторов или регрессоров, они могут прогнозировать сродство связывания белка с лигандом и классифицировать лиганды по набору их признаков. Отметим, что ОФ с ГО и ОФ, основанные на физических знаниях, и эмпирические ОФ активно применяются в научных экспериментах. Такие ОФ были успешно включены в популярное программное обеспечение для докинга, такое как DOCK, SYBYL, Discovery Studio, Schrodinger, Autodock, AutodockVina и многие другие. Вместе с тем у каждого ОФ есть очевидные недостатки, программное обеспечение для докинга не может работать идеально во всех аспектах. ОФ, основанные на физических процессах, являются мощными инструментами прогнозирования свободной энергии, однако работают достаточно медленно.

Анализ научных публикаций показал, что методы ГО значительно лучше классических методов для 6 наборов данных, а именно CAPRI Scoreset v2022 (как сложных, так и простых комплексов), BM4, Dockground, PDB-2023 и MaSIF-test. Программный инструмент Protein Interfaces with Transformer Network (PIsToN) показал самое высокое значение Area Under The Curve (площадь под кривой) для четырех из 6 наборов данных, в то время как GNN-DOVE (Graph Neural Network-based DOcking decoy eValuation scorE)

превзошел своих конкурентов для Dockground. Почти для всех наборов данных методы, основанные на ГО, превзошли конкурирующие классические методы. Для наборов данных CAPRI Score Refined, BM4 и PDB-2023 у PIsToN было наибольшее количество результатов в верхних рейтингах для прогнозов top 1, top 10, top 25, top 100 и top 200, за исключением одного значения, когда HADDOCK (High Ambiguity Driven DOCKing) [20] показал себя лучше других методов с 94 % попаданий для прогнозов top 25 только с набором данных BM4. Заметим, что PIsToN уступает своему конкуренту dMaSIF с точки зрения измерения времени выполнения из-за этапа предварительной обработки. PIsToN предварительно вычисляет характеристики сетки поверхности двух белков, участвующих в создании комплекса, чтобы зафиксировать топологические взаимосвязи и геометрию их поверхностей.

Несмотря на то, что ГО справляется с большими и сложными молекулярными данными, оно имеет ряд критических недостатков при моделировании QSAR. К ним можно отнести следующие: склонность к переобучению, экстремальная зависимость от огромных объемов высококачественных данных и проблема «черного ящика», которая скрывает механизмы прогнозирования связи между структурой и активностью. Модели ГО крайне чувствительны к шумам, ошибкам и неравномерному распределению классов в обучающей выборке. В базах скрининга «неактивных» молекул всегда значительно больше, чем «активных». Глубокие сети легко перенимают этот дисбаланс и начинают предсказывать только мажоритарный класс, что требует сложных методов аугментации данных. Прогнозы надежны только для тех соединений, которые входят в химическое пространство, использованное при обучении. Поскольку ГО фиксирует сложные нелинейные связи, оно может выдавать нереалистичные прогнозы для молекул с принципиально новыми каркасами или редкими комбинациями функциональных групп. Обучение глубоких архитектур требует больших мощностей и долгой настройки множества гиперпараметров (скорость обучения, размер батча, количество слоев). Процессы итеративной разработки моделей и кросс-валидации становятся медленными и дорогими по сравнению с традиционными статистическими методами [33, 34].

Гибридные мультикомпонентные оценочные функции. Учитывая, что классические ОФ имеют свои преимущества и ограничения, для повышения точности ОФ обычно используется объединение различных функций оценки (гибридные консенсусные ОФ). Примерами гибридных ОФ являются MultiScore, GFscore [35], SeleX-CS [36], VoteDock [37]. Гибридные ОФ объединяют различные индивидуальные оценки с помощью стратегий консенсуса, таких как голосование по числу, число по числу, ранг по числу, средний ранг, линейная комбинация и т. д. Можно предположить, что гибридные ОФ будут работать лучше, чем автономные единичные ОФ. Данное предположение связано с тем, что объединение разнородных типов ОФ может компенсировать недостатки каждого из них. Однако такие функции подсчета баллов, очевидно, требуют больше времени. По этой причине исследования функций оцен-

ки на основе машинного обучения стали в настоящее время актуальной областью исследований. Точность прогнозирования для ОФ, основанных на машинном обучении, можно повысить, увеличив набор обучающих данных. Поскольку из академических и промышленных областей будет поступать все больше структурных и интерактивных данных, можно допустить, что ОФ машинного обучения имеют потенциал доминировать как ОФ будущего. PIsToN является одним из самых успешных с точки зрения точности и эффективности и новейшим среди современных инструментов на основе ГО. PIsToN вырезает и изолирует пары участков из интерфейса, по одному для каждого взаимодействующего белка. Затем информация с каждого участка преобразуется в многоканальные изображения, где каждый канал представляет одну из характеристик взаимодействия, включая относительную доступную площадь поверхности, кривизну, заряд, электростатику и гидрофобность. Использование модуля Vision Transformer (ViT), отличает его от других методов. Встраивание, произведенное ViT, затем объединяется с «гибридными» эмпирическими энергетическими терминами, включая Ван-дер-Ваальса, электростатику и десольватацию, что позволяет ему изучать характеристики сложных энергий связи.

Модель множественного внимания группирует энергетические характеристики и свойства интерфейса в несколько групп, каждая из которых содержит независимую сеть трансформаторов. Эти независимые сети объединяются в преобразователь-кодер для оценки комплексов как собственных или не собственных ложных сигналов.

Графовые нейронные сети (ГНН) — новый метод обработки данных, представляющий собой набор архитектур ГО для работы с данными, которые можно представить в виде графа. Подавляющее большинство ГНН делятся на три категории (сверточные, внимания и передачи сообщений).

Граф (V, E) состоит из наборов вершин $v_i \in V$ и ребер $e_{ij} \in E$, соединяющих вершины. Характеристики x связаны с вершинами (и, необязательно, ребрами), и такие характеристики впоследствии обновляются следующим образом [38]:

$$h_u = \Phi(x_u, \bigoplus_{v \in \mathcal{N}_u} \psi(x_u, x_v)),$$

где Φ и ψ — обучаемые функции (часто обучаемые аффинные преобразования с функциями активации); $\bigoplus$ — инвариантная к перестановкам функция, позволяющая агрегировать признаки (такие как сумма, среднее значение и максимум по соседству $\mathcal{N}_u$ узла u). При этом: ψ — функция передачи сообщений (которая может быть обобщена для включения также признаков ребер), а Φ — функция обновления вершин. Заметим, что можно изучать признаки ребер, введя скрытое представление характеристик-признаков h_{uv} для ребер [38].

Фактически молекулы можно естественным образом представить в виде графов, где узлы графов являются различными атомами, а ребра графа представляют химические связи между атомами, ГНН хорошо подходят для применения в области химии [38].

Консенсусная оценка и переоценка

Известно, что каждая функция оценки имеет свои сильные и слабые стороны. Способом улучшения результатов сверх уровня, который может быть получен с помощью одной ОФ, является использование методологии, называемой консенсусной оценкой.

ОФ консенсуса объединяют различные ОФ, присваивая разные веса вкладу различных альтернатив, чтобы улучшить различение активных и неактивных элементов в виртуальном скрининге. Ряд исследований продемонстрировали, что применение ОФ консенсуса может уменьшить количество ложноположительных и ложноотрицательных результатов, а также виртуального скрининга.

В результате повторной оценки позиций (конформации), сгенерированные ОФ программой докинга, оцениваются с целью поиска благоприятных конформаций и ранжируются для выбора наиболее удачных совпадений. Однако расчет сродства связывания является неопределенным из-за внутренних проблем, связанных с упрощенными условиями оценки. Один из способов решения этой проблемы — использование более строгих энергетических расчетов после процесса докинга, хотя это ограничено чрезмерными вычислительными затратами для больших химических библиотек. Таким образом, результаты, полученные после докинга, можно переоценить и/или отфильтровать с использованием различных ОФ, которые оценивают различные характеристики.

Программное обеспечение XScore и DrugScore eXtended (DSX) обычно используется для переоценки моделей МД: XScore¹ вычисляет сродство связывания заданных молекул лиганда, связанных с целевым белком, с помощью эмпирической функции, которая включает взаимодействия Ван-дер-Ваальса, водородные связи, а также гидрофобные и деформационные члены. DSX — компьютерная программа, имеющая ОФ, основанную на знаниях, позволяющая оценивать интересующие комплексы белок-лиганд и визуализировать вклады оценок на атом, что является интуитивно понятным способом узнать о различиях между предполагаемыми геометриями лигандов или о важности определенных областей связывания. Для определения того, какие соединения, по прогнозам, будут связываться с мишенью, можно использовать как первичные, так и пересчитанные баллы.

В работе [39] выполнена оценка четырех ОФ для повторного контроля результатов докинга с использованием высоконадежного набора данных скрининга, охватывающего несколько фармацевтических целей. Было обнаружено, что переоценка, основанная на упрощенных ОФ на знаниях, например, измерение образов взаимодействия, работает лучше, чем современные методы машинного обучения, что показывает важность использования методов переоценки для адекватного обнаружения наиболее мощных связующих веществ.

Дальнейшие исследования показали, что использование подходов машинного обучения для переоценки результатов докинга значительно повышает производительность структурных моделей, а ансамбли функций переоценки повышают точность прогнозирования. В [39] был сделан вывод, что использование эмпирических данных для оценки прогнозов докинга является ключевым фактором для улучшения прогнозирования взаимодействий белок-лиганд при разработке лекарственных препаратов.

Таким образом, методы переоценки могут повысить точность результатов докинга. Однако переоценка результирующих позиций, полученных в AutoDock4 [40–42] с помощью XScore [43, 44] или DSX [45], улучшила способность модели различать связывающих и не связывающих людей. В нескольких исследованиях использование стратегий консенсуса увеличило площадь под ROC-кривой (Receiver Operating Characteristic, рабочая характеристика приемника), что указывает на улучшение дискриминации активных соединений по сравнению с ложными, неактивными. Методы нормализованного отношения баллов, ранжирования по числу, ранжирования по рангу и Z-оценки превосходят любое индивидуальное программное обеспечение для докинга.

Общая эффективность различных ОФ сопоставима при дифференцировке активных и ложных целей. Однако результаты могут существенно различаться в зависимости от типа мишени и лиганда. Например, ChemScore [46] дает значительно лучшие результаты, за ним следует Vina [47, 48]. Из-за недостаточной точности функций подсчета баллов лучшие решения типичной кампании виртуального скрининга включают в себя большое количество соединений, которые при экспериментальном тестировании не связываются прочно с мишенью (т. е. дают ложноположительные результаты). В целом было показано, что точность различных программ докинга и ОФ зависит от конкретного изучаемого белка-мишени и типа лигандов. Связывающие карманы в разных белках различаются по размеру, полярности, доступности растворителя, составу аминокислотных остатков и, что наиболее важно, по количеству, типу и ориентации специфических доступных групп взаимодействия. Лиганды могут существенно различаться по размеру, длине, полярности и гидрофобности, количеству вращающихся связей, количеству возможных доноров или акцепторов водорода и т. д. Сочетание этих особенностей приводит к тому, что перед программой докинга могут стоять совершенно разные задачи, и лучшее решение для решения одной проблемы не обязательно будет лучшим решением для решения другой проблемы. Среди всех классических алгоритмов и инструментов оценки моделей докинга белков Atomic Potential Protein Interfaces, Surfaces and Assemblies (AP-PISA) оказался наиболее эффективным методом во всех восьми наборах данных. В отличие от других классических методов, которые в основном рассматривают потенциалы на уровне остатков, AP-PISA переоценивает уточненные модели докинга, используя комбинацию атомных и остаточных потенциалов, что помогает более точно определять почти нативные позиции.

¹ [Электронный ресурс]. Режим доступа: https://www.ics.uci.edu/~dock/manuals/xscore1.1_manual/intro.html (дата обращения: 21.07.2026).

Однако AP-PISA модель опирается исключительно на потенциальные характеристики, но включение энергетических характеристик может повысить производительность за счет использования эмпирических и основанных на знаниях характеристик.

По метрике AUC (Area Under The Curve, площадь под кривой) метод Coarse-grained Potentials — Protein Interaction Energy (CP-PIE) оказался вторым лучшим классическим методом и лучшим по ранжирующей способности. CP-PIE эффективно отфильтровывает маловероятные позиции докинга перед выполнением ресурсоемких вычислений. Он отбрасывает неправильно стыкованные структуры, измеряя область перекрытия между белками в каждой модели докинга. Один из выводов, который можно сделать, заключается в том, что на этапе уточнения удаление позиций, которые далеки от правильных, может иметь решающее значение для точной оценки моделей докинга [15].

Заключение

Не существует универсальной идеальной оценочной функции, которая могла бы хорошо работать для любого вычисления молекулярного докинга. Таким образом, исследователям приходится разрабатывать оценочную функцию для конкретной цели, используя арсенал возможностей современной математики, биофизики и кибернетики. Новые классы характеристик (межмолекулярные характеристики, характеристики, связанные только с лигандами и белками) могут быть использо-

ваны для улучшения производительности оценочной функции. Эффективность скрининга многозадачной оценочной функции существенно выше, чем у однозадачной нейронной сети. Разрыв в производительности между двумя функциями гораздо меньше для задач докинга и подсчета очков, с небольшим преимуществом при традиционном подходе с одной задачей. Рост производительности будет гораздо более значительным, если данные оценки, скрининга и докинга будут более разнообразными и объемными.

Важным фактором, влияющим на различия в производительности, могут быть инструменты докинга и итоговое качество моделей докинга, используемых для обучения моделей. Различные методы применяют разные инструменты докинга для создания моделей докинга для своих инструментов оценки. Характеристики состава и качества данных играют решающую роль в определении общей эффективности модели, поскольку они напрямую влияют на способность модели улавливать значимые закономерности и делать точные прогнозы.

Поскольку структурная биоинформатика и молекулярное моделирование продолжают развиваться благодаря технологиям искусственного интеллекта и глубокому обучению, в частности, крайне важно оценивать как классические, так и подходы глубокого обучения с использованием различных наборов данных и конкретных предварительно выбранных целей, чтобы в итоге обеспечить последовательное, эффективное и справедливое сравнение результатов моделирования.

Литература

1. Ugbaja S.C., Mushebenge A.G., Kumalo H., Ngcobo M., Gqaleni N. Potential benefits of in silico methods: a promising alternative in natural compound's drug discovery and repurposing for HBV therapy // *Pharmaceuticals*. 2025. V. 18. N 3. P. 419. doi: 10.3390/ph18030419
2. Tanoli Z., Fernandez-Torres A., Ozcan U.O., Kushnir A., Nader K.M., Gadiya Y., et al. Computational drug repurposing: approaches, evaluation of in silico resources and case studies // *Nature Reviews Drug Discovery*. 2025. V. 24. N 7. P. 521–542. doi: 10.1038/s41573-025-01164-x
3. Kalsi P., Jain P., Goyal G., Sharma H. In silico drug repurposing approach to predict most effective HAART for HIV drug resistance variants prevalent in the Indian HIV-positive population // *AIDS Reviews*. 2024. V. 26. N 3. P. 93–101. doi: 10.24875/AIDSRev.24000010
4. Hussein D., Almatrafi A., Gomaa M., Alhowsawi A., Almustafa S., Alsaihaty H., et al. In silico drug repurposing using molecular docking and dynamics to target the protein interaction between the SARS-CoV-2 S-glycoprotein and the ACE2 receptor // *F1000Research*. 2023. V. 12. P. 1452. doi: 10.12688/f1000research.131508.2
5. Bhaskar V., Kumar S., Nair A.S., Gokul S., Krishnendu P.R., Benny S., et al. In silico development of potential InhA inhibitors through 3D-QSAR analysis, virtual screening and molecular dynamics // *Journal of Biomolecular Structure and Dynamics*. 2025. V. 43. N 3. P. 1329–1351. doi: 10.1080/07391102.2023.2291549
6. Moulishankar A., Sankaranarayanan M., Thirugnanasambandam S., Dhamotharan J., Mohanradja D., Sivakumar P.M. Identification of novel DNA gyrase inhibitor by combined pharmacophore modeling, QSAR analysis, molecular docking, molecular dynamics, ADMET and DFT approaches // *Acta Tropica*. 2024. V. 260. P. 107460. doi: 10.1016/j.actatropica.2024.107460
7. Agarwal S., Verma E., Kumar V., Lall N., Sau S., Iyer A.K., et al. An integrated computational approach of molecular dynamics

References

1. Ugbaja S.C., Mushebenge A.G., Kumalo H., Ngcobo M., Gqaleni N. Potential benefits of in silico methods: a promising alternative in natural compound's drug discovery and repurposing for HBV therapy. *Pharmaceuticals*, 2025, vol. 18, no. 3, pp. 419. doi: 10.3390/ph18030419
2. Tanoli Z., Fernandez-Torres A., Ozcan U.O., Kushnir A., Nader K.M., Gadiya Y., et al. Computational drug repurposing: approaches, evaluation of in silico resources and case studies. *Nature Reviews Drug Discovery*, 2025, vol. 24, no. 7, pp. 521–542. doi: 10.1038/s41573-025-01164-x
3. Kalsi P., Jain P., Goyal G., Sharma H. In silico drug repurposing approach to predict most effective HAART for HIV drug resistance variants prevalent in the Indian HIV-positive population. *AIDS Reviews*, 2024, vol. 26, no. 3, pp. 93–101. doi: 10.24875/AIDSRev.24000010
4. Hussein D., Almatrafi A., Gomaa M., Alhowsawi A., Almustafa S., Alsaihaty H., et al. In silico drug repurposing using molecular docking and dynamics to target the protein interaction between the SARS-CoV-2 S-glycoprotein and the ACE2 receptor. *F1000Research*, 2023, vol. 12, pp. 1452. doi: 10.12688/f1000research.131508.2
5. Bhaskar V., Kumar S., Nair A.S., Gokul S., Krishnendu P.R., Benny S., et al. In silico development of potential InhA inhibitors through 3D-QSAR analysis, virtual screening and molecular dynamics. *Journal of Biomolecular Structure and Dynamics*, 2025, vol. 43, no. 3, pp. 1329–1351. doi: 10.1080/07391102.2023.2291549
6. Moulishankar A., Sankaranarayanan M., Thirugnanasambandam S., Dhamotharan J., Mohanradja D., Sivakumar P.M. Identification of novel DNA gyrase inhibitor by combined pharmacophore modeling, QSAR analysis, molecular docking, molecular dynamics, ADMET and DFT approaches. *Acta Tropica*, 2024, vol. 260, pp. 107460. doi: 10.1016/j.actatropica.2024.107460
7. Agarwal S., Verma E., Kumar V., Lall N., Sau S., Iyer A.K., et al. An integrated computational approach of molecular dynamics

- simulations, receptor binding studies and pharmacophore mapping analysis in search of potent inhibitors against tuberculosis // *Journal of Molecular Graphics and Modelling*. 2018. V. 83. P. 17–32. doi: 10.1016/j.jmgm.2018.04.019
8. Blanes-Mira C., Fernández-Aguado P., de Andrés-López J., Fernández-Carvajal A., Ferrer-Montiel A., Fernández-Ballester G. Comprehensive survey of consensus docking for high-throughput virtual screening // *Molecules*. 2022. V. 28. N 1. P. 175. doi: 10.3390/molecules28010175
 9. Drakontaeidi A., Pontiki E. A Review on molecular docking on HDAC isoforms: novel tool for designing selective inhibitors // *Pharmaceuticals*. 2023. V. 16. N 12. P. 1639. doi: 10.3390/ph16121639
 10. Scior T., Bender A., Tresadern G., Medina-Franco J.L., Martinez-Mayorga K., et al. Recognizing pitfalls in virtual screening: a critical review // *Journal of Chemical Information and Modeling*. 2012. V. 52. N 4. P. 867–881. doi: 10.1021/ci200528d
 11. Lavecchia A., Di Giovanni C. Virtual screening strategies in drug discovery: a critical review // *Current Medicinal Chemistry*. 2013. V. 20. N 23. P. 2839–2860. doi: 10.2174/09298673113209990001
 12. Murugan N.A., Podobas A., Gadioli D., Vitali E., Palermo G., Markidis S. A Review on parallel virtual screening softwares for high-performance computers // *Pharmaceuticals*. 2022. V. 15. N 1. P. 63. doi: 10.3390/ph15010063
 13. Liu J., Wang R. Classification of current scoring functions // *Journal of Chemical Information and Modeling*. 2015. V. 55. N 3. P. 475–482. doi: 10.1021/ci500731a
 14. Balamurugan S., Krishnan A.T., Goyal D., Chandrasekaran B., Pandi B. *Computation in Bioinformatics: Multidisciplinary Applications*. Wiley-Scrivener, 2021. 323 p.
 15. Shirali A., Stebliankin V., Karki U., Shi J., Chapagain P., Narasimhan G. A comprehensive survey of scoring functions for protein docking models // *BMC Bioinformatics*. 2025. V. 26. N 1. P. 25. doi: 10.1186/s12859-024-05991-4
 16. Li J., Fu A., Zhang L. An Overview of scoring functions used for protein-ligand interactions in molecular docking // *Interdisciplinary Sciences: Computational Life Sciences*. 2019. V. 11. N 2. P. 320–328. doi: 10.1007/s12539-019-00327-w
 17. Boeckler F.M., Bauer M.R., Ibrahim T.M., Vogel S.M. Use of DEKOIS 2.0 to gain insights for virtual screening // *Journal of Cheminformatics*. 2014. V. 6. Suppl 1. P. O24. doi: 10.1186/1758-2946-6-S1-O24
 18. Cleves A.E., Jain A.N. Structure- and Ligand-Based Virtual screening on DUD-E(+): performance dependence on approximations to the binding pocket // *Journal of Chemical Information and Modeling*. 2020. V. 60. N 9. P. 4296–4310. doi: 10.1021/acs.jcim.0c00115
 19. Guedes I.A., Pereira F.S., Dardenne L.E. Empirical scoring functions for structure-based virtual screening: applications, critical aspects, and challenges // *Frontiers in Pharmacology*. 2018. V. 9. P. 1089. doi: 10.3389/fphar.2018.01089
 20. Shirali A., Stebliankin V., Karki U., Shi J., Chapagain P., Narasimhan G. A comprehensive survey of scoring functions for protein docking models // *BMC Bioinformatics*. 2025. V. 26. N 1. P. 25. doi: 10.1186/s12859-024-05991-4
 21. de Angelo R.M., de Sousa D.S., da Silva A.P., Chiari L.P., da Silva A.B., Honorio K.M. Molecular docking: State-of-the-art scoring functions and search algorithms // *Computer-Aided Drug Discovery and Design*. 2025. V. 3. P. 163–198. doi: 10.1007/978-3-031-76718-0_7
 22. Eldridge M.D., Murray C.W., Auton T.R., Paolini G.V., Mee R.P. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes // *Journal of Computer-Aided Molecular Design*. 1997. V. 11. N 5. P. 425–445. doi: 10.1023/a:1007996124545
 23. Murray C.W., Auton T.R., Eldridge M.D. Empirical scoring functions. II. The testing of an empirical scoring function for the prediction of ligand-receptor binding affinities and the use of Bayesian regression to improve the quality of the model // *Journal of Computer-Aided Molecular Design*. 1998. V. 12. N 5. P. 503–519. doi: 10.1023/a:1008040323669
 24. Wang R., Lai L., Wang S. Further development and validation of empirical scoring functions for structure-based binding affinity prediction // *Journal of Computer-Aided Molecular Design*. 2002. V. 16. N 1. P. 11–26. doi: 10.1023/a:1016357811882
 25. de Azevedo W.F., Dias R. Evaluation of ligand-binding affinity using polynomial empirical scoring functions // *Bioorganic and Medicinal*
 - simulations, receptor binding studies and pharmacophore mapping analysis in search of potent inhibitors against tuberculosis. *Journal of Molecular Graphics and Modelling*, 2018, vol. 83, pp. 17–32. doi: 10.1016/j.jmgm.2018.04.019
 8. Blanes-Mira C., Fernández-Aguado P., de Andrés-López J., Fernández-Carvajal A., Ferrer-Montiel A., Fernández-Ballester G. Comprehensive survey of consensus docking for high-throughput virtual screening. *Molecules*, 2022, vol. 28, no. 1, pp. 175. doi: 10.3390/molecules28010175
 9. Drakontaeidi A., Pontiki E. A Review on molecular docking on HDAC isoforms: novel tool for designing selective inhibitors. *Pharmaceuticals*, 2023, vol. 16, no. 12, pp. 1639. doi: 10.3390/ph16121639
 10. Scior T., Bender A., Tresadern G., Medina-Franco J.L., Martinez-Mayorga K., et al. Recognizing pitfalls in virtual screening: a critical review. *Journal of Chemical Information and Modeling*, 2012, vol. 52, no. 4, pp. 867–881. doi: 10.1021/ci200528d
 11. Lavecchia A., Di Giovanni C. Virtual screening strategies in drug discovery: a critical review. *Current Medicinal Chemistry*, 2013, vol. 20, no. 23, pp. 2839–2860. doi: 10.2174/09298673113209990001
 12. Murugan N.A., Podobas A., Gadioli D., Vitali E., Palermo G., Markidis S. A Review on parallel virtual screening softwares for high-performance computers. *Pharmaceuticals*, 2022, vol. 15, no. 1, pp. 63. doi: 10.3390/ph15010063
 13. Liu J., Wang R. Classification of current scoring functions. *Journal of Chemical Information and Modeling*, 2015, vol. 55, no. 3, pp. 475–482. doi: 10.1021/ci500731a
 14. Balamurugan S., Krishnan A.T., Goyal D., Chandrasekaran B., Pandi B. *Computation in Bioinformatics: Multidisciplinary Applications*. Wiley-Scrivener, 2021. 323 p.
 15. Shirali A., Stebliankin V., Karki U., Shi J., Chapagain P., Narasimhan G. A comprehensive survey of scoring functions for protein docking models. *BMC Bioinformatics*, 2025, vol. 26, no. 1, pp. 25. doi: 10.1186/s12859-024-05991-4
 16. Li J., Fu A., Zhang L. An Overview of scoring functions used for protein-ligand interactions in molecular docking. *Interdisciplinary Sciences: Computational Life Sciences*, 2019, vol. 11, no. 2, pp. 320–328. doi: 10.1007/s12539-019-00327-w
 17. Boeckler F.M., Bauer M.R., Ibrahim T.M., Vogel S.M. Use of DEKOIS 2.0 to gain insights for virtual screening. *Journal of Cheminformatics*, 2014, vol. 6, suppl 1, pp. O24. doi: 10.1186/1758-2946-6-S1-O24
 18. Cleves A.E., Jain A.N. Structure- and Ligand-Based Virtual screening on DUD-E(+): performance dependence on approximations to the binding pocket. *Journal of Chemical Information and Modeling*, 2020, vol. 60, no. 9, pp. 4296–4310. doi: 10.1021/acs.jcim.0c00115
 19. Guedes I.A., Pereira F.S., Dardenne L.E. Empirical scoring functions for structure-based virtual screening: applications, critical aspects, and challenges. *Frontiers in Pharmacology*, 2018, vol. 9, pp. 1089. doi: 10.3389/fphar.2018.01089
 20. Shirali A., Stebliankin V., Karki U., Shi J., Chapagain P., Narasimhan G. A comprehensive survey of scoring functions for protein docking models. *BMC Bioinformatics*, 2025, vol. 26, no. 1, pp. 25. doi: 10.1186/s12859-024-05991-4
 21. de Angelo R.M., de Sousa D.S., da Silva A.P., Chiari L.P., da Silva A.B., Honorio K.M. Molecular docking: State-of-the-art scoring functions and search algorithms. *Computer-Aided Drug Discovery and Design*, 2025, vol. 3, pp. 163–198. doi: 10.1007/978-3-031-76718-0_7
 22. Eldridge M.D., Murray C.W., Auton T.R., Paolini G.V., Mee R.P. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *Journal of Computer-Aided Molecular Design*, 1997, vol. 11, no. 5, pp. 425–445. doi: 10.1023/a:1007996124545
 23. Murray C.W., Auton T.R., Eldridge M.D. Empirical scoring functions. II. The testing of an empirical scoring function for the prediction of ligand-receptor binding affinities and the use of Bayesian regression to improve the quality of the model. *Journal of Computer-Aided Molecular Design*, 1998, vol. 12, no. 5, pp. 503–519. doi: 10.1023/a:1008040323669
 24. Wang R., Lai L., Wang S. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *Journal of Computer-Aided Molecular Design*, 2002, vol. 16, no. 1, pp. 11–26. doi: 10.1023/a:1016357811882
 25. de Azevedo W.F., Dias R. Evaluation of ligand-binding affinity using polynomial empirical scoring functions. *Bioorganic and Medicinal*

- Chemistry. 2008. V. 16. N 20. P. 9378–9382. doi: 10.1016/j.bmc.2008.08.014
26. Dias R., Timmers L.F., Caceres R.A., de Azevedo W.F. Evaluation of molecular docking using polynomial empirical scoring functions // *Current Drug Targets*. 2008. V. 9. N 12. P. 1062–1070. doi: 10.2174/138945008786949450
27. Korb O., Stutzle T., Exner T.E. Empirical scoring functions for advanced protein-ligand docking with PLANTS // *Journal of Chemical Information and Modeling*. 2009. V. 49. N 1. P. 84–96. doi: 10.1021/ci800298z
28. Liu Z., Malone B., Yuan C. Empirical evaluation of scoring functions for Bayesian network model selection // *BMC Bioinformatics*. 2012. V. 13. Suppl 15. P. S14. doi: 10.1186/1471-2105-13-S15-S14
29. Zulhafiz N.A., Teoh T.C., Chin A.V., Chang S.W. Drug repurposing using artificial intelligence, molecular docking, and hybrid approaches: A comprehensive review in general diseases vs Alzheimer's disease // *Computer Methods and Programs in Biomedicine*. 2025. V. 261. P. 108604. doi: 10.1016/j.cmpb.2025.108604
30. Fujimoto K.J., Minami S., Yanai T. Machine-learning- and knowledge-based scoring functions incorporating ligand and protein fingerprints // *ACS Omega*. 2022. V. 7. N 22. P. 19030–19039. doi: 10.1021/acsomega.2c02822
31. Guedes I.A., Barreto A.M.S., Marinho D., Krempser E., Kuenemann M.A., Sperandio O., et al. New machine learning and physics-based scoring functions for drug discovery // *Scientific Reports*. 2021. V. 11. N 1. P. 3198. doi: 10.1038/s41598-021-82410-1
32. Bitencourt-Ferreira G., Rizzotto C., de Azevedo Junior W.F. Machine learning-based scoring functions, development and applications with SAnDReS // *Current Medicinal Chemistry*. 2021. V. 28. N 9. P. 1746–1756. doi: 10.2174/0929867327666200515101820
33. Ghasemi F., Mehridehnavi A., Pérez-Garrido A., Pérez-Sánchez H. Neural network and deep-learning algorithms used in QSAR studies: merits and drawbacks // *Drug Discovery Today*. 2018. V. 23. N 10. P. 1784–1790. doi: 10.1016/j.drudis.2018.06.016
34. Matsuzaka Y., Uesawa Y. A Deep learning-based quantitative structure-activity relationship system construct prediction model of agonist and antagonist with high performance // *International Journal of Molecular Sciences*. 2022. V. 23. N 4. P. 2141. doi: 10.3390/ijms23042141
35. Betzi S., Suhre K., Chetrit B., Guerlesquin F., Morelli X. GFScore: a general nonlinear consensus scoring function for high-throughput docking // *Journal of Chemical Information and Modeling*. 2006. V. 46. N 4. P. 1704–1712. doi: 10.1021/ci0600758
36. Bar-Haim S., Aharon A., Ben-Moshe T., Marantz Y., Senderowitz H. SeleX-CS: a new consensus scoring algorithm for hit discovery and lead optimization // *Journal of Chemical Information and Modeling*. 2009. V. 49. N 3. P. 623–633. doi: 10.1021/ci800335j
37. Plewczynski D., Lazniewski M., von Grothuss M., Rychlewski L., Ginalski K. VoteDock: consensus docking method for prediction of protein-ligand interactions // *Journal of Computational Chemistry*. 2011. V. 32. N 4. P. 568–581. doi: 10.1002/jcc.21642
38. Meli R., Morris G.M., Biggin P.C. Scoring functions for protein-ligand binding affinity prediction using structure-based deep learning: a review // *Frontiers in Bioinformatics*. 2022. V. 2. P. 885983. doi: 10.3389/fbinf.2022.885983
39. Tran-Nguyen V.-K., Bret G., Rognan D. True accuracy of fast scoring functions to predict high-throughput screening data from docking poses: the simpler the better // *Journal of Chemical Information and Modeling*. 2021. V. 61. N 6. P. 2788–2797. doi: 10.1021/acs.jcim.1c00292
40. Morris G.M., Huey R., Lindstrom W., Sanner M.F., Belew R.K., Goodsell D.S., et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility // *Journal of Computational Chemistry*. 2009. V. 30. N 16. P. 2785–2791. doi: 10.1002/jcc.21256
41. Bitencourt-Ferreira G., Pintro V.O., de Azevedo W.F. Docking with AutoDock4 // *Methods in Molecular Biology*. 2019. V. 2053. P. 125–148. doi: 10.1007/978-1-4939-9752-7_9
42. Santos-Martins D., Solis-Vasquez L., Tillack A.F., Sanner M.F., Koch A., Forli S. Accelerating AutoDock4 with GPUs and gradient-based local search // *Journal of Chemical Theory and Computation*. 2021. V. 17. N 2. P. 1060–1073. doi: 10.1021/acs.jctc.0c01006
43. Obiol-Pardo C., Rubio-Martinez J. Comparative evaluation of MMPBSA and XSCORE to compute binding free energy in XIAP-peptide complexes // *Journal of Chemical Information and Modeling*. 2007. V. 47. N 1. P. 134–142. doi: 10.1021/ci600412z
- Chemistry, 2008, vol. 16, no. 20, pp. 9378–9382. doi: 10.1016/j.bmc.2008.08.014
26. Dias R., Timmers L.F., Caceres R.A., de Azevedo W.F. Evaluation of molecular docking using polynomial empirical scoring functions. *Current Drug Targets*, 2008, vol. 9, no. 12, pp. 1062–1070. doi: 10.2174/138945008786949450
27. Korb O., Stutzle T., Exner T.E. Empirical scoring functions for advanced protein-ligand docking with PLANTS. *Journal of Chemical Information and Modeling*, 2009, vol. 49, no. 1, pp. 84–96. doi: 10.1021/ci800298z
28. Liu Z., Malone B., Yuan C. Empirical evaluation of scoring functions for Bayesian network model selection. *BMC Bioinformatics*, 2012, vol. 13, suppl 15, pp. S14. doi: 10.1186/1471-2105-13-S15-S14
29. Zulhafiz N.A., Teoh T.C., Chin A.V., Chang S.W. Drug repurposing using artificial intelligence, molecular docking, and hybrid approaches: A comprehensive review in general diseases vs Alzheimer's disease. *Computer Methods and Programs in Biomedicine*, 2025, vol. 261, pp. 108604. doi: 10.1016/j.cmpb.2025.108604
30. Fujimoto K.J., Minami S., Yanai T. Machine-learning- and knowledge-based scoring functions incorporating ligand and protein fingerprints. *ACS Omega*, 2022, vol. 7, no. 22, pp. 19030–19039. doi: 10.1021/acsomega.2c02822
31. Guedes I.A., Barreto A.M.S., Marinho D., Krempser E., Kuenemann M.A., Sperandio O., et al. New machine learning and physics-based scoring functions for drug discovery. *Scientific Reports*, 2021, vol. 11, no. 1, pp. 3198. doi: 10.1038/s41598-021-82410-1
32. Bitencourt-Ferreira G., Rizzotto C., de Azevedo Junior W.F. Machine learning-based scoring functions, development and applications with SAnDReS. *Current Medicinal Chemistry*, 2021, vol. 28, no. 9, pp. 1746–1756. doi: 10.2174/0929867327666200515101820
33. Ghasemi F., Mehridehnavi A., Pérez-Garrido A., Pérez-Sánchez H. Neural network and deep-learning algorithms used in QSAR studies: merits and drawbacks. *Drug Discovery Today*, 2018, vol. 23, no. 10, pp. 1784–1790. doi: 10.1016/j.drudis.2018.06.016
34. Matsuzaka Y., Uesawa Y. A Deep learning-based quantitative structure-activity relationship system construct prediction model of agonist and antagonist with high performance. *International Journal of Molecular Sciences*, 2022, vol. 23, no. 4, pp. 2141. doi: 10.3390/ijms23042141
35. Betzi S., Suhre K., Chetrit B., Guerlesquin F., Morelli X. GFScore: a general nonlinear consensus scoring function for high-throughput docking. *Journal of Chemical Information and Modeling*, 2006, vol. 46, no. 4, pp. 1704–1712. doi: 10.1021/ci0600758
36. Bar-Haim S., Aharon A., Ben-Moshe T., Marantz Y., Senderowitz H. SeleX-CS: a new consensus scoring algorithm for hit discovery and lead optimization. *Journal of Chemical Information and Modeling*, 2009, vol. 49, no. 3, pp. 623–633. doi: 10.1021/ci800335j
37. Plewczynski D., Lazniewski M., von Grothuss M., Rychlewski L., Ginalski K. VoteDock: consensus docking method for prediction of protein-ligand interactions. *Journal of Computational Chemistry*, 2011, vol. 32, no. 4, pp. 568–581. doi: 10.1002/jcc.21642
38. Meli R., Morris G.M., Biggin P.C. Scoring functions for protein-ligand binding affinity prediction using structure-based deep learning: a review. *Frontiers in Bioinformatics*, 2022, vol. 2, pp. 885983. doi: 10.3389/fbinf.2022.885983
39. Tran-Nguyen V.-K., Bret G., Rognan D. True accuracy of fast scoring functions to predict high-throughput screening data from docking poses: the simpler the better. *Journal of Chemical Information and Modeling*, 2021, vol. 61, no. 6, pp. 2788–2797. doi: 10.1021/acs.jcim.1c00292
40. Morris G.M., Huey R., Lindstrom W., Sanner M.F., Belew R.K., Goodsell D.S., et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry*, 2009, vol. 30, no. 16, pp. 2785–2791. doi: 10.1002/jcc.21256
41. Bitencourt-Ferreira G., Pintro V.O., de Azevedo W.F. Docking with AutoDock4. *Methods in Molecular Biology*, 2019, vol. 2053, pp. 125–148. doi: 10.1007/978-1-4939-9752-7_9
42. Santos-Martins D., Solis-Vasquez L., Tillack A.F., Sanner M.F., Koch A., Forli S. Accelerating AutoDock4 with GPUs and gradient-based local search. *Journal of Chemical Theory and Computation*, 2021, vol. 17, no. 2, pp. 1060–1073. doi: 10.1021/acs.jctc.0c01006
43. Obiol-Pardo C., Rubio-Martinez J. Comparative evaluation of MMPBSA and XSCORE to compute binding free energy in XIAP-peptide complexes. *Journal of Chemical Information and Modeling*, 2007, vol. 47, no. 1, pp. 134–142. doi: 10.1021/ci600412z

44. Pencheva T., Soumana O.S., Pajeva I., Miteva M.A. Post-docking virtual screening of diverse binding pockets: comparative study using DOCK, AMMOS, X-Score and FRED scoring functions // *European Journal of Medicinal Chemistry*. 2010. V. 45. N 6. P. 2622–2628. doi: 10.1016/j.ejmech.2009.12.025
45. Neudert G., Klebe G. DSX: a knowledge-based scoring function for the assessment of protein-ligand complexes // *Journal of Chemical Information and Modeling*. 2011. V. 51. N 10. P. 2731–2745. doi: 10.1021/ci200274q
46. Wang W.J., Huang Q., Zou J., Li L.L., Yang S.Y. TS-Chemscore, a target-specific scoring function, significantly improves the performance of scoring in virtual screening // *Chemical Biology and Drug Design*. 2015. V. 86. N 1. P. 1–8. doi: 10.1111/cbdd.12470
47. Eberhardt J., Santos-Martins D., Tillack A.F., Forli S. AutoDock Vina 1.2.0: New docking methods, expanded force field, and python bindings // *Journal of Chemical Information and Modeling*. 2021. V. 61. N 8. P. 3891–3898. doi: 10.1021/acs.jcim.1c00203
48. Lu H., Wei Z., Wang C., Guo J., Zhou Y., Wang Z., et al. Redesigning Vina@QNLM for ultra-large-scale molecular docking and screening on a sunway supercomputer // *Frontiers in Chemistry*. 2021. V. 9. P. 750325. doi: 10.3389/fchem.2021.750325
49. Pencheva T., Soumana O.S., Pajeva I., Miteva M.A. Post-docking virtual screening of diverse binding pockets: comparative study using DOCK, AMMOS, X-Score and FRED scoring functions. *European Journal of Medicinal Chemistry*, 2010, vol. 45, no. 6, pp. 2622–2628. doi: 10.1016/j.ejmech.2009.12.025
50. Neudert G., Klebe G. DSX: a knowledge-based scoring function for the assessment of protein-ligand complexes. *Journal of Chemical Information and Modeling*, 2011, vol. 51, no. 10, pp. 2731–2745. doi: 10.1021/ci200274q
51. Wang W.J., Huang Q., Zou J., Li L.L., Yang S.Y. TS-Chemscore, a target-specific scoring function, significantly improves the performance of scoring in virtual screening. *Chemical Biology and Drug Design*, 2015, vol. 86, no. 1, pp. 1–8. doi: 10.1111/cbdd.12470
52. Eberhardt J., Santos-Martins D., Tillack A.F., Forli S. AutoDock Vina 1.2.0: New docking methods, expanded force field, and python bindings. *Journal of Chemical Information and Modeling*, 2021, vol. 61, no. 8, pp. 3891–3898. doi: 10.1021/acs.jcim.1c00203
53. Lu H., Wei Z., Wang C., Guo J., Zhou Y., Wang Z., et al. Redesigning Vina@QNLM for ultra-large-scale molecular docking and screening on a sunway supercomputer. *Frontiers in Chemistry*, 2021, vol. 9, pp. 750325. doi: 10.3389/fchem.2021.750325

Авторы

Спринджук Матвей Владимирович — кандидат технических наук, старший научный сотрудник, Объединенный институт проблем информатики Национальной академии наук Беларуси, Минск, 220072, Беларусь, <https://orcid.org/0000-0001-9500-2954>, stepanenkomatvei@yandex.ru

Кукарцев Владислав Викторович — кандидат технических наук, доцент, Российский государственный аграрный университет — МСХА имени К.А. Тимирязева, Москва, 127434, Российская Федерация; ведущий научный сотрудник, Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет), Москва, 105005, Российская Федерация, <https://orcid.org/0000-0001-6382-1736>, vlad_saa_2000@mail.ru

Ноздрачева Анна Валерьевна — кандидат медицинских наук, руководитель лабораторией, Национальный исследовательский центр эпидемиологии и микробиологии имени Н.Ф. Гамалеи, Москва, 123098, Российская Федерация, <https://orcid.org/0000-0002-8521-1741>, nozdracheva0506@gmail.com

Кончиц Андрей Петрович — кандидат биологических наук, ведущий научный сотрудник, Институт леса Национальной академии наук Беларуси, Гомель, 246050, Беларусь, <https://orcid.org/0000-0002-8823-0630>, konchits@yandex.ru

Владыко Александр Станиславович — доктор медицинских наук, внештатный научный консультант, Научно-исследовательский институт гигиены, токсикологии, эпидемиологии, вирусологии и микробиологии, государственного учреждения «Республиканский центр гигиены, эпидемиологии и общественного здоровья», Минск, 220114, Беларусь, <https://orcid.org/0009-0001-2188-1893>, vladalex1703@gmail.com

Берник Василий Иванович — доктор физико-математических наук, профессор, главный научный сотрудник, Институт математики Национальной академии наук Беларуси, Минск, 220072, Беларусь, <https://orcid.org/0000-0001-8096-1079>, bernik@im.bas-net.by

Сукач Елена Ивановна — кандидат технических наук, доцент, доцент, Гомельский государственный университет, Гомель, 246050, Беларусь, <https://orcid.org/0000-0002-3693-1494>, konchits@yandex.ru

Authors

Matvey V. Sprindzuk — PhD, Senior Researcher, State Scientific Institution “United Institute of Informatics Problems” of the National Academy of Sciences of Belarus, Minsk 220072, Belarus, <https://orcid.org/0000-0001-9500-2954>, stepanenkomatvei@yandex.ru

Vladislav V. Kukartsev — PhD, Associate Professor, Russian State Agrarian University — Moscow Timiryazev Agricultural Academy, Moscow, 127434, Russian Federation; Leading Researcher, Bauman Moscow State Technical University, Moscow, 127434, Russian Federation, <https://orcid.org/0000-0001-6382-1736>, vlad_saa_2000@mail.ru

Anna V. Nozdracheva — PhD (Medicine), Head of Laboratory, The National Research Center for Epidemiology and Microbiology named after Honorary Academician N.F. Gamaleya, Moscow, 123098, Russian Federation, <https://orcid.org/0000-0002-8521-1741>, nozdracheva0506@gmail.com

Andrey P. Konchits — PhD (Biology), Leading Researcher, Forest Institute of the National Academy of Sciences of the Republic of Belarus, Gomel, 246050, Belarus, <https://orcid.org/0000-0002-8823-0630>, konchits@yandex.ru

Alexander S. Vladiko — D.Sc. (Medicine), Freelance Scientific Consultant, Scientific Research Institute of Hygiene, Toxicology, Epidemiology, Virology and Microbiology, State Institution “Republican Center of Hygiene, Epidemiology and Public Health”, Minsk, 220114, Belarus, <https://orcid.org/0009-0001-2188-1893>, vladalex1703@gmail.com

Vassily I. Bernik — D.Sc. (Physics & Mathematics), Professor, Chief Researcher, Institute of Mathematics of the National Academy of Sciences of Belarus, Minsk, 220072, Belarus, <https://orcid.org/0000-0001-8096-1079>, bernik@im.bas-net.by

Elena I. Sukach — PhD, Associate Professor, Associate Professor, Gomel State University of Belarus, Gomel, 246050, Belarus, <https://orcid.org/0000-0002-3693-1494>, konchits@yandex.ru

Статья поступила в редакцию 02.04.2026
Одобрена после рецензирования 06.06.2026
Принята к печати 26.07.2026

Received 02.04.2026
Approved after reviewing 06.06.2026
Accepted 26.07.2026



Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»