

УДК 615.065:544.165

<https://doi.org/10.30895/2312-7821-2023-11-4-372-389>

Обзорная статья / Review



Оценка безопасности фармакологически активных веществ *in silico* с применением методов машинного обучения: обзор

В.В. Поройков¹✉, А.В. Дмитриев¹, Д.С. Дружиловский¹, С.М. Иванов^{1,2}, А.А. Лагунин^{1,2},
П.В. Погодин¹, А.В. Рудик¹, П.И. Савосина¹, О.А. Тарасова¹, Д.А. Филимонов¹

¹ Федеральное государственное бюджетное научное учреждение
«Научно-исследовательский институт биомедицинской химии имени В.Н. Ореховича»,
Погодинская ул., д. 10, стр. 8, Москва, 119121, Российская Федерация

² Федеральное государственное автономное образовательное учреждение высшего образования
«Российский национальный исследовательский медицинский университет им. Н.И. Пирогова»
Министерства здравоохранения Российской Федерации,
ул. Островитянова, д. 1, Москва, 117997, Российская Федерация

✉ Контактное лицо: Поройков Владимир Васильевич vladimir.poroikov@ibmc.msk.ru

РЕЗЮМЕ

Актуальность. В исследованиях и разработках новых лекарственных препаратов повсеместно используются методы машинного обучения (ML). Их применение особенно актуально для оценки безопасности фармакологически активных веществ на ранних стадиях исследований, что существенно снижает риски получения отрицательных результатов в дальнейшем.

Цель. Обзор основных информационных и прогностических ресурсов, которые могут быть использованы для оценки безопасности фармакологически активных веществ с применением методов *in silico*.

Обсуждение. Использование новых методов ML позволяет на основе анализа зависимостей «структура–активность» оценивать наиболее вероятные молекулярные мишени, с которыми может взаимодействовать конкретное соединение, потенциальные фармакотерапевтические и побочные эффекты, острую и специфическую токсичность, метаболизм и другие фармакодинамические, фармакокинетические и токсикологические характеристики изучаемых веществ. Получение этой информации дает возможность определить приоритетные направления экспериментального тестирования биологической активности на ранних стадиях исследований и отобрать соединения с низкой вероятностью проявления побочных и токсических эффектов. В настоящем обзоре рассмотрены свободно доступные в сети Интернет информационные и прогностические ресурсы, которые позволяют оценивать безопасность соединения на основе его структурной формулы с применением подходов ML. Особое внимание уделено отечественным компьютерным разработкам, представленным на платформе Way2Drug (<https://www.way2drug.com/dr/>).

Выводы. Современные методы компьютерной оценки свойств фармакологически активных веществ на основе анализа зависимостей «структура–активность» методами ML обеспечивают получение информации о различных характеристиках безопасности этих соединений и позволяют отбирать наиболее перспективные кандидаты для углубленных доклинических и клинических исследований.

Ключевые слова: фармакологически активные вещества; безопасность; исследования *in silico*; анализ зависимость «структура–активность»; компьютерное конструирование лекарственных средств; машинное обучение; платформа Way2Drug

Для цитирования: Поройков В.В., Дмитриев А.В., Дружиловский Д.С., Иванов С.М., Лагунин А.А., Погодин П.В., Рудик А.В., Савосина П.И., Тарасова О.А., Филимонов Д.А. Оценка безопасности фармакологически активных веществ *in silico* с применением методов машинного обучения: обзор. *Безопасность и риск фармакотерапии*. 2023;11(4):372–389. <https://doi.org/10.30895/2312-7821-2023-11-4-372-389>

© В.В. Поройков, А.В. Дмитриев, Д.С. Дружиловский, С.М. Иванов, А.А. Лагунин, П.В. Погодин, А.В. Рудик, П.И. Савосина, О.А. Тарасова, Д.А. Филимонов, 2023

Финансирование. Работа выполнена в рамках гранта Российского научного фонда, проект № 19-15-00396, <https://rscf.ru/project/19-15-00396/>

Конфликт интересов. Все авторы являются разработчиками ресурсов на платформе Way2Drug.

In Silico Estimation of the Safety of Pharmacologically Active Substances Using Machine Learning Methods: A Review

Vladimir V. Poroikov^{1,✉}, Alexander V. Dmitriev¹, Dmitry S. Druzhilovskiy¹, Sergey M. Ivanov^{1,2}, Alexey A. Lagunin^{1,2}, Pavel V. Pogodin¹, Anastasiya V. Rudik¹, Polina I. Savosina¹, Olga A. Tarasova¹, Dmitry A. Filimonov¹

¹ V.N. Orekhovich Research Institute of Biomedical Chemistry,
10/8 Pogodinskaya St., Moscow 119121, Russian Federation

² N.I. Pirogov Russian National Research Medical University,
1 Ostrovityanov St., Moscow 117997, Russian Federation

✉ Corresponding author: **Vladimir V. Poroikov** vladimir.poroikov@ibmc.msk.ru

ABSTRACT

Scientific relevance. Currently, machine learning (ML) methods are widely used in the research and development of new pharmaceuticals. ML methods are particularly important for assessing the safety of pharmacologically active substances early in the research process because such safety assessments significantly reduce the risk of obtaining negative results in the future.

Aim. This study aimed to review the main information and prediction resources that can be used for the assessment of the safety of pharmacologically active substances *in silico*.

Discussion. Novel ML methods can identify the most likely molecular targets for a specific compound to interact with, based on structure–activity relationship analysis. In addition, ML methods can be used to search for potential therapeutic and adverse effects, as well as to study acute and specific toxicity, metabolism, and other pharmacodynamic, pharmacokinetic, and toxicological characteristics of investigational substances. Obtained at early stages of research, this information helps to prioritise areas for experimental testing of biological activity, as well as to identify compounds with a low probability of producing adverse and toxic effects. This review describes free online ML-based information and prediction resources for assessing the safety of pharmacologically active substances using their structural formulas. Special attention is paid to the Russian computational products presented on the Way2Drug platform (<https://www.way2drug.com/dr/>).

Conclusions. Contemporary approaches to the assessment of pharmacologically active substances *in silico* based on structure–activity relationship analysis using ML methods provide information about various safety characteristics and allow developers to select the most promising candidates for further in-depth preclinical and clinical studies.

Keywords: pharmacologically active substances; safety; *in silico* studies; structure–activity relationship; SAR; computer-aided drug design; machine learning; Way2Drug

For citation: Poroikov V.V., Dmitriev A.V., Druzhilovskiy D.S., Ivanov S.M., Lagunin A.A., Pogodin P.V., Rudik A.V., Savosina P.I., Tarasova O.A., Filimonov D.A. *In silico* estimation of the safety of pharmacologically active substances using machine learning methods: a review. *Safety and Risk of Pharmacotherapy*. 2023;11(4):372–389. <https://doi.org/10.30895/2312-7821-2023-11-4-372-389>

Funding. This study was carried out within the framework of Russian Science Foundation project No. 19-15-00396, <https://rscf.ru/project/19-15-00396/>

Disclosure. The authors are developers of the resources presented on the Way2Drug platform.

Введение

Целью доклинических исследований лекарственных средств для медицинского применения является оценка их безопасности, качества и эффективности¹. Традиционно изучение безопасности фармакологически активных веществ² осуществляется на экспериментальных животных в соответствии с правилами надлежащей лабораторной практики (Good Laboratory Practice, GLP)³. На основании экспертизы пакета документов, содержащего информацию о результатах изучения фармакодинамических, фармакокинетических и токсикологических характеристик фармакологически активных веществ, уполномоченным федеральным органом исполнительной власти выносится решение о выдаче разрешения на проведение клинических исследований⁴.

Методы компьютерного конструирования лекарственных средств (Computer-Aided Drug Design, CADD), широко применяемые в последние 30 лет для поиска перспективных лекарственно-подобных химических соединений (drug-like compounds) и оптимизации их свойств, стали использоваться на всех стадиях исследований и разработок новых лекарственных препаратов [1]. Предпосылкой для этого явилось накопление свободно доступных в сети Интернет сведений о различных видах биологической активности веществ разных химических классов [2]. В совокупности веб-ресурсы содержат огромный массив гетерогенных химических и биомедицинских данных, которые по своему объему, разнообразию и скорости накопления соответствуют понятию «большие данные» (Big Data) [3]. Доступные данные после предварительной обработки в соответствии с современными требованиями [4] используются в качестве обучающих выборок для построения количественных и классификационных моделей зависимостей «структура–активность» (structure–activity relationship, SAR) с применением различных методов машинного обучения (machine learning, ML)⁵. С помощью этих моделей осуществляется оценка наиболее вероятных

молекулярных мишеней, с которыми может взаимодействовать анализируемое соединение, потенциальных фармакотерапевтических и побочных эффектов, острой и специфической токсичности, биодоступности, метаболизма в организме человека и др. [5].

В последние годы методы ML, используемые для анализа зависимостей «структура–активность», нередко называют методами искусственного интеллекта (artificial intelligence, AI) [6]. Однако, как видно из опубликованных аналитических обзоров [7–9], методы искусственного интеллекта успешно применяются в решении задач распознавания образов (например, рентгеновских снимков), а не прогноза свойств новых химических соединений. В задачах анализа SAR ничего выходящего за рамки ML не применяют (даже если авторы называют эти методы AI), поэтому в нашем обзоре мы используем термин «машинное обучение».

Применение методов *in silico* для оценки безопасности потенциальных лекарственных препаратов недавно одобрено официально: в США принят закон, согласно которому тестирование лекарственных средств на животных перед тем, как начать испытания на людях, больше не является обязательным для получения одобрения Управления по контролю за качеством продуктов питания и лекарственных средств (Food and Drug Administration, FDA) на проведение клинических испытаний [10]. Согласно этому закону «... термин «доклиническое испытание» означает испытание, проводимое *in vitro*, *in silico* или *in chemico*, или испытание *in vivo* без участия человека, которое проводится до или во время клинического исследования безопасности и эффективности лекарственного средства и может включать тесты на животных или методы испытаний, основанные на биологии человека или модельных системах, таких как клеточные тест-системы, микрофизиологические системы, биопринтинговые или компьютерные модели».

Цель работы – обзор основных информационных и прогностических ресурсов, которые

¹ Федеральный закон Российской Федерации от 12.04.2010 № 61-ФЗ «Об обращении лекарственных средств».

² Здесь и далее термином «фармакологически активное вещество» обозначены органические соединения, изучаемые с целью создания лекарственных препаратов независимо от стадии исследований (тестирование биологической активности в экспериментах *in vitro* или *in vivo*, доклинические исследования, клинические испытания), а также разрешенные к медицинскому применению фармацевтические субстанции, побочное действие препаратов на основе которых изучается в процессе широкомасштабного применения в клинике, либо исследуемые с целью расширения первоначальных показаний к применению.

³ Приказ Министерства здравоохранения Российской Федерации от 01.04.2016 № 199н «Об утверждении Правил надлежащей лабораторной практики».

⁴ Федеральный закон Российской Федерации от 12.04.2010 № 61-ФЗ «Об обращении лекарственных средств».

⁵ Здесь и далее термин «активность» означает фармакотерапевтические эффекты, молекулярные механизмы действия, специфическую токсичность, цитотоксичность по отношению к разным клеточным линиям, побочное действие, влияние на метаболизм и другие фармакодинамические и фармакокинетические характеристики лекарственно-подобных соединений.

могут быть использованы для оценки безопасности фармакологически активных веществ с применением методов *in silico*.

Информационные ресурсы

Информация о наиболее важных свободно доступных в сети Интернет базах данных (БД), содержащих сведения о структуре, биологической активности, побочных эффектах, токсичности, метаболизме и других характеристиках фармакологически активных веществ⁶, приведена в *таблице 1*.

PubChem⁷ — это открытая химическая БД Национальных институтов здоровья США (National Institutes of Health, NIH), в которую пользователи могут загружать собственные данные с целью последующего некоммерческого использования. Актуальная версия БД содержит информацию о структуре 115,7 млн уникальных соединений, 1,6 млн тест-систем, 292 млн записей данных о биологической активности, 113,7 тыс. мишеней и др.

ChEMBL⁸ — аннотируемая вручную БД по биологически активным веществам, имеющим лекарственно-подобные характеристики. В ней содержится информация о химической структуре и биологической активности химических соединений, а также геномные данные, что должно способствовать трансляции результатов исследования геномов в создание новых лекарственных средств. Актуальная версия БД содержит информацию о 2,4 млн уникальных соединений, 14 тыс. лекарственных средств, 6,7 тыс. механизмах действия, 1,6 млн тест-систем, 20,3 млн записей данных о биологической активности и др.

DrugBank⁹ содержит информацию о лекарственных средствах и фармакологических мишенях. БД (версия 5.1.10) содержит 15 858 записей о лекарственных средствах, включая информацию о разрешенных к медицинскому применению 2750 малых молекулах, 1597 биопрепаратах (белки, пептиды, вакцины и аллергены), 134 пищевых добавках и сведения о 6721 фармакологически активном веществе, изучаемых с целью создания лекарственных

средств. Кроме того, в DrugBank содержится информация о 5296 различных белках (мишени, ферменты, транспортеры, переносчики), связанная с записями о лекарственных средствах.

DrugCentral¹⁰ содержит информацию о 4773 лекарственных препаратах для терапии заболеваний человека (2331 зарегистрированных FDA и 456 — Европейским агентством по лекарственным средствам (European Medicines Agency, EMA)), а также о 396 ветеринарных препаратах, зарегистрированных FDA; всего о 142 303 торговых наименованиях. В БД представлены сведения о клинических показаниях, фармакологических эффектах, молекулярных мишенях, механизмах действия и др.

WWAD (World-Wide Approved Drugs)¹¹ содержит информацию об активных фармацевтических субстанциях, зарегистрированных в 50 странах мира, включая названия, структурные формулы, молекулярные мишени, фармакотерапевтические группы, показания к применению, страны, в которых препарат впервые введен в медицинскую практику, ссылки на соответствующие регуляторные агентства, идентификаторы пространственной структуры соединения в комплексе с белком в банке данных трехмерных структур белков (Protein Data Bank, PDB)¹².

Pharos¹³ содержит информацию из базы знаний, создаваемой в рамках объединенной программы NIH Illuminating the Druggable Genome. Целью данной программы является разработка всеобъемлющей интегрированной базы знаний по фармакологическим мишеням, включая слабоаннотированные участки генома. Особое внимание уделено трем наиболее часто встречающимся семействам белков: рецепторы, сопряженные с G-белками, ионные каналы и киназы.

Open Targets¹⁴ — платформа для доступа к агрегированным общедоступным знаниям о связи молекулярных мишеней и заболеваний с целью приоритизации исследований и разработок. Предоставлены агрегированные данные по генетике, соматическим мутациям, анализу геной экспрессии, лекарственным средствам, фармакологическим моделям и ссылки на литературу. Связь мишени с заболеванием

⁶ Информация о содержании баз данных приведена по состоянию на август 2023 г.

⁷ <https://pubchem.ncbi.nlm.nih.gov>

⁸ <https://www.ebi.ac.uk/chembl>

⁹ <https://go.drugbank.com>

¹⁰ <https://drugcentral.org>

¹¹ https://www.way2drug.com/dr/ww_drug_approved.php

¹² <https://www.rcsb.org/>

¹³ <https://pharos.nih.gov>

¹⁴ <https://www.opentargets.org>

Таблица 1. Информационные ресурсы, которые могут использоваться при подготовке обучающих выборок для анализа зависимостей «структура–активность» химических соединений

Table 1. Information resources that may be used to prepare training sets for structure–activity relationship analysis of chemical compounds

Название <i>Name</i>	Краткая характеристика <i>Brief description</i>	Интернет-адрес <i>Internet address</i>
PubChem	Крупнейший свободно доступный ресурс с информацией о структуре и свойствах известных химических соединений <i>The largest free resource on the structure and properties of known chemical compounds</i>	https://pubchem.ncbi.nlm.nih.gov
ChEMBL	Информация о биологически активных соединениях <i>Information about biologically active compounds</i>	https://www.ebi.ac.uk/chembl
DrugBank	Информация о лекарственных средствах и фармакологических мишенях <i>Information about medicines and pharmacological targets</i>	https://go.drugbank.com
DrugCentral	Информация об активных фармацевтических субстанциях, зарегистрированных в США, Европейском Союзе, Японии <i>Information about active pharmaceutical substances registered in USA, EU, and Japan</i>	https://drugcentral.org
WWAD	Информация об активных фармацевтических субстанциях, зарегистрированных в 50 странах мира <i>Information about active pharmaceutical substances approved in 50 countries</i>	https://www.way2drug.com/drug_approved.php
Pharos	Информация о мишенях и лекарственных средствах, отобранная в рамках Программы Национальных институтов здравоохранения США (National Institutes of Health, NIH) "Illuminating the Druggable Genome" <i>Information about targets and medicines collected in the framework of the Illuminating the Druggable Genome programme of the National Institutes of Health (NIH)</i>	https://pharos.nih.gov
Open Targets	Информация о связи молекулярных мишеней и заболеваний для приоритизации исследований <i>Information about molecular target–disease relationships for study prioritisation</i>	https://www.opentargets.org
SIDER	Информация о нежелательных реакциях при применении лекарственных препаратов <i>Information about adverse drug reactions associated with medicinal products</i>	http://sideeffects.embl.de
CTD	Информация о взаимодействиях химических веществ с генами/белками и взаимосвязях между веществами, генами/белками, заболеваниями и фенотипами <i>Information about interactions of chemical substances with genes/proteins and relationships between chemicals, genes/proteins, diseases, and phenotypes</i>	https://ctdbase.org
HMDB	Информация о метаболоме человека <i>Information about the human metabolome</i>	https://hmdb.ca
FoodB	Информация о компонентах пищевых продуктов <i>Information about food ingredients</i>	https://foodb.ca
RISCTOX	Информация о широко используемых опасных химических веществах <i>Information about commonly used hazardous chemicals</i>	https://risctox.istas.net/en
TOXRIC	Детальная информация о токсикологических характеристиках свыше 100 тысяч химических соединений <i>Detailed information about toxicological characteristics of over 100,000 chemical compounds</i>	https://toxric.bioinformai.tech
Tox21	Информация о структуре и токсичности около 10 тысяч химических соединений <i>Information on the structure and toxicity of about 10,000 chemical compounds</i>	https://tripod.nih.gov/tox21/pubdata
OnSIDES	Информация о побочном действии лекарственных препаратов, извлеченная из инструкций по медицинскому применению <i>Information on side effects of medicinal products obtained from product labels</i>	https://nsides.io/

Таблица составлена авторами / The table is prepared by the authors

оценивается на основе расчетного интегрального показателя.

SIDER¹⁵ содержит информацию о 1430 разрешенных к медицинскому применению лекарственных средствах и 5868 зарегистрированных нежелательных реакциях при их использовании в клинике (всего 139756 пар записей «лекарственное средство – нежелательная реакция»). Эта информация извлечена из общедоступных документов и инструкций по медицинскому применению лекарственных препаратов. Представленная в SIDER информация включает классификацию лекарственных средств и нежелательных реакций, частоту развития нежелательных реакций, а также ссылки на дополнительную информацию, например о взаимосвязях между действующими фармацевтическими субстанциями и фармакологическими мишенями.

CTD (Comparative Toxicogenomics Database)¹⁶ – информационный ресурс, предоставляющий собранную вручную информацию о взаимодействиях химических веществ с генами и белками, взаимосвязях между химическими веществами и болезнями и между генами и болезнями. Эти данные объединяются с данными о функциях генов и белков и данными о регуляторных сигнальных путях, что позволяет формировать гипотезы о механизмах, лежащих в основе заболеваний, связанных с неблагоприятным воздействием окружающей среды на организм человека.

HMDB¹⁷ содержит подробную информацию о низкомолекулярных метаболитах, обнаруженных в организме человека, и может использоваться для исследований в метаболомике, медицинской химии, поиске биомаркеров, а также для образовательных целей. БД содержит информацию о трех типах данных и взаимосвязях между ними: 1) химические данные, 2) клинические данные и 3) данные молекулярной биологии/биохимии. Всего имеется 220945 записей о метаболитах (как водорастворимых, так и жирорастворимых), а также данные о 8610 белковых последовательностях (ферментах и переносчиках), связанных с записями о метаболитах. Примерно 2/3 информации относится к химическим/клиническим данным, а 1/3 посвящена

ферментативным или биохимическим данным. Имеются гиперссылки на другие базы данных (KEGG, PubChem, MetaCyc, ChEBI, PDB, UniProt и GenBank).

FoodB¹⁸ – наиболее полный ресурс по компонентам пищевых продуктов, включая макро- и микроэлементы, а также соединения, которые придают пище цвет, текстуру, вкус и аромат. Для каждого химического соединения представлены данные о названии, структуре, химическом классе, физико-химических свойствах, его источнике (источниках) в пищевых продуктах, цвете, аромате, вкусе, физиологическом действии, предполагаемом влиянии на здоровье (по опубликованным результатам исследований) и сведения о концентрации в различных пищевых продуктах.

RISCTOX¹⁹ содержит в систематизированной форме краткую информацию об используемых в быту и промышленности опасных химических веществах (канцерогенах, мутагенах, эндокринных деструкторах, нейротоксикантах, сенсibilизаторах и др.), применение которых сопряжено с рисками для здоровья и окружающей среды.

TOXRIC²⁰ содержит токсикологическую информацию о 113720 химических соединениях, охарактеризованных по 13 категориям токсичности, включая 275 параметров (endpoints) и 38 типов дескрипторов (структурные, транскриптомные, метаболические данные и др.). Различные наборы данных могут быть загружены в формате *.csv и использованы для построения моделей SAR с применением ML.

Tox21²¹ содержит около 10 тысяч записей, включая структурные формулы и данные о 38 различных видах токсичности химических соединений. Информация из этой БД широко используется для валидации различных методов ML.

OnSIDES²² содержит информацию о нежелательных реакциях почти 2000 однокомпонентных и комбинированных препаратов, извлеченную из разделов «Boxed Warnings» и «Adverse Reactions» инструкций по медицинскому применению.

Необходимо подчеркнуть, что в настоящее время в свободном доступе нет единого

¹⁵ <http://sideeffects.embl.de>

¹⁶ <https://ctdbase.org>

¹⁷ <https://hmdb.ca>

¹⁸ <https://foodb.ca>

¹⁹ <https://risctox.istas.net/en>

²⁰ <https://toxric.bioinformai.tech>

²¹ <https://tripod.nih.gov/tox21/pubdata>

²² <https://nsides.io/>

информационного ресурса, который содержал бы исчерпывающую информацию, необходимую для оценки безопасности лекарственно-подобных химических соединений. И хотя при формировании перечисленных выше информационных ресурсов их авторами проведена определенная работа по стандартизации представления данных, для обеспечения хорошего качества моделей (количественных) зависимостей «структура–активность» ((quantitative) structure–activity relationships, (Q)SAR), наряду с необходимостью извлечения информации из различных источников и ее агрегации, требуется тщательная предварительная обработка данных в соответствии с современными рекомендациями [4].

Прогностические ресурсы

В таблице 2 приведена информация о некоторых свободно доступных в сети Интернет прогностических веб-ресурсах, позволяющих получать оценки различных характеристик биологической активности лекарственно-подобных

химических соединений на основе предварительно построенных моделей (Q)SAR.

SwissTargetPrediction²³ прогнозирует наиболее вероятные взаимодействия химических соединений с более чем 3000 белков человека, мыши и крысы на основе оценки структурного сходства с 370 000 биологически активных соединений.

SuperPred²⁴ прогнозирует наиболее вероятные взаимодействия химических соединений с 2353 молекулярными мишенями и их принадлежность к 4403 анатомо-терапевтическо-химическим классам (в соответствии с классификацией Всемирной организации здравоохранения) на основе оценки структурного сходства с известными фармацевтическими субстанциями.

ADMETlab²⁵ прогнозирует характеристики химических соединений, связанные с поступлением в организм, распределением, метаболизмом и выделением, а также токсическими или побочными эффектами (Absorption, Distribution, Metabolism, Excretion, and Toxicity, ADMET). Прогноз осуществляется на основе

Таблица 2. Прогностические ресурсы, которые могут использоваться для оценки активности химических соединений

Table 2. Prediction resources that may be used to estimate the activity of chemical compounds

Название <i>Name</i>	Краткая характеристика <i>Brief description</i>	Интернет-адрес <i>Internet address</i>
SwissTargetPrediction	Прогноз взаимодействия химических соединений с 3 тысячами молекулярных мишеней на основе структурного сходства <i>Prediction of the interaction of chemical compounds with 3000 molecular targets based on structural similarity</i>	http://www.swisstargetprediction.ch/
SuperPred	Прогноз принадлежности химических соединений к различным классам (в соответствии с анатомо-терапевтическо-химической классификацией лекарственных средств), и их взаимодействия с молекулярными мишенями <i>Prediction of chemical compounds' classes (according to the Anatomical Therapeutic Chemical classification system) and interactions with molecular targets</i>	https://prediction.charite.de/
ADMETlab	Прогноз ADMET* характеристик химических соединений <i>Prediction of ADMET* characteristics of chemical compounds</i>	https://admet.scbdd.com/
SwissADME	Прогноз ADME характеристик химических соединений <i>Prediction of ADME characteristics of chemical compounds</i>	http://www.swissadme.ch/
ProTox-II	Прогноз различных видов токсичности химических соединений <i>Prediction of various types of toxicity of chemical compounds</i>	https://tox-new.charite.de/prottox_II/
NCATS Predictor	Прогноз различных видов биологической активности и токсичности химических соединений <i>Prediction of various types of biological activity and toxicity of chemical compounds</i>	https://predictor.ncats.io/predictor/

Таблица составлена авторами / The table is prepared by the authors

* ADMET — абсорбция, распределение, метаболизм, выведение из организма и токсичность.

* ADMET: Adsorption, Distribution, Metabolism, Excretion, and Toxicity.

²³ <http://www.swisstargetprediction.ch/>

²⁴ <https://prediction.charite.de/>

²⁵ <https://admet.scbdd.com/>

структурной формулы с применением моделей (Q)SAR, построенных с использованием обучающей выборки, содержащей 288 967 записей; при этом прогнозируется 31 характеристика, связанная с ADMET.

SwissADME²⁶ прогнозирует лекарственное подобие (drug-likeness), различные физико-химические и ADME характеристики химических соединений.

ProTox-II²⁷ прогнозирует острую токсичность, гепатотоксичность, цитотоксичность, канцерогенность, мутагенность, иммунотоксичность, а также влияние на мишени и регуляторные сигнальные пути, ассоциированные с проявлением токсичности, в соответствии с информацией из БД Tox21.

NCATS Predictor²⁸ позволяет осуществлять прогноз 1180 видов биологической активности на основе структурных формул химических соединений с использованием моделей (Q)SAR, построенных путем анализа зависимостей «структура–активность» для обучающей выборки из 80 000 соединений.

Аналогично рассмотренной выше ситуации с информационными источниками в свободном доступе нет единого веб-ресурса, который обеспечивал бы прогноз всего комплекса характеристик, необходимых для оценки безопасности фармакологически активных веществ. Кроме того, сравнение различных прогностических веб-ресурсов между собой показывает, что их точность и предсказательная способность существенно различаются [11–14]. Поэтому, прежде чем приступать к практическому использованию конкретных веб-ресурсов, следует протестировать их пригодность для получения расчетных оценок в конкретной предметной области (например, выполнив прогноз характеристик веществ из изучаемого химического класса, для которых соответствующие параметры надежно определены в эксперименте).

Прогностические ресурсы на платформе Way2Drug

PASS (Prediction of Activity Spectra for Substances) – одна из наиболее известных в мире компьютерных программ, обеспечивающая прогнозирование нескольких тысяч видов активности лекарственно-подобных химических соединений на основе их структурных формул [15–18].

Разработка этой программы осуществляется нашим коллективом в течение более чем 30 лет (рис. 1). Установление зависимостей «структура–активность» в современной версии PASS 2022 осуществляется с применением оригинального классификатора, основанного на наивном байесовском подходе, разработанных нами дескрипторов множественных атомных окрестностей (multilevel neighbourhoods of atoms, MNA) и анализе обучающей выборки, содержащей информацию о структуре и биологической активности свыше 1,6 млн лекарственно-подобных химических соединений. Для более 8,5 тыс. прогнозируемых видов биологической активности (механизмы действия, фармакотерапевтические эффекты, побочные и токсические эффекты, взаимодействие с нежелательными мишенями, взаимодействие с ферментами лекарственного метаболизма, взаимодействие с белками-транспортерами, влияние на изменение экспрессии отдельных генов) среднее значение инвариантной точности прогноза при скользящем контроле с исключением по одному выше 0,93.

Нами реализован свободно доступный в сети Интернет веб-ресурс PASS Online²⁹, который обеспечивает прогнозирование профилей биологической активности лекарственно-подобных химических соединений на основе их структурных формул [16, 17]. PASS Online в настоящее время используют 50 000 исследователей из 106 стран мира с целью отбора наиболее перспективных соединений для синтеза и определения приоритетных направлений тестирования их биологической активности. Точность и предсказательная способность веб-ресурса PASS Online превосходят таковые у доступных в интернете аналогов [12]. В соответствии со значениями чувствительности для прогноза исходных показаний лекарственных препаратов прогностические веб-ресурсы ранжируются в следующем порядке убывания: PASS Online (1,00), ChemProt (0,82), TargetHunter (0,82), Similarity Ensemble Approach (0,80), SuperPred (0,76), SwissTargetPrediction (0,72), TarPred (0,64). Для перепрофилированных показаний порядок следования и значения чувствительности несколько отличаются: PASS Online (0,98), ChemProt (0,84), TargetHunter (0,82), Similarity Ensemble Approach (0,78), SwissTargetPrediction (0,72), SuperPred (0,64), TarPred (0,64).

²⁶ <http://www.swissadme.ch/>

²⁷ https://tox-new.charite.de/protox_II/

²⁸ <https://predictor.ncats.io/predictor/>

²⁹ <https://www.way2drug.com/all/>

Вводя в качестве входной информации структурную формулу лекарственно-подобного химического соединения, представленную в формате MOL или SDF, исследователь получает список прогнозируемых видов биологической активности с двумя оценками вероятностей: P_a — вероятность наличия и P_i — вероятность отсутствия активности. Использование прогноза PASS Online позволяет сократить необходимые объемы экспериментальных исследований в десятки раз. При этом оптимальная стратегия — это последовательные испытания прогнозируемых активностей в порядке убывания значений P_a (либо $P_a - P_i$). Детальное описание реализованного в PASS метода и интерпретации результатов прогноза представлено в работе³⁰ [17].

GUSAR (General Unrestricted Structure-Activity Relationships) — разработанная нами компьютерная программа для построения регрессионных и классификационных (Q)SAR моделей и последующего прогноза биологической активности и других свойств органических соединений по их структурным формулам [18]. Преимущества GUSAR по отношению к другим методам (CoMFA, CoMSIA, HQSAR и др.) продемонстрированы в сравнительных вычислительных экспериментах [18]. Использование в качестве независимых переменных прогнозируемых PASS спектров биологической активности позволяет определять вероятные механизмы возникновения токсических эффектов [19].

В 2022 г. нами разработан метод построения классификационных зависимостей, основанный на логистической и экспоненциальной самосогласованной классификации [20]. Сопоставление с результатами, полученными на основе самосогласованной регрессии, метода опорных векторов, искусственных нейронных сетей и классификатора PASS, показало, что новый метод обеспечивает сопоставимую с этими методами точность при существенно меньшем числе независимых переменных, что свидетельствует о более высоком качестве классификации³¹. (Q)SAR модели, построенные с использованием этого классификатора на основе обучающих выборок Tox21, могут быть использованы для оценки

безопасности лекарственно-подобного химического соединения [21].

Реализованные нами компьютерные методы обеспечивают высокую точность анализа (Q)SAR и хорошую предсказательную способность построенных моделей, поэтому они были применены для создания специализированных компьютерных программ, прогнозирующих по структурным формулам веществ характеристики, связанные с отдельными видами биологической активности, метаболизмом, токсичностью, межлекарственными взаимодействиями и др. В процессе разработки этих программ использовались специально подготовленные обучающие выборки, а для построенных (Q)SAR моделей проводилась валидация точности и предсказательной способности. Кроме того, при этом были усовершенствованы методические подходы с учетом особенностей конкретной предметной области и специфики решаемых задач.

Краткое описание некоторых специализированных программ, реализованных в виде прогностических веб-ресурсов на платформе Way2Drug, приведено ниже.

Прогнозирование отдельных классов биологической активности

PASS Targets³² прогнозирует взаимодействия лекарственно-подобных химических соединений с молекулярными мишенями [22]. KinScreen³³ прогнозирует взаимодействия лекарственно-подобных химических соединений с киномом человека. CLC-Pred 2.0³⁴ прогнозирует цитотоксичность лекарственно-подобных химических соединений в отношении опухолевых и неопухолевых клеточных линий [23]. DIGEP-Pred³⁵ прогнозирует влияние лекарственно-подобных химических соединений на генную экспрессию [24].

Прогнозирование побочных эффектов и токсичности

ADVER-Pred прогнозирует побочное действие лекарственно-подобных химических соединений на сердечно-сосудистую и гепатобилиарную системы [25]. hERG-Pred³⁶ прогнозирует ингибирование лекарственно-подобными химическими

³⁰ <http://bmc-rm.org/index.php/BMCRM/article/view/4/3>

³¹ http://ibmc.msk.ru/content/thesisDocs/StolbovLA_autoref.pdf

³² <https://www.way2drug.com/passtargets/>

³³ <https://www.way2drug.com/KinScreen/>

³⁴ <https://www.way2drug.com/CLC-pred>

³⁵ <https://www.way2drug.com/ge/>

³⁶ <https://way2drug.com/hERG/>

соединениями hERG-каналов. ROSC-Pred³⁷ прогнозирует органоспецифичную канцерогенность лекарственно-подобных химических соединений [26]. Acute rat toxicity³⁸ прогнозирует острую токсичность для крыс при четырех путях введения лекарственно-подобного химического соединения³⁹ [27]. Antitarget prediction⁴⁰ прогнозирует взаимодействие лекарственно-подобных химических соединений с нежелательными мишенями [28]. DDI-Pred⁴¹ прогнозирует межлекарственные взаимодействия лекарственно-подобных химических соединений [29].

Прогнозирование метаболизма лекарственно-подобных химических соединений и их проникновения через гематоэнцефалический барьер

Metabolic Stability⁴² прогнозирует метаболическую стабильность лекарственно-подобных химических соединений [30]. SOMP⁴³ прогнозирует сайты метаболизма лекарственно-подобных химических соединений [31]. RA⁴⁴ прогнозирует вероятные реакции и сайты биотрансформации [32]. SMP⁴⁵ прогнозирует специфичность субстратов и метаболитов по отношению к ферментам биотрансформации [33]. MetaTox⁴⁶ прогнозирует токсичность лекарственно-подобных химических соединений с учетом их метаболизма [34, 35]. MetaPASS⁴⁷ прогнозирует профили биологической активности исходной фармацевтической субстанции и ее метаболитов [36]. BBB permeability⁴⁸ прогнозирует проникновение лекарственно-подобных химических соединений через гематоэнцефалический барьер [37].

Представленные на платформе Way2Drug прогностические веб-ресурсы обеспечивают возможность оценки как отдельных фармакодинамических и фармакокинетических характеристик органических соединений, изучаемых с целью создания лекарственных средств (прогноз острой токсичности, побочного действия, сайтов метаболизма, метаболической стабильности и др.), так

и интегральных оценок токсичности (MetaTox) [34, 35] и биологической активности (MetaPASS) веществ с учетом их метаболизма [36].

Отбор потенциально наиболее безопасных кандидатов на основе оценок *in silico*

В созданных нами компьютерных программах и веб-ресурсах для оценки различных фармакодинамических и фармакокинетических характеристик лекарственно-подобного соединения входной информацией является его структурная формула, поэтому эти программы могут быть использованы для установления приоритетных направлений экспериментальных исследований на ранних стадиях поиска и разработки новых лекарственных препаратов (даже для виртуальных, сгенерированных *in silico*, структур химических соединений). При этом можно осуществить выбор наиболее безопасных кандидатов из числа изучаемых производных конкретного химического класса. Это чрезвычайно важно, поскольку затраты на проведение доклинических и клинических исследований с учетом требований надлежательной исследовательской практики намного превосходят стоимость теоретических и экспериментальных оценок такого рода на ранних стадиях изучения фармакологически активных веществ. Очевидно, что в случае неудачи на поздних стадиях проекта неизбежны значительные финансовые потери.

Показательной в этом отношении является оценка общего числа побочных и токсических эффектов, прогнозируемых программой PASS для 276 лекарственных препаратов, которые были введены в медицинскую практику, но отозваны с рынка вследствие недостаточной безопасности. Актуальная на момент проведения расчета (2017 г.) версия PASS прогнозировала 494 вида побочных и токсических эффектов. Если принять в качестве эмпирического порогового значения в прогнозе 108 таких эффектов, прогнозируемых PASS, то около 88%

³⁷ <https://www.way2drug.com/ROSC/>

³⁸ <https://www.way2drug.com/gusar/acutoxpredict.html/>

³⁹ Реализована также локальная версия программы, которая прогнозирует острую токсичность для мышей при четырех путях введения фармакологически активного вещества.

⁴⁰ <https://www.way2drug.com/gusar/antitargets.html>

⁴¹ <https://www.way2drug.com/ddi/>

⁴² <https://www.way2drug.com/metastab/>

⁴³ <http://www.way2drug.com/SOMP/>

⁴⁴ <http://way2drug.com/RA/>

⁴⁵ <http://www.way2drug.com/SMP/>

⁴⁶ <http://way2drug.com/mq2/>

⁴⁷ <http://way2drug.com/MetaPASS/>

⁴⁸ <http://www.way2drug.com/geb/>

отозванных с фармацевтического рынка лекарственных препаратов удовлетворяет данному критерию (рис. 1).

По-видимому, совместное использование «больших данных» (информация по 494 побочным и токсическим эффектам, представленным в обучающей выборке программы PASS 2017 114755 записями, характеризующими парные ассоциации «химическое соединение — нежелательный эффект») позволяет компенсировать неполноту доступной информации для каждого отдельного эффекта и существенно повысить качество интегральной токсикологической оценки вещества *in silico*⁴⁹. Например, прогноз побочных и токсических эффектов препарата фенспирид, отозванного с рынка 08.02.2019 вследствие возникновения аритмий у принимавших его пациентов, указывает на вероятность проявления 283 из 494 прогнозируемых эффектов, что намного выше порогового значения в 108 вероятных видов побочных и токсических эффектов (рис. 2).

Таким образом, нами показано, что интегральная оценка прогнозируемых с использованием программы PASS побочных и токсических

эффектов во многих случаях позволяет отсеять потенциально опасные химические соединения на ранних стадиях исследований и отобрать более перспективные препараты-кандидаты для дальнейших исследований.

Определение возможных механизмов развития нежелательных реакций при применении лекарственных препаратов

В последнее десятилетие разработано большое число методов для выявления *in silico* различных химических и биологических характеристик (например, структурные фрагменты, белки-мишени, регуляторные сигнальные сети и др.), коррелирующих с развитием в организме нежелательных реакций при применении лекарственных препаратов и объясняющих возможные механизмы их возникновения. Ранее нами были детально изучены опубликованные в данной области работы [38]. Разработанный нами подход, основанный на применении методов системной фармакологии, рассмотрим на примере лекарственно-индуцированной желудочковой тахикардии (ventricular tachycardia, VT) [39]. Он включает следующие этапы (рис. 3).

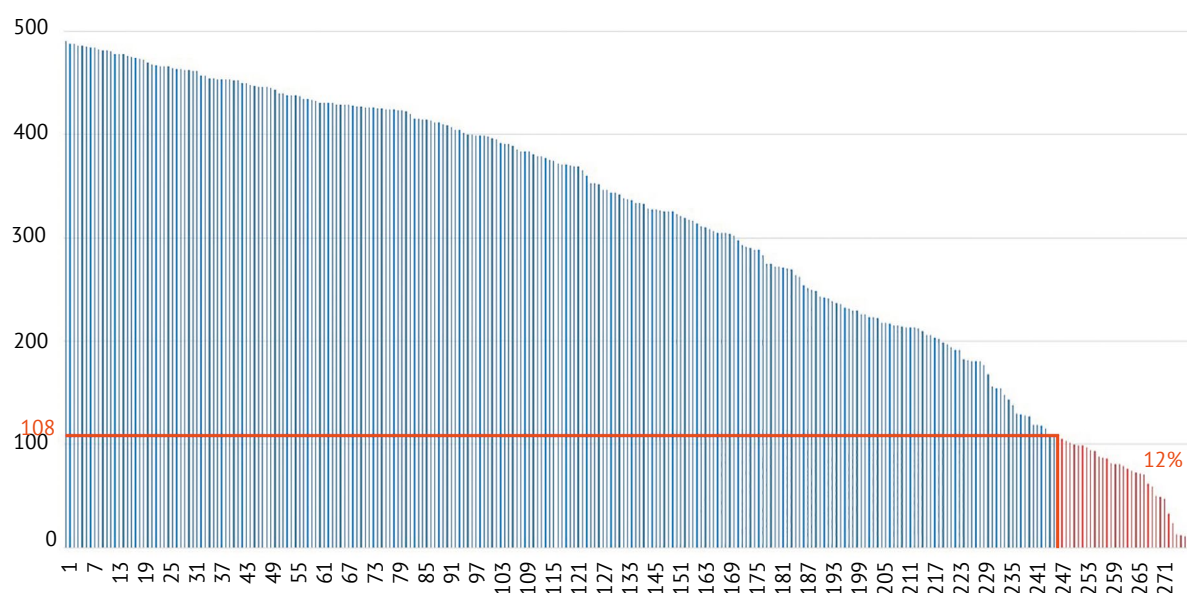


Рисунок подготовлен авторами по собственным данным / The figure is prepared by the authors using their own data

Рис. 1. Число побочных и токсических эффектов, прогнозируемых программой PASS для 276 лекарственных препаратов, отозванных с фармацевтического рынка вследствие недостаточной безопасности

Fig. 1. Number of adverse and toxic effects predicted by PASS for 276 medicinal products withdrawn from the market due to safety issues

⁴⁹ Современная версия программы, PASS 2022 прогнозирует 686 побочных и токсических эффектов на основе анализа 145711 парных ассоциаций «химическое соединение — нежелательный эффект».

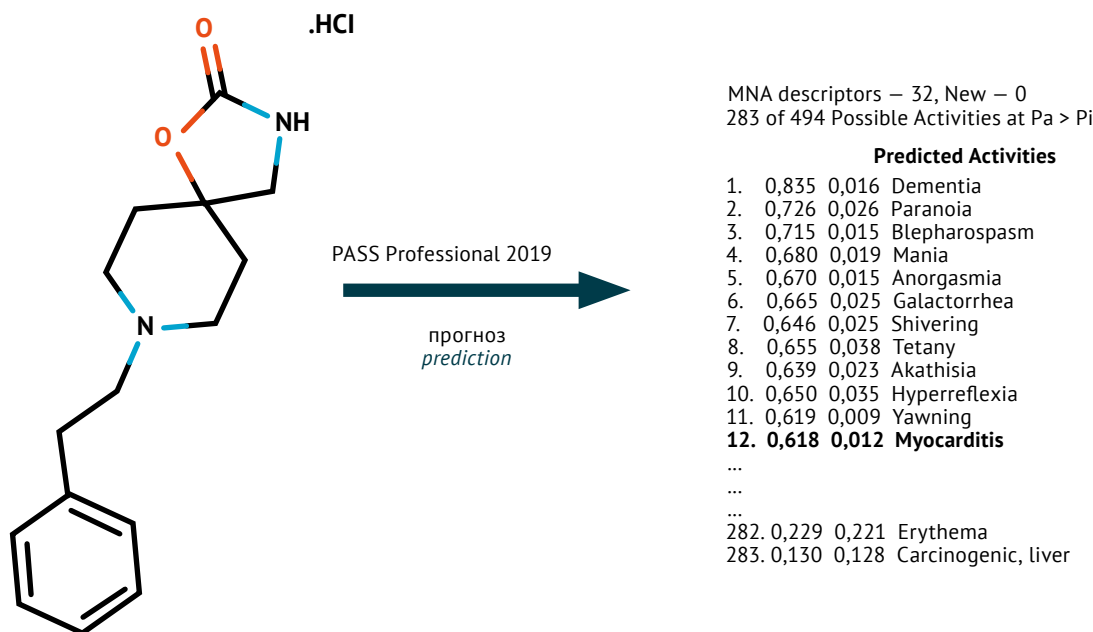


Рисунок подготовлен авторами по собственным данным / The figure is prepared by the authors using their own data

Рис. 2. Прогноз побочных и токсических эффектов для препарата фенспирид, выполненный программой PASS

Fig. 2. PASS prediction of adverse and toxic effects for fenspiride

1) Создание выборки структур лекарственно-подобных химических соединений с информацией об их способности вызывать VT, полученной из общедоступных баз данных и литературы.

2) Прогноз профилей взаимодействия соединений выборки с 1738 белками-мишенями при помощи программы PASS и поиск корреляций между предсказанным действием на мишени и способностью этих соединений вызывать VT. На данном этапе нами было идентифицировано 203 белка-мишени, которые могут быть ассоциированы с индукцией VT.

3) Идентификация сигнальных и регуляторных путей, ассоциированных с выявленными белками. Соответствующие пути были найдены при помощи анализа «обогащения», который позволяет идентифицировать пути, включающие в себя большое количество исследуемых белков. В результате мы обнаружили, что выявленные 203 белка-мишени участвуют в регуляции потенциала действия кардиомиоцитов, функций митохондрий, секреции и ответе клетки на действие инсулина, регуляции вегетативной нервной системы, межклеточных контактов, апоптоза кардиомиоцитов, регуляции электролитного баланса. Нарушение всех найденных процессов способно повышать риск развития VT.

4) Построение регуляторной сети кардиомиоцита с использованием информации о найденных путях, баз данных по белок-белковым взаимодействиям и данных литературы. Сеть включала в себя 1026 вершин, представляющих собой белки, их комплексы, гены, вторичные мессенджеры и ионы, и 2952 ребра двух типов: активация и ингибирование.

5) Дискретное моделирование поведения сети при условии ингибирования ее вершин. Разработанный нами ранее метод дихотомического моделирования [40] был применен для поиска вершин, «ингибирование» которых ассоциировано с изменением состояния «ключевых» вершин, являющихся маркерами описанных выше функций кардиомиоцита. На данном этапе было выявлено 119 белков-мишеней, ассоциированных с VT.

6) Агрегация полученных результатов, ретроспективная валидация на основе анализа данных литературы, классификация выявленных белков по степени достоверности связи с VT. Всего в ходе анализа было идентифицировано 312 уникальных белков-мишеней. Анализ литературы позволил отнести их к одной из трех категорий достоверности по их связи с индукцией VT. Связь 36 белков-мишеней с индукцией VT хорошо известна и является высоко достоверной, например связь с ацетилхолинэстеразой,

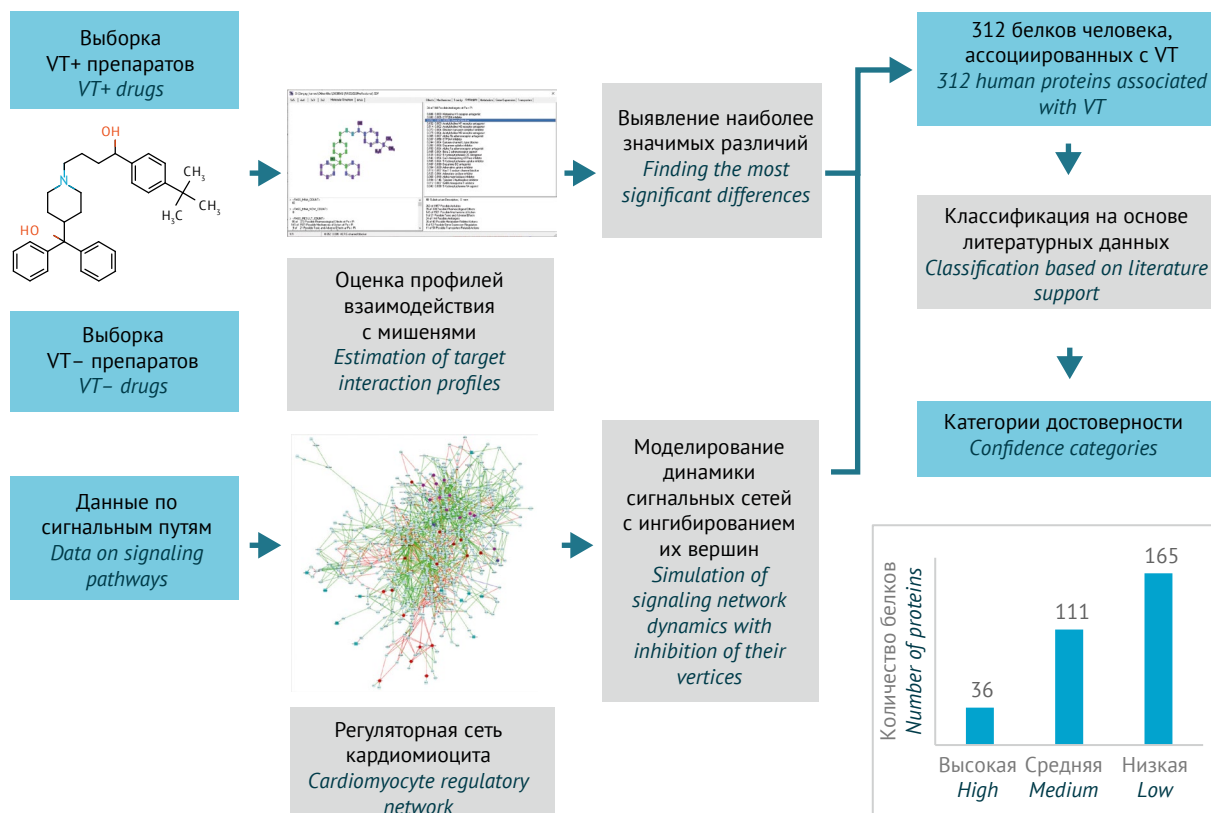


Рисунок подготовлен авторами по собственным данным / The figure is prepared by the authors using their own data

Рис. 3. Схема разработанного подхода для оценки белков-мишеней лекарственных препаратов, связанных с индукцией желудочковой тахикардии (ventricular tachycardia, VT). «VT+» — препараты, вызывающие VT; «VT-» — препараты, не вызывающие VT

Fig. 3. Scheme of the approach developed to evaluate target proteins associated with the induction of ventricular tachycardia (VT). VT+, drugs that cause VT. VT-, drugs that do not cause VT

адренорецепторами, каннабиноидными рецепторами, ионными каналами, включая HERG-каналы. Связь 111 белков-мишеней с индукцией VT не описана в литературе, но известно их участие в вышеописанных процессах, нарушение которых приводит к увеличению риска развития VT. Соответствующие мишени отнесены к категории средней степени достоверности, например киназы CAMK2D, MAPK8, GSK3B, PRKAA1 и др., рецепторы для факторов роста EPHB4, FGFR1-2, IGF1R, фосфолипазы PLD1 и PLD2. Оставшиеся 165 белков-мишеней были отнесены к категории низкой степени достоверности, их связь с VT маловероятна.

Анализ профилей взаимодействия лекарственных препаратов с их мишенями показал, что в 44% случаев связь с развитием VT можно объяснить только взаимодействием препаратов с белками, выявленными в нашем исследовании.

Другие примеры проведенного нами анализа нежелательных реакций, развивающихся

при применении лекарственных препаратов и их комбинаций, представлены в работах [41–45].

Перспективы развития *in silico* подходов

Поиск и разработка новых фармакологических активных веществ — динамично развивающаяся область, постоянно претерпевающая как количественные, так и качественные изменения [46–48]. Более глубокое понимание связей между молекулярными механизмами действия и вызываемыми ими фармакотерапевтически эффектами привело к появлению и развитию «сетевой фармакологии» [49].

Развитие омических технологий и их применение в биомедицинских исследованиях способствует накоплению «больших данных», анализ которых дает возможность уточнить механизмы патологических процессов, что, по-видимому, в перспективе приведет к формированию новой классификации заболеваний [50].

Углубление знаний о механизмах патологических процессов приводит к необходимости разработки новых тест-систем для исследования биологической активности лекарственных-подобных химических соединений. При этом неизбежно возникает проблема сопоставимости данных, полученных в разных экспериментальных условиях. Поэтому необходимо формирование онтологии, характеризующей особенности различных тест-систем, что повысит качество агрегации доступных данных из научных публикаций. Нами было показано, что предварительная обработка данных о биологической активности с целью повышения их однородности усиливает точность и предсказательную способность моделей (Q)SAR, построенных на этих данных [51].

Возможности оценок фармакодинамических и фармакокинетических характеристик лекарственных-подобных химических соединений *in silico* с применением методов ML ограничены уровнем современного состояния развития биомедицинской науки и наличием высококачественных данных, пригодных для создания обучающих выборок [52, 53]. Необходимы постоянные усилия, направленные не только на то, чтобы пополнять обучающие выборки новыми данными о структуре, биологической активности и других свойствах химических соединений, но и существенным образом уточнять понятийный аппарат описания химико-биологических взаимодействий в ряду «лиганд — мишень — биологический процесс — болезнь». Для частичной автоматизации процесса поиска, отбора и анализа релевантной информации будут полезны разрабатываемые нами методы интеллектуального анализа текстов с применением ML [54–56].

Заключение

Проведенный анализ сведений об основных информационных и прогностических ресурсах показал, что в настоящее время в сети Интернет доступны базы данных, агрегированные сведения из которых могут быть использованы для построения методами машинного обучения (Q)SAR моделей с целью расчета различных характеристик фармакологически активных веществ, а также ряд прогностических веб-ресурсов, предоставляющих возможность получать *in silico* оценки параметров, связанных с их безопасностью. Входной информацией для получения прогноза является структурная формула химического соединения, поэтому соответствующие

расчеты могут быть осуществлены на ранних стадиях его изучения, что существенно снижает риск получения отрицательных результатов в процессе дальнейших доклинических и клинических исследований.

Опыт показывает, что высокие характеристики точности и прогностической способности построенных методами ML (Q)SAR моделей достигаются при условии тщательной предварительной обработки данных, полученных из разных источников информации. Сравнение точности и предсказательной способности прогностических веб-ресурсов показывает значительные различия. С целью повышения надежности получаемых оценок можно рекомендовать предварительно протестировать конкретные прогностические ресурсы на примерах, относящихся к интересующей исследователя предметной области.

Установлено, что в настоящее время в свободном доступе нет как единого информационного ресурса, содержащего исчерпывающую информацию о фармакодинамике, фармакокинетике и токсикологии химических соединений, так и единого прогностического ресурса, который обеспечивает прогноз всего комплекса характеристик, необходимых для оценки безопасности фармакологически активных веществ. Наиболее детальный набор характеристик химических соединений может быть получен *in silico* с использованием прогностических ресурсов, представленных на платформе Way2Drug. Использование этих ресурсов позволяет значительно сократить необходимые объемы экспериментальных исследований и осуществить выбор наиболее безопасных кандидатов из числа изучаемых веществ конкретного химического класса.

Применение методов интеллектуального анализа текстов с использованием ML повышает эффективность поиска релевантной информации с целью создания высококачественных обучающих выборок. Однако наряду с необходимостью пополнения обучающих выборок новыми данными о структуре, биологической активности и других свойствах химических соединений, необходимо постоянное уточнение понятийного аппарата описания химико-биологических взаимодействий в ряду «лиганд — мишень — биологический процесс — болезнь». Накопление и анализ «больших данных» открывают возможность уточнять механизмы патологических процессов, что в перспективе может привести к созданию новой классификации заболеваний.

Литература / References

1. Jorgensen WL. The many roles of computation in drug discovery. *Science*. 2004;303(5665):1813–8. <https://doi.org/10.1126/science.1096361>
2. Беженцев ВМ, Дружиловский ДС, Иванов СМ, Филимонов ДА, Sastry GN, Пороиков ВВ. Веб-ресурсы для поиска и разработки новых лекарственных препаратов. *Химико-фармацевтический журнал*. 2017;51(2):3–11. <https://doi.org/10.30906/0023-1134-2017-51-2-3-11>
3. Bajorath J, Overington J, Jenkins JL, Walters P. Drug discovery and development in the era of Big Data. *Future Med Chem*. 2016;8(15):1807–13. <https://doi.org/10.4155/fmc-2014-0081>
4. Fourches D, Muratov E, Tropsha A. Trust, but verify II: a practical guide to chemogenomics data curation. *J Chem Inf Model*. 2016;56(7):1243–52. <https://doi.org/10.1021/acs.jcim.6b00129>
5. Muratov EN, Bajorath J, Sheridan RP, Tetko IV, Filimonov D, Poroikov V, et al. QSAR without borders. *Chem Soc Rev*. 2020;49(11):3525–64. <https://doi.org/10.1039/D0CS00098A>
6. Pun FW, Ozerov IV, Zhavoronkov A. AI-powered therapeutic target discovery. *Trends Pharmacol Sci*. 2023;44(9):561–72. <https://doi.org/10.1016/j.tips.2023.06.010>
7. Bender A, Cortés-Ciriano I. Artificial intelligence in drug discovery: what is realistic, what are illusions? Part 1: Ways to make an impact, and why we are not there yet. *Drug Discov Today*. 2021;26(2):511–24. <https://doi.org/10.1016/j.drudis.2020.12.009>
8. Bender A, Cortes-Ciriano I. Artificial intelligence in drug discovery: what is realistic, what are illusions? Part 2: a discussion of chemical and biological data. *Drug Discov Today*. 2021;26(4):1040–52. <https://doi.org/10.1016/j.drudis.2020.11.037>
9. Hasselgren C, Oprea TI. Artificial intelligence for drug discovery: are we there yet? *Annu Rev Pharmacol Toxicol*. 2024;64:12023. <https://doi.org/10.1146/annurev-pharmtox-040323-040828>
10. Wadman M. FDA no longer has to require animal testing for new drugs. *Science*. 2023;379(6628):127–8. <https://doi.org/10.1126/science.adg6276>
11. Luo M, Wang XS, Tropsha A. Comparative analysis of QSAR-based vs. chemical similarity based predictors of GPCRs binding affinity. *Mol Inform*. 2016;35(1):36–41. <https://doi.org/10.1002/minf.201500038>
12. Murtazaliev KA, Druzhilovskiy DS, Goel RK, Sastry GN, Poroikov VV. How good are publicly available web services that predict bioactivity profiles for drug repurposing? *SAR QSAR Environ Res*. 2017;28(10):843–62. <https://doi.org/10.1080/1062936X.2017.1399448>
13. Forouzes A, Samadi Foroushani S, Forouzes F, Zand E. Reliable target prediction of bioactive molecules based on chemical similarity without employing statistical methods. *Front Pharmacol*. 2019;10:835. <https://doi.org/10.3389/fphar.2019.00835>
14. Ji KY, Liu C, Liu ZQ, Deng YF, Hou TJ, Cao DS. Comprehensive assessment of nine target prediction web services: which should we choose for target fishing? *Brief Bioinform*. 2023;24(2):bbad014. <https://doi.org/10.1093/bib/bbad014>
15. Буров ЮВ, Корольченко ЛВ, Пороиков ВВ. Государственная система регистрации и биологических испытаний химических соединений: возможности для изыскания новых лекарственных препаратов. *Бюллетень Всесоюзного научного центра по безопасности биологически активных веществ*. 1990;(1):4–25. Burov YuV, Korolchenko LV, Poroikov VV. State system of registration and biological testing of chemical compounds: opportunities for finding new drugs. *Bulletin of the All-Union Scientific Center for the Safety of Biologically Active Substances*. 1990;(1):4–25 (In Russ.).
16. Lagunin A, Stepanchikova A, Filimonov D, Poroikov V. PASS: prediction of activity spectra for biologically active substances. *Bioinformatics*. 2000;16(8):747–8. <https://doi.org/10.1093/bioinformatics/16.8.747>
17. Филимонов ДА, Дружиловский ДС, Лагунин АА, Глоризова ТА, Рудик АВ, Дмитриев АВ и др. Компьютерное прогнозирование спектров биологической активности химических соединений: возможности и ограничения. *Biomedical Chemistry: Research and Methods*. 2018;1(1):e00004. Filimonov DA, Druzhilovskiy DS, Lagunin AA, Glorizova TA, Rudik AV, Dmitriev AV, et al. Computer-aided prediction of biological activity spectra for chemical compounds: opportunities and limitations. *Biomedical Chemistry: Research and Methods*. 2018;1(1):e00004 (In Russ.). <https://doi.org/10.18097/bmcr00004>
18. Filimonov DA, Zakharov AV, Lagunin AA, Poroikov VV. QNA based “Star Track” QSAR approach. *SAR QSAR Environ Res*. 2009;20(7–8):679–709. <https://doi.org/10.1080/10629360903438370>
19. Lagunin A, Zakharov A, Filimonov D, Poroikov V. QSAR modelling of rat acute toxicity on the basis of PASS prediction. *Mol Inform*. 2011;30(2–3):241–50. <https://doi.org/10.1002/minf.201000151>
20. Stolbov LA, Filimonov DA, Poroikov VV. SAR based on self-consistent classifier. *SAR QSAR Environ Res*. 2022;33(10):793–804. <https://doi.org/10.1080/1062936X.2022.2139751>
21. Sakamuru S, Huang R, Xia M. Use of Tox21 screening data to evaluate the COVID-19 drug candidates for their potential toxic effects and related pathways. *Front Pharmacol*. 2022;13:935399. <https://doi.org/10.3389/fphar.2022.935399>
22. Pogodin PV, Lagunin AA, Filimonov DA, Poroikov VV. PASS Targets: ligand-based multi-target computational system based on public data and Naïve Bayes

- approach. *SAR QSAR Environ Res.* 2015;26(10):783–93.
<https://doi.org/10.1080/1062936X.2015.1078407>
23. Lagunin AA, Rudik AV, Pogodin PV, Savosina PI, Tarasova OA, Dmitriev AV, et al. CLC-Pred 2.0: a freely available web application for *in silico* prediction of human cell line cytotoxicity and molecular mechanisms of action for druglike compounds. *Int J Mol Sci.* 2023;24(2):1689.
<https://doi.org/10.3390/ijms24021689>
24. Lagunin A, Ivanov S, Rudik A, Filimonov D, Poroikov V. DIGEP-Pred: web-service for *in silico* prediction of drug-induced expression profiles based on structural formula. *Bioinformatics.* 2013;29(16):2062–63.
<https://doi.org/10.1093/bioinformatics/btt322>
25. Ivanov SM, Lagunin AA, Rudik AV, Filimonov DA, Poroikov VV. ADVERPred – web service for prediction of adverse effects of drugs. *J Chem Inf Model.* 2018;58(1):8–11.
<https://doi.org/10.1021/acs.jcim.7b00568>
26. Lagunin A, Rudik A, Filimonov D, Druzhilovskiy D, Poroikov V. ROSC-Pred: web-service for rodent organ-specific carcinogenicity prediction. *Bioinformatics.* 2018;34(4):710–12.
<https://doi.org/10.1093/bioinformatics/btx678>
27. Lagunin A, Zakharov A, Filimonov D, Poroikov V. QSAR modelling of rat acute toxicity on the basis of PASS prediction. *Mol Inform.* 2011;30(2–3):241–50.
<https://doi.org/10.1002/minf.201000151>
28. Zakharov AV, Lagunin AA, Filimonov DA, Poroikov VV. Quantitative prediction of antitarget interaction profiles for chemical compounds. *Chem Res Toxicol.* 2012;25(11):2378–85.
<https://doi.org/10.1021/tx300247r>
29. Dmitriev AV, Filimonov DA, Rudik AV, Pogodin PV, Karasev DA, Lagunin AA, Poroikov VV. Drug-drug interaction prediction using PASS. *SAR QSAR Environ Res.* 2019;30(9):655–64.
<https://doi.org/10.1080/1062936X.2019.1653966>
30. Короткевич ЕИ, Рудик АВ, Дмитриев АВ, Лагунин АА, Филимонов ДА. Прогноз метаболической стабильности ксенобиотиков программами PASS и GUSAR. *Биомедицинская химия.* 2021;67(3):295–9.
Korotkevich EI, Rudik AV, Dmitriev AV, Lagunin AA, Filimonov DA. Predict of metabolic stability of xenobiotics by the PASS and GUSAR programs. *Biomeditsinskaya Khimiya.* 2021;67(3):295–9 (In Russ.).
<https://doi.org/10.18097/PBMC20216703295>
31. Rudik A, Dmitriev A, Lagunin A, Filimonov D, Poroikov V. SOMP: web-service for *in silico* prediction of sites of metabolism for drug-like compounds. *Bioinformatics.* 2015;31(12):2046–8.
<https://doi.org/10.1093/bioinformatics/btv087>
32. Rudik AV, Dmitriev AV, Lagunin AA, Filimonov DA, Poroikov VV. Prediction of reacting atoms for the major biotransformation reactions of organic xenobiotics. *J Cheminform.* 2016;8:68.
<https://doi.org/10.1186/s13321-016-0183-x>
33. Rudik AV, Dmitriev AV, Lagunin AA, Filimonov DA, Poroikov VV. Metabolism sites prediction based on xenobiotics structural formulae and PASS prediction algorithm. *J Chem Inf Model.* 2014;54(2):498–507.
<https://doi.org/10.1021/ci400472j>
34. Rudik AV, Bezhentsev VM, Dmitriev AV, Druzhilovskiy DS, Lagunin AA, Filimonov DA, Poroikov VV. MetaTox: web application for predicting structure and toxicity of xenobiotics' metabolites. *J Chem Inf Model.* 2017;57(4):638–42.
<https://doi.org/10.1021/acs.jcim.6b00662>
35. Rudik A, Bezhentsev V, Dmitriev A, Lagunin A, Filimonov D, Poroikov V. MetaTox – web application for generation of metabolic pathways and toxicity estimation. *J Bioinform Comput Biol.* 2019;17(1):1940001.
<https://doi.org/10.1142/S0219720019400018>
36. Rudik A, Dmitriev A, Lagunin A, Filimonov D, Poroikov V. MetaPASS: a web application for analyzing the biological activity spectrum of organic compounds taking into account their biotransformation. *Mol Inform.* 2021;40(4):2000231.
<https://doi.org/10.1002/minf.202000231>
37. Раевский ОА, Солодова СЛ, Лагунин АА, Поройков ВВ. Компьютерное моделирование проницаемости физиологически активных веществ через гематоэнцефалический барьер. *Биомедицинская химия.* 2014;60(2):161–81.
Raevsky OA, Solodova SL, Lagunin AA, Poroikov VV. Computer modeling of blood–brain barrier permeability of physiologically active compounds. *Biomeditsinskaya Khimiya.* 2014;60(2):161–81 (In Russ.).
<https://doi.org/10.18097/PBMC20146002161>
38. Ivanov SM, Lagunin AA, Poroikov VV. *In silico* assessment of adverse drug reactions and associated mechanisms. *Drug Discov Today.* 2016;21(1):58–71.
<https://doi.org/10.1016/j.drudis.2015.07.018>
39. Ivanov SM, Lagunin AA, Pogodin PV, Filimonov DA, Poroikov VV. Identification of drug targets related to the induction of ventricular tachyarrhythmia through systems chemical biology approach. *Toxicol Sci.* 2015;145(2):321–36.
<https://doi.org/10.1093/toxsci/kfv054>
40. Koborova ON, Filimonov DA, Zakharov AV, Lagunin AA, Ivanov SM, Kel A, Poroikov VV. *In silico* method for identification of promising anticancer drug targets. *SAR QSAR Environ Res.* 2009;20(7–8):755–66.
<https://doi.org/10.1080/10629360903438628>
41. Ivanov SM, Lagunin AA, Pogodin PV, Filimonov DA, Poroikov VV. Identification of drug-induced myocardial infarction-related protein targets through the prediction of drug–target interactions and analysis of biological processes. *Chem Res Toxicol.* 2014;27(7):1263–81.
<https://doi.org/10.1021/tx500147d>
42. Поройков ВВ, Филимонов ДА, Глориезова ТА, Лагунин АА, Дружилловский ДС, Рудик АВ и др. Компьютерный прогноз спектров биологической активности органических соединений: возможности и ограничения. *Известия Академии наук. Серия химическая.* 2019;(12):2143–54. EDN: YQLMTT
Poroikov VV, Filimonov DA, Gloriozova TA, Lagunin AA, Druzhilovskiy DS, Rudik AV, et al. Compu-

- ter-aided prediction of biological activity spectra for organic compounds: the possibilities and limitations. *Russ Chem Bull.* 2019;(12):2143–54.
<https://doi.org/10.1007/s11172-019-2683-0>
43. Ivanov S, Lagunin A, Filimonov D, Poroikov V. Assessment of the cardiovascular adverse effects of drug–drug interactions through a combined analysis of spontaneous reports and predicted drug–target interactions. *PLoS Comput Biol.* 2019;15(7):e1006851.
<https://doi.org/10.1371/journal.pcbi.1006851>
44. Ivanov S, Lagunin A, Filimonov D, Poroikov V. Relationships between the structure and severe drug-induced liver injury for low, medium and high doses of drugs. *Chem Res Toxicol.* 2022;35(3):402–11.
<https://doi.org/10.1021/acs.chemrestox.1c00307>
45. Сухачёв ВС, Иванов СМ, Дмитриев АВ, Прогнозирование неблагоприятных эффектов межлекарственных взаимодействий на сердечно-сосудистую систему на основе анализа связей «структура–активность». *Биохимия.* 2023;88(5):773–84.
<https://doi.org/10.31857/S0320972523050068>
Sukhachev VS, Ivanov SM, Dmitriev AV. Prediction of adverse effects of drug–drug interactions on cardiovascular system based on the analysis of structure–activity relationships. *Biochemistry (Moscow).* 2023;88(5):630–39.
<https://doi.org/10.1134/S0006297923050061>
46. Irurzun-Arana I, Rackauckas C, McDonald TO, Trocóniz IF. Beyond deterministic models in drug discovery and development. *Trends Pharmacol Sci.* 2020;41(11):882–95.
<https://doi.org/10.1016/j.tips.2020.09.005>
47. Blanco MJ, Gardinier KM, Namchuk MN. Advancing new chemical modalities into clinical studies. *ACS Med Chem Lett.* 2022;13(11):1691–8.
<https://doi.org/10.1021/acsmedchemlett.2c00375>
48. Bonner S, Barrett IP, Ye C, Swiers R, Engkvist O, Bender A, Hoyt CT, Hamilton WL. A review of biomedical datasets relating to drug discovery: a knowledge graph perspective. *Brief Bioinform.* 2022;23(6):bbac404.
<https://doi.org/10.1093/bib/bbac404>
49. Hopkins AL. Network pharmacology: the next paradigm in drug discovery. *Nat Chem Biol.* 2008;4(11):682–90.
<https://doi.org/10.1038/nchembio.118>
50. Loscalzo J, Kohane I, Barabasi AL. Human disease classification in the postgenomic era: a complex systems approach to human pathobiology. *Mol Syst Biol.* 2007;3:124.
<https://doi.org/10.1038/msb4100163>
51. Tarasova OA, Urusova AF, Filimonov DA, Nicklaus MC, Zakharov AV, Poroikov VV. QSAR modeling using large-scale databases: case study for HIV-1 reverse transcriptase inhibitors. *J Chem Inf Model.* 2015;55(7):1388–99.
<https://doi.org/10.1021/acs.jcim.5b00019>
52. Alharbi E, Gadiya Y, Henderson D, Zaliani A, Delfin-Rossaro A, Cambon-Thomsen A, et al. Selection of data sets for FAIRification in drug discovery and development: which, why, and how? *Drug Discov Today.* 2022;27(8):2080–5.
<https://doi.org/10.1016/j.drudis.2022.05.010>
53. Перфилова ВН. Возможности и перспективы доклинической оценки лекарственной безопасности с использованием альтернативных методов: опыт реализации программы «Токсикология в XXI веке» в США. *Безопасность и риск фармакотерапии.* 2023. Perfilova VN. Opportunities and prospects for pre-clinical drug safety assessment using alternative methods: experience from the Toxicology in the 21st Century (Tox21) programme in the USA. *Safety and Risk of Pharmacotherapy.* 2023 (In Russ).
<https://doi.org/10.30895/2312-7821-2023-379>
54. Tarasova OA, Biziukova NYu, Filimonov DA, Poroikov VV, Nicklaus MC. Data mining approach for extraction of useful information about biologically active compounds from publications. *Journal of Chemical Information and Modeling.* 2019;59(9):3635–44.
<https://doi.org/10.1021/acs.jcim.9b00164>
55. Tarasova OA, Biziukova NYu, Rudik AV, Dmitriev AV, Filimonov DA, Poroikov VV. Extraction of data on parent compounds and their metabolites from texts of scientific abstracts. *J Chem Inf Model.* 2021;61(4):1683–90.
<https://doi.org/10.1021/acs.jcim.0c01054>
56. Tarasova OA, Rudik AV, Biziukova NYu, Filimonov DA, Poroikov VV. Chemical named entity recognition in the texts of scientific publications using the Naïve Bayes classifier approach. *J Cheminform.* 2022;14:55.
<https://doi.org/10.1186/s13321-022-00633-4>

Вклад авторов. Все авторы подтверждают соответствие своего авторства критериям ICMJE. Наибольший вклад распределен следующим образом: В.В. Поройков — идея и концепция исследования, систематизация данных литературы, написание и редактирование текста рукописи, формулировка выводов, утверждение окончательной версии рукописи для публикации; А.В. Дмитриев, Д.С. Дружиловский, С.М. Иванов, А.А. Лагунин, П.В. Погодин, А.В. Рудик, П.И. Савосина, О.А. Тарасова, Д.А. Филимонов — сбор, анализ и систематизация данных литературы и собственных данных, написание и редактирование текста рукописи.

Authors' contributions. All the authors confirm that they meet the ICMJE criteria for authorship. The most significant contributions were as follows. Vladimir V. Poroikov elaborated the study idea and concept, collated literature data, drafted and edited the manuscript, formulated the conclusions, and approved the final version of the manuscript for publication. Alexander V. Dmitriev, Dmitry S. Druzhilovskiy, Sergey M. Ivanov, Alexey A. Lagunin, Pavel V. Pogodin, Anastasiya V. Rudik, Polina I. Savosina, Olga A. Tarasova, Dmitry A. Filimonov collected, analysed, and collated literature and own data, drafted and edited the manuscript.

ОБ АВТОРАХ / AUTHORS

Поройков Владимир Васильевич, член-корреспондент РАН, профессор, д-р биол. наук, канд. физ.-мат. наук

ORCID: <https://orcid.org/0000-0001-7937-2621>
vladimir.poroikov@ibmc.msk.ru

Дмитриев Александр Викторович, канд. биол. наук
ORCID: <https://orcid.org/0000-0002-2431-3429>
a.v.dmitriev@mail.ru

Дружиловский Дмитрий Сергеевич, канд. биол. наук
ORCID: <https://orcid.org/0000-0001-9024-1331>
dmitry.druzhilovsky@ibmc.msk.ru

Иванов Сергей Михайлович, канд. биол. наук
ORCID: <https://orcid.org/0000-0002-3177-6237>
smivanov7@gmail.com

Лагунин Алексей Александрович, д-р биол. наук, профессор РАН
ORCID: <https://orcid.org/0000-0003-1757-8004>
alexey.lagunin@ibmc.msk.ru

Погодин Павел Викторович, канд. биол. наук
ORCID: <https://orcid.org/0000-0003-1843-7668>
pogodinpv@gmail.com

Рудик Анастасия Владимировна, канд. биол. наук
ORCID: <https://orcid.org/0000-0002-8916-9675>
rudik_anastassia@mail.ru

Савосина Полина Игоревна
ORCID: <https://orcid.org/0000-0001-7066-7925>
polina.savosina@ibmc.msk.ru

Тарасова Ольга Александровна, канд. биол. наук
ORCID: <https://orcid.org/0000-0002-3723-7832>
olga.a.tarasova@gmail.com

Филимонов Дмитрий Алексеевич, канд. физ.-мат. наук
ORCID: <https://orcid.org/0000-0002-0339-8478>
dmitry.filimonov@ibmc.msk.ru

Поступила 07.09.2023

После доработки 13.11.2023

Принята к публикации 17.11.2023

Vladimir V. Poroikov, Corresponding Member of the Russian Academician of Science (RAS), Professor, Dr. Sci. (Biol.), Cand. Sci. (Phys.-Math.)

ORCID: <https://orcid.org/0000-0001-7937-2621>
vladimir.poroikov@ibmc.msk.ru

Alexander V. Dmitriev, Cand. Sci. (Biol.)
ORCID: <https://orcid.org/0000-0002-2431-3429>
a.v.dmitriev@mail.ru

Dmitry S. Druzhilovskiy, Cand. Sci. (Biol.)
ORCID: <https://orcid.org/0000-0001-9024-1331>
dmitry.druzhilovsky@ibmc.msk.ru

Sergey M. Ivanov, Cand. Sci. (Biol.)
ORCID: <https://orcid.org/0000-0002-3177-6237>
smivanov7@gmail.com

Alexey A. Lagunin, Dr. Sci. (Biol.), Professor of the Russian Academy of Sciences (RAS)
ORCID: <https://orcid.org/0000-0003-1757-8004>
alexey.lagunin@ibmc.msk.ru

Pavel V. Pogodin, Cand. Sci. (Biol.)
ORCID: <https://orcid.org/0000-0003-1843-7668>
pogodinpv@gmail.com

Anastasiya V. Rudik, Cand. Sci. (Biol.)
ORCID: <https://orcid.org/0000-0002-8916-9675>
rudik_anastassia@mail.ru

Polina I. Savosina
ORCID: <https://orcid.org/0000-0001-7066-7925>
polina.savosina@ibmc.msk.ru

Olga A. Tarasova, Cand. Sci. (Biol.)
ORCID: <https://orcid.org/0000-0002-3723-7832>
olga.a.tarasova@gmail.com

Dmitry A. Filimonov, Cand. Sci. (Phys.-Math.)
ORCID: <https://orcid.org/0000-0002-0339-8478>
dmitry.filimonov@ibmc.msk.ru

Received 7 September 2023

Revised 13 November 2023

Accepted 17 November 2023