

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ НАУЧНОЕ
УЧРЕЖДЕНИЕ «НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ ИНСТИТУТ
БИОМЕДИЦИНСКОЙ ХИМИИ ИМЕНИ В.Н. ОРЕХОВИЧА

На правах рукописи

Ионов Никита Сергеевич

РАЗРАБОТКА ИНФОРМАЦИОННО-ВЫЧИСЛИТЕЛЬНОЙ ПЛАТФОРМЫ ДЛЯ
ОЦЕНКИ ФАРМАКОЛОГИЧЕСКОГО ПОТЕНЦИАЛА ФИТОКОМПОНЕНТОВ
ЛЕКАРСТВЕННЫХ РАСТЕНИЙ

1.5.8. – Математическая биология, биоинформатика

Диссертация на соискание ученой степени
кандидата биологических наук

Научный руководитель:
доктор биологических наук, профессор,
член-корреспондент РАН
Поройков Владимир Васильевич

Москва 2024

СОДЕРЖАНИЕ

СПИСОК СОКРАЩЕНИЙ	5
ВВЕДЕНИЕ.....	7
ГЛАВА 1. ОБЗОР ЛИТЕРАТУРЫ.....	13
1.1 Низкомолекулярные химические соединения природного происхождения и их роль в разработке лекарств.....	13
1.2 Анализ химического пространства природных соединений	19
1.3 Разработка лекарственных субстанций на основе природных соединений...	23
1.3.1 Подходы и проблемы, связанные с источниками химических соединений	26
1.3.2 Случайный скрининг библиотек экстрактов.....	27
1.3.3 Подход, основанный на знаниях традиционной медицины	28
1.3.4 Экологический подход.....	29
1.3.5 Хемотаксономический подход.....	29
1.3.6 Подходы к тестированию веществ природного происхождения с целью выявления их биологической активности	30
1.3.7 Компьютерный подход	31
1.4 Репозитории структур природных соединений	35
1.5 Заключение	41
ГЛАВА 2. МАТЕРИАЛЫ И МЕТОДЫ	44
2.1 Процедура извлечения данных	44
2.1.1 Подготовка набора видовых названий.....	44
2.1.2 Подготовка набора структурных формул фитокомпонентов и их характеристик	44
2.2 Методы работы с базами данных	46
2.2.1 База данных PubChem.....	46

2.2.2 База данных ChEMBL.....	48
2.3 Компьютерный прогноз спектров биологической активности	49
2.3.1 Программа PASS	49
2.3.2 Программа PharmaExpert.....	51
2.4 Хранение и обработка данных.....	51
2.5 Разработка веб-интерфейса информационно-вычислительной платформы..	51
2.6 Поисковые инструменты	52
2.6.1 Поиск по текстовым запросам	52
2.6.2 Алгоритмы поиска по сходству химических структур	52
2.7 Оценка химического пространства, образованного химическими соединениями из баз данных традиционной медицины	56
2.7.1 «Отпечатки пальцев» с расширенной связностью (ECFP)	56
2.7.2 UMAP.....	57
2.7.3 DBSCAN	58
2.7.4 Кластеризация данных методом k-средних	59
2.8 Валидация информационно-вычислительной платформы Phyto4Health	60
2.8.1 Ретроспективная валидация.....	61
2.8.2 Проспективная валидация.....	62
ГЛАВА 3. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ.....	69
3.1 Логическая структура данных	69
3.2 Извлечение и агрегация данных	72
3.3 Создание доступного в сети Интернет веб-интерфейса информационно-вычислительной платформы	76
3.3.1 Поиск по текстовым запросам	77
3.3.2 Поиск по структурному сходству	78

3.3.3 Результаты выполнения поисковых запросов	81
3.4 Сопоставление с данными о фитокомпонентах традиционной медицины других стран.....	84
3.5 Ретроспективная валидация информационно-вычислительной платформы .	93
3.6 Проспективная валидация.....	98
3.6.1 Оценка цитотоксического действия в отношении клеток рака мочевого пузыря двух вариантов препарата ФЛА.....	99
3.6.2 Оценка спектров биологической активности соединений для фитокомпонентов <i>Rubus chamaemorus</i> L.	105
3.6.3 Исследование влияния мажорных компонентов <i>Cichorium intybus</i> L. на активность цитохрома P450 и глутатионтрансферазы	106
3.6.4 БД SistemатX	108
3.7 Компьютерный анализ фармакологического потенциала фармакопейных растений.....	109
ЗАКЛЮЧЕНИЕ	119
ВЫВОДЫ	121
СПИСОК ОПУБЛИКОВАННЫХ РАБОТ ПО ТЕМЕ ДИССЕРТАЦИИ	122
БЛАГОДАРНОСТИ	124
ФИНАНСИРОВАНИЕ РАБОТЫ	125
СПИСОК ЛИТЕРАТУРЫ	126

СПИСОК СОКРАЩЕНИЙ

МС – масс-спектрометрия

МФА – «Мультифитоадаптоген»

НМИЦ – Национальный медицинский исследовательский центр

ОЛРР – официальные лекарственные растения России

ПС – природные соединения

СПХФУ – Санкт-Петербургский государственный химико-фармацевтический университет

ВИЛАР –Всероссийский научно-исследовательский институт лекарственных и ароматических растений

ФЛА – «Фитоадаптоген»

ЯМР – ядерно-магнитный резонанс

ACD – каталог доступных химических веществ (Available Chemicals Directory)

AlogP – значение коэффициента распределения на границе раздела фаз «октанол-вода»

BDNP – база данных биологически активных природных соединений (Biologically Active Natural Products Database)

CSS – Формальный язык декорирования веб-страниц (Cascading Style Sheets)

DBSCAN - Основанная на плотности пространственная кластеризация для приложений с шумами (Density-based spatial clustering of applications with noise)

DNP – словарь природных соединений (Dictionary of Natural Products)

ECFP – молекулярные «отпечатки пальцев» с расширенной связностью (Extended-connectivity fingerprints)

FDA – Управление по санитарному надзору за качеством пищевых продуктов и медикаментов США (U.S. Food and Drug Administration)

Fsp3 – отношение количества sp³-гибризованных атомов углерода к общему числу атомов углерода

GNPS – база знаний с открытым доступом для организации и обмена необработанными, обработанными или аннотированными данными

фрагментарной масс-спектрометрии (Global Natural Products Social Molecular Networking)

HBA – количество доноров водородных связей

HBD – количество акцептов водородных связей

HMDB – база данных метаболома человека (Human Metabolome Database)

HTML – Язык гипертекстовой разметки (Hyper Text Markup Language)

MNA – дескрипторы многоуровневых атомных окрестностей (Multilevel Neighborhoods of Atoms)

MoNA – база масс-спектральных данных Северной Америки (MassBank of North America)

MW – молекулярная масса

NCBI – Национальный центр биотехнологической информации

PASS – Prediction of Activity Spectra for Substances

PSA – полярная площадь поверхности

QNA – дескрипторы количественных атомных окрестностей (Quantitative Neighborhoods of Atoms)

(Q)SAR – количественные зависимости «структура-свойство» ((Quantitative) Structure-Activity Relationship)

RTB – количество связей, допускающих вращение

SDF – файл с данными о химических структурах и их характеристиках (Structure Data File)

TCMD – база данных традиционных китайских лекарственных средств (Traditional Chinese Medicines Database)

tPSA – топологическая полярная площадь поверхности

UMAP – метод унифицированной аппроксимации и проекции многообразия (Uniform Approximation and Projection)

YMDB – база данных метаболомов дрожжей (Yeast Metabolome Database)

ВВЕДЕНИЕ

Актуальность работы

Высокое разнообразие химических структур веществ природного происхождения по сравнению с соединениями синтетической природы обуславливает более широкий спектр потенциальных видов биологической активности и, соответственно, расширение возможностей для поиска и создания новых лекарственных препаратов [1]. Однако, ещё до начала лабораторного тестирования биологической активности разработка фармацевтических субстанций на основе природных соединений может быть сопряжена с проблемами наработки опытных образцов, идентификацией и выделением целевых соединений, ограничениями на защиту интеллектуальной собственности, а также выбором наиболее перспективных направлений исследования биологической активности [2, 3].

В настоящее время накоплен обширный опыт применения лекарственных растений в рамках традиционной медицины различных стран. Кроме того, в научной литературе представлено множество информации, включающей в себя данные о методиках выделения и стандартизации индивидуальных веществ, о потенциальных токсических и фармакологических эффектах, а также сведения об экономической обоснованности их исследования и дальнейшего применения для терапии различных патологий [4].

Таким образом, проведенная на основе анализа накопленных сведений оценка фармакологического потенциала фитокомпонентов лекарственных растений, включенных в Государственную Фармакопею РФ, позволит минимизировать описанные выше сложности. Для проведения такой оценки необходим информационный ресурс, интегрирующий накопленные данные.

На территории Российской Федерации располагается ряд климатических зон, что обуславливает высокое видовое разнообразие произрастающих растений. Поэтому набор их фитокомпонентов, синтезируемых для обеспечения выживания в условиях внешней среды, характеризуется структурным разнообразием.

Методы виртуального скрининга широко используются для предварительной оценки биологической активности и определения перспективных направлений экспериментальных исследований [5]. Полноценное применение компьютерных методов для оценки фармакологического потенциала официальных (включенных в Государственную фармакопею) лекарственных растений России (ОЛРР), также ограничено отсутствием массива структур фитокомпонентов, представленных в надлежащем для *in silico* анализа виде.

Целью диссертационной работы является разработка и валидация информационно-вычислительной платформы, обеспечивающей хранение, обработку и анализ информации о фитохимическом составе лекарственных растений России и характеристик отдельных фитокомпонентов для обогащения знаний о их фармакологическом потенциале.

Для достижения указанной цели были поставлены следующие задачи:

1. Агрегация известных данных о структуре и биологической активности фитокомпонентов официальных лекарственных растений России.
2. Создание информационно-вычислительной платформы, предоставляющей в сети Интернет сведения о структуре, биологической активности и других характеристиках фитокомпонентов официальных лекарственных растений России, а также вычислительные инструменты для анализа их фармакологического потенциала.
3. Анализ структурного разнообразия фитокомпонентов официальных лекарственных растений России и сопоставление их с соответствующими данными о фитокомпонентов растений других регионов (Индия, Бразилия, Мексика, Северная Африка, Европейские страны).
4. Ретроспективная и проспективная валидация функциональности информационно-вычислительной платформы как «инструмента» для оценки фармакологического потенциала лекарственных растений России.

Научная новизна

Впервые разработана информационно-вычислительная платформа Phyto4Health, содержащая сведения о структуре, биологической активности и

других характеристиках фитокомпонентов ОЛРР. Phyto4Health предоставляет исследователям вычислительные инструменты для прогнозирования фармакологических эффектов, механизмов действия, нежелательных побочных эффектов и цитотоксического действия в отношении опухолевых и нормальных клеточных линий. Таким образом, информационный компонент платформы Phyto4Health позволяет получить сведения о химической структуре, молекулярных свойствах и результатах лабораторных экспериментов, а вычислительный компонент позволяет оценить ранее неизвестные виды биологической активности отдельных фитокомпонентов ОЛРР, а также провести поиск структурных аналогов. Созданная нами платформа свободно доступна для академических исследователей в сети Интернет (<http://www.way2drug.com/phyto4health/>).

Ретроспективная валидация результатов компьютерного прогноза на основе данных из крупных репозиторий химических структур показала, что результаты компьютерного прогноза биологической активности согласуются с данными лабораторных экспериментов. Совместно с исследователями из других организаций проведена апробация информационно-вычислительной платформы, в результате чего выявлено новое направление лабораторного исследования фитокомпозиции «Фитоладаптоген» (ФЛА) как противоопухолевого препарата для лечения рака мочевого пузыря, впервые установлен антиагрегантный эффект фитокомпонентов из листьев морошки (*Rubus chamaemorus* L.), а также выявлены и подтверждены в эксперименте ранее неизвестные механизмы гепатопротекторного действия фитокомпонентов надземной части цикория обыкновенного (*Cichorium intybus* L.).

Научно-практическая значимость

Разработанная нами информационно-вычислительная платформа обеспечивает доступ широкому кругу исследователей к сведениям о фитохимическом составе ОЛРР, а также их характеристикам, включая различные представления структурных формул, молекулярные свойства, результаты лабораторного тестирования их взаимодействия с молекулярными мишенями человека. Инструменты для компьютерной оценки наиболее вероятных видов

биологической активности позволяют установить наиболее перспективные направления исследований с целью расширения фармакотерапевтического применения отдельных фитокомпонентов ОЛРР и фармацевтических композиций, созданных на их основе. С применением разработанной нами информационно-вычислительной платформы установлены фармакотерапевтические эффекты фитокомпозиции «Фитоладаптоген» и фитокомпонентов из листьев морозники, а также механизмы гепатопротекторного действия фитокомпонентов цикория обыкновенного.

Личный вклад автора

Автором проведён сбор сведений о фитохимическом составе, молекулярных свойствах и результатах тестирования фитокомпонентов ОЛРР. Разработана логическая структура данных, осуществлена агрегация сведений, а также реализован специализированный веб-интерфейс информационно-вычислительной платформы. В рамках ретроспективной валидации, выполнен анализ результатов компьютерного прогноза с использованием данных из крупных репозиториях химических структур. В рамках проспективной валидации, автором получен прогноз спектров биологической активности, проведены интерпретация результатов и сопоставление с экспериментальными данными, полученными в НМИЦ онкологии им. Н.Н. Блохина, СПХФУ и ВИЛАР.

Положения, выносимые на защиту:

1. Собран массив данных, включающий в себя сведения о фитохимическом составе 233 официальных лекарственных растений России, химических структурах, молекулярных свойствах и результатах лабораторного исследования взаимодействия с молекулярными мишенями человека для 3128 уникальных фитокомпонентов.
2. Веб-интерфейс информационно-вычислительной платформы обеспечивает доступ к базе данных и представляет расширенные опции поиска сведений и оценки *in silico* биологической активности отдельных фитокомпонентов с учётом конкретных целей исследования потенциальных пользователей.

3. Сопоставление представленных в БД фитоконпонентов с фитоконпонентами растений других регионов, включая Бразилию, Индию, Центральную Европу и Северную Африку, показало, что содержащаяся в БД информация во многих случаях является уникальной и отражает широкое химическое разнообразие фитоконпонентов официальных лекарственных растения России.
4. В результате ретроспективной и проспективной валидации продемонстрирована возможность практического использования разработанной информационно-вычислительной платформы для анализа фармакологического потенциала официальных лекарственных растений России и выявления новых направлений их исследований.

Апробация работы

Основные положения диссертационной работы были представлены на российских и международных научных конференциях и симпозиумах: ACS Spring National Meeting (Orlando, 2019), Markovnikov Congress on Organic Chemistry (Казань, 2019) Научно-методической конференции «IV Гаммермановские чтения» (Санкт-Петербург, 2019), XXVII Симпозиуме «Биоинформатика и компьютерное конструирование лекарств» (онлайн, 2021), XXVIII Symposium «Bioinformatics and Computer-Aided Drug Discovery» (онлайн, 2022), Международной научной конференции «От биохимии растений к биохимии человека» (Москва, 2022), «Персонализированная медицина: технологии профилактики и ранней диагностики заболеваний» (Орёл, 2023), Научно-практической конференции «Гаммермановские чтения» (Пермь, 2023).

Публикации

По материалам диссертации опубликовано 11 работ в российских и международных научных изданиях. Среди них 6 статей в рецензируемых научных журналах, из которых 3 – в международных журналах (2 - Q1, 1 – Q2) и входящих в перечень, рекомендованный ВАК, а также 5 публикаций в трудах конференций. Получено свидетельство Роспатента о государственной регистрации базы данных Phyto4Health № 2023622658 от 14.07.2023.

Объём и структура диссертации

Диссертационная работа состоит из введения, обзора литературы, материалов и методов исследования, результатов и обсуждения, заключения, выводов и списка литературы, включающего 191 источник. Работа изложена на 141 странице, содержит 38 рисунков и 18 таблиц.

ГЛАВА 1. ОБЗОР ЛИТЕРАТУРЫ

1.1 Низкомолекулярные химические соединения природного происхождения и их роль в разработке лекарств

В настоящее время существует множество разнообразных определений для химических соединений, синтезируемых живыми организмами. В частности, это связано с происхождением и функциями этих веществ. Например, в научных публикациях встречаются разночтения, связанные с разделением химических соединений на продукты первичного и вторичного метаболизма. При этом не учитывают продукты конечного метаболизма (этиловый спирт, метан и т.д.). Другие возможные варианты определений связаны с установленными функциями химических веществ. Например, в качестве химических веществ природного происхождения рассматриваются продукты вторичного метаболизма, выполняющие преимущественно регуляторную функцию, и исключаются вторичные метаболиты, выполняющие другие функции (например, защитную и репродуктивную). Чтобы избежать возможных разночтений в рамках данной работы в настоящей работе мы используем наиболее общее определение, приведенное ниже.

Низкомолекулярные химические соединения природного происхождения составляют группу химических соединений, синтезируемых живыми организмами с целью адаптации и выживания в условиях окружающей среды. Природные соединения (ПС) являются результатом длительной эволюции биологических систем, а с учётом общности происхождения и схожести молекулярных структур, не вызывает удивления тот факт, что эти соединения проявляют биологическую активность в других организмах, в том числе и в организме человека [6]. ПС длительное время применяются в традиционной медицине, а разработанные на основе ПС лекарства широко используются в современной фармакотерапии множества нозологий [7].

В периодических обзорах Д. Ньюмена и соавт., публикуемых с 1997 по 2019 года, используется классификация лекарственных препаратов по происхождению

их действующих веществ [8]. В представленной классификации действующие вещества разделяются на 8 групп. К пяти из них могут быть отнесены низкомолекулярные химические соединения. Эти группы приведены в таблице 1.

Таблица 1 – Классификация лекарств на основе ПС по происхождению действующих веществ

№	Название группы	Примеры действующих веществ лекарств и их фармакологические эффекты/показания к применению
1	Лекарства, разработанные на основе структур ПС в неизменном виде	Нетилмицина сульфат (антибиотик), меглутол (гиполипидемический эффект), лентинан (иммуностимулянт)
2	Одобрённые к применению растительные препараты	Экстракт <i>Serenoa repens</i> (доброкачественная гипертрофия предстательной железы), Крофелемер (противодиарейный эффект), IZN-6D4 (местный противоязвенный)
3	Лекарства, разработанные на основе производных ПС	Цефоперазон натрия (антибактериальный), тестостерон ундеканат (гипогонадизм), терпенон (противоязвенный)
4	Лекарства, разработанные на основе химических соединений, полученных путем органического синтеза	Кетоконазол (противогрибковый), оксатомид (противоаллергический), бутконазол (противогрибковый)
5	Лекарства, разработанные на основе химических соединений, полученных путем органического синтеза на основе фармакофоров ПС	Целипролола гидрохлорид (антигипертензивный), эноцитабин (противоопухолевый), митоксантрона гидрохлорид (противоопухолевый)

Важно заметить, что в обзорах Д. Ньюмена и соавт. дополнительно приводятся сведения о вакцинах и биологических макромолекулах, одобренных к применению Управлением по санитарному надзору за качеством пищевых продуктов и медикаментов США (англ. - U.S. Food and Drug Administration (FDA)), однако в рамках данного раздела они учитываться не будут. Кроме того, авторами обзора [8] рассматривается подкатегория веществ, являющихся аналогами ПС по механизмам действия (NM).

Согласно литературным данным, в период с 1981 по 2019 гг. FDA было одобрено к применению около 1880 лекарств по более чем 253 терапевтическим показаниям [8]. Эти препараты охватывают все категории, перечисленные в таблице

1, а также вакцины и биологические макромолекулы. Число препаратов, разработанных на основе низкомолекулярных химических соединений, составляет 1398. Распределение одобренных лекарств в период с 1981 по 2019 годы по категориям таблицы 1, приведено на рисунке 1.

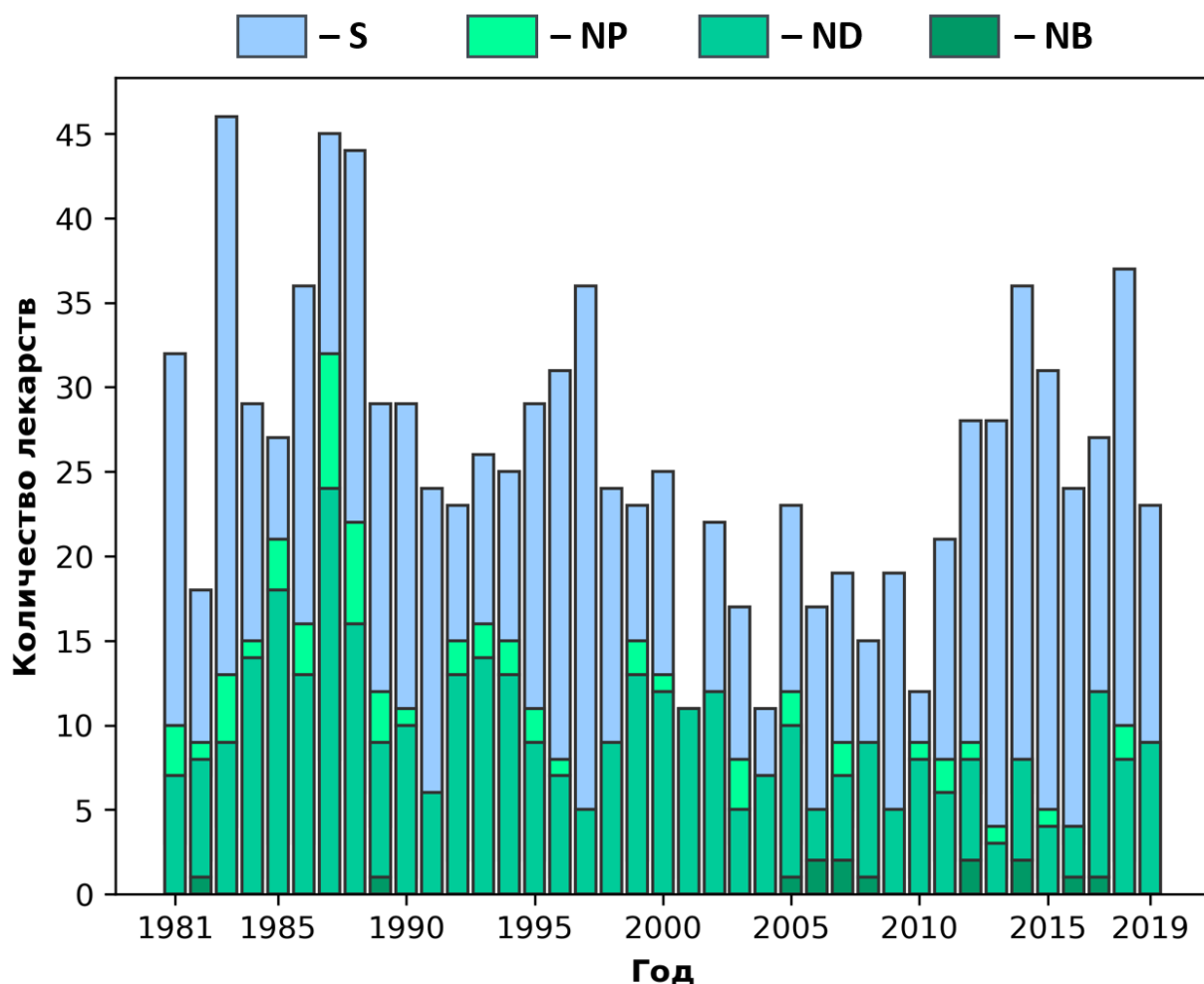


Рисунок 1 – Число лекарств, разрешенных к медицинскому применению с период с 1981 по 2019 гг. «S» – лекарства, разработанные на основе химических соединений, полученных путем органического синтеза (соответствует категориям 4 и 5 из таблицы 1); «NP» – лекарства, разработанные на основе структур ПС в неизменном виде (соответствует категории 1 из таблицы 1); «ND» – лекарства, разработанные на основе структур модифицированных структур ПС (соответствует категории 3 из таблицы 1); «NB» – лекарственные растительные препараты (соответствует категории 2 таблицы 1). Адаптировано из [8]

Как видно из приведенных на рисунке 1 данных, количество лекарств, разработанных на основе ПС, не превышает количество лекарств, полученных на основе синтетических молекул. Доля одобренных лекарств, разработанных на

основе действующих веществ синтетической природы (S), составляет 50 % от общего количества лекарств и 68 % от лекарств, соответствующих категориям, приведенным в таблице 1. Доля одобренных лекарств, разработанных на основе действующих веществ природного происхождения (NP, ND, NB), составляет 24 % от общего количества лекарств и 31 % от лекарств, соответствующих категориям, приведенным в таблице 1. Важно отметить, что Д. Ньюмен и соавт. указывают, что доля лекарств, одобренных в применении FDA и разработанных на основе соединений природного происхождения, составляет свыше 70 %. Однако к таковым соединениям авторами были отнесены также биологические макромолекулы, вакцины, а также соединения, соответствующие категории 5 из таблицы 1, и соединения синтетической природы, имитирующие механизм действия ПС.

Согласно литературным данным, известные структуры ПС характеризуются более высоким химическим разнообразием по сравнению с соединениями, получаемыми методами органического синтеза [9]. Это является причиной более широкого разнообразия проявляемых ими видов биологической активности [10]. Поэтому не вызывает удивления тот факт, что наибольшее количество “first-in-class” лекарств (инновационные лекарственные препараты, первые в своем классе) было разработано на основе веществ природного происхождения [11]. На рисунке 2 приведена количественная характеристика “first-in-class” лекарств, одобренных FDA в период с 1983 по 2018 гг.

Как видно из представленных на рисунке 2 данных, из 188 лекарств 73 (39%) имеют природное происхождение. Кроме того, ряд лекарств, активные фармацевтические субстанции которых получают методами органического синтеза, были разработаны на основе фармакофоров природных аналогов (37) или имитируют их механизм действия (36). Таким образом, при разработке 146 (77%) из всех одобренных к применению «first-in-class» лекарств были использованы данные о ПС.

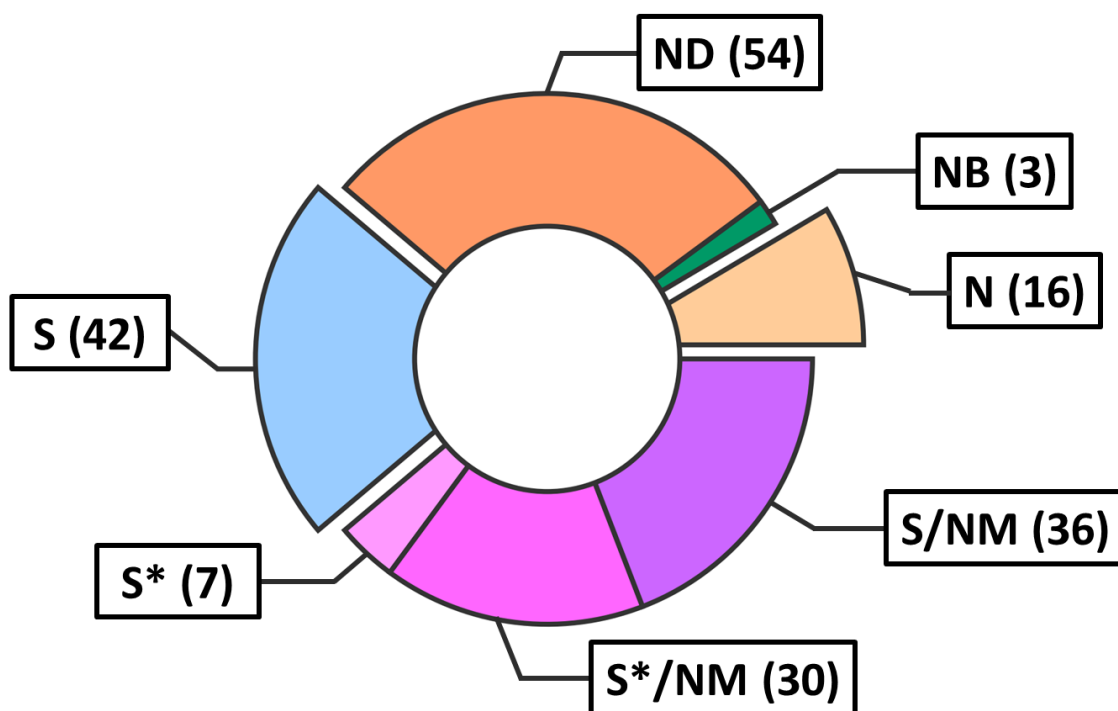


Рисунок 2 – Количество “first-in-class” лекарств, одобренных FDA в период с 1983 по 2018 годы. «S» – лекарства, разработанные на основе химических соединений, полученных путем лабораторного синтеза (соответствует категориям 4 и 5 из таблицы 1); «N» – лекарства, разработанные на основе структур ПС в неизменном виде (соответствует категории 1 из таблицы 1); «ND» – лекарства, разработанные на основе структур модифицированных структур ПС (соответствует категории 3 из таблицы 1); «NB» – лекарственные растительные препараты (соответствует категории 2 таблицы 1); S* – Лекарства, разработанные на основе химических соединений, полученных путем тотального синтеза на основе фармакофоров ПС; «NM» – указание на то, что механизм действия лекарства имитирует механизм действия природного аналога. Для построения графика использовались данные о происхождении химических структур из работы [8] и данные о соответствии лекарства категории «first-in-class» из публикации [12]

Исходя из вышеуказанного, использование информации о структуре и свойствах ПС играет важную роль в разработке лекарств на протяжении всего развития фармацевтической промышленности. Соотношение лекарств, разработанных на основе веществ, полученных методами органического синтеза, и не модифицированных структур ПС приведено на рисунке 3.

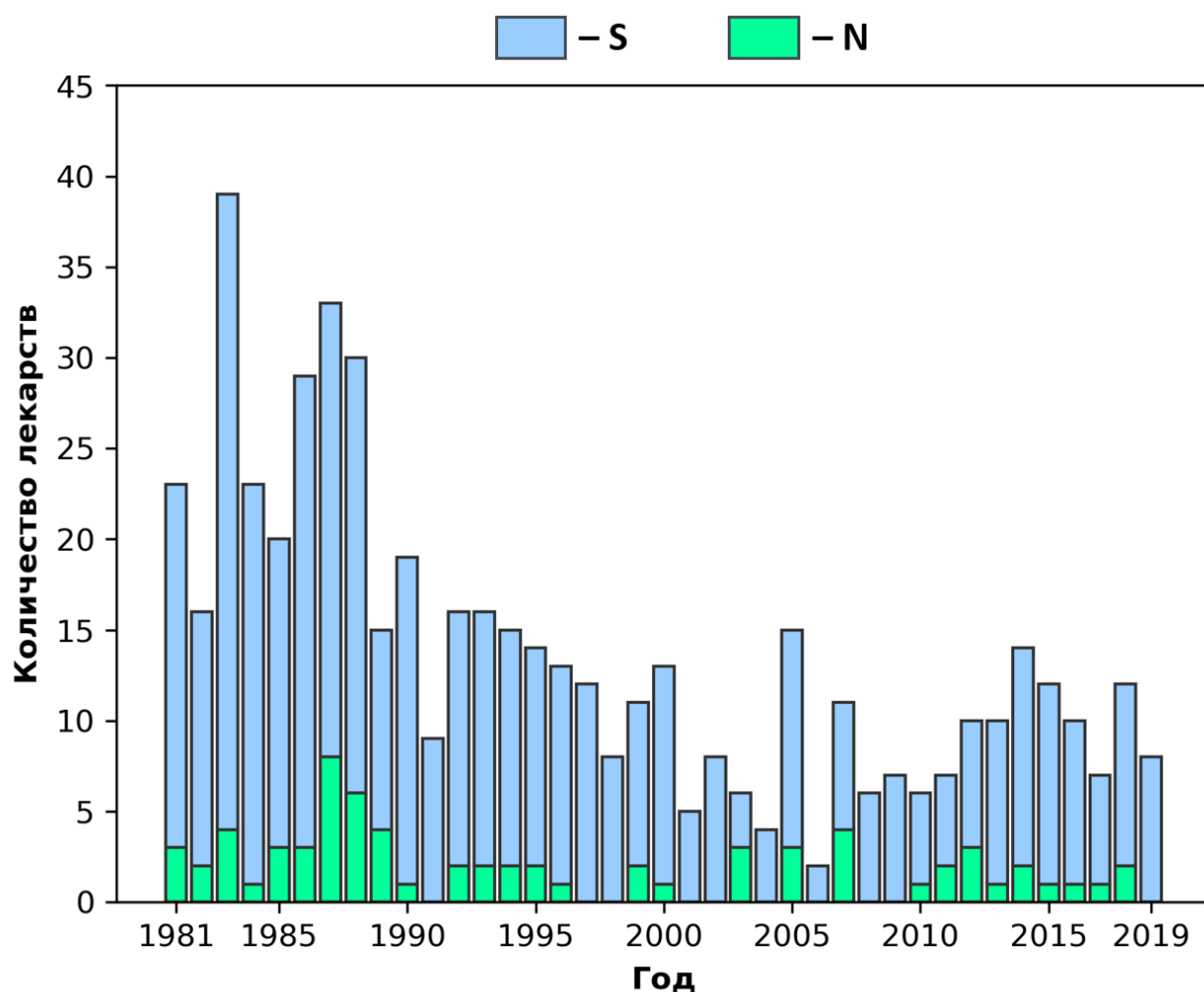


Рисунок 3 – Соотношение лекарств, действующие вещества которых получены методами органического синтеза («S»), и лекарств, разработанных на основе не модифицированных структур ПС («N»). Адаптировано из [8]

Как видно из рисунка 3, количество лекарств, разработанных на основе не модифицированных структур соединений природного происхождения, значительно меньше лекарств синтетической природы и составляет 3% от всех вышедших на рынок лекарств. В то же время, количество лекарств, разработанных на основе соединений синтетической природы, составляет 24%. Из приведенных на рисунке 2 данных видно, что в 1991, 1997, 1998, 2001, 2002, 2004, 2006, 2008, 2009 и 2019 годах не было выпущено на рынок ни одного лекарства, разработанного на основе не модифицированных ПС.

Сравнительно низкое число лекарств, разработанных на основе не модифицированных структур природного происхождения связано с несколькими обстоятельствами. Во-первых, исходные ПС часто не обладают оптимальными

свойствами, связанными с биодоступностью и безопасностью для разработки новых лекарств [5]. Поэтому для получения молекул, которые потенциально могут быть одобрены для использования в качестве лекарств, применяются методы функционально-ориентированного синтеза, то есть идентифицируется и синтезируется отдельно фрагмент молекулы, который влияет на проявление желаемой активности [13]. Во-вторых, важным критерием, характеризующим ПС как объект для исследований, являются финансовые затраты на выделение и очистку. Стоимость сырья и его дальнейшей обработки для природных биологически активных веществ может значительно превышать стоимость органического синтеза более простых молекул. Однако развитие биотехнологии, в частности, методов геномной инженерии, позволяющих повысить выход изучаемого химического соединения из исходного сырья, открывает новые перспективы использования природных биологически активных молекул в качестве источника соединений-кандидатов для дальнейшей разработки [14].

Таким образом, процесс разработки лекарств на основе ПС может оказаться длительным и трудоемким. Отчасти с этим связано снижение количества лекарств, разработанных на основе ПС и разрешенных к применению. Однако ПС остаются перспективным источником разнообразных химических структур с богатым потенциалом для обнаружения новых структур, которые потенциально могут стать эффективными препаратами для лечения различных социально-значимых заболеваний человека.

1.2 Анализ химического пространства природных соединений

В настоящее время не существует устоявшегося определения концепции химического пространства, несмотря на повсеместное использование данного термина в научных публикациях. В первой части учебного пособия «Введение в хемоинформатику: Компьютерное представление химических структур» Маджидовым и соавт. предлагается следующая формулировка: «Химическое пространство – это набор химических объектов, для которых определено отношение, описывающие их сходство друг с другом» [15]. В рамках данного

определения, химическое пространство характеризуется двумя важными аспектами – способом представления химической структуры (дескрипторы) и методикой расчёта сходства химических объектов (расстояние между объектами в химическом пространстве). Для построения и оценки химического пространства обычно применяют физико-химические или топологические дескрипторы, а также молекулярные «отпечатки пальцев» (англ. fingerprints) [16, 17]. В качестве метрики расстояния могут использоваться как индексы оценки сходства между объектами (Танимото [18], Дайса [19], Оchiaи [20] и т.д.), так и некоторые метрики для оценки расстояния (Евклидово расстояние [21], расстояние городских кварталов [22] и т.д.). На заключительных этапах может применяться процедура кластеризации [23]. Кроме того, зачастую применяются более сложные методики построения химических пространств. Например, методы самоорганизующихся карт Кохонена [24], генеративных топографических карт [25] или «тримаппинг» (англ. tree map, TMAP) [26].

Исходя из основной парадигмы QSAR (Quantitative Structure-Activity Relationship) – «Схожие структуры имеют схожие свойства», в научных публикациях проводится анализ химического пространства ПС относительно химических пространств других типов молекул [27]. К таким молекулам могут относиться соединения, доступные для синтеза или одобренные к применению лекарства.

В работе Сингха и соавт. был проведен сравнительный анализ химических пространств ПС, представленных в базах данных Dictionary of Natural Products (DNP, 78 318 химических структур) и Biologically Active Natural Products Database (BDNP, 28 432 химических структур), с химическим пространством синтетических структур, представленных в базах данных Available Chemicals Directory (ACD, 182 822 химических структуры) и набором химических соединений, полученных из химических библиотек фармацевтических компаний (например, Bayer AG) и включающих в себя химические структуры одобренных к применению лекарств и активных соединений (14 596 химических структур) [28]. В результате было показано, что ПС охватывают более широкое химическое пространство, чем

соединения, представленные в синтетических и комбинаторных библиотеках. Химические структуры ПС содержат больше хиральных центров и конденсированных колец чем синтетические структуры. Физико-химические параметры ПС, обладающих пероральной биодоступностью, часто выходят за пределы, установленные правилом Липински для перорально доступных лекарственных соединений [29].

Гу и соавт. провели сравнительный анализ химического пространства ПС, полученных из баз данных Reaxys, Chinese Natural Product Database (CNPD) [30], Traditional Chinese Medicines Database (TCMD) [31], CHDD [32] и одобренных к применению FDA лекарств, загруженных из DrugBank [33]. Для описания химических структур авторами использовался набор значений физико-химических дескрипторов, включающих в себя количество доноров и акцепторов водородных связей (HBD, HBA), площадь полярной поверхности (PSA), коэффициент липофильности (AlogP) и др. В результате было установлено, что ПС охватывают большую область исследуемого химического пространства и демонстрируют большее разнообразие химических структур, по сравнению с одобренными FDA лекарствами [34].

Чен и соавт., наряду с библиотеками ПС и синтетических молекул, использовали также выборку ПС – метаболитов человека. Авторами установлено, что структуры метаболитов человека и ПС охватывают различные области химического пространства по сравнению с синтетическими соединениями. Кроме того, установлено, что ПС имеют наибольшую сложность организации ароматических и алифатических колец, а также ПС и метаболиты человека обладают большим разнообразием пространственной организации химической структуры [35].

Эль-Элима и соавт. провели сравнение химического пространства ПС, выделенных из нитчатых грибов (105 химических структур), цианобактерий (75 химических структур) и тропических растений (163 химических структуры) с химическим пространством, охватываемым противоопухолевыми препаратами, одобренными к применению FDA (96 химических структур) [36]. Пространство

было задано физико-химическими дескрипторами (молекулярный вес, число хиральных центров, число связей, допускающих вращение и др.). Снижение размерности пространства физико-химических дескрипторов проводилось с использованием метода главных компонент. Визуализация химического пространства приведена на рисунке 4.

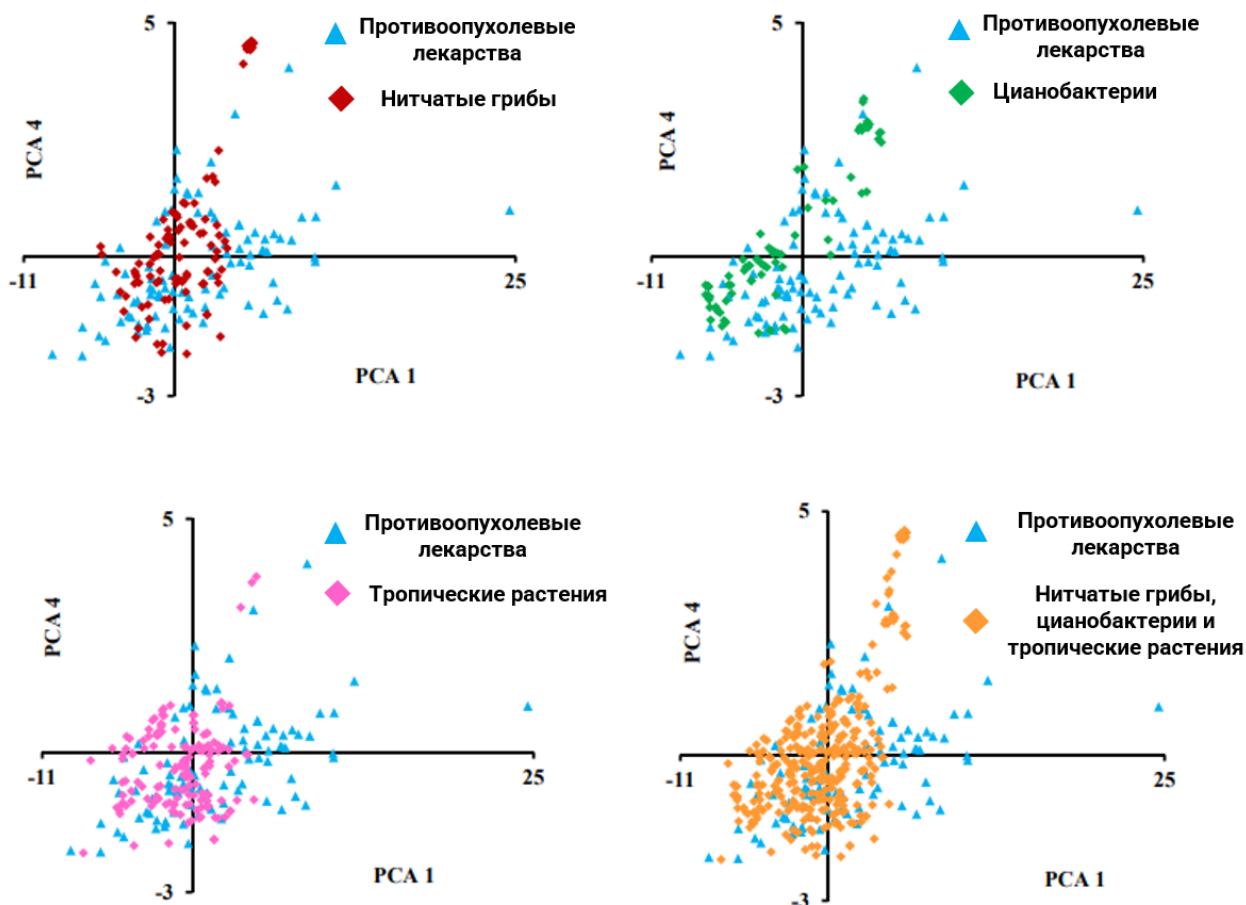


Рисунок 4 – Визуализация химического пространства ПС, выделенных из нитчатых грибов, цианобактерий, тропических растений. Адаптировано из [36]

Как видно из представленных на рисунке 4 данных, на каждом из четырех графиков наблюдается частичное перекрывание областей химического пространства представителями 4-х различных источников, а также наблюдаются области, заполненные только одним из представителей. Таким образом, основываясь на близости химических соединений в пространстве физико-химических параметров, авторы утверждают, что используемые ПС, близкие к

противоопухолевым лекарствам, могут оказаться перспективными для дальнейшего исследования в качестве противоопухолевых веществ.

Исследование со схожей методикой, но с использованием более объемных наборов данных, было проведено Чен и соавт. [1]. В рамках данного исследования использовались структуры ПС, полученные из различных источников, включая базы данных традиционной китайской медицины (TCMID [37], HIM [38], HIT [39] и др.), БД ПС отдельных географических регионов (AfroDB [40], NPCARE [41] и др.), а также крупные репозитории данных о ПС (DNP [42], UNDP [43]). Кроме того, авторами использовались данные, полученные из коммерческих библиотек ПС (Ambinter and Greenpharma NPs [44], InterBioScreen NPs [45]). В качестве синтетических объектов для сравнения использовались выборка по одобренным лекарствам из работы Ньюмен и соавт. [8], а также данные из DrugBank [33]. Ключевой особенностью данной работы является анализ химического пространства ПС, которые в настоящее время доступны для использования. Авторами показано, что доступные для использования ПС охватывают гораздо более широкое химическое пространство, чем разрешенные к применению лекарственные препараты, и при этом значительное число ПС расположено в тех областях химического пространства, которые также заполнены лекарственными препаратами. Кроме того, авторы показали, что доступные для использования ПС, охватывают все основные классы природных соединений.

1.3 Разработка лекарственных субстанций на основе природных соединений

Химическое разнообразие синтетических структур потенциально бесконечно и ограничивается возможностями методов органического синтеза, доступных в настоящее время. В свою очередь, химическое разнообразие ПС ограничено метаболомом организмов-продуцентов. Исключением могут являться соединения, продуцируемые микроорганизмами. Это связано с тем, что гены, ответственные за синтез вторичных метаболитов в геноме микроорганизмов, зачастую объединены в

кластеры, что, в свою очередь, открывает широкие возможности для изменения структуры вторичных метаболитов [46].

В сравнении с химическими соединениями, которые в настоящее время доступны для синтеза, ПС обладают несколькими особенностями, которые могут как позитивно, так и негативно влиять на процесс разработки на их основе новых лекарств. Во-первых, ПС могут характеризоваться опытом применения в рамках практик традиционной медицины различных стран, что позволяет получить представления об их эффективности и безопасности [4]. Во-вторых, разнообразие структур ПС охватывает большую область химического пространства и, как следствие они проявляют более широкий набор биологических активностей в сравнении с соединениями, доступными для получения методами органического синтеза [47].

Дополнительно, структуры ПС обладают рядом отличий от синтетических структур, в числе которых большая молекулярная масса (MW), преобладание sp^3 -гибридизованных атомов, большее число хиральных центров, содержание большего числа атомов кислорода и меньшего числа атомов азота и серы [1]. Однако эти данные получены при сравнении комбинаторных библиотек синтетических соединений и некоммерческих баз данных ПС (например, [48]). Существенным недостатком данного подхода является отсутствие достаточной теоретической базы для дальнейшей разработки лекарств на основе таких соединений. При попытке внедрения этих соединений в практику исследователи могут столкнуться с рядом проблем, включая:

- многостадийный и сложный синтез в промышленных масштабах, что делает эти соединения экономически не рентабельными для большинства фармацевтических компаний;
- низкие показатели целевой биологической активности;
- низкое содержание целевого ПС в его природном источнике.

Принимая во внимание указанные недостатки, полезным может оказаться сравнение уже одобренных к применению лекарств, которые были разработаны на основе ПС, с соединениями, которые в настоящее время могут быть доступны для

синтеза. Такое сравнение было проведено в работе [49]. Авторами рассматривались физико-химические показатели, используемые в рамках «правил подобия» перорально доступным лекарственным препаратам (англ. drug-likeness rules) – «Правило 5 (Липински)» и «Правило 2 (Вебера)» [49]. Физико-химические параметры, а также характеристики используемых правил приведены в таблице 2.

Таблица 2 – Характеристики теоретических правил для отбора перорально доступных лекарств

Название правила	Диапазоны значений физико-химических параметров	Теоретическое обоснование	Ссылка
Правило Липински (англ. «Rule of 5»)	HBA<5 HBD<10 AlogP<5 MW<500	Основано на статистических данных о большинстве перорально доступных лекарствах [50]	[29]
Правило Вебера (англ. «Rule of 2»)	PSA<140Å RTB (количество связей, допускающих вращение)>10	Выведено на основе данных, полученных при оценке концентрации введенных пероральных препаратов в крови у крыс [51]	[51]

Как показано авторами работы [52], препараты, разработанные на основе ПС (соответствующие категориям 1 и 3 из таблицы 1), в сравнении с препаратами, разработанными на основе соединений, полученных методом органического синтеза (соответствующие категории 4 из таблицы 1), имеют более высокие значения некоторых физико-химических характеристик, представленных в таблице 2 (HBA, HBD, MW, PSA и RTB) и более низкие значения AlogP. Кроме того, для препаратов, разработанных на основе ПС, характерно большее количество атомов кислорода и стереоцентров.

В процессе разработки лекарств на основе ПС, выделяют ряд сложностей, косвенно или напрямую связанных с описанными особенностями. Условно их можно разделить на 2 группы.

1. Проблемы, связанные с источниками получения ПС.

2. Проблемы тестирования веществ природного происхождения с целью выявления их биологической активности.

1.3.1 Подходы и проблемы, связанные с источниками химических соединений

Традиционно основными источниками структур природного происхождения являлись растения, грибы и бактерии. Одной из ключевых проблем, которая может существенно повлиять на стратегию разработки лекарственной субстанции является доступность сырья для производства, особенно в случаях, когда целевая структура не может быть получена экономически целесообразными методами химического синтеза. Причем это может быть связано как с недостаточностью сырьевой базы, так и низким содержанием целевого соединения. Один из наиболее известных примеров связан с противоопухолевым препаратом таксол (торговое название Паклитаксел) [53]. Данное соединение относится к классу таксанов, а его основным природным источником является кора тиса тихоокеанского (лат. *Taxus brevifolia*) [54]. Впервые паклитаксел был одобрен для терапии рака яичников. Ситуация, которая в научной литературе получила название «Кризис поставок таксола», возникла в связи с возросшими клиническими потребностями в данном препарате, обусловленными его успешным репозиционированием на другие типы опухолей. Производство таксола из его природного источника оказалось длительным и трудоемким процессом, а сам процесс выделения и очистки таксола из растительного источника для получения индивидуального соединения мог занимать до двух лет. В связи с этим были инициированы исследования по выявлению новых источников, а также по разработке методик синтеза [55]. В результате таких исследований было обнаружено множество новых растений-источников, включая *Taxus baccata* [56], *Taxus cuspidata* [57], а также разработан подход к синтезу таксола на основе его соединения-предшественника 10-дезацетилбаккатина [58].

К другим классам проблем, связанных с природными источниками фармакологических веществ, относят сложности, связанные с культивированием.

Часто, организм-продуцент, извлечённый из естественной среды обитания, перестаёт продуцировать целевое соединение, а в некоторых случаях культивация организма-продуцента вне естественной среды обитания может оказаться невозможной [5].

Еще одна проблема носит экологический характер и, вероятно, является следствием из описанных ранее. По разным оценкам, темпы вымирания растительных видов превышают естественные фоновые показатели в 100-1000 раз, и в настоящее время более 15 000 видов лекарственных растений находятся под угрозой исчезновения [59].

Помимо описанных выше общих проблем существует ряд сложностей, которые могут возникнуть в зависимости от подхода, выбранного для определения потенциального источника для выявления ПС, проявляющих необходимую биологическую активность. В литературе выделяют пять подходов [60]:

- подход, основанный на скрининге библиотек экстрактов, сформированных случайным образом (случайный скрининг);
- подход, основанный на знаниях традиционной медицины;
- хемотаксономический подход;
- экологический подход;
- компьютерный (вычислительный) подход.

1.3.2 Случайный скрининг библиотек экстрактов

Основным критерием при формировании библиотек экстрактов и отдельных соединений является доступность сырья [61]. Преимущества такого подхода проявляются при использовании образцов из регионов с высоким биоразнообразием, что, в свою очередь, обеспечивает широкое химическое разнообразие структур ПС. Применение данного подхода повышает шанс обнаружения ранее неизвестных видов биологической активности [62]. В качестве главных ограничений данного подхода выделяют его высокую стоимость, низкую вероятность обнаружения требуемой биологической активности, а также

временные издержки, связанные с недостаточно высокой производительностью аналитических методов [63].

1.3.3 Подход, основанный на знаниях традиционной медицины

С целью минимизации недостатков случайного скрининга зачастую применяются подходы, основанные на известных данных о организме-продуcente, в числе которых подход, построенный на опыте применения растений и их смесей в рамках традиционной медицины различных стран.

Несмотря на высокий шанс обнаружения на этой основе химических соединений с требуемой биологической активностью, применение таких подходов также сопряжено с рядом сложностей [60]. Во-первых, растение, выбранное на основе данных традиционной медицины, должно быть охарактеризовано с точки зрения среды его обитания и доступности (достаточных размеров сырьевой базы). Кроме того, необходимо знание точной ботанической классификации, а также сведений о том, отнесено ли растение к подверженным вымиранию видам [64]. Во-вторых, большинство фармакологических эффектов фармацевтических композиций связано с синергетическим/аддитивным действием комбинации компонентов входящих в их состав [65].

Тем не менее, существует множество успешных примеров разработки лекарств на основе данных традиционной медицины (см. рисунок 5).

На рисунке 5 приведены примеры структур лекарственных субстанций, которые были разработаны на основе лекарственных растений, используемых в рамках традиционной медицины различных стран. Среди них папаверин – алкалоид растения *Papaver somniferum* L., обладающий гипотензивным и спазмолитическим эффектом [66]; хеллин – мажорный фитокомпонент *Ammi visnaga* (L.) Lam., применяемый для лечения бронхиальной астмы [67]; галегин – алкалоид *Galega officinalis* L. применяемый как антидиабетический препарат [68] и хинин – алкалоид коры хинного дерева, применяемый для лечения малярии [69].

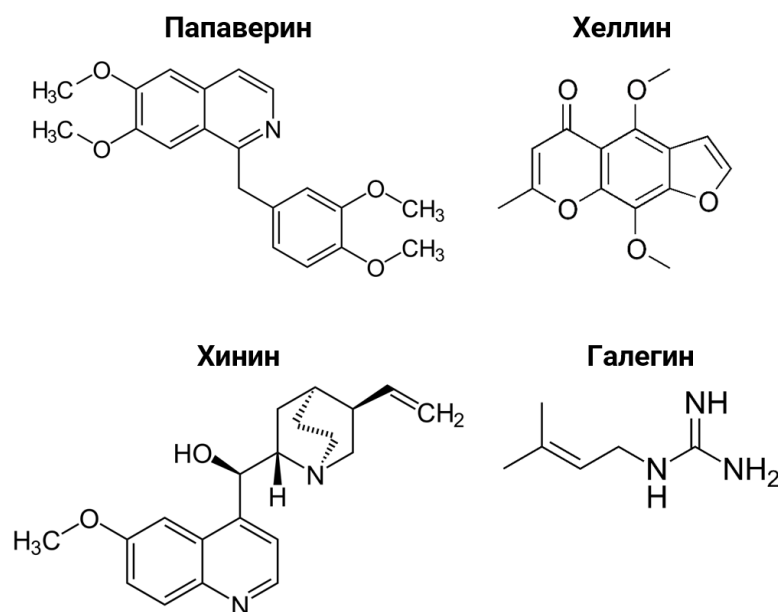


Рисунок 5 – Структуры лекарственных субстанций, разработанных на основе опыта традиционной медицины

1.3.4 Экологический подход

В основе экологического подхода лежат сведения, полученные в результате наблюдения за взаимодействием организма-продуцента с окружающей средой. Например, соединения, синтезируемые растениями для защиты от микробных патогенов, потенциально могут стать отправными точками для разработки противомикробных препаратов, а соединения, синтезируемые с целью обеспечения защитной функции от травоядных через нейротоксическую активность, могут послужить отправной точкой для разработки седативных препаратов [70].

1.3.5 Хемотаксономический подход

Для расширения диапазона возможных химических структур, проявляющих целевую активность, используется хемотаксономический подход. Данный подход основан на изучении химических групп, присутствующих в растениях, с учетом их таксономических характеристик [71]. Основным недостатком данного подхода является то, что фитохимический состав родственных растений не является решающим в разработке лекарственных препаратов и, зачастую, не позволяет судить об их безопасности и эффективности [72].

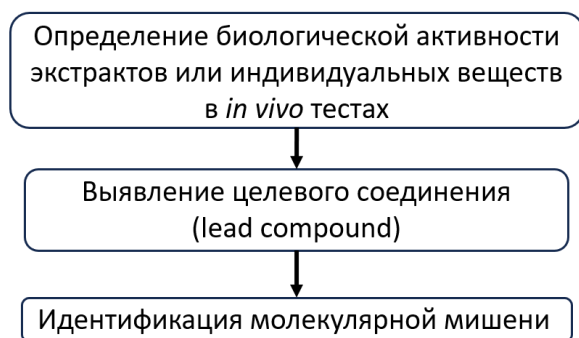
1.3.6 Подходы к тестированию веществ природного происхождения с целью выявления их биологической активности

В настоящее время существует два подхода к исследованиям веществ природного происхождения:

- «прямая» фармакология;
- «обратная» фармакология.

Данные подходы схематично представлены на рисунке 6.

Прямой фармакологический подход



Обратный фармакологический подход



Рисунок 6 – Схематическая иллюстрация прямого и обратного фармакологического подходов. Адаптировано из [60]

Изначально исследование экстрактов и индивидуальных веществ природного происхождения проводилось только с использованием прямого фармакологического подхода. Множество фракций, выделенных из природного источника, тестировалось на модельных животных, тканях и клетках с последующей идентификацией индивидуального соединения и дальнейшим установлением механизмов действия [73]. Однако, применение данного подхода к экстрактам осложняется риском обнаружения дубликатов – уже известных соединений, которые невозможно запатентовать [74]. Поэтому с целью преодоления этой проблемы, в качестве подхода к анализу растительных экстрактов и индивидуальных соединений природного происхождения часто применяется обратный фармакологический подход. На первом этапе проводится исследование

взаимодействия ПС из химических библиотек с предварительно охарактеризованными молекулярными мишенями. В результате такого тестирования происходит выявление «хитов», которые затем изучаются в тестах *in vivo* с целью подтверждения их биологической активности [75]. В рамках каждого из подходов применяются одни и те же методы тестирования, основные различия заключаются в порядке их использования. Основные метода анализа биологической активности приведены в таблице 3.

Таблица 3 – Основные методы тестирования ПС с целью выявления их биологической активности, а также их ограничения

Тип анализа	Достоинства	Ограничения
Биохимический анализ <i>in vitro</i> с использованием рекомбинантных белков [76]	Высокая производительность анализа. Сравнительная простота (отсутствие клеточных линий и модельных животных)	Высокий риск ложноположительных результатов [76]
Анализ с использованием клеточных тест-систем [76]	Средняя производительность эксперимента, заранее известна целевая мишень	Трудоемкий процесс, связанный с использованием клеточных культур.
Анализы <i>in situ</i> / <i>ex vivo</i> с изолированными тканями или органами [78, 79]	Высокая патофизиологическая релевантность; некоторые из приложений позволяют сократить количество используемых животных и обеспечивают более высокую пропускную способность по сравнению с моделями на грызунах [78, 79]	Низкая производительность экспериментов, этические проблемы
Модели с использованием лабораторных животных (грызунов)	Демонстрация биологической активности на уровне целого организма	Низкая производительность экспериментов, этические проблемы, видовые различия

1.3.7 Компьютерный подход

Один из современных подходов основан на применении компьютерных методов для виртуального скрининга. Методы компьютерного конструирования лекарств базируются на данных о пространственной структуре биологической мишени или на данных о структуре лигандов, взаимодействующих с биологическими мишенями [80].

К наиболее часто используемым методам, основанным на данных о пространственной структуре биологической мишени, относится молекулярный докинг. Он позволяет предсказать наиболее выгодное расположение лиганда в комплексе с мишенью, и получать количественные оценки энергии связывания (scoring function) в активном центре белка-мишени [81].

Для проведения виртуального скрининга, основанного на знании о структуре лигандов, применяется несколько групп методов:

- методы, основанные на поиске по молекулярному подобию;
- методы, основанные на построении моделей зависимостей «структура-активность» ((Q)SAR ((Quantitative) Structure-Activity Relationship)).

К первым работам по прогнозированию биологической активности соединений относят работы Хэнча 1963 года, в которых была представлена эмпирическая модель связи «структура-активность», основанная на линейной зависимости свободной энергии исследуемого процесса от физико-химических параметров соединения, рассматриваемых как независимые переменные [82]. Важную роль в развитии методов компьютерного конструирования лекарств сыграли работы Авидона и соавт. 70–х годов [83]. В них с использованием массива данных, содержащего 7200 соединений, были исследованы возможности прогнозирования биологической активности на основе подструктурного подхода и анализа сходства, при этом сходство рассчитывалось по близости двух структур соединений в пространстве дескрипторов ФСКП. По мере развития компьютерных методов стало возможным проводить анализ большого количества переменных, характеризующих особенности структуры и её физико-химических свойств, а также выявлять факторы, определяющие наличие у соединения целевой биологической активности [84]. Кроме того, значительный вклад в развитие (Q)SAR методов внесли работы Фри и Вильсона, описывающие одноименную модель (аддитивная модель Фри-Вильсона). Согласно методу Фри-Вильсона, биологическая активность определяется как сумма вкладов, вносимых каждым заместителем в каждом положении, без учёта физико-химических свойств молекулы [85].

(Q)SAR модели. (Q)SAR модели описывают математическую зависимость между структурой молекулы и её биологическими свойствами [86]. Эти модели нашли широкое применение в области медицинской химии, где они используются для прогнозирования биологической активности веществ. Это связано с тем, что различные виды биологической активности молекулы напрямую зависят от его химической структуры.

Для построения (Q)SAR моделей необходимы четыре ключевых компонента [87]:

1. данные о химической структуре и биологической активности соединений (обучающая выборка);
2. машиночитаемые представления химических структур;
3. оценки биологической активности (качественные, количественные);
4. методы машинного обучения, необходимые для установления зависимости между структурой и биологической активностью.

Общие этапы построения моделей можно представить следующим образом:

- подготовка данных обучающей выборки, их нормализация и стандартизация;
- расчет дескрипторов, которыми, в частности, могут служить различные характеристики химической структуры молекул (в том числе квантово-химические параметры), структурные фрагменты, физико-химические свойства (гидрофильность, липофильность, растворимость) и др.;
- анализ зависимостей с применением методов машинного обучения, таких как множественная линейная регрессия, нейронные сети и др.;
- оценка характеристик качества модели (R^2 , RMSE, и др.).

Для проведения валидации (Q)SAR моделей в 2004 году были утверждены основные принципы организации экономического сотрудничества и развития (OECD) [88]. К ним относятся:

- четко определенная цель создания (Q)SAR модели;
- «прозрачность» алгоритма построения зависимости «структура-активность»;

- определенная область применимости. Четко установлен тип и свойства химических структур, для которых модель позволяет получать надежные оценки;
- надежные критерии точности и предсказательной способности модели;
- интерпретация механизмов действия (соответствие реальным механизмам прогнозируемых взаимодействий). Данный пункт является желательным, но не обязательным.

В зависимости от данных, которые используются в исследовании, методы (Q)SAR можно разделить на две группы. В том случае, когда влияние стереохимических факторов на исследуемое свойство сведено к минимуму или отсутствует вовсе, применяют методы 2D (Q)SAR, в которых используется структурная формула, которая представляется как молекулярный граф [89]. Для описания хиральных структур используют методы 3D (Q)SAR. При таком подходе пространственная структура молекулы учитывается в явном виде [90].

Для некоторых классов соединений различные стереоизомеры, неразличимые в 2D пространстве, обладают разной биологической активностью. С целью решения данной проблемы, в работе [91], чтобы предсказать возможную биологическую активность стероидов, для которых характерно наличие нескольких хиральных центров, применено 2,5D-симплексное представление молекулярной структуры. Данный способ описания молекулярной структуры реализуется следующим образом. При описании хиральной структуры для каждого асимметрического атома углерода и четырёх атомов его ближайшего окружения рассчитывается набор связанных 3D-симплексов; затем оценивается стереохимическая конфигурация каждого из хиральных симплексов. Подобный подход дает возможность оценить непосредственно вклад хирального фрагмента, как обуславливающего различие в биологической активности. Полученные наборы дескрипторов используются вместе с 2D-дескрипторами, отсюда и название – «2,5D». Однако, несмотря на полученные авторами обнадеживающие результаты, расширение и апробация предложенного метода для других видов биологической активности осложняется ограниченностью информации о биологической

активности различных стереоизомеров, что не позволяет создать необходимые обучающие выборки.

1.4 Репозитории структур природных соединений

Как видно из представленной выше информации, методы *in silico* широко применяются для оценки фармакологического потенциала и механизмов действия. Наряду с этим, виртуальный скрининг может быть применен для выявления химических соединений, ответственных за проявление биологической активности в составе экстрактов или других смесей соединений природного происхождения.

В настоящее время наблюдается рост публикаций, описывающих репозитории (базы данных) структур природного происхождения, которые могут быть использованы для изучения потенциальных биологических активностей ПС методами виртуального скрининга. Как сообщается в обзоре [92], в период с 2000 по 2020 годы было опубликовано около 120 работ, посвященных созданию баз данных, содержащих структуры веществ природного происхождения. Рост количества новых баз данных в области ПС продолжается до сих пор. Так, например, с начала 2020 года было опубликовано 9 статей, содержащих описание разработанных авторами новых баз данных [93–101].

Все существующие базы данных условно можно разделить на несколько классов:

- крупные репозитории данных о структурах и свойствах химических соединений;
- базы данных для дерепликации (определение термина приведено ниже) ПС;
- базы данных, характеризующие химическое разнообразие отдельных географических регионов;
- базы данных, характеризующие отдельные таксономические группы.

Крупные репозитории данных о структуре и свойствах химических соединений. К основным репозиториям, в которых содержится информация о структуре и данные о результатах лабораторного тестирования, которые постоянно

поддерживаются разработчиками, в настоящее время относятся такие базы данных, как ChEMBL [102], ChEBI [103], COCONUT [104], LOTUS [105], Super Natural 3.0 [106]. Краткая характеристика этих баз данных приведена в таблице 4.

Таблица 4 – Характеристики наиболее популярных репозиториев структур природного происхождения

Название репозитория	Число структур природного происхождения	Характеристика	Дата последнего обновления
ChEMBL (актуальная версия – 34)	Приблизительно 2000 структур	База данных содержит сведения о химических структурах, результатах тестирования химических соединений на биологическую активность, биологических мишенях и лекарствах.	Март 2024
ChEBI (актуальная версия – 226)	>55 000	База содержит записи о более чем 55 000 аннотированных вручную, на основе общедоступной литературы, химических соединений, в число которых входят как природные, так и синтетические молекулы. Объекты, содержащиеся в базе данных, классифицированы в соответствии с их видовой принадлежностью и структурными особенностями.	Октябрь 2023
COCONUT (актуальная версия – 1)	406 076	Набор структур химических соединений природного происхождения, собранный из множества других баз данных.	Январь 2022
LOTUS (актуальная версия - 1)	276 518	Набор структур из COCONUT, которые были проаннотированы с точки зрения их представленности в природных источниках.	Август 2023
Super Natural (актуальная версия – 3)	449 058	База данных содержит структуры ПС и их производных, а также сведения о взаимодействии	Октябрь 2023

Название репозитория	Число структур природного происхождения	Характеристика	Дата последнего обновления
		с молекулярными мишенями человека, токсичности и информацию о поставщиках.	

Кроме перечисленных в таблице 4 баз данных, известны и другие крупные репозитории химических структур природного происхождения. К ним относятся, DNP (>153 000 структур ПС) и UNDP (>225 000 структур ПС) [92]. Из них в настоящее время DNP является коммерчески-доступной базой данных, в том числе и для академического использования, а набор данных UNDP недоступен, как и сам веб-ресурс [107].

Базы данных для дерепликации ПС. Дерепликация – процесс распознавания и исключения из набора анализируемых экстрактов те экстракты, которые содержат известные биологически активные соединения. Является одним из важнейших этапов при разработке лекарств на основе природных источников, поскольку позволяет избежать повторного выделения и многократной характеристики уже известных молекул [5]. По типу содержащихся сведений выделяют два типа баз данных:

- базы данных, содержащие аннотированные результаты масс-спектрометрии (МС);
- базы данных, содержащие аннотированные результаты спектров ядерного магнитного резонанса (ЯМР).

Базы данных, содержащие аннотированные результаты МС. В современной научной литературе представлено 7 баз данных, содержащих указанные сведения:

- MassBank of North America (MoNA) [108];
- European MassBank [109];
- Japanese MSSJ MassBank [110];
- METLIN [111];
- Human Metabolome Database (HMDB) [112];

- Yeast Metabolome Database (YMDB) [113];
- Global Natural Products Social Molecular Networking (GNPS) [114].

Приведенные базы данных содержат эталонные и экспериментальные масс-спектры метаболитов, а также обширные метаданные.

Базы данных, содержащие ЯМР-спектры. В данную категорию входит 3 базы данных:

- NMRshiftDB [115];
- NMRdata [116];
- NAPROC-13 [117].

Представленные базы данных содержат экспериментальные ЯМР-спектры химических соединений, а веб-интерфейс таких баз данных позволяет проводить поисковые процедуры для дерепликации ПС.

Базы данных, характеризующие химическое разнообразие отдельных географических регионов. Базы данных, содержащие сведения о химическом разнообразии отдельных географических регионов являются самой объемной группой информационных ресурсов. Примеры таких баз данных приведены на рисунке 7.

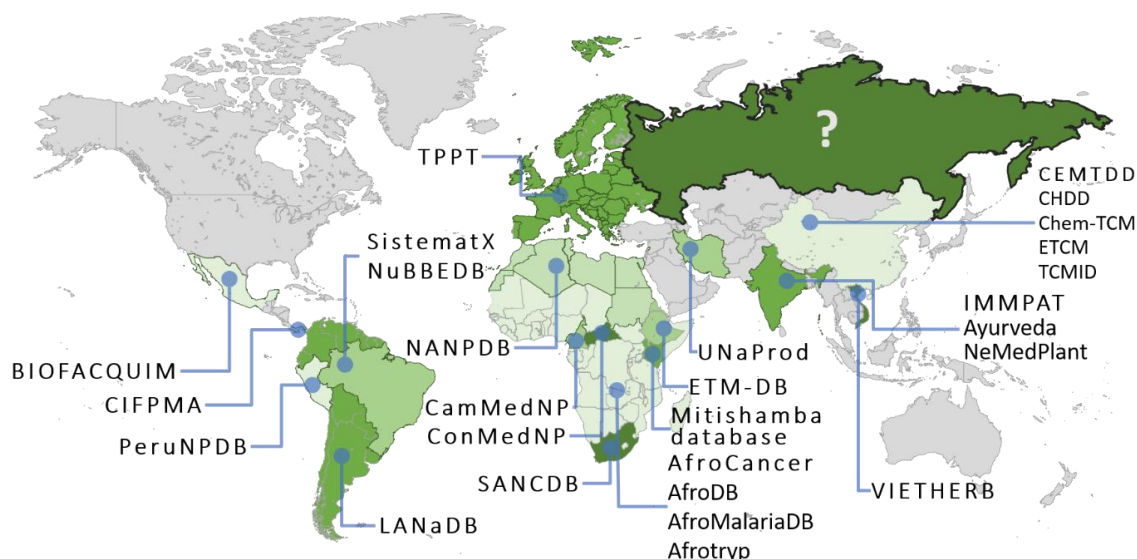


Рисунок 7 – Примеры баз данных и соответствующих им географических регионов

Как видно из рисунка 7, к информационным источникам, характеризующим отдельные географические регионы, относится более 28 баз данных. Указанные базы данных содержат сведения о химическом разнообразии биологических источников (преимущественно растений) Северной и Южной Америки, Африки, Европы и Азии. Часть из этих баз данных позиционируются как базы данных растений традиционной медицины отдельных регионов. Например, Ayurveda – база данных фитокомпонентов традиционной индийской медицины [118] или TCMID – база данных фитокомпонентов традиционной китайской медицины [37]. Важно заметить, что не для всех приведенных на рисунке 7 баз данных информация была опубликована в рецензируемых научных журналах. Характеристика опубликованных баз данных приведена в таблице 5.

Таблица 5 – Характеристика баз данных характеризующих химическое разнообразие отдельных географических регионов

Название БД [Ссылка]	Географический регион (страна)	Количество структур	Представленные данные	Позиционирование
BIOFACQUIM [119]	Северная Америка (Мексика)	423	Структурные формулы, биологические источники, географическое положение, биологическая активность	Химическое разнообразие региона
PeruNPDB [99]	Южная Америка (Перу)	280	Структурные формулы, биологические источники, классификация ПС	Химическое разнообразие региона
CIFPMA [120]	Центральная Америка (Панама)	354	Структурные формулы	Химическое разнообразие региона
SistematX [121]	Южная Америка (Бразилия)	>8 000	Структурные формулы, биологические источники, географическое положение, биологическая активность, ЯМР- спектры	Химическое разнообразие региона

Название БД [Ссылка]	Географический регион (страна)	Количество структур	Представленные данные	Позиционирование
NuBBEDB [48]	Южная Америка (Бразилия)	2215	Структурные формулы, биологические источники, географическое положение, биологическая активность, ЯМР- спектры	Химическое разнообразие региона
TPPT [122]	Центральная Европа	1583	Структурные формулы, токсичность	Химическое разнообразие региона
NANPDB [123]	Северная Африка	6832	Структурные формулы, биологические источники	Химическое разнообразие региона
CamMedNP [124]	Центральная Африка (Камерун)	>2500	Структурные формулы, биологические источники	Химическое разнообразие региона
ConMedNP [125]	Центрально- африканская Республика	>3200	Структурные формулы, биологические источники	Химическое разнообразие региона
AfroDB [40]	Африка	954	Структурные формулы, биологические источники	Химическое разнообразие региона
UNaProd [94]	Иран	3413	-	Традиционная медицина
ETM-DB	Африка (Эфиопия)	4285	Структурные формулы, биологические источники	Традиционная медицина
Mitishamba database [126]	Африка (Кения)	1102	Структурные формулы	Химическое разнообразие региона
IMPPAT [127]	Индия	9596	Структурные формулы	Традиционная медицина
Ayurveda [118]	Индия	950	Структурные формулы	Традиционная медицина
VIETHERB [128]	Вьетнам	>10 000	Структурные формулы, биологические источники, терапевтические эффекты	Традиционная медицина

Название БД [Ссылка]	Географический регион (страна)	Количество структур	Представленные данные	Позиционирование
TCMID [37]	Китай	12 549	Структурные формулы, биологические источники, биологические активности	Традиционная медицина

Большинство баз данных, включая географические базы данных и базы данных традиционной медицины, пересекаются по содержащейся в них информации. Кроме того, показано, что представленные сведения содержат большое количество ошибок, включая некорректные структуры химических соединений и неточные ссылки на оригинальные источники, что требует проведения дополнительной процедуры ручной аннотации для их практического использования [92].

Базы данных, характеризующие отдельные таксономические группы. В рамках данной категории баз данных зачастую выделяют 2 подкатегории:

- ПС, синтезируемые растениями;
- ПС, синтезируемые микроорганизмами.

В число баз данных, относящихся к первой подкатегории, входит большинство баз данных ПС. Ко второй подкатегории относятся следующие базы данных:

- StreptomeDB – коллекция ПС синтезируемых бактериями рода *Streptomyces*, которые играют важную в разработке антибиотиков и противоопухолевых лекарств [129];
- ProCarDB – база данных каротиноидов, синтезируемых бактериями [130];
- PAMDB – база данных, содержащая сведения о метаболоме бактерий рода *Pseudomonas* [131].

1.5 Заключение

Территория Российской Федерации охватывает несколько климатических зон, что объясняет высокое биоразнообразие эндемичных растений. Высокое

биоразнообразие растений ведет к разнообразию химических структур синтезируемых ими вторичных метаболитов. Это делает вторичные метаболиты (далее, фитокомпоненты) этих растений перспективной отправной точкой для поиска и разработки новых фармакологических веществ. Особую роль играют растения, используемые в традиционной медицине, для получения растительного сырья которых разработаны и апробированы методы анализа и стандартизации. Перечень таких растений приведен в Государственной Фармакопее Российской Федерации [132]. Кроме того, перечень биологически активных веществ растительного происхождения, которые в том числе могут быть обнаружены в растениях России, описан в трёх томах монографии [133]. Однако эти данные представлены только в печатных изданиях, что осложняет их компьютерный анализ.

Применение методов виртуального скрининга для оценки потенциальной биологической активности фитокомпонентов растений может быть ограничено недостатком данных об их фитохимическом составе и отсутствием данных о структуре фитокомпонентов, представленных в надлежащем для *in silico* анализа виде. Несмотря на обилие зарубежных баз данных, содержащих информацию о структурах и свойствах ПС, их использование для анализа фитохимического состава российских лекарственных растений может быть ограничено несколькими факторами. Во-первых, не во всех базах данных упоминается природное происхождение соединений со ссылкой на тематические публикации. Во-вторых, наличие видов растений, уникальных для флоры России, ограничивает использование баз данных, характеризующих фитокомпонентный состав растений, произрастающих в других странах. В большинстве существующих баз данных записи описывают содержание фитокомпонентов во всем растении без указания его конкретной части (частей), что может ввести исследователей в заблуждение, поскольку в лабораторных экспериментах, как правило, изучаются конкретные части отдельных растений (листья, корни и т.д.). Кроме того, базы данных часто содержат ошибки и неточности, что обуславливает необходимость

предварительной ручной аннотации (по-сути, перепроверку всей представленной в конкретной базе данных информации).

Учитывая вышеперечисленные недостатки, создание базы данных и разработка на её основе информационно-вычислительной платформы, предоставляющей пользователям сведения о структурах фитокомпонентов лекарственных растений России и известную информацию об их биологической активности, является актуальной задачей. Наличие в составе ИВП инструментов для оценки различных видов биологической активности *in silico*, а также поиск по химическому сходству, позволяют расширить возможности использования скрытого фармакологического потенциала, доступного на территории РФ растительного сырья, благодаря выявлению наиболее перспективных направлений экспериментальных исследований отдельных химических соединений и фармацевтических композиций.

ГЛАВА 2. МАТЕРИАЛЫ И МЕТОДЫ

2.1 Процедура извлечения данных

2.1.1 Подготовка набора видовых названий

Набор видовых названий растений, используемых на территории России в качестве растительного сырья для производства лекарственных средств, был нами составлен на основе Государственной Фармакопеи Российской Федерации (14-го издания) [134]. Для каждого видового названия был произведен поиск синонимов. Формирование первичного массива синонимов было реализовано с использованием веб-ресурса The Plant List [135]. Отбор синонимов проводился вручную на основе результатов оценок достоверности видовых названий (англ. Confidence level) синонимов. В первичный массив синонимов видовых названий вошли синонимы, которые по результатам аннотации данных ресурса The Plant List были отмечены максимальной оценкой уровня достоверности. Данные этого ресурса аннотируются вручную исследователями, однако, несмотря на это, возможно появление видовых названий, синонимы которых могут быть отнесены к двум и более растениям. При подготовке первичного массива синонимов видовых названий подобные записи удалялись. Дальнейший анализ видовых названий и их синонимов проводили с учетом их представленности в текстах тематических научных публикаций.

На заключительном этапе подготовки набора видовых названий растений из базы данных NCBI Taxonomy были извлечены таксономические характеристики растений [136]. Таксономические характеристики включали в себя названия семейств и подцарств.

2.1.2 Подготовка набора структурных формул фитокомпонентов и их характеристик

Данные о структуре фитокомпонентов, представленных в фитохимическом составе частей растений, были подготовлены на основе анализа текстов

тематических научных публикаций и свободно доступных баз данных структур ПС. Доступ к текстам научным публикаций, а также поиск информации о существующих базах данных производился с использованием информационных систем Google Scholar [137] и PubMed [138].

Данные о структуре фитокомпонентов были представлены в следующих форматах:

- Канонические SMILES [139];
- InChi [140];
- InChiKey (хешированная версия линейного представления InChi) [141];
- MOL (V2000) [142].

Для каждого фитокомпонента был получен набор значений его молекулярных характеристик. Соответствующие параметры были извлечены из крупных репозиториях химических структур (PubChem [143], ChEMBL [102]). Важно заметить, что характеристики, представленные в указанных БД, не охватывают все молекулярные свойства, используемые в качестве критериев для отбора потенциальных «кандидатов» с целью создания лекарственных средств. Поэтому для подготовки данных о значениях молекулярных характеристик отобранных фитокомпонентов был также использован свободно доступный в сети Интернет веб-сервис ADMETlab [144]. Перечень отобранных молекулярных характеристик представлен в таблице 6.

Таблица 6 – Молекулярные характеристики и источники информации

Молекулярное свойство	База данных/веб-сервис
MW	PubChem
AlogP	ChEMBL
HBA	ChEMBL
HBD	ChEMBL
RTB	ChEMBL
Топологическая полярная площадь поверхности (tPSA)	PubChem

Молекулярное свойство	База данных/веб-сервис
Число ароматических колец	ChEMBL
Отношение количества sp ³ -гибризованных атомов углерода к общему числу атомов углерода (Fsp ³)	ADMETlab 2.0
Число гетероатомов	ADMETlab 2.0
Число циклов	ADMETlab 2.0

Выбор этих молекулярных параметров обусловлен использованием их комбинаций в молекулярных фильтрах для отбора потенциальных лекарственных-подобных соединений, широко применяемых в медицинской химии [145].

Результаты экспериментального тестирования взаимодействия химических соединений с молекулярными мишенями в организме человека были извлечены из баз данных PubChem и ChEMBL, а также из текстов научных публикаций.

2.2 Методы работы с базами данных

2.2.1 База данных PubChem

База данных PubChem – информационный ресурс Национального центра биотехнологической информации США (NCBI), предоставляющий широкому кругу исследователей сведения о химических соединениях, включая их химические структуры в различных форматах, молекулярные свойства, а также результаты тестирования биологической активности. Кроме того, база данных содержит ссылки на внешние источники, такие как DrugBank, ChEBI и LOTUS.

Извлечение информации из базы данных PubChem осуществлялось с использованием программного интерфейса для доступа к базе данных – PUG REST [146]. PUG REST позволяет проводить выгрузку данных из базы данных PubChem путем конструирования специализированных URL адресов по заданным шаблонам.

URL-адрес для загрузки данных имел следующий вид: «[https://pubchem.ncbi.nlm.nih.gov/rest/pug/compound/name/{Название химического соединения}/SDF](https://pubchem.ncbi.nlm.nih.gov/rest/pug/compound/name/{Название_химического_соединения}/SDF)». На данном этапе в качестве названия химического соединения использовалось название, извлеченное из текста научной публикации или базы данных. Указанный URL-адрес инициирует загрузку сведений о химическом соединении, представленных в SD-файле. При выполнении таких запросов поиск проводится не только по тривиальным названиям соединений, но и по представленным в PubChem синонимам. Поэтому при совпадении названий у разных соединений запрос вернет SD-файл, содержащий информацию о всех обнаруженных соединениях с таким названием. В таких случаях проводилось сопоставление загруженных данных и данных в первоначальном источнике и использовались только взаимно согласованные данные.

Используя данную процедуру, нами были сгенерированы запросы для каждого названия фитокомпонента и осуществлена загрузка данных. На следующем этапе, из загруженных SD-файлов были извлечены тривиальные названия и идентификаторы PubChem, которые впоследствии были вручную проаннотированы на предмет соответствия химическим соединениям из первичного источника. В случаях, когда с использованием запроса не удавалось обнаружить необходимые сведения, поиск проводили вручную с использованием веб-интерфейса PubChem.

Для предоставления расширенной информации о химическом соединении, а также для дополнительной валидации соответствующих сведений пользователями, нами были добавлены ссылки на крупные внешние репозитории химических структур (ChEMBL, ChEBI и PubChem).

Для подготовки массива внешних ссылок на базы данных ChEBI и ChEMBL нами были использованы списки известных синонимов названий химических соединений, представленных в базе данных PubChem. Для осуществления загрузки списков известных синонимов был сконструирован URL-адрес, который позволяет загрузить список синонимов с использованием уникальных идентификаторов соединений в PubChem (CID). URL-адреса генерировались

согласно схеме: `«https://pubchem.ncbi.nlm.nih.gov/rest/pug/compound/cid/{Уникальный идентификатор соединения}/synonyms/TXT»`.

URL-адреса были сгенерированы для каждого идентификатора PubChem из ранее сформированного набора. Генерацию URL-адресов, а также дальнейшая обработка полученных данных проводили с использованием языка программирования Python 3.8.8, а также его библиотек, включая requests и pandas [147–149]. Наборы синонимов PubChem включают в себя идентификаторы во внешних базах данных, в том числе ChEMBL и ChEBI. Эти идентификаторы были отобраны и проаннотированы вручную на предмет соответствия химическим соединениям из первоначального источника.

2.2.2 База данных ChEMBL

Европейский институт биоинформатики (EBI) позволяет получать с официального сайта базы данных ChEMBL всю содержащуюся в ней информацию в виде SQL-дампа (файла, позволяющего воссоздать точную копию базы данных из минимального набора данных при помощи автоматически выполняющихся SQL-запросов). Это дает возможность, используя СУБД MySQL, работать со сформированной нами локальной версией этой базы данных.

Взаимодействия с локальной версией ChEMBL осуществлялись с использованием языка программирования Python 3.8.8 и его специализированных библиотек (pandas, mysql-connector-python [150]).

Для получения и фильтрации необходимых сведений использовались внешние идентификаторы химических соединений, полученные из базы данных PubChem (процесс формирования списка внешних идентификаторов описан в разделе 2.2.1).

2.3 Компьютерный прогноз спектров биологической активности

2.3.1 Программа PASS

С целью расширения известных знаний о фармакологическом потенциале доступного на территории РФ растительного сырья веб-интерфейс платформы был дополнен инструментами для расчёта вероятностных профилей биологической активности отдельных фитокомпонентов. Данные вычисления осуществляются с использованием современной версии компьютерной программы PASS (Свидетельство Роспатента № 2006613275 от 15.09.06).

Профиль биологической активности представляет собой набор фармакологических эффектов и механизмов действия, а также ряда других категорий активностей, которые может проявить вещество при взаимодействии с биологическим объектом.

Алгоритм PASS основан на модифицированном наивном байесовском классификаторе (формула 1) и представлении структурных формул в виде MNA-дескрипторов [83].

$$\ln \left[\frac{P(A|C)}{1-P(A|C)} \right] \cong \ln \left[\frac{P(A)}{1-P(A)} \right] + \sum_{i=1}^m \left\{ \ln \left[\frac{P(A|D_i)}{1-P(A|D_i)} \right] - \ln \left[\frac{P(A)}{1-P(A)} \right] \right\} \quad (1),$$

где $P(A|C)$ условная вероятность структуры C при условии, что химическое соединение принадлежит к классу A ; $P(A)$ априорная вероятность принадлежности химического соединения к классу A ; $P(C)$ – априорная вероятность структуры C ; $P(A|D_i)$ – условная вероятность дескриптора D_i при условии, что химическое соединение принадлежит к классу A ;

В ходе обучения создается база знаний SAR Base, содержащая в себе информацию о биологической активности, наборы MNA-дескрипторов (MNA-дескрипторы и процедура их генерации описаны в разделе 2.6), а также зависимости «структура-активность», установленные в процессе обучения [151].

Дальнейший прогноз активности для соединений осуществляется с использованием полученной таким образом SAR Base.

При проведении прогноза оцениваются средние значения вкладов отдельных дескрипторов в вероятность проявления активности исследуемого соединения. По итогам анализа дескрипторного состава рассчитываются оценки вероятности наличия (P_a) или вероятности отсутствия (P_i) для анализируемой активности. Для регуляризации и решения проблемы нулевой вероятности используется арксинусное преобразование Фишера [83].

Базы знаний по зависимостям «структура-активность» прогноза.

Характеристики баз знаний «структура-активность», используемых для выполнения компьютерного прогноза приведены в таблице 7.

Таблица 7 – Базы знаний, используемые для выполнения компьютерного прогноза на информационно-вычислительной платформе Phyto4Health

База знаний	Число видов биологической активности	Среднее значение IAP*
Фармакологические эффекты (PASS 2022 Refined)	358	0.94±0.04
Механизмы действия (PASS 2022 Refined)	1495	0.98±0.02
Нежелательные побочные эффекты [152]	5	0.81±0.05
Цитотоксический эффект в отношении нормальных и опухолевых клеточных линий [153]	441	0.92±0.06

* IAP – инвариантная точность прогноза.

Как видно из таблицы 7, для выполнения компьютерного прогноза в рамках Phyto4Health используется четыре базы знаний, характеризующих зависимости «структура-активность». Общее число прогнозируемых видов биологической активности – 2299. Процесс интеграции инструментов для вычисления профилей биологической активности рассмотрен ниже.

2.3.2 Программа PharmaExpert

В рамках вычислительных экспериментов, проводимых с целью выявления наиболее перспективных направлений исследований экстрактов, выделенных из растений, профили биологической активности PASS были проанализированы с применением компьютерной программы PharmaExpert.

PharmaExpert – компьютерная программа, которая позволяет выполнять анализ интегрального действия сочетания фитокомпонентов (фитокомпозиций, экстрактов и т.д.). Входной информацией для PharmaExpert являются результаты компьютерного прогноза PASS, представленные в формате SDF. Анализ PharmaExpert позволяет получить комплексное представление о взаимосвязи прогнозируемых биологических активностей и, соответственно, оценить возможное аддитивное/синергетическое действие и потенциальные межлекарственные взаимодействия комбинаций различных фитокомпонентов. В основе PharmaExpert лежит база знаний, содержащая в себе информацию о более чем 15 000 известных ассоциаций между механизмами действия и фармакологическими эффектами.

2.4 Хранение и обработка данных

Хранение и обработка собранных сведений о структуре, молекулярных характеристиках, а также биологической активности отдельных фитокомпонентов осуществлялись с использованием системы управления базами данных MySQL [154].

2.5 Разработка веб-интерфейса информационно-вычислительной платформы

Для осуществления взаимодействия пользователя с информационно-вычислительной платформой Phyto4Health реализован пользовательский веб-интерфейс. Веб-интерфейс был разработан с использованием языка гипертекстовой разметки (HTML), формального языка декорирования веб-страниц (CSS), а также

языка программирования JavaScript [155-157]. Серверная часть веб-приложения была реализована на языке программирования PHP [158].

2.6 Поисковые инструменты

Для осуществления доступа к данным Phyto4Health были реализованы два поисковых инструмента:

- Поиск по текстовым запросам.
- Поиск по сходству химических структур.

2.6.1 Поиск по текстовым запросам

Поиск по текстовым запросам реализован с использованием DataTables [159]. Данный инструмент является jQuery-плагином для создания динамических HTML-таблиц. Процедура поиска осуществляется путем фильтрации данных в таблице в веб-браузере. В качестве поисковых запросов могут использоваться названия химических соединений или растений, а сортировка данных происходит динамически (без обновления страницы).

2.6.2 Алгоритмы поиска по сходству химических структур

Для реализации поиска по химическому сходству между структурами в базе данных и использованной пользователем в качестве «запроса» структурой были реализованы два алгоритма оценки. В основе первого алгоритма лежит расчёт коэффициента Танимото между структурами, представленными в формате MNA дескрипторов.

MNA-дескрипторы (Multilevel Neighborhoods of Atoms – многоуровневые атомные окрестности) – компьютерное представление химических структур, в котором учитываются валентности и заряды атомов, а также явно указываются все атомы водорода, но не учитываются типы связей [160].

Схема генерации MNA-дескрипторов имеет вид:

- MNA-дескриптор нулевого уровня – метка самого атома (X). Если атом не включен в цикл, то к его обозначению добавляется символ «←».

- MNA-дескриптор любого следующего уровня представляет собой конкатенацию дескриптора нулевого уровня и лексикографически упорядоченного списка дескрипторов предыдущего уровня его ближайших соседей указываемого в круглых скобках.

Пример MNA-дескрипторов первого и второго уровней, сгенерированных для химической структуры кверцетина представлен на рисунке 8.

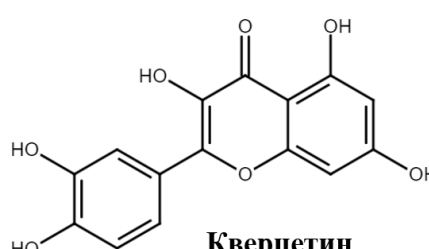
	MNA/1	MNA/2
 Кверцетин	HC	C(C(CCC)C(CC-H)O(CC))
	HO	C(C(CCC)C(CC-H)-H(C))
	CHCC	C(C(CCC)C(CC-H)-O(C-H))
	CCCC	C(C(CCC)C(CC-O)O(CC))
	CCCO	C(C(CCC)C(CC-O)-H(C))
	ОНС	C(C(CCC)C(CC-O)-O(C))
	OC	C(C(CCO)C(CC-H)C(CC-H))
	OCC	C(C(CCO)C(CC-O)C(CC-O))
		C(C(CCO)C(CC-O)-H(C))
		C(C(CCO)C(CC-O)-O(C-H))
		C(C(CC-H)C(CC-H)-O(C-H))
		C(C(CC-H)C(CC-O)-H(C))
		C(C(CC-H)C(CC-O)-O(C-H))
		C(C(CC-O)C(CC-O)-H(C))
		O(C(CCO)C(CCO))
		-H(C(CC-H))
		-H(-O(C-H))
		-O(C(CC-O))
		-O(C(CC-O)-H(-O))

Рисунок 8 – Химическая структура кверцетина и набор MNA-дескрипторов 1-го и 2-го уровней

Важно отметить, что генерация MNA-дескрипторов может быть продолжена до любого уровня [160].

Формула для расчёта коэффициента Танимото между двумя структурами, представленными в виде MNA-дескрипторов, представлена ниже:

$$\text{sim}(A, B) = \frac{\sum_{i=1}^M \min[A(i), B(i)]}{\sum_{i=1}^M A(i) + \sum_{i=1}^M B(i) - \sum_{i=1}^M \min[A(i), B(i)]} \quad (2),$$

где $A(i)$ и $B(i)$ – количество дескрипторов i в молекулах A и B ; M – общее количество дескрипторов в массиве [160].

Второй алгоритм оценки сходства основан на использовании QNA-дескрипторов.

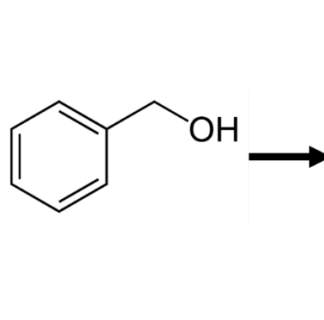
QNA-дескрипторы рассчитываются на основе матрицы связности (C), стандартных значений потенциала ионизации (IP) и сродства к электрону (EA) [161]. Для каждого атома в молекуле QNA-дескрипторы рассчитываются в соответствии с формулами 3 и 4.

$$P_i = B_i \sum_k (\text{Exp}(-\frac{1}{2}C))_{ik} B_k \quad (3)$$

$$Q_i = B_i \sum_k (\text{Exp}(-\frac{1}{2}C))_{ik} B_k A_k \quad (4),$$

где $A_k = \frac{1}{2}(IP_k + EA_k)$, $B_k = (IP_k - EA_k)^{-\frac{1}{2}}$. Значения EA и IP агрегированы из различных источников [161].

Пример QNA-дескрипторов, сгенерированных для структуры бензилового спирта приведен на рисунке 9.



P	Q
0.2236	-1.0943
0.2340	-0.2705
-2.0864	0.4080
-0.6349	0.2723
-0.5079	0.1183
-0.5228	0.1489
0.6879	0.6894
1.1436	-0.0051
0.8425	0.0193
0.8306	0.0427
0.8317	0.0393

Рисунок 9 – Пример генерации QNA – дескрипторов для структуры бензилового спирта

Важной особенностью QNA-дескрипторов, является то, что при таком представлении структурной формулы, описывается каждый из атомов молекулы, но каждое из значений P и Q зависит от всего состава структуры молекулы.

Поскольку в данном компьютерном описании структура химического соединения представлена как массив пар взаимно зависимых значений, оценка сходства с применением традиционных коэффициентов затруднена. Поэтому нами использован подход, предложенный Р. Тодескини [162], в котором рассчитывается максимальное сходство элементов одного множества с элементами другого множества. Максимальные вклады суммируются и усредняются по общему числу элементов в обоих множествах. В качестве меры сходства используется коэффициент Танимото. В данном случае формула для расчета имеет следующий вид:

$$F(A, B) = \frac{n(A \cap B)}{n(A) + n(B) - n(A \cap B)} \quad (5),$$

где $n(A) = N_A$ и $n(B) = N_B$ – количества пар P и Q дескрипторов в молекулах A и B , соответственно. $n(A \cap B)$ рассчитывается по формуле 6.

$$n(A \cap B) = \frac{1}{2} \left(\sum_A \max_{b \in B} [S_{ab}] + \sum_B \max_{a \in A} [S_{ab}] \right) \quad (6)$$

$$S_{ab} = \text{Exp}(-12N_B((P_a - P_b)^2 + (Q_a - Q_b)^2)) \quad (7)$$

$$S_{ba} = \text{Exp}(-12N_A((P_a - P_b)^2 + (Q_a - Q_b)^2)) \quad (8),$$

где S_{ab} и S_{ba} – попарные оценки сходства между QNA-дескрипторов атома a в молекуле A и QNA-дескрипторов атома b в молекуле B ; P_a и Q_a – QNA дескрипторы атома a в молекуле A ; P_b и Q_b – QNA-дескрипторы атома a в молекуле B ; Множители $12N_B$ и $12N_A$ были подобраны эмпирически [163].

2.7 Оценка химического пространства, образованного химическими соединениями из баз данных традиционной медицины

С целью оценки распределения химических соединений из различных баз данных в едином химическом пространстве мы выполнили процедуру кластеризации с использованием двух алгоритмов. Первый алгоритм, основанный на плотности пространственной кластеризации для приложений с шумами (англ. Density-based Spatial Clustering of Applications with Noise, DBSCAN), был использован для выявления выбросов в анализируемых данных, а окончательные выводы были сделаны на основе результатов разбиения на кластеры методом *k*-средних. Оптимальное количество кластеров для разбиения было подобрано на основе рассчитанных коэффициентов силуэта. Для выполнения вышеописанных процедур нами были использованы представления химических структур в виде молекулярных «отпечатков пальцев» с расширенной связностью (англ. Extended-Connectivity Fingerprints, ECFP). В качестве оценок сходства использовались коэффициенты Танимото. Снижение размерности пространства оценок сходства было выполнено с применением алгоритма равномерной аппроксимации и проекции (англ. Uniform Approximation and Projection, UMAP).

2.7.1 «Отпечатки пальцев» с расширенной связностью (ECFP)

ECFP представляют недавно разработанный класс молекулярных отпечатков пальцев, для расчёта которых используется модификация алгоритма Моргана [164]. Для генерации ECFP используется следующий алгоритм. На первой стадии каждому атому присваивается целочисленный идентификатор. На следующих этапах, формируется набор признаков, учитывающий соседние атомы и далее по удалению от стартового атома. Данная стадия схожа с аналогичной стадией алгоритма канонизации Моргана, но в данном случае количество итераций устанавливается пользователем. На последней стадии происходит формирование битовой карты (англ. bit array) [165]. Пример признаков, сгенерированных для структуры бутирамида приведены на рисунке 10.

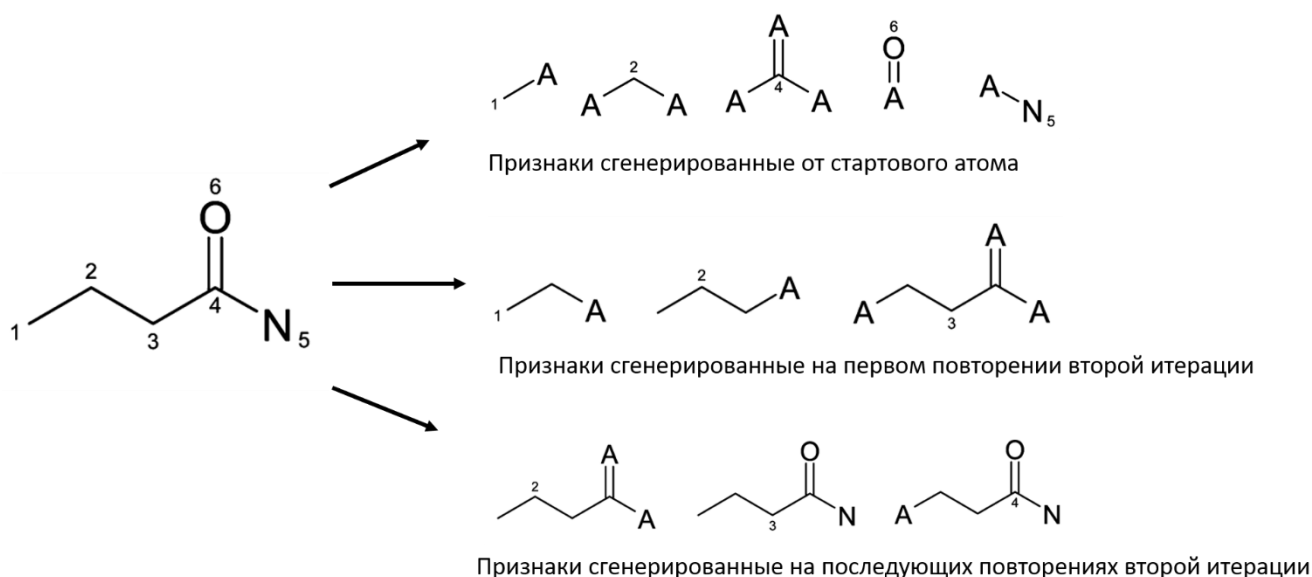


Рисунок 10 – Сгенерированные для химической структуры бутирамида признаки. Адаптировано из [165]

Как видно на рисунке 10, на первой итерации от стартового атома было сгенерировано 5 признаков, на второй итерации – 3 признака и ещё 3 признака были сгенерированы для последующих повторений второй итерации.

Для генерации ECFP-дескрипторов была использована библиотека RDkit для языка программирования Python [166].

2.7.2 UMAP

Для снижения размерности полученных оценок сходства был использован алгоритм UMAP. Основным преимуществом данного алгоритма является его устойчивость к выбросам, что крайне важно для анализа химического пространства природных соединений.

Алгоритм работы UMAP основан на построении взвешенного ориентированного графа, в котором ребрами соединены объекты, которые являются ближайшими соседями. В таком графе вес ребра интерпретируется как вероятность существования данного ребра. Затем аналогичный граф создается в целевом низкоразмерном пространстве, и осуществляется минимизация различий между двумя графами с использованием стохастического градиентного спуска [167].

Снижение размерности оценок сходства было выполнено с помощью библиотеки UMAP для языка программирования Python [168].

2.7.3 DBSCAN

В нашей работе для выявления потенциальных выбросов в данных мы использовали алгоритм DBSCAN [169]. DBSCAN позволяет группировать точки на основе их плотности, что в свою очередь позволяет выявлять выбросы и шум в данных. DBSCAN основан на следующем принципе: на начальном этапе задается минимальное количество точек, которые находятся в заданном радиусе от рассматриваемой точки. Далее выбирается произвольная точка, для которой оценивается количество ближайших точек в ранее заданном радиусе. Точка помечается как выброс, если количество точек в заданном радиусе меньше заданного на начальном этапе количества точек. Если у точки количество соседей превышает заданное значение, она обозначается как «корневая точка». Аналогичным образом обрабатываются точки, находящиеся в заданном радиусе от «корневой точки». Если им не был присвоен кластер, присваивается кластер «корневой точки». Таким образом, кластер увеличивается до тех пор, пока в радиусе корневой точки не останется не проанализированных точек. Эта процедура повторяется для каждой точки, не входящей в радиус корневой точки, до тех пор, пока не будут проанализированы все точки.

Как видно из описания алгоритма, DBSCAN не требует предварительного задания кластеров, но предполагает установку двух параметров, что, в рамках нашей задачи, может затруднить объективную оценку качества кластеризации. Поэтому данный алгоритм был применен нами только для оценки выбросов, а для финальной кластеризации был использован метод k-средних.

Для выполнения расчетов была использована функция DBSCAN из библиотеки scikit-learn языка программирования Python [170].

2.7.4 Кластеризация данных методом k-средних

Для разбиения химических соединений на заданное количество кластеров нами был использован метод k-средних. Алгоритм работы данного метода основан на минимизации суммарного квадратичного отклонения точек от центров кластеров (формула 9) [171].

$$V = \sum_{i=1}^k \sum_{x \in S_i} (x - \mu_i)^2 \quad (9)$$

где k – число кластеров; S_i – полученные кластеры; $i = 1, 2, \dots, k$; μ_i – центры масс векторов x из кластера S_i .

Как видно из формулы 9, перед началом работы алгоритма число кластеров необходимо задать вручную. Для выявления оптимального количества кластеров нами был использован метод оценки силуэтов [172]. Этот метод позволяет получить представление о качестве кластеризации, основываясь на том, насколько хорошо объект «похож» на объекты своего кластера по сравнению с объектами из других кластеров. Значение силуэта может находиться в пределах от -1 до +1. Если значение силуэта равно +1, это указывает на то, что объекты хорошо разделены на кластеры с высокой плотностью.

Для расчета значения силуэта для каждого объекта кластера вычисляется его среднее значение расстояния между ним и объектами его кластера (формула 10), а также расстояние между ним и объектами других кластеров (формула 11).

$$a_i = \frac{1}{|C_I| - 1} \sum_{j \in C_I, i \neq j} d(i, j) \quad (10)$$

где $|C_I|$ – число объектов в кластере; $d(i, j)$ – расстояние между объектами i и j кластера I .

$$b_i = \min_{I \neq J} \frac{1}{|C_J|} \sum_{j \in C_J} d(i, j) \quad (11)$$

где $j \in C_J$ – объект из не I -го кластера.

Значение силуэта для i -го объекта рассчитывается по формуле 12.

$$s_i = \frac{b_i - a_i}{\max(a_i, b_i)} \quad (12)$$

На заключительном этапе значения s_i усредняются для объектов.

Для расчёта коэффициентов силуэта и кластеризации методом k -средних были использована функции `silhouette_score` и `KMeans` из библиотеки `scikit-learn` языка программирования `python` [173, 174].

2.8 Валидация информационно-вычислительной платформы Phyto4Health

Известно, что разнообразие химических соединений природного происхождения охватывает бóльшую область химического пространства по сравнению с химическими соединениями, синтезируемыми в лабораторных условиях [1]. Поэтому использование моделей машинного обучения, разработанных на основе обучающих выборок, содержащих синтетические соединения с целью оценки различных свойств молекул природного происхождения, может оказаться малоэффективным [175].

В настоящей работе для оценки профилей биологической активности используется компьютерная программа `PASS`, при подготовке обучающих выборок которой разделение соединений по их происхождению на природные и синтетические не проводилось. Поскольку в нашей работе рассматриваются именно ПС, важно оценить качество прогноза `PASS` для данной группы соединений. Для оценки возможности расширения сведений о потенциальных видах биологической активности ПС с использованием информационно-

вычислительной платформы Phyto4Health нами были выполнены процедуры ретроспективной и проспективной валидации.

2.8.1 Ретроспективная валидация

В качестве массива структур природного происхождения использовалась БД LOTUS. БД LOTUS была выбрана в связи с тем, что, в отличие от других крупных репозиториях химических структур (например, COCONUT), предоставляет данные об организмах, из которых потенциально может быть выделено химическое соединение. В настоящем исследовании рассматривались только химические соединения, которые могут быть выделены из растений.

Для подготовки массива структур с официального сайта LOTUS БД был загружен полный массив информации, представленный в формате BSON (англ. Binary JavaScript Object Notation). Полученный файл был преобразован в JSON (англ. JavaScript Object Notation). Из массива были извлечены структуры химических соединений, представленные в формате InChiKey, для которых в таксономии источников в качестве царства указано «Plantae».

Для подготовки набора данных, содержащих информацию о результатах лабораторного тестирования ПС, была произведена выгрузка релевантных данных из БД PubChem. Выгрузка осуществлена с использованием API PUG REST путем отправки специализированных запросов на сервер PubChem. Запрос для выгрузки результатов лабораторного тестирования имел следующий вид:

«<https://pubchem.ncbi.nlm.nih.gov/rest/pug/compound/inchikey/{InChiKey}/assaysummary//CSV>».

Для подготовки списка молекулярных мишеней, взаимодействие с которыми исследовано в различных тест-системах, была произведена повторная выгрузка данных. Запрос для выгрузки сведений о молекулярных мишенях имел следующий вид:

«<https://pubchem.ncbi.nlm.nih.gov/rest/pug/assay/aid/{Идентификатор карточки с результатами лабораторного тестирования}/targets/ProteinName,ProteinAccession/JSON>».

В приведенном запросе «*ProteinName*» – название молекулярной мишени, а «*ProteinAccession*» – поле, содержащее идентификатор мишени (в том числе и идентификатор UniProt)).

На основе идентификаторов UniProt, из механизмов действия PASS Standard 2022 (8566 типов биологической активности) были отобраны только механизмы действия, ассоциированные с ранее отобранными молекулярными мишенями.

«Активными» соединениями считались соединения, для которых в БД PubChem, в соответствующей карточке была пометка «Active». Соединения, для которых была другая пометка (например «Unspecified» или «Inactive») считались неактивными. «Активными» по результатам компьютерного прогноза PASS считались соединения, для которых конкретный вид биологической активности прогнозируется при пороге $P_a > P_i$.

На основе сформированного таким образом набора данных были рассчитаны чувствительность (формула 9), специфичность (формула 10) и сбалансированная точность (формула 11).

$$\text{Чувствительность (SENS)} = \frac{TP}{TP + FN} \quad (9)$$

$$\text{Специфичность (SPEC)} = \frac{TN}{TN + FP} \quad (10)$$

$$\text{Сбалансированная точность (BA)} = \frac{1}{2} \left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP} \right) \quad (11)$$

где TP – число истинно положительных результатов; TN – число истинно отрицательных результатов; FP – число ложноположительных результатов; FN – число ложноположительных результатов.

2.8.2 Проспективная валидация

Нами была осуществлена экспериментальная валидация применимости инструментов компьютерного прогноза для анализа фармакологического потенциала индивидуальных химических соединений из базы данных Phyto4Health и комплексных фармацевтических композиций, содержащих также

дополнительные фитокомпоненты. Прогноз PASS был сопоставлен с результатами лабораторных экспериментов, проведенных коллегами из других организаций с использованием многокомпонентных смесей и отдельных ПС, выделенных из различных лекарственных растений. Информация об этих исследованиях представлена ниже:

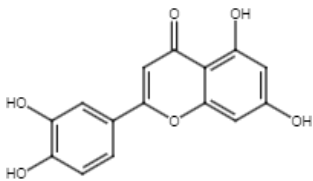
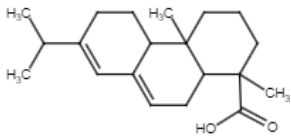
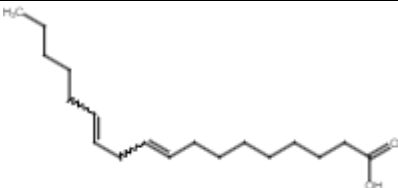
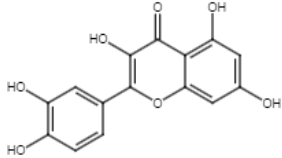
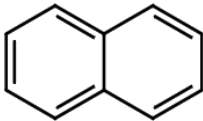
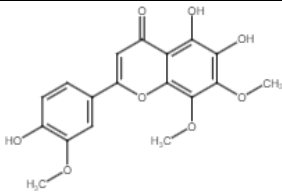
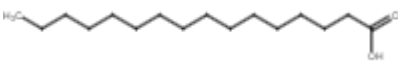
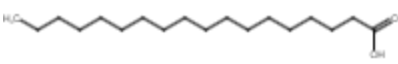
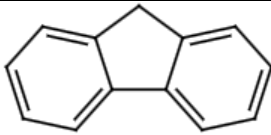
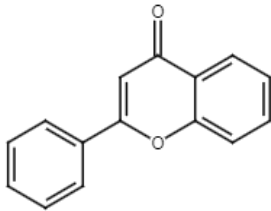
– Лаборатория иммунофармакологии (НМИЦ онкологии им. Н.Н. Блохина). Оценка цитотоксического действия в отношении клеток рака мочевого пузыря для двух вариантов препарата «Фитоладаптоген».

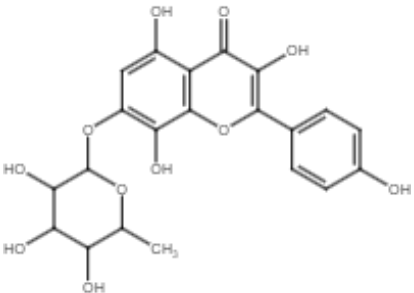
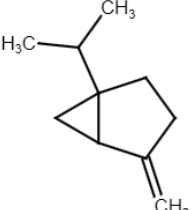
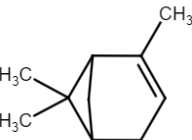
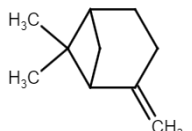
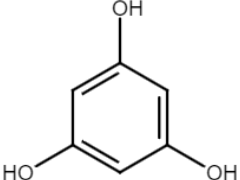
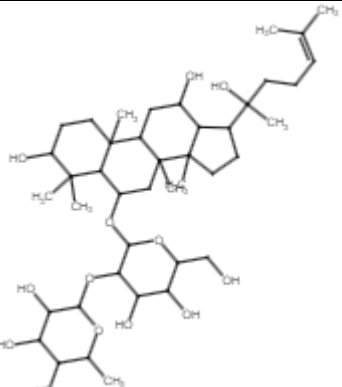
– Санкт-Петербургский химико-фармацевтический университет МЗ РФ совместно с ПГФА МЗ РФ, БГМУ МЗ, ИЭФБ РАН и Сеченовским университетом МЗ РФ. Оценка антиагрегантного и антикоагулянтного действия вторичных метаболитов морошки (*Rubus chamaemorus* L.).

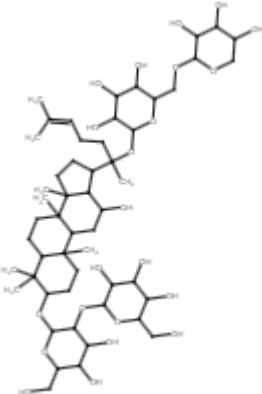
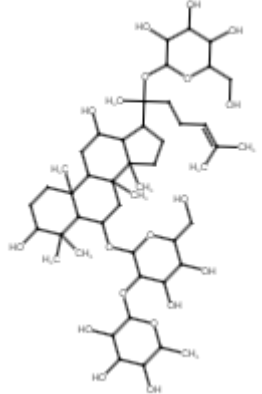
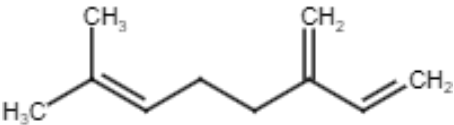
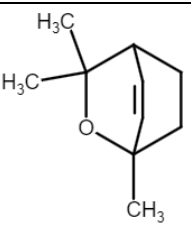
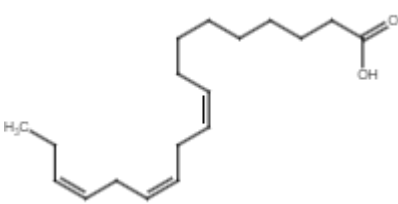
– Центр доклинических исследований Всероссийский НИИ лекарственных и ароматических растений (ВИЛАР). Оценка гепатопротекторного эффекта цикориевой и хлорогеновой кислот из цикория обыкновенного (*Cichorium intybus* L.) и их влияния на активность цитохрома P450 и глутатионтрансферазы.

Оценка цитотоксического действия в отношении клеток рака мочевого пузыря для двух вариантов препарата ФЛА. Работа по подготовке образцов препарата ФЛА, идентификации индивидуальных химических веществ в составе отдельных экстрактов, также лабораторному тестированию ФЛА на клеточной культуре RT112 была выполнена сотрудниками лаборатории иммунофармакологии НМИЦ онкологии им. Н.Н. Блохина (заведующий лабораторией д.б.н., профессор Бочарова Ольга Алексеевна). Идентификация мажорных биологически активных соединений проводилась методом жидкостной хроматографии и тандемной масс-спектрометрии (ВЭЖХ-МС). Перечень идентифицированных химических соединений приведен в таблице 8.

Таблица 8 – Структурные формулы фитокомпонентов входящих в мажорный и минорный варианты фармацевтической композиции ФЛА

Структурная формула	Название	Вариант фитокомпозиции
	Лютеолин	Минорный/Мажорный составы
	Абиетиновая кислота	Минорный/Мажорный составы
	Линолевая кислота	Минорный/Мажорный составы
	Кверцетин	Минорный/Мажорный составы
	Нафталин	Минорный/Мажорный составы
	Тимонин	Минорный/Мажорный составы
	Пальмитиновая кислота	Минорный/Мажорный составы
	Стеариновая кислота	Минорный состав Мажорный состав
	Флуорен	Минорный/Мажорный составы
	Олеиновая кислота	Минорный/Мажорный составы
	Флавоон	Минорный/Мажорный составы

Структурная формула	Название	Вариант фитокомпозиции
	Родионин	Минорный/Мажорный составы
	Сабинен	Мажорный состав
	α-Пинен	Мажорный состав
	β-Пинен	Мажорный состав
	Флороглюцинол	Мажорный состав
	Гинзенозид Rg2	Мажорный состав

Структурная формула	Название	Вариант фитокомпозиции
	Гинзенозид В2	Мажорный состав
	Гинзенозид Rb2	Мажорный состав
	β -Мирцен	Мажорный состав
	1,8-цинеол	Мажорный состав
	α -линолевая кислота	Минорный/Мажорный составы

Исследование цитотоксичности двух фармацевтических композиций было выполнено на клеточной линии рака мочевого пузыря человека линии RT112. Для оценки цитотоксичности использовался МТТ-тест [176]. Механизм клеточной гибели был определен с использованием метода двойного окрашивания аннексином V и йодидом пропидия. Детектирование активной каспазы-3 проводилось с применением набора FITC Active Caspase-3 Apoptosis Kit (“BD Pharmingen”, США).

Оценка антиагрегантного и антикоагулянтного действия вторичных метаболитов *Rubus chamaemorus* L. На базе Санкт-Петербургского химико-фармацевтического университета были выделены индивидуальные химические соединения из листьев *Rubus chamaemorus* L. Работа выполнена сотрудниками кафедры фармакогнозии под руководством к.ф.н. Андрея Кеннета Уэйли и к.б.н. Владимира Геннадьевича Лужанина. Структуры идентифицированных индивидуальных компонентов приведены на рисунке 11.

Выделенные метаболиты растений были подвергнуты лабораторному тестированию с целью выявления их антиагрегантной и антикоагуляционной активности. Эксперименты были выполнены на крови здоровых доноров мужского пола в возрасте от 18 до 24 лет (34 человека). В ходе экспериментов с применением стандартных методик были исследованы функциональная активность тромбоцитов (агрегометр «АТ-02», Россия) и антикоагулянтная активность (Solar CGL 2110, Россия). Препараты сравнения – ацетилсалициловая кислота, пентоксифиллин и гепарин натрия.

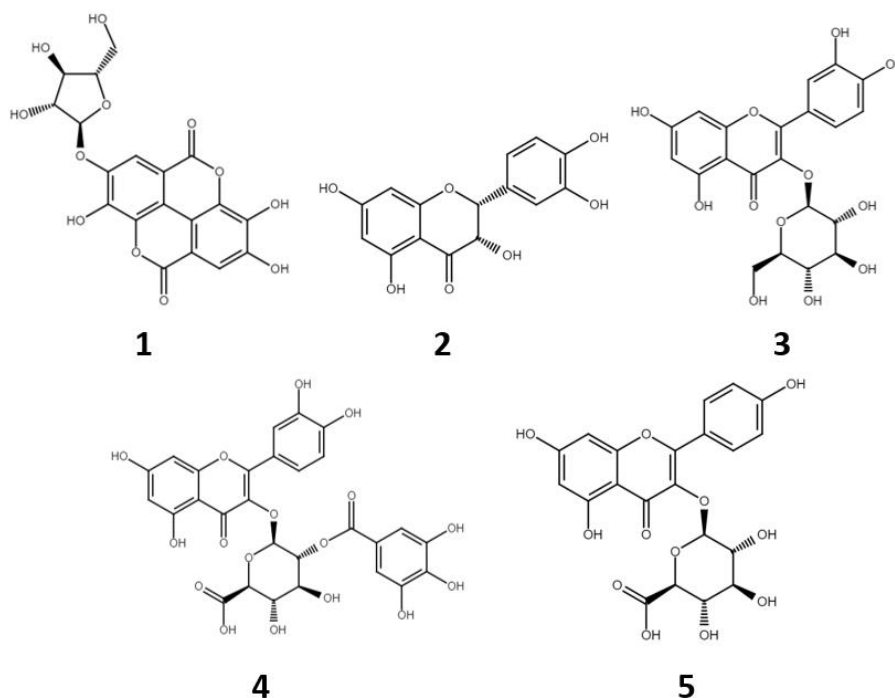


Рисунок 11 – Химические структуры индивидуальных фитокомпонентов *Rubus chamaemorus* L. 1 – 4-О-α-L-арабинофурано-зилэллаговая кислота; 2 – (–)-Эпикатехин; 3 – Кверцетин-3-О-β-D-глюкуронид; 4 – Кверцетин-3-О-β-D-2''-галлоилглюкуронид; 5 – Кемпферол-3-О-β-D-глюкуронид

Определение влияния цикориевой и хлорогеновой кислот из *Cichorium intybus* L. на активность цитохрома P450 и глутатион трансферазы. Определение индивидуальных веществ в составе травы цикория обыкновенного было проведено сотрудниками центра доклинических исследований ФГБНУ «Всероссийский НИИ лекарственных и ароматических растений» под руководством к.б.н. Лупановой Ирины Александровны. В качестве объектов исследования были выбраны цикориевая и хлорогеновая кислоты (рисунок 12). Индивидуальные соединения были выделены из водно-спиртового экстракта травы цикория обыкновенного (*Cichorium intybus* L.) с использованием избирательной жидкостной экстракции и колоночной хроматографии.

В качестве источника CYP450 и GST использовали суспензию микросом, полученную из печени интактных белых нелинейных крыс методом дифференциального центрифугирования. Содержание белка в суспензии было определено методом Лоури [187]. Содержание CYP450 определено по методике предложенной Омура и Сато [188]. Скорость монооксигеназных реакций, катализируемых CYP450, и реакций конъюгации 2,4-динитрохлорбензола, катализируемой GST, определяли с использованием спектрофотометра MPS-2000 (Shimadzu, Япония).

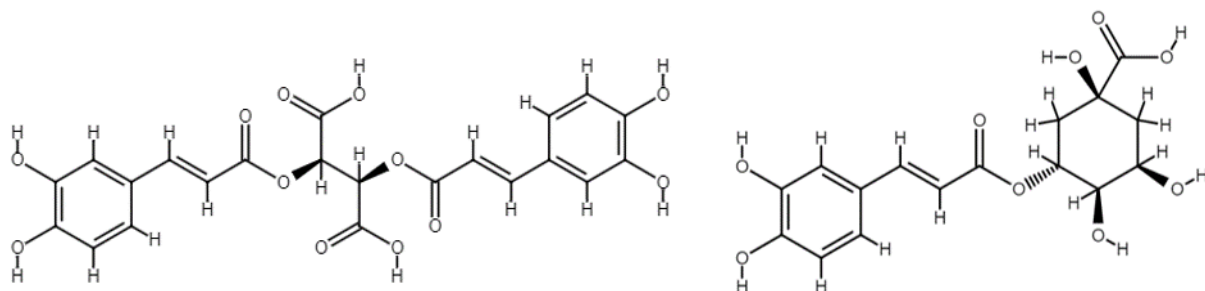


Рисунок 12 – Химические структуры фитокомпонентов *Cichorium intybus* L. (цикориевая кислота – слева; хлорогеновая кислота – справа)

ГЛАВА 3. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

3.1 Логическая структура данных

Для обеспечения хранения и обработки информации по фитокомпонентам официальных лекарственных растений Российской Федерации нами была разработана логическая структура данных (рисунок 13).

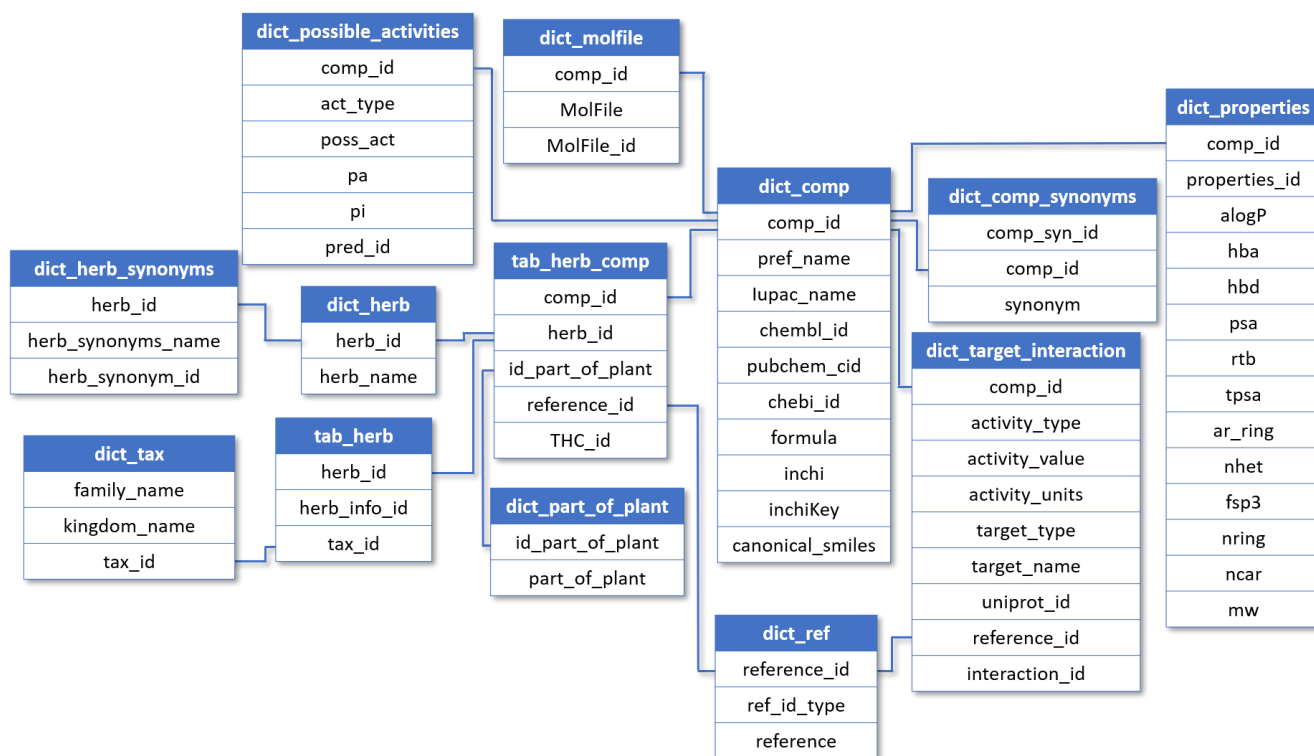


Рисунок 13 – Логическая структура данных

Как видно из рисунка 13, схема базы данных содержит 13 таблиц, объединённых между собой посредством внешних ключей. Важно отметить, что на приведенной схеме не представлены служебные таблицы, необходимые для взаимодействия с веб-ресурсом. Описания каждой из таблиц приведены ниже (таблица 9).

Таблица 9 – Описание схемы базы данных Phyto4Health

Название таблицы	Содержание таблицы	Поле таблицы
dict_comp	Идентификаторы химических соединений Phyto4Health	comp_id
	Тривиальные названия химических соединений	pref_name
	Идентификаторы в крупных репозиториях химических данных (PubChem, ChEMBL, ChEBI)	pubchem_cid, chembl_id, chebi_id
	Линейные нотации химических соединений (брутто формула, InChi, InChi Key, канонические SMILES)	formula, inchi, inchikey, canonical_smiles
dict_comp_synonyms	Идентификаторы синонимов	comp_syn_id
	Идентификаторы химических соединений Phyto4Health	comp_id
	Синонимы названий химических соединений включая линейные нотации и идентификаторы в крупных репозитории химических структур	Synonym
dict_molfile	Идентификаторы химических соединений Phyto4Health	comp_id
	Структура в формате MOL	Molfile
	Идентификатор структуры MolFile_id	Molfile_id
dict_properties	Идентификаторы химических соединений Phyto4Health	comp_id
	Идентификатор записи в БД, содержащей значений молекулярных свойств	properties_id
	Значение коэффициента распределения в системе октанол-вода	Alogp
	Число акцепторов водородных связей	Hba
	Число доноров водородных связей	Hbd
	Полярная площадь поверхности	Psa
	Топологическая полярная площадь поверхности	Tpsa
	Число ароматических колец	ar_ring
	Число гетероатомов	Nhet
	Доля sp ³ гибридизованных атомов	fsp3
	Число колец в химической структуре	Nring
	Число атомов углерода	Ncar
	Молекулярный вес	Mw

Название таблицы	Содержание таблицы	Поле таблицы
dict_target_interaction	Идентификаторы химических соединений Phyto4Health	comp_id
	Указание типа оценки взаимодействия с молекулярными мишенями человека (IC50, Ki и тд.)	activity_type
	Единицы измерения взаимодействия	activity_units
	Значение оценки взаимодействия	activity_units
	Тип молекулярной мишени	target_type
	Название молекулярной мишени	target_name
	Идентификатор молекулярной мишени в базе данных UniProt	uniport_id
	Идентификатор ссылки на источник информации (вторичный ключ для таблицы dict_ref)	reference_id
dict_possible_activities	Идентификатор записи, содержащей сведения о взаимодействии с молекулярными мишенями человека.	Interaction_id
	Идентификаторы химических соединений Phyto4Health	comp_id
	Тип биологической активности PASS	act_type
	Название вида биологической активности	poss_act
	Значение оценки «Pa»	Pa
	Значение оценки «Pi»	Pi
	Идентификатор записи о результатах компьютерного прогноза	pred_id
dict_herb	Предпочтительные названия лекарственных растений.	Herb_name
	Идентификатор записи с названием лекарственного растения	herb_id
dict_herb_synonyms	Синонимы названий лекарственных растений.	Herb_synonyms_name
	Идентификатор лекарственного растения	herb_id
	Идентификатор синонима	herb_synonym_id
dict_tax	Название семейства	family_name
	Название царства	kingdom_name
	Идентификатор записи с таксономией	tax_id
tab_herb	Идентификатор лекарственного растения	herb_id
	Идентификатор записи	herb_id_info

Название таблицы	Содержание таблицы	Поле таблицы
	Идентификатор записи с таксономией	tax_id
dict_part_of_plant	Названий частей растения.	Part_of_plant
	Идентификатор названий частей растения	id_part_of_plant
tab_herb_comp	Идентификаторы химических соединений Phyto4Health	comp_id
	Идентификатор записи с названием лекарственного растения	herb_id
	Идентификатор названий частей растения	id_part_of_plant
	Идентификатор ссылки на источник информации (вторичный ключ для таблицы dict_ref)	reference_id
	Ключ для сводных данных таблицы	THC_id
dict_ref	Идентификатор ссылки на источник информации	reference_id
	Тип ссылки на источник (идентификатор pubmed, doi или ссылка на набор данных)	ref_id_type
	Ссылка на источник данных	Reference

Как видно из рисунка 13 и представленных в таблице 9 сведений, каждое лекарственное растение в разработанной нами логической структуре данных охарактеризовано с точки зрения его таксономии, синонимов видового названия, а также фитохимического состава его частей. Кроме того, логическая структура данных содержит таблицы для хранения сведений о взаимодействии отдельных фитокомпонентов с молекулярными мишенями человека, а также результаты оценки *in silico* потенциальных фармакологических эффектов, механизмов действия и потенциальных токсических эффектов.

3.2 Извлечение и агрегация данных

Из 14-го издания Государственной Фармакопеи Российской Федерации было извлечено 268 видовых названий официальных лекарственных растений. Для указанных видовых названий был осуществлен поиск известных синонимов с применением веб-ресурса The Plant List. Полученный список синонимов содержал

свыше 1000 наименований. Важно отметить, что основной целью создания списка синонимов является оптимизация пользовательских поисковых запросов. По-видимому, такой объем данных является избыточным для достижения данной цели и может ввести в заблуждение конечного пользователя веб-ресурса. Поэтому из представленного набора были отобраны только синонимы, широко используемые в научной литературе. В результате проведенного нами анализа исходного списка синонимов удалось сформировать массив, содержащий 533 известных синонима.

Для собранных растений путем ручной аннотации текстов тематических публикаций и анализа свободно доступных баз данных был составлен набор из 3 128 структур уникальных фитокомпонентов, входящих в фитохимический состав растений. Количество записей в базе данных, характеризующих содержание в части растения фитокомпонента, составило 9 489. Видовые названия двадцати пяти растений с наибольшими количествами известных фитокомпонентов приведены на рисунке 14.

Как видно из рисунка 14, в сформированном нами наборе данных наиболее детально охарактеризованным растением является *Panax ginseng* (женьшень обыкновенный). Для данного растения в базе данных представлено 320 химических соединений, входящих в его фитохимический состав, а также 523 записи, указывающих на содержание этих соединений в отдельных его частях. Для фитохимического состава растения *Humulus lupulus* (хмель обыкновенный) представлено 204 химических соединения, а количество записей «растения-часть растения-фитокомпонент» составило 226. Для растения *Vitis vinifera* (виноград культурный) представлено 198 фитокомпонентов, а количество записей «растение-часть растения-фитокомпонент» составило 210.

Записи о химических соединениях были дополнены значениями молекулярных свойств, включая MW, AlogP, HBA/HBD, RTB, tPSA, число ароматических колец, Fsp3, число гетероатомов, общим числом колец и числом атомов углерода. Распределения значений молекулярных свойств приведены на рисунке 15.

Структуры химических соединений представлены в форматах MOL, SMILES, InChi и InChi Key. С целью оптимизации поисковых запросов из базы данных PubChem были извлечены синонимы названий химических соединений. Из сформированного набора были удалены неуникальные термины. Общее количество отобранных синонимов составило 166 263. В этот набор в том числе входят линейные нотации и идентификаторы из различных баз данных (ChEMBL, ChEBI, LOTUS и т.д.).

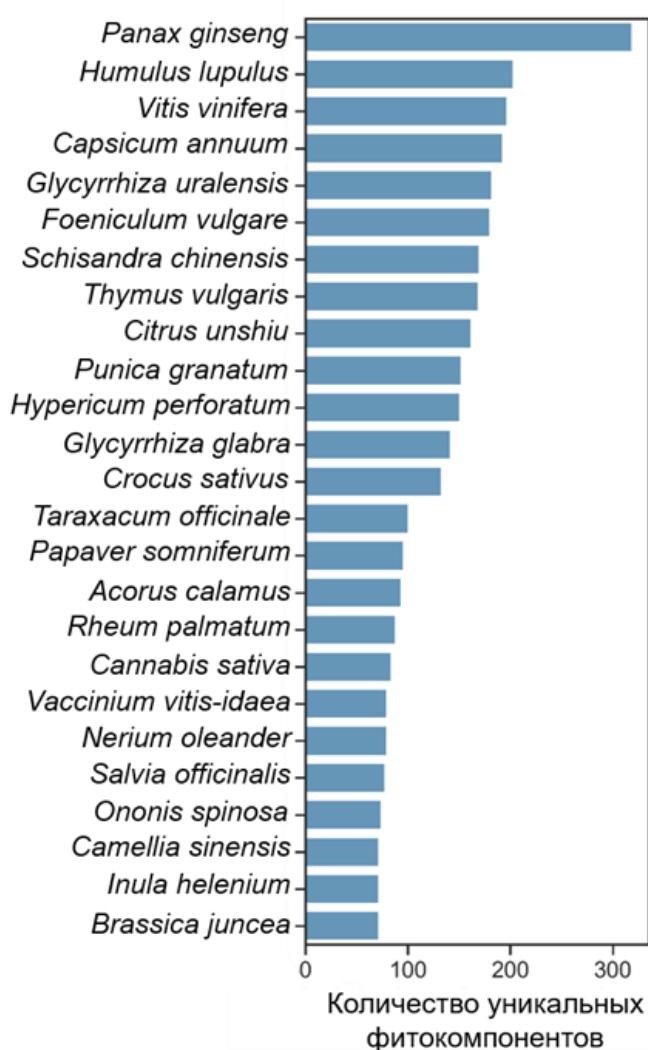


Рисунок 14 – Наименования двадцати пяти растений с наибольшим количеством известных фитокомпонентов

Данные о взаимодействии химических соединений с молекулярными мишенями человека были извлечены из баз данных PubChem и ChEMBL, а также из текстов тематических научных публикаций. В результате агрегации доступных

сведений был подготовлен массив данных, содержащий информацию о 13 688 взаимодействиях с 802 различными молекулярными мишенями человека. Каждая запись в массиве содержит наименование биологической мишени, идентификатор мишени в базе данных UniProt [179], тип и значение оценки взаимодействия (IC50, Ki, EC50), а также ссылки на источники данных.

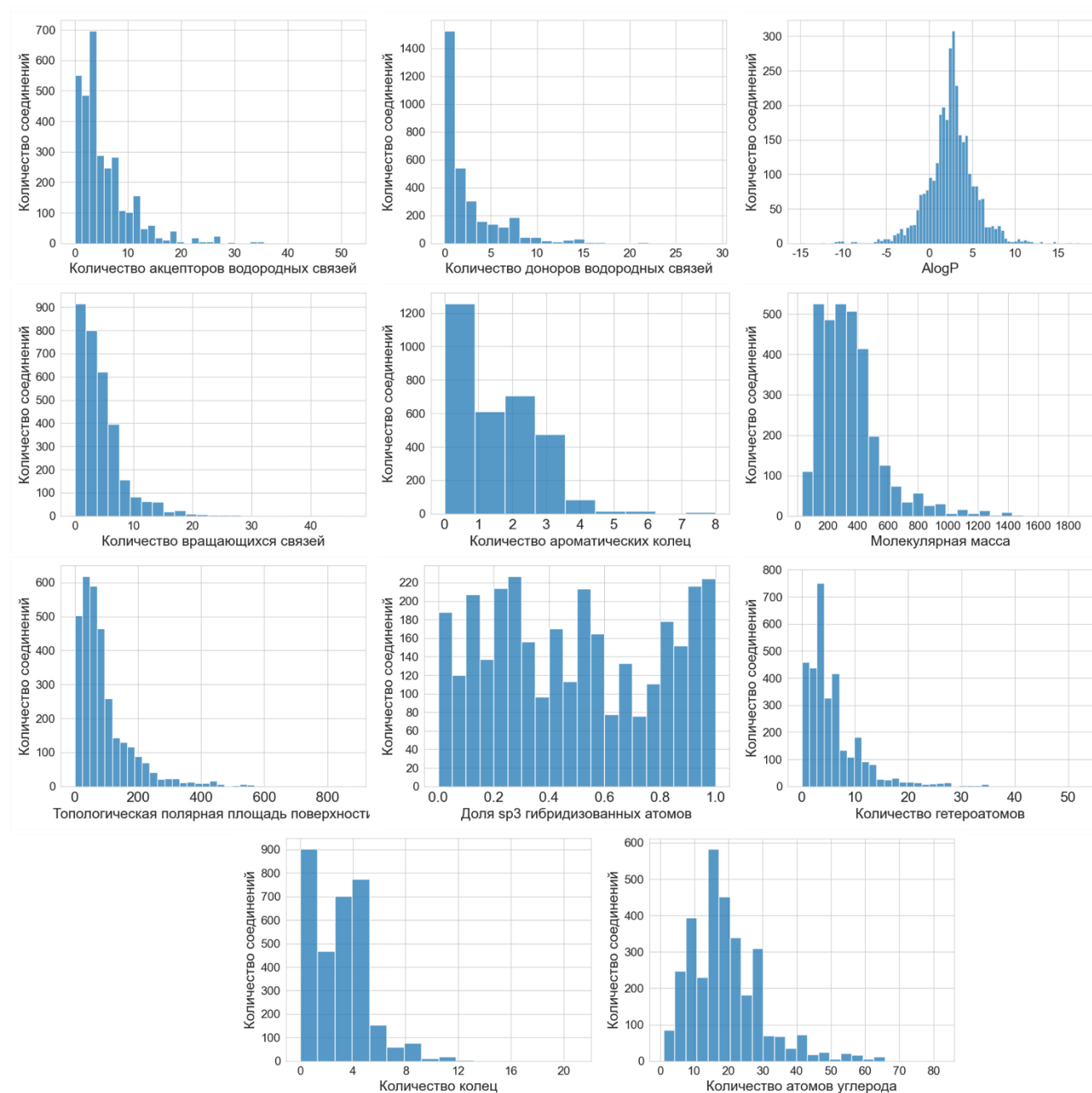


Рисунок 15 – Распределения значений молекулярных свойств соединений Phyto4Health

3.3 Создание доступного в сети Интернет веб-интерфейса информационно-вычислительной платформы

Для осуществления доступа к сведениям, представленным в базе данных, был разработан доступный в сети Интернет информационный ресурс (<http://www.way2drug.com/phyto4health/>). Важно отметить, что с момента публикации работы, посвященной первоначальной версии платформы, осуществлялось постоянное развитие ресурса. Изменения были реализованы с учётом комментариев пользователей, а также данных, полученных с применением сервиса Яндекс.Метрика. С описанием первоначальной версии веб-интерфейса можно ознакомиться в публикации [180].

Главная страница информационно-вычислительной платформы Phyto4Health приведена на рисунке 16.

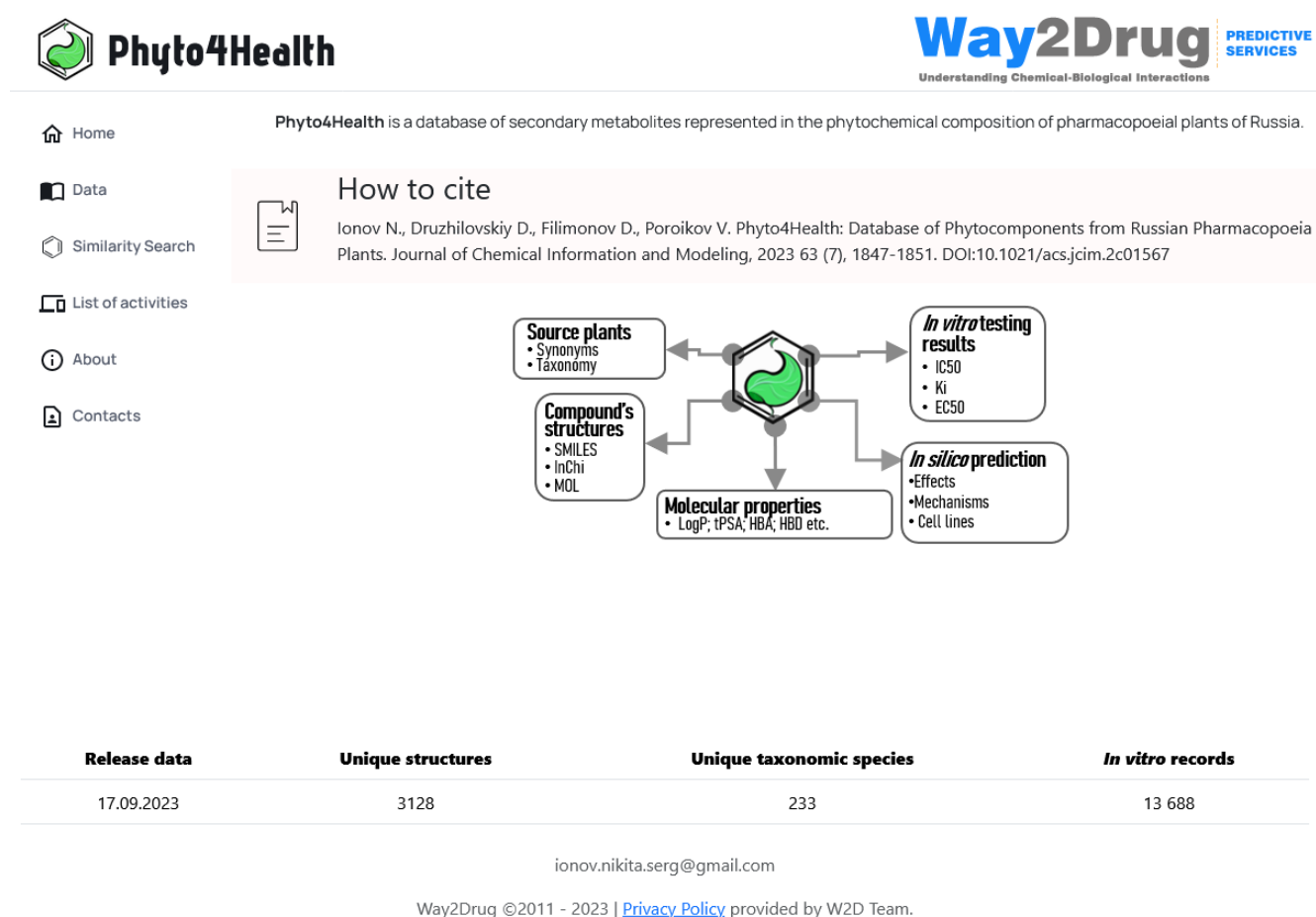
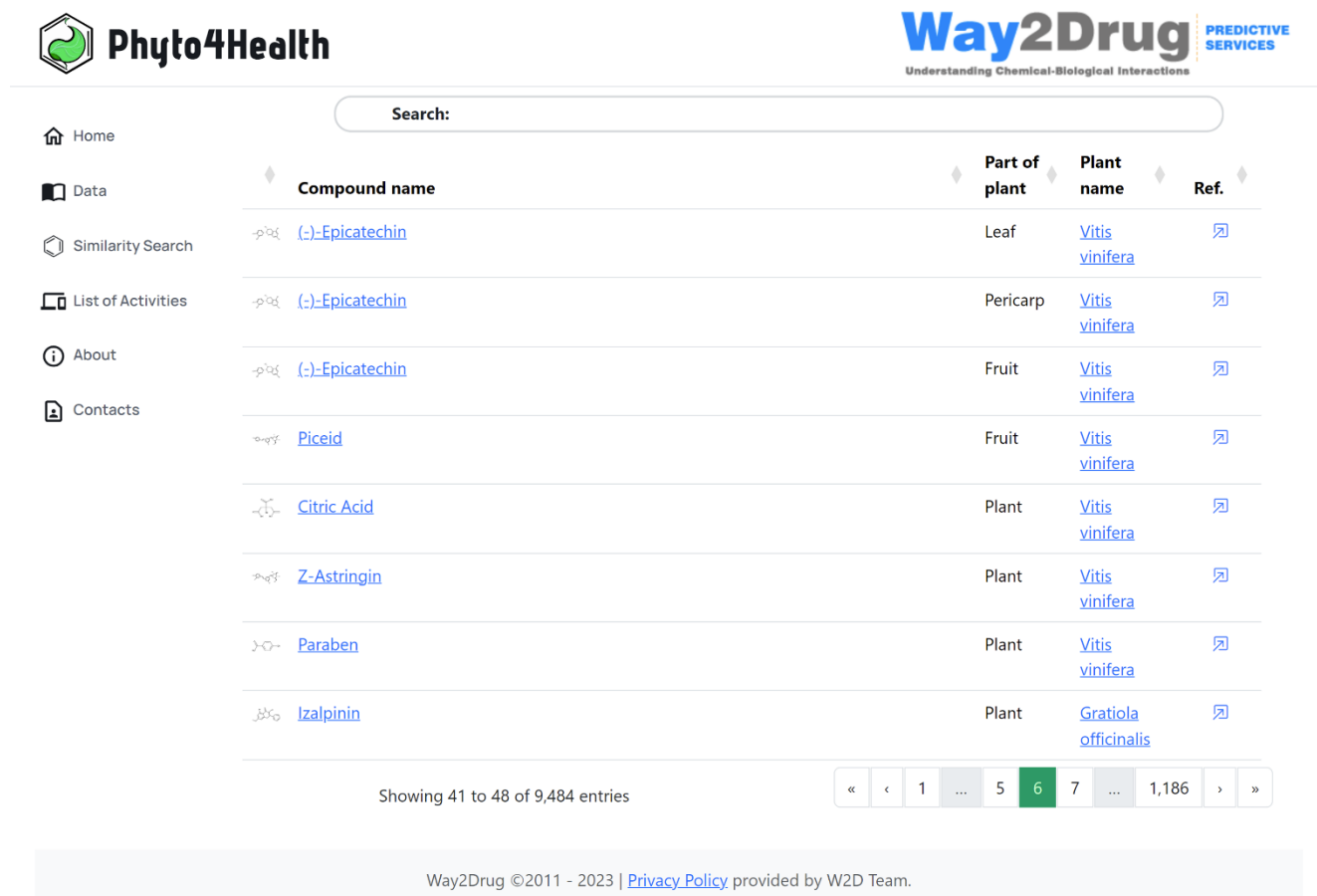


Рисунок 16 – Главная страница ИВП Phyto4Health

Поиск релевантной информации среди сведений, представленных в БД, возможен двумя способами: (1) поиск по текстовым запросам (вкладка «Data»); (2) поиск по сходству химических структур (вкладка «Similarity Search»).

3.3.1 Поиск по текстовым запросам

Поиск по текстовым запросам выполняется в таблице, которая становится доступной при переходе во вкладку «Data» (рисунок 17).



The screenshot displays the Way2Drug Phyto4Health web application. The top navigation bar includes the Way2Drug logo and the text 'PREDICTIVE SERVICES' and 'Understanding Chemical-Biological Interactions'. A search bar is located at the top center. On the left side, there is a sidebar with navigation links: Home, Data (selected), Similarity Search, List of Activities, About, and Contacts. The main content area shows a table of search results for the query '(-)-Epicatechin'. The table has four columns: Compound name, Part of plant, Plant name, and Ref. The results list various compounds and their associated plant parts and names. At the bottom of the table, it indicates 'Showing 41 to 48 of 9,484 entries' and a pagination control showing the current page (6) and total pages (1,186).

Compound name	Part of plant	Plant name	Ref.
(-)-Epicatechin	Leaf	Vitis vinifera	Ref.
(-)-Epicatechin	Pericarp	Vitis vinifera	Ref.
(-)-Epicatechin	Fruit	Vitis vinifera	Ref.
Piceid	Fruit	Vitis vinifera	Ref.
Citric Acid	Plant	Vitis vinifera	Ref.
Z-Astringin	Plant	Vitis vinifera	Ref.
Paraben	Plant	Vitis vinifera	Ref.
Izalpinin	Plant	Gratiola officinalis	Ref.

Showing 41 to 48 of 9,484 entries

Way2Drug ©2011 - 2023 | [Privacy Policy](#) provided by W2D Team.

Рисунок 17 – Представление сведений из БД для осуществления поиска по текстовым запросам

Как видно на рисунка 17, таблица содержит в себе следующие колонки:

- Тривиальное название фитокомпонента, при нажатии на которые осуществляется переход на страницу с описанием информации о соединении.
- Часть растения.
- Название растения, при нажатии на которые осуществляется переход на страницу растения.

– Ссылка на источник данных.

В данном случае, в качестве поискового запроса может выступать название соединения или растения. Отбор релевантных данных осуществляется путем фильтрации таблицы (без обновления страницы).

3.3.2 Поиск по структурному сходству

Кроме поиска по текстовым запросам, была реализована процедура поиска по сходству химических структур. Данная поисковая процедура даёт возможность пользователям проводить поиск в БД, используя в качестве запроса структурную формулу химического соединения и, таким образом, выявить его близкие структурные аналоги. В частности, таким путем могут быть идентифицированы природные соединения, являющиеся аналогами новых (впервые выделенных или синтезированных) соединений, для которых выявлены в эксперименте перспективные виды фармакологической активности.

Для осуществления поиска по сходству химических структур нами был использован специализированный химический редактор Marvin Sketch, разработанный компанией Chemaxon [181]. Marvin Sketch позволяет нарисовать химическую структуру либо загрузить файл, содержащий химическую структуру в формате MOL или SMILES. Пример ввода информации о химической структуре приведен на рисунке 18.

Оценка сходства между структурой, введённой пользователем в качестве запроса, и структурами фитокомпонентов, представленными в базе данных, осуществляется с применением двух алгоритмов. Первый алгоритм производит оценку сходства по Танимото между структурами представленным в виде MNA-дескрипторов. Второй производит оценку сходства с использованием подхода Тодескини и структурами, представленными в виде QNA-дескрипторов. Детальное описание этих алгоритмов приведено в разделе «Материалы и методы». Результат выполнения поискового запроса для соединения кверцетин (первые 5 наиболее похожих структур) приведен на рисунке 19.

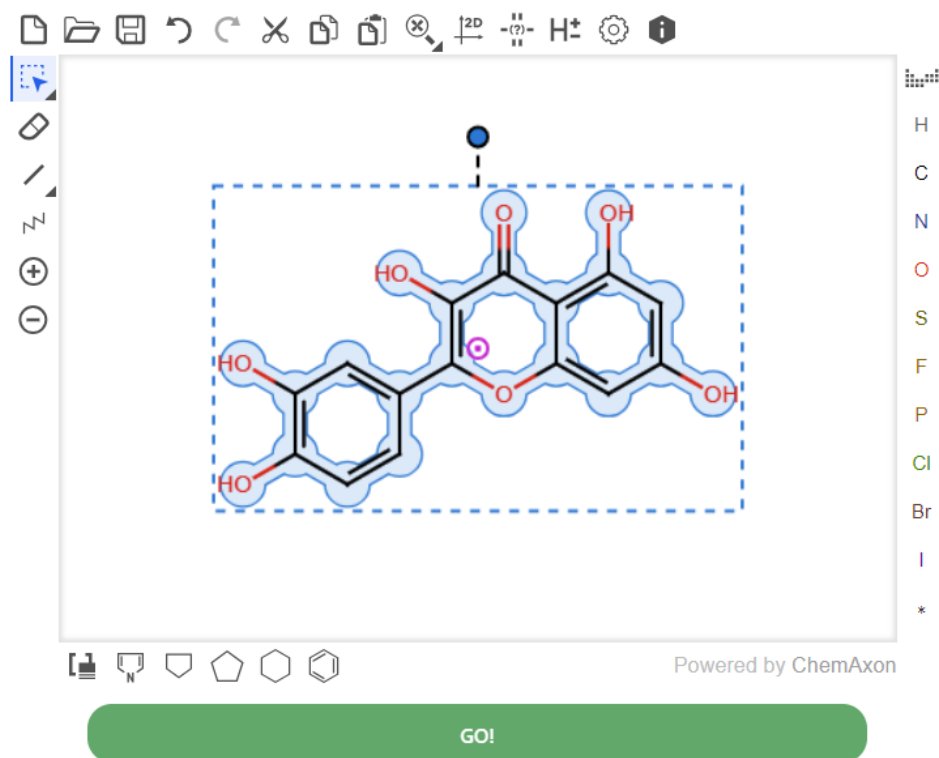


Рисунок 18 – Панель для ввода химических структур с использованием MarvinJS (представленная структура – кверцетин)

Search:				
Compound name	PubChem Link	MNA similarity value	QNA similarity value	
 Quercetin	5280343	1.000	1.000	
 Kaempferol	5280863	0.926	0.987	
 Myricetin	5281672	0.893	0.987	
 Herbacetin	5280544	0.889	0.986	
 Gossypetin	5280647	0.889	0.989	

Showing 1 to 5 of 2,983 entries

« < 1 2 3 4 5 ... 597 > »

Рисунок 19 – Результат выполнения поискового запроса по структуре соединения кверцетин

Как видно из рисунка 19, в результате поиска по структурному сходству для соединения кверцетин выявлен сам кверцетин, а также его близкие структурные аналоги – фисетин, госсипетин, гербацетин, мирицетин и кемпферол.

Как упомянуто выше, поиск по сходству химических структур может использоваться для выявления информации о структурных аналогах новых соединений. Данная процедура помогает исследователям установить (при наличии) структурные аналоги синтетических молекул среди ПС и определить перспективные направления исследования как новой молекулы, так и её установленных структурных аналогов.

Например, результаты оценки сходства для противовирусного препарата нирматрелвир (структурная формула приведена на рисунке 20) – ингибитора протеазы 3CLpro, приведены на рисунке 21.

Как видно из представленных на рисунке данных, среди пяти соединений с наиболее высокими оценками сходства, не выявлено близких структурных аналогов нирматрелвира. Оценки сходства, рассчитанных по алгоритму MNA/Танимото, равны 0, а при оценке сходства, рассчитанного с использованием дескрипторов QNA по алгоритму Тодекскини, максимальные оценки сходства получены для фитокомпонентов *Origanum majorana*. Таким образом, можно заключить, что соединение нирматрелвир не имеет структурных аналогов среди фитокомпонентов ОЛРР, представленных в базе данных Phyto4Health.

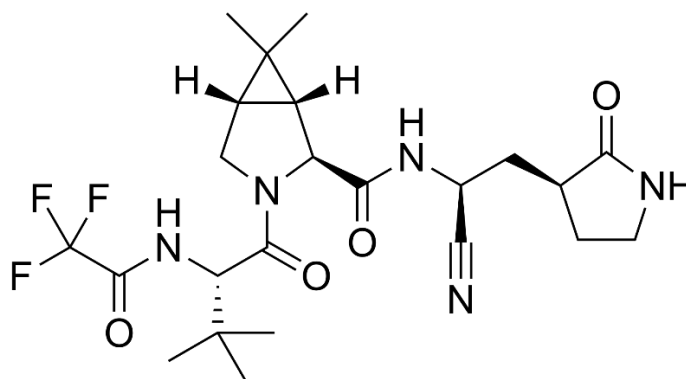


Рисунок 20 – Структурная формула препарата нирматрелвир

Compound name	PubChem Link	MNA similarity value	QNA similarity value
 Methyl (1S,4As,5As,6R,10As)-1-Methyl-2'-Oxospiro[1.4A,5,5A,7,8,10,10A-Octahydropyrano[3,4-F]Indolizine-6,3'-1H-Indole]-4-Carboxylate	10429112	0	0.189
 Methyl (1S,4As,5As,6S,10As)-1-Methyl-2'-Oxospiro[1.4A,5,5A,7,8,10,10A-Octahydropyrano[3,4-F]Indolizine-6,3'-1H-Indole]-4-Carboxylate	9885603	0	0.189
 Methyl (1S,4As,5As,6R,10Ar)-1-Methyl-2'-Oxospiro[1.4A,5,5A,7,8,10,10A-Octahydropyrano[3,4-F]Indolizine-6,3'-1H-Indole]-4-Carboxylate	94160	0	0.189
 (3S,3'as,6's,6As,6Bs,7'ar,9R,11As,11Br)-3-Hydroxy-3',6',10,11B-Tetramethylspiro[1,2,3,4,6,6A,6B,7,8,11A-Decahydrobenzo[A]Fluorene-9,2'-3A,4,5,6,7,7A-Hexahydro-3H-Furo[3,2-B]Pyridine]-11-One	6093180	0	0.160
 Dimethylamine	674	0	0.155

Showing 1 to 5 of 169 entries (filtered from 864 total entries)

« < 1 2 3 4 5 ... 34 > »

Рисунок 21 – Результат выполнения поискового запроса для соединения нирматрелвир

3.3.3 Результаты выполнения поисковых запросов

В результате выполнения поисковых запросов пользователь получает таблицу, содержащую релевантные химические соединения. Далее, выбрав необходимое соединение, пользователь может перейти к соответствующей записи, содержащей сведения о его идентификаторах во внешних базах данных, химической структуре, представленной в различных форматах и значениях молекулярных свойств. Пример такой записи приведен на рисунке 22.

Как видно из рисунка 22, данные представлены в 5 секциях:

– «Representations & DB's id» – содержит информацию о структуре химического соединения и его идентификаторах в химических базах данных.

- «Molecular properties» – содержит значения молекулярных свойств фитокомпонента.
- «Plant sources» – содержит сведения о представленности химического соединения в фитохимическом составе частей лекарственных растений.
- «*In vitro* data» – содержат сведения о результатах экспериментального тестирования.
- «*In silico* data» – данная секция позволяет выполнить компьютерный прогноз для соединения с использованием выбранной пользователем базы знаний зависимостей «структура-активность». Результаты выполнения прогноза фармакологических эффектов для кверцетина приведены на рисунке 23.

Phyto4Health **Way2Drug** PREDICTIVE SERVICES
Understanding Chemical-Biological Interactions

Home Data Similarity Search List of activities About Contacts

Quercetin

Oc1cc(O)c2c(c1)oc(c(c2=O)O)c1ccc(c(c1)O)O

Representations & DB's id		Molecular properties	
ChEBI:	CHEBI:16243	AlogP:	1.99
ChEMBL:	CHEMBL50	Hydrogen bonds acceptors:	7
PubChem:	5280343	Hydrogen bonds donors:	5
IUPAC:	2-(3,4-dihydroxyphenyl)-3,5,7-trihydroxychromen-4-one	Rotatable bonds:	1
Standard InChI:	InChI=1S/C15H10O7/c16-7-4-10(19)12-11(5-7)22-15(14(21)13(12)20)6-1-2-8(17)h1-5,16-19,21H	Number of rings:	3
Standard InChI Key:	REFJWTPEDVJJIV-UHFFFAOYSA-N	Molecular Weight:	302.04
SMILES:	<chem>Oc1cc(O)c2c(c1)oc(c(c2=O)O)c1ccc(c(c1)O)O</chem>	Topological polar surface area:	127
		Number of aromatic rings:	3
		Fsp3:	0
		Number of carbons:	15

Plant sources

Search:

Part of plant	Plant name	Ref.
Flos	Tussilago farfara	
Flower	Taraxacum officinale	
Flower	Taraxacum officinale	
Flower	Crocus sativus	
Fructus	Punica granatum	

Showing 1 to 5 of 53 entries

In vitro data

Search:

Activity Type	Activity Value	Target Name	UniProt ID	Ref.
EC50	100000 nM	Serotonin 1a (5-HT1a) receptor	P08908	
EC50	8020 nM	G-protein coupled receptor 35	Q9HC97	
EC50	5930 nM	G-protein coupled receptor 35	Q9HC97	
EC50	35000 nM	Tyrosine-protein kinase SYK	P43405	
IC50	190000 nM	Epidermal growth factor receptor erbB1	P00533	

Showing 1 to 5 of 284 entries

In silico data

Pharmacological effects Get prediction

Way2Drug ©2011 - 2023 | [Privacy Policy](#) provided by W2D Team.

Рисунок 22 – Результат выполнения поискового запроса на примере соединения кверцетин (англ. quercetin)

In silico data

Pharmacological effects ▾

Get prediction

Search:

Pa	Pi	Pharmacological Effect
0,958	0,001	Antimutagenic
0,923	0,003	Antioxidant
0,884	0,004	Chemoprotective
0,875	0,008	Apoptosis agonist
0,845	0,004	Cardioprotectant
0,794	0,002	Antihelmintic (Trematoda)

Showing 1 to 6 of 174 entries

« < 1 2 3 4 5 ... 29 > »

Рисунок 23 – Пример результатов компьютерного прогноза фармакологических эффектов для соединения кверцетин

Из рисунка 23 видно, что результатом выполнения прогноза, является таблица, содержащая в себе значения P_a и P_i , а также наименования биологической активности.

На веб-сайте Phyto4Health приводится список видов биологической активности (вкладка «List of activities»). Выдержка из представляемого списка приведена на рисунке 24.



Phyto4Health



PREDICTIVE SERVICES

Understanding Chemical-Biological Interactions

Home

Data

Similarity Search

List of activities

About

Contacts

Search:

Activity	ROC AUC, LOO CV	Number of Compounds	Type
Calu-1	0.85	85	Cell Line
CAL-27	0.85	113	Cell Line
A2780S	0.85	13	Cell Line
Autoimmune disorders treatment	0.84	34030	Pharmacological effect
Apoptosis agonist	0.84	11822	Pharmacological effect
Antineoplastic (myeloid leukemia)	0.84	951	Pharmacological effect
Apoptosis agonist	0.84	11822	Mechanism of action
RCC4	0.84	3	Cell Line
NCI-H520	0.84	112	Cell Line
NCI-H292	0.84	142	Cell Line

Showing 2,231 to 2,240 of 2,299 entries

<<

<

1

...

223

224

225

...

230

>

>>

Showing 2,231 to 2,240 of 2,299 entries

« < 1 ... 223 224 225 ... 230 > »

Рисунок 24 – Список видов биологических активностей, для которых может быть выполнен компьютерный прогноз

Кроме визуализации списка видов биологической активности, при нажатии на название, пользователю предоставляется перечень фитокомпонентов, для которых данная активность прогнозируется при пороге $P_a > P_i$.

3.4 Сопоставление с данными о фитокомпонентах традиционной медицины других стран

Для сопоставления представленности в химическом пространстве собранного нами набора данных с соединениями из различных БД мы сравнили данные Phyto4Health с фитокомпонентами из БД традиционной медицины других стран.

Среди БД, рассмотренных в обзоре [92], а также опубликованных в рецензируемых научных журналах, начиная с 2020 года, нами были отобраны БД, соответствующие следующим критериям:

- БД должна содержать структуры ПС, которые были обнаружены в фитохимическом составе различных образцов растений.
- Веб-интерфейс БД должен предоставлять возможность свободной выгрузки структур химических соединений в надлежащем для проведения компьютерного анализа виде (SDF, коды SMILES и т.д.).
- Перечень растений в БД должен относиться к флоре отдельной страны или географического региона.

Кроме того, среди недавно опубликованных БД, удалось обнаружить публикацию, в которой описана БД ПС, синтезируемых эндофитами – микроорганизмами, населяющими ткани растений [182]. В рамках нашей работы интерес к данной группе веществ вызван тем, что зачастую взаимодействие эндофитов и растений является симбиотическим [183]. Известно, что данная группа микроорганизмов способна выделять вещества, влияющие на рост растения, а также вещества, оказывающие противомикробное действие, тем самым, способствуя выживаемости отдельных видов в условиях изменяющейся окружающей среды [183].

Руководствуясь указанными критериями, в качестве объектов для сопоставления сведений из созданной нами Phyto4Health с БД фитокомпонентов, входящих в состав растений других географических регионов, было отобрано 5 БД. Их характеристики представлены в таблице 10, а число пересекающихся соединений между рассматриваемыми БД приведено в таблице 11. Оценка количества пересекающихся соединений была осуществлена на основе сопоставления их структурных формул, представленных в формате SMILES и MNA-дескрипторами.

Таблица 10 – Характеристики БД, отобранных для оценки сходства с Phyto4Health

Название базы данных	Географический регион	Количество уникальных химических структур	Ссылка
NANPDB	Северная Африка	6328	[123]
SistematX	Бразилия	8940	[121]
Ayurveda	Индия	2102	[118]
BIOFACQUIM	Мексика	423	[119]
TPPT	Центральная Европа	1586	[122]
EMNPD	*	6632	[184]
Phyto4Health	Россия	3128	

Примечание - БД EMNPD содержит сведения о ПС, синтезируемые эндофитами.

Таблица 11 – Количества пересекающихся соединений в рассматриваемых БД

База данных	Ayurveda	BIOFACQUIM	EMNPD	NANPDB	P4H	SistematX	TPPT
Ayurveda	2102	19	67	146	187	70	69
BIOFACQUIM	7	423	65	95	113	64	34
EMNPD	35	54	6632	435	174	180	83

База данных	Ayurveda	BIOFACQUIM	EMNPD	NANPDB	P4H	SistematX	TPPT
NANPDB	66	75	436	6328	183	717	303
P4H	88	92	399	115	3128	160	122
SistematX	23	48	153	315	119	8940	212
TPPT	43	27	66	254	105	146	1568

Примечание - P4H – сокращенное наименование Phyto4Health. Над диагональю таблицы приведены оценки, полученные на основе MNA-дескрипторов, под диагональю приведены оценки, полученные на основе SMILES.

Из приведенных в таблице 11 данных видно, что наибольшие количества соединений из Phyto4Health пересекаются с контентом БД NANPDB, TPPT, SistematX и EMNPD. При этом число пересекающихся соединений между отдельными БД невелико, в большинстве случаев составляет не более 13%. Отдельный интерес вызывает то, что наибольшее пересечение по количеству общих химических структур обнаружено между БД Phyto4Health и БД ПС, синтезируемых эндофитами (EMNPD). Этот вопрос, очевидно, требует специального сопоставления генетических особенностей соответствующих лекарственных растений и эндофитов, обуславливающих синтез вторичных метаболитов, что, однако, выходит за рамки настоящей работы.

Для оценки различий между группами соединений, представленными в анализируемых БД, на следующем этапе мы оценили средние значения попарного сходства между входящими в их состав химическими соединениями.

Оценку сходства проводили с использованием коэффициента Танимото, рассчитанного между структурами, представленными MNA дескрипторами. Для анализа сходства множеств химических соединений нами был применён подход, предложенный в [162]. Результаты анализа попарного сходства структур химических соединений из БД Phyto4Health с веществами из отобранных баз данных приведены в таблице 12.

Таблица 12 – Результаты оценки попарного сходства структур фитокомпонентов в анализируемых БД

	Ayurveda	BIOFACQUIM	EMNPD	NANPDB	P4H	SistematX	TPPT
Ayurveda	1,00	0,52	0,51	0,61	0,65	0,52	0,55
BIOFACQUIM	0,52	1,00	0,46	0,54	0,57	0,49	0,47
EMNPD	0,51	0,46	1,00	0,60	0,60	0,51	0,45
NANPDB	0,61	0,54	0,60	1,00	0,71	0,66	0,62
P4H	0,65	0,57	0,60	0,71	1,00	0,62	0,65
SistematX	0,52	0,49	0,51	0,66	0,62	1,00	0,55
TPPT	0,55	0,47	0,45	0,62	0,65	0,55	1,00

Примечание - P4H – сокращенное наименование Phyto4Health.

Как видно из представленных в таблице 12 данных, средние значения оценок попарного сходства для различных БД не превышают 0,71. Среднее значение оценок сходства соединений в Phyto4Health и соединений, представленных в БД Ayurveda, составляет 0,65; 0,57 - с соединениями, представленными в BIOFACQUIM; 0,71 - с соединениями NANPDB; 0,62 - с соединениями SistematX, 0,60 - с соединениями TPPT и с соединениями EMNPD. Эти оценки, по-видимому, являются несколько завышенными, поскольку MNA дескрипторы не учитывают стереоизомеры, а в БД Phyto4Health число стереоизомеров составляет около 25%. Более того, как было показано в работе [185], при величине коэффициента Танимото, превышающей 0,85, лишь 30% «схожих» соединений проявляют одинаковые виды активности. Таким образом, можно сделать заключение, что сходство между структурами химических соединений и их фармакологическим потенциалом в анализируемых БД сравнительно невелико.

Для визуальной оценки различий состава химических соединений в анализируемых БД мы рассчитали коэффициенты сходства Танимото на основе молекулярных «отпечатков пальцев» ECFP. Снижение размерности пространства оценок сходства было выполнено с применением алгоритма UMAP (Uniform Approximation and Projection). Кроме того, чтобы выделить отдельные группы

соединений с высоким сходством химических структур, мы выполнили кластеризацию полученных оценок сходства.

Очевидно, что структурное разнообразие исследуемых наборов химических соединений может быть сопряжено с большим количеством выбросов при оценке их распределения в едином химическом пространстве. Поэтому перед проведением основного этапа кластеризации мы выполнили предварительный анализ с применением алгоритма DBSCAN (англ. Density-Based Spatial Clustering of Applications with Noise). Данный алгоритм позволяет выполнить предварительную оценку количества кластеров и выбросов. Алгоритм был настроен таким образом, чтобы идентифицировать близкие группы точек, включающие в себя минимум 100 соединений. В результате проведенного нами расчета установлено, что на основе плотности данные разделяются на 23 кластера. Точки, представленные в образованных таким образом кластерах, были использованы для последующего анализа, а идентифицированные выбросы в дальнейшем не рассматривались.

На следующем этапе, для определения наиболее оптимального количества кластеров, нами был выполнен расчёт коэффициентов силуэта при разбиении на кластеры в диапазоне от 2 до 30. Результаты расчета коэффициента силуэта для используемого диапазона количества кластеров приведены на рисунке 25.

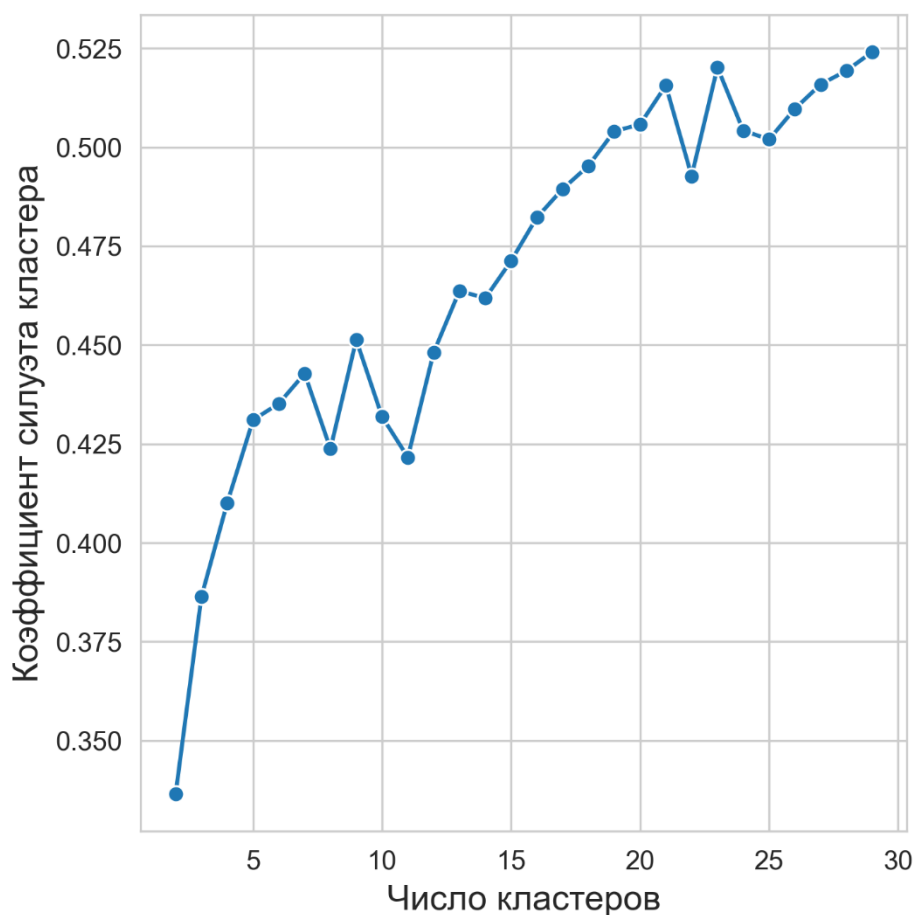


Рисунок 25 – Значения коэффициента силуэта в зависимости от количества кластеров

При детальном анализе разделения соединений из БД Phyto4Health на кластеры в общем химическом пространстве установлено, что 98% из них представлены только в 8 кластерах (таблица 13).

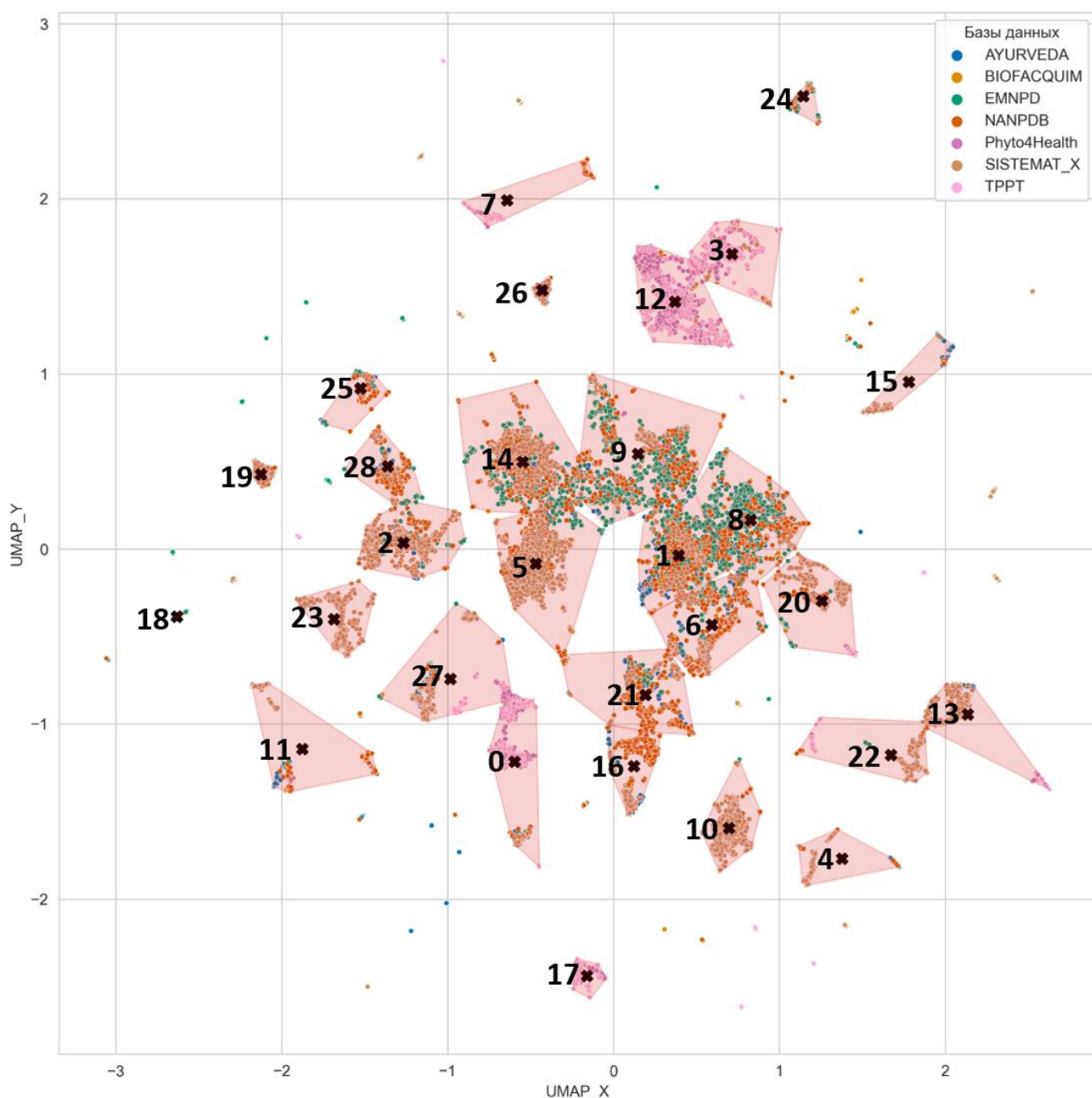


Рисунок 26 – Визуализация химического пространства соединений из отобранных баз данных. Крестики – центры образованных кластеров

Таблица 13 – Доли от общего количества соединений в анализируемых базах данных представленные в отдельных кластерах

Номер кластера	AYURVEDA	BIOFACQUIM	EMNPD	NANPDB	P4h	SISTEMAT_X	TPPT
0	0,07	0,03	0,00	0,15	0,00	0,04	0,00
1	0,00	0,03	0,00	0,10	0,00	0,17	0,00
2	0,05	0,18	0,23	0,07	0,00	0,01	0,00
3	0,00	0,00	0,00	0,00	0,42	0,00	0,29
4	0,08	0,08	0,03	0,07	0,00	0,02	0,00

Номер кластера	AYURVEDA	BIOFACQUIM	EMNPD	NANPDB	P4h	SISTEMAT_X	TPPT
5	0,01	0,01	0,00	0,01	0,03	0,10	0,01
6	0,07	0,02	0,00	0,02	0,00	0,02	0,00
7	0,03	0,01	0,01	0,03	0,01	0,04	0,04
8	0,00	0,02	0,07	0,02	0,00	0,01	0,00
9	0,08	0,00	0,00	0,00	0,00	0,03	0,00
10	0,00	0,00	0,00	0,00	0,08	0,00	0,07
11	0,05	0,00	0,00	0,00	0,00	0,03	0,04
12	0,00	0,00	0,00	0,02	0,04	0,00	0,12
13	0,07	0,09	0,10	0,12	0,00	0,05	0,00
14	0,00	0,00	0,03	0,01	0,00	0,00	0,00
15	0,19	0,20	0,15	0,08	0,00	0,02	0,00
16	0,01	0,02	0,01	0,01	0,00	0,09	0,00
17	0,00	0,00	0,00	0,00	0,00	0,10	0,00
18	0,03	0,02	0,02	0,04	0,00	0,11	0,00
19	0,01	0,03	0,00	0,02	0,00	0,02	0,00
20	0,02	0,00	0,00	0,01	0,24	0,00	0,14
21	0,01	0,02	0,02	0,02	0,02	0,06	0,05
22	0,00	0,00	0,00	0,01	0,16	0,01	0,25
23	0,02	0,07	0,09	0,03	0,00	0,01	0,00
24	0,01	0,00	0,02	0,01	0,01	0,00	0,00
25	0,13	0,12	0,02	0,08	0,00	0,03	0,00
26	0,01	0,02	0,00	0,02	0,00	0,02	0,00
27	0,04	0,01	0,15	0,04	0,00	0,01	0,00
28	0,00	0,00	0,02	0,00	0,00	0,00	0,00

Примечание - P4H – сокращенное наименование Phyto4Health.

Как видно из приведенных в таблице 13 данных, в результате кластеризации были сформированы кластеры, содержащие преимущественно соединения из БД Phyto4Health. К ним относятся кластеры под номерами 17 (8%) и 12 (41%). Кроме того, 20% соединений Phyto4Health попали в кластер номер 0, 16% попали в кластер номер 3. Для оценки различий между химическими структурами, представленными в отдельных кластерах, нами были использованы евклидовы расстояния между их центроидами. Результаты расчёта евклидовых расстояний приведены на рисунке 27.

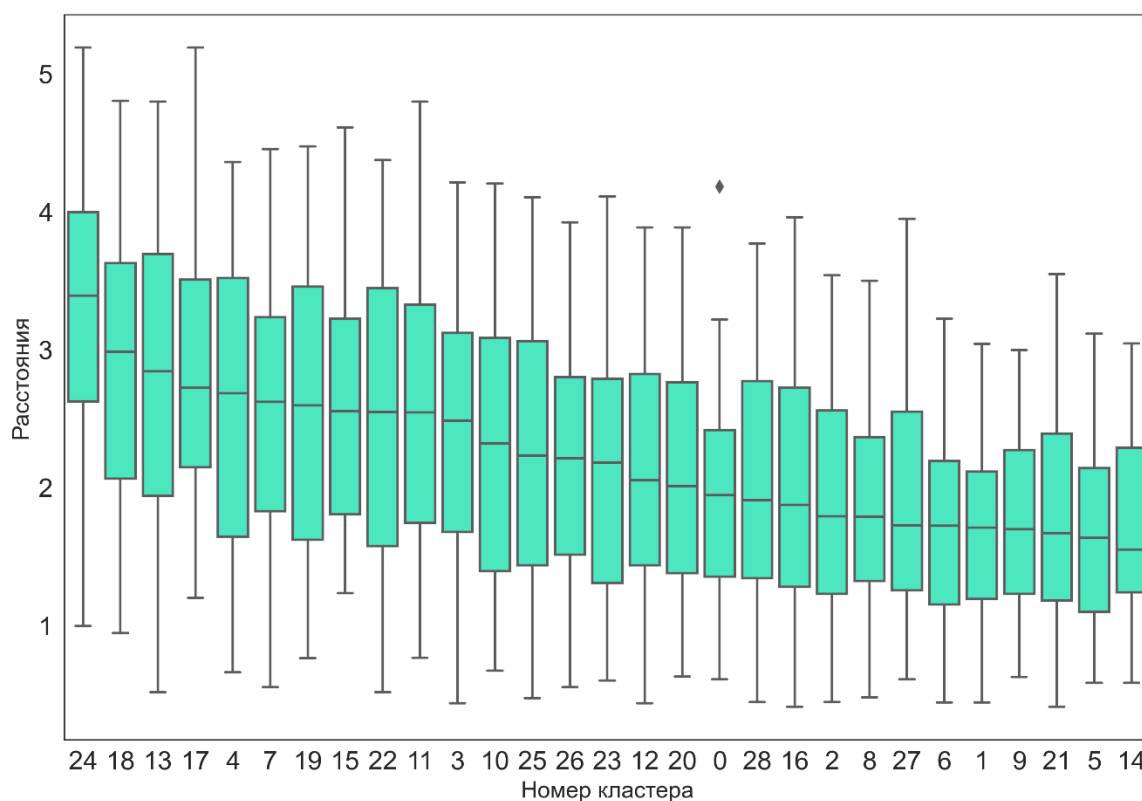


Рисунок 27 – Евклидовы расстояния между центроидами отдельных кластеров. Верхняя и нижняя границы линий, выходящих из прямоугольника, соответствуют верхней и нижней границе в данных. Верхняя и нижняя границы прямоугольника, соответствуют верхнему и нижнему квартилям. Точка на графике – выброс. Данные отсортированы в порядке убывания медианного расстояния центроидов кластеров

Как видно из данных, приведенных на рисунке 27, медианные оценки эвклидова расстояния центроидов кластеров, преимущественно содержащих соединения из БД Phyto4Health (кластеры номер 12, 0, 3 и 7) превышают таковые оценки для других кластеров. Для кластера 12 медианное значение эвклидова расстояния выше, чем таковое для 44% кластеров, для кластера 0 медианное значение эвклидова расстояния превышает таковое для 37% кластеров, для кластера 7 медианное значение эвклидова расстояния превышает таковое для 80% кластеров.

Таким образом можно заключить, что сформированный нами набор данных в БД Phyto4Health охватывает отдельную область химического пространства по сравнению с соединениями других БД. Кроме того, поскольку соединения в БД Phyto4Health представлены только в восьми из 29 кластеров, можно прийти к заключению, что соединения из БД Phyto4Health характеризуются меньшим

разнообразием химических структур, но охватывают отдельную область химического пространства.

3.5 Ретроспективная валидация информационно-вычислительной платформы

Валидация результатов компьютерного прогноза была проведена с использованием структур ПС, извлеченных из БД LOTUS. Из 290 000 структур, представленных в БД LOTUS, было отобрано 191 886 структур, которые потенциально могут быть выделены из растений. В БД PubChem была обнаружена информация о результатах тестирования биологической активности для 27 497 ПС. После сопоставления с видами биологической активности в PASS было отобрано 15 477 пересекающихся записей. Данный набор содержит 624 различных механизма действия и 2 732 структуры различных органических соединений. Для повышения надежности оценки были отобраны механизмы действия, в отношении которых было протестировано более 100 уникальных химических соединений. Общее количество отобранных видов биологической активности составило 83. Результаты расчета метрик качества прогноза, а также отобранные мишени и число проанализированных соединений приведены в таблице 14.

Таблица 14 – Результаты оценки качества прогноза на основе экспериментальных данных БД PubChem

Название	SPEC	SENS	BA	N	<i>In silico</i> (активные/ неактивные)	<i>In vitro</i> (активные/ неактивные)
Vascular endothelial growth factor 1 antagonist	1,00	0,98	0,99	244	7/237	1/243
Vasopressin 1A antagonist	1,00	0,98	0,99	243	9/234	0/243
Dopamine D4 antagonist	1,00	0,96	0,98	243	25/218	0/243
Dopamine D3 antagonist	1,00	0,96	0,98	242	22/220	12/230
Platelet activating factor antagonist	1,00	0,94	0,97	243	15/228	1/242
Acetylcholine M2 receptor antagonist	1,00	0,90	0,95	254	40/214	16/238

Название	SPEC	SENS	BA	N	<i>In silico</i> (активные/ неактивные)	<i>In vitro</i> (активные/ неактивные)
Histamine H1 receptor antagonist	1,00	0,90	0,95	244	30/214	5/239
Acetylcholine M4 receptor antagonist	1,00	0,88	0,94	246	34/212	6/240
Histamine H2 receptor antagonist	1,00	0,87	0,94	242	34/208	4/238
Alpha 2b adrenoreceptor antagonist	1,00	0,86	0,93	242	50/192	18/224
Acetylcholine M3 receptor antagonist	1,00	0,84	0,92	248	48/200	10/238
Alpha 2a adrenoreceptor antagonist	1,00	0,84	0,92	244	58/186	22/222
Alpha 1d adrenoreceptor antagonist	0,88	0,94	0,91	244	28/216	16/228
5 Hydroxytryptamine uptake inhibitor	0,90	0,90	0,90	319	102/217	89/230
Glucocorticoid antagonist	0,93	0,86	0,90	261	48/213	15/246
5 Hydroxytryptamine 6 antagonist	0,86	0,94	0,90	254	39/215	29/225
5 Hydroxytryptamine 2A antagonist	0,85	0,93	0,89	247	33/214	20/227
Beta 2 adrenoreceptor antagonist	1,00	0,79	0,89	245	57/188	7/238
Alpha 2c adrenoreceptor antagonist	1,00	0,79	0,89	243	63/180	15/228
Adrenaline uptake inhibitor	0,86	0,89	0,88	292	76/216	58/234
Dopamine D1 antagonist	1,00	0,77	0,88	255	71/184	15/240
Opioid kappa receptor antagonist	1,00	0,76	0,88	253	68/185	11/242
Beta 1 adrenoreceptor antagonist	1,00	0,75	0,88	244	64/180	4/240
5 Hydroxytryptamine 2B antagonist	0,85	0,92	0,88	243	35/208	20/223
Angiotensin AT2 receptor antagonist	0,86	0,89	0,88	242	8/234	0/242
Bradykinin B2 receptor antagonist	0,86	0,89	0,88	242	2/240	0/242
Dopamine uptake inhibitor	0,93	0,80	0,87	300	102/198	58/242

Название	SPEC	SENS	BA	N	<i>In silico</i> (активные/ неактивные)	<i>In vitro</i> (активные/ неактивные)
Opioid mu receptor antagonist	0,90	0,84	0,87	270	66/204	30/240
Dopamine D2 antagonist	0,82	0,92	0,87	259	37/222	22/237
Cathepsin G inhibitor	1,00	0,73	0,87	246	66/180	1/245
Endothelin A receptor antagonist	0,93	0,80	0,87	245	16/229	0/245
Beta 3 adrenoreceptor antagonist	1,00	0,73	0,87	242	66/176	2/240
HMG CoA reductase inhibitor	0,91	0,80	0,86	254	58/196	11/243
Opioid delta receptor antagonist	0,82	0,86	0,84	278	54/224	22/256
Acetylcholine M5 receptor antagonist	0,80	0,88	0,84	245	37/208	10/235
Neuropeptide Y2 antagonist	0,80	0,88	0,84	242	64/178	0/242
Dopamine transporter inhibitor	0,81	0,86	0,83	300	81/219	58/242
MAO A inhibitor	1,00	0,63	0,82	402	201/201	83/319
Phosphodiesterase 5A inhibitor	0,71	0,93	0,82	272	35/237	24/248
5 Hydroxytryptamine 2C antagonist	0,74	0,91	0,82	250	41/209	27/223
Estrogen receptor alpha antagonist	1,00	0,61	0,80	384	229/155	128/256
Estrogen receptor beta antagonist	1,00	0,58	0,79	357	209/148	103/254
HERG channel blocker	0,65	0,93	0,79	314	48/266	43/271
Cyclooxygenase 1 inhibitor	0,96	0,58	0,77	497	285/212	141/356
Acetylcholine M1 receptor antagonist	0,93	0,57	0,75	250	114/136	14/236
Cyclooxygenase 2 inhibitor	0,89	0,60	0,74	639	347/292	192/447
Adenosine A2a receptor antagonist	0,50	0,94	0,72	242	16/226	4/238
Adenosine A1 receptor antagonist	0,50	0,92	0,71	243	21/222	4/239
Bile salt export pump inhibitor	1,00	0,43	0,71	186	113/73	15/171
Acetylcholinesterase inhibitor	0,92	0,46	0,69	693	457/236	212/481
Melanocortin 3 antagonist	0,92	0,46	0,69	242	0/242	0/242

Название	SPEC	SENS	BA	N	<i>In silico</i> (активные/ неактивные)	<i>In vitro</i> (активные/ неактивные)
Melanocortin 4 antagonist	0,92	0,46	0,69	242	0/242	0/242
Melanocortin 5 antagonist	0,92	0,46	0,69	242	0/242	0/242
5-Lipoxygenase inhibitor	0,99	0,34	0,66	209	176/33	114/95
Cannabinoid CB1 receptor antagonist	0,33	0,96	0,65	257	15/242	15/242
Butyrylcholinesterase inhibitor	0,98	0,30	0,64	214	174/40	89/125
Metalloproteinase-9 inhibitor	0,36	0,88	0,62	275	36/239	14/261
Interleukin 1 beta converting enzyme inhibitor	0,50	0,72	0,61	247	71/176	4/243
Secretase beta 1 inhibitor	0,50	0,70	0,60	158	59/99	56/102
MAO B inhibitor	0,95	0,25	0,60	134	109/25	43/91
Matrix metalloproteinase 1 inhibitor	0,25	0,93	0,59	267	22/245	12/255
Xanthine dehydrogenase inhibitor	0,96	0,19	0,58	103	88/15	28/75
Proto-oncogene tyrosine-protein kinase Fyn inhibitor	0,20	0,93	0,57	253	19/234	10/243
MAP kinase 3 inhibitor	0,33	0,81	0,57	242	46/196	3/239
Aromatase inhibitor	0,98	0,16	0,57	192	174/18	91/101
Breast cancer-resistant protein inhibitor	0,95	0,19	0,57	140	126/14	88/52
Sigma 1 receptor antagonist	0,13	0,97	0,55	254	9/245	15/239
Neurokinin 1 antagonist	0,13	0,97	0,55	242	10/232	0/242
Protein-tyrosine phosphatase 1B inhibitor	0,98	0,09	0,54	622	583/39	252/370
Lck kinase inhibitor	0,10	0,97	0,54	283	9/274	10/273
Xanthine oxidase inhibitor	0,96	0,11	0,54	103	94/9	28/75
Amyloid beta precursor protein antagonist	0,90	0,17	0,53	163	139/24	49/114

Название	SPEC	SENS	BA	N	<i>In silico</i> (активные/ неактивные)	<i>In vitro</i> (активные/ неактивные)
P-glycoprotein 1 inhibitor	0,99	0,05	0,52	273	265/8	149/124
MAP kinase 14 inhibitor	0,00	1,00	0,50	248	0/248	3/245
ErbB-2 antagonist	0,00	0,99	0,50	246	2/244	4/242
Protein phosphatase 2B inhibitor	0,00	1,00	0,50	242	0/242	0/242
Monophenol monooxygenase inhibitor	1,00	0,00	0,50	122	122/0	45/77
Epidermal growth factor antagonist	0,00	0,98	0,49	317	7/310	25/292
Neurokinin 2 antagonist	0,00	0,98	0,49	242	4/238	4/238
Cholecystokinin A antagonist	0,00	0,98	0,49	242	6/236	1/241
Epidermal growth factor receptor kinase inhibitor	0,00	0,97	0,48	317	10/307	25/292
ErbB-1 antagonist	0,08	0,89	0,48	317	35/282	25/292
Beta amyloid protein antagonist	0,61	0,35	0,48	163	104/59	49/114
CC chemokine 4 receptor antagonist	0,00	0,94	0,47	248	16/232	1/247
CC chemokine 5 receptor antagonist	0,00	0,94	0,47	245	4/241	0/245
CXC chemokine 2 receptor antagonist	0,00	0,94	0,47	242	18/224	0/242
Carbonic anhydrase II inhibitor	0,00	0,85	0,43	252	37/215	1/251
Thromboxane synthase inhibitor	0,14	0,73	0,43	245	66/179	7/238
CC chemokine 2 receptor antagonist	0,00	0,85	0,43	242	8/234	0/242
Lymphocyte kinase inhibitor	0,1	0,72	0,41	283	78/205	10/273
MAP kinase 1 inhibitor	0,00	0,80	0,40	242	48/194	3/239

Примечание - SPEC – чувствительность; SENS – специфичность; BA – сбалансированная точность; N – количество соединений.

Как видно из таблицы 14, в ряде случаев соотношение активных и неактивных для результатов *in vitro* и *in silico* экспериментов не сбалансировано (количество отрицательных примеров значительно выше положительных).

Отметим, что нами также не проведена процедура переобучения PASS с удалением используемых структур из обучающих выборок. Описанные процедуры не проводились, поскольку в рамках данного раздела тестируется не алгоритм работы PASS, а информационно-вычислительная платформа, которая будет доступна пользователям.

В результате анализа установлено, что среднее значение чувствительности составило 0,8, специфичности 0,6, а среднее значение сбалансированной точности прогноза составляет 0,7. Проведенный анализ, позволяет заключить что результаты компьютерного прогноза согласуются с данными, полученными в ходе *in vitro* экспериментов по определению биологической активности ПС. Это показывает адекватность и целесообразность использования компьютерных оценок биологической активности, рассчитываемых в рамках платформы Phyto4Health.

3.6 Проспективная валидация

Проспективная валидация информационно-вычислительной платформы Phyto4Health с целью возможности оценки фармакологического потенциала фитокомпонентов лекарственных растений России была выполнена совместно с исследователями из НМИЦ онкологии им. Н.Н. Блохина, ФГБНУ «Всероссийский НИИ лекарственных и ароматических растений» и Кафедры фармакогнозии ФГБОУ ВО «Санкт-Петербургский государственный химико-фармацевтический университет» Министерства здравоохранения РФ. Методики проведения экспериментов подробно рассмотрены в разделе 2.7.2 главы «Материалы и методы».

Информация о большинстве исследуемых индивидуальных фитокомпонентов содержится в разработанной нами базе данных Phyto4Health, однако в составе комплексных фармацевтических композиций были также вторичные метаболиты лекарственных растений, не входящих в Государственную Фармакопею РФ.

3.6.1 Оценка цитотоксического действия в отношении клеток рака мочевого пузыря двух вариантов препарата ФЛА

Нами проанализированы предложенные исследователями из НМИЦ онкологии им. Н.Н. Блохина два варианта фармацевтической композиции ФЛА¹. Фитохимический состав мажорного варианта содержал 22 соединения, а минорный вариант – 13 соединений. Структуры фитокомпонентов, установленных в фитохимическом составе фармацевтической композиции ФЛА, приведены в таблице 7 главы «Материалы и методы».

В результате анализа данных, представленных в Phyto4Health для фитокомпонентов «ФЛА», установлено, что число активностей, прогнозируемых при условии, что вероятность проявления существенно превышает вероятность отсутствия активности ($P_a - P_i > 0,5$), для минорного набора составляет 273, для мажорного набора — 324. Для 9 соединений мажорного набора, не входящих в минорный, при заданном пороге прогнозируется 165 биологических активностей. Соответствующие результаты приведены на рисунке 28.

Из рисунка 29 также видно, что для мажорного состава ФЛА прогнозируется более 1400 видов биологических активностей. Широкий набор потенциальных видов биологической активности, в данном случае, может быть обусловлен высоким разнообразием структур химических соединений, определяемых в составе вариантов ФЛА, что косвенно подтверждается оценками сходства между соединениями. Попарные оценки сходства между структурами фитокомпонентов ФЛА приведены на рисунке 29.

¹Работа выполнена в сотрудничестве с лабораторией иммунофармакологии НМИЦ онкологии им. Н.Н. Блохина (зав. лабораторией – д.б.н., профессор О.А. Бочарова).

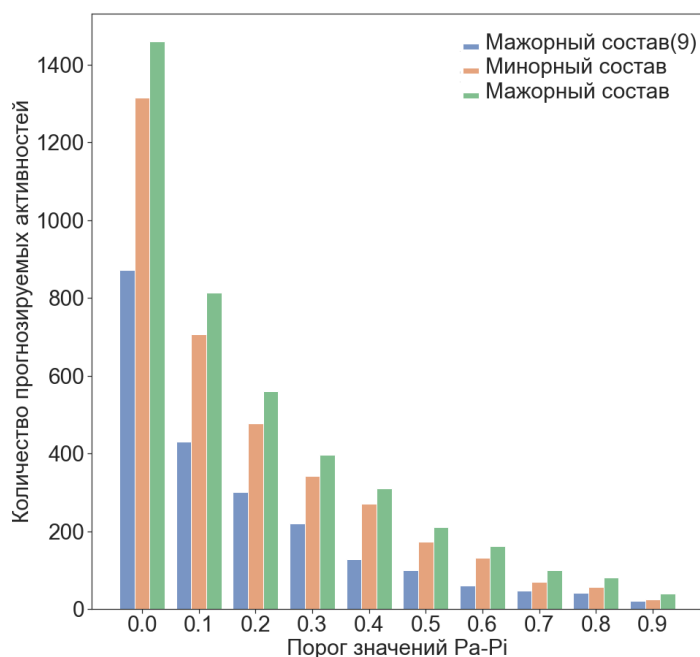


Рисунок 28 – Число уникальных активностей, прогнозируемых для мажорного и минорного состава фитокомпонентов при различных порогах отсечения значений Pa-Pi

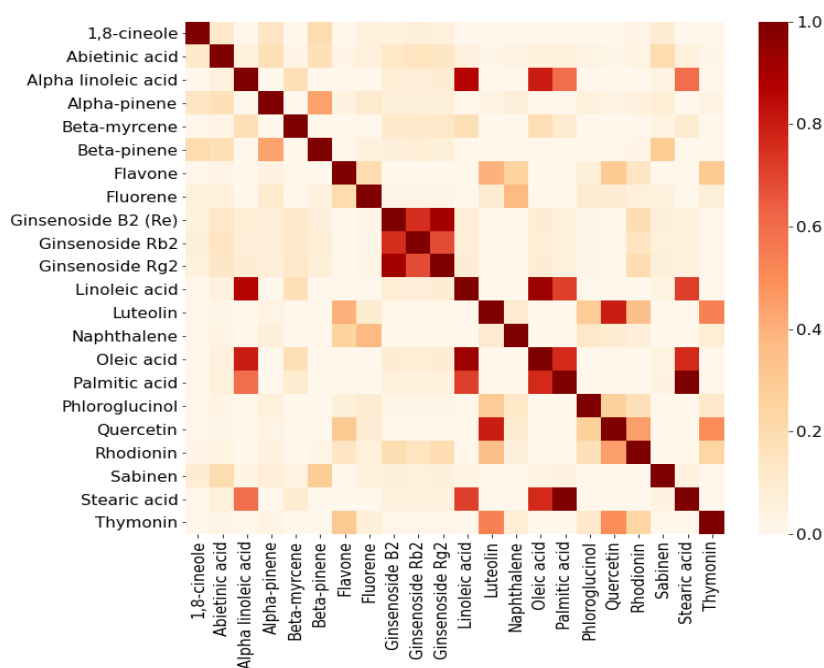


Рисунок 29 – Тепловая карта оценок сходства между уникальными соединениями мажорного и минорного вариантов фармкомпозиции ФЛА, полученных на основе MNA-дескрипторов и коэффициента Танимото

На тепловой карте оценок сходства (рисунок 29) видно, что представленные в ФЛА фитокомпоненты, в большинстве случаев характеризуются низкими оценками сходства между собой (среднее значение оценок сходства составляет

0,09 ± 0,08). Однако, среди попарных оценок присутствуют оценки, превышающие 0,6. Такие оценки получены для органических кислот (линоленовой, альфа-линоленовой, олеиновой, стеариновой и пальмитиновой) и представителей класса гинзенозидов (гинзенозид B2 (Re), гинзенозид Rb2 и гинзенозид Rg2).

Исследуемая фармацевтическая композиция разрабатывается как противоопухолевый препарат, поэтому в дальнейшем мы рассматривали только механизмы, которые, согласно базе знаний программы PharmaExpert, связаны с противоопухолевым эффектом. Тепловая карта вероятностных оценок для противоопухолевых механизмов действия, полученные для соединений фармацевтической композиции, приведены на рисунке 30.

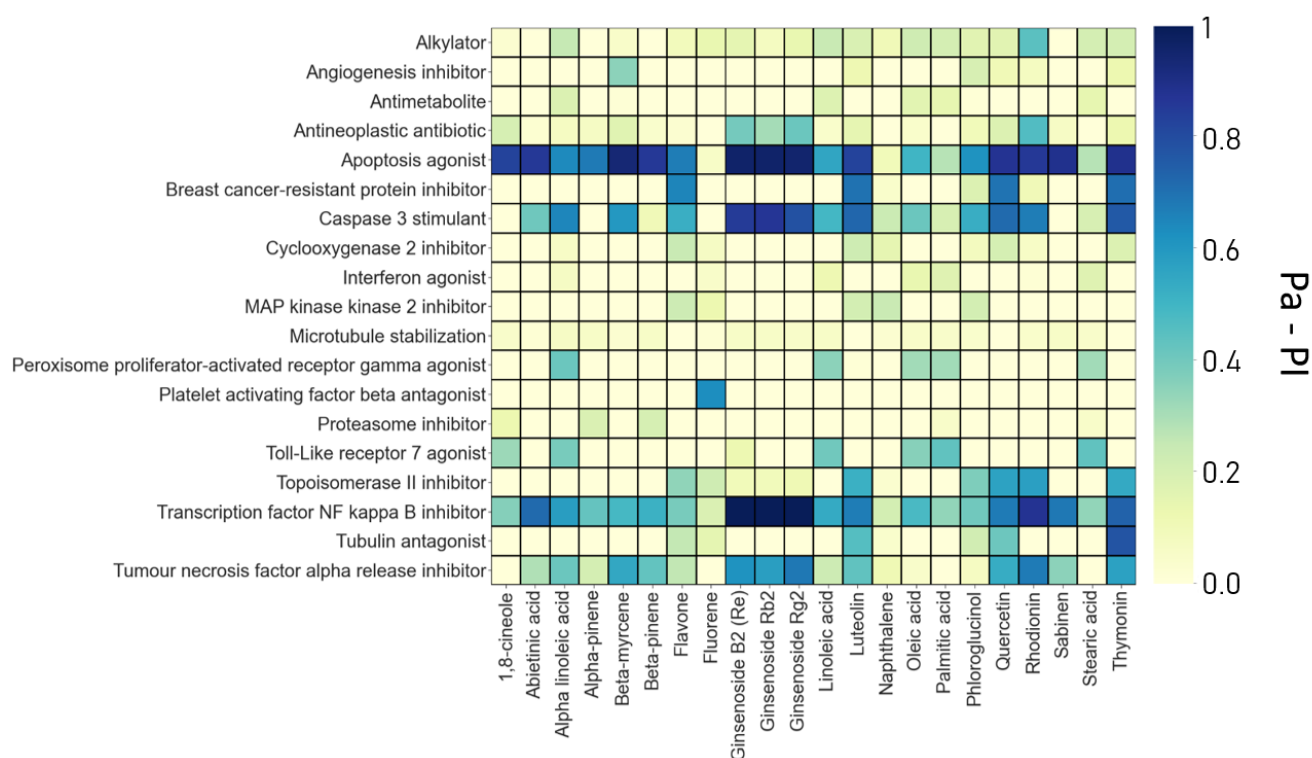


Рисунок 30 – Предсказанные значения оценок Ra-Pi для механизмов действия, ассоциированных с противоопухолевым эффектом. Отрицательные оценки сходства сведены к нулю

Как видно из рисунка 30, в результате компьютерного прогноза PASS, наивысшие оценки Ra-Pi были получены для трёх механизмов действия, ассоциированных с противоопухолевым эффектом. К ним относятся агонистический эффект по отношению к апоптозу (Apoptosis agonist), ингибирование транскрипционного фактора NF-kB (Transcription factor NF kappa B

inhibitor) и стимулирование каспазы-3 (Caspase 3 stimulant). Диаграммы рассеивания оценок Pa-Pi, полученных для указанных механизмов действия и исследуемых соединений, приведены на рисунке 31.

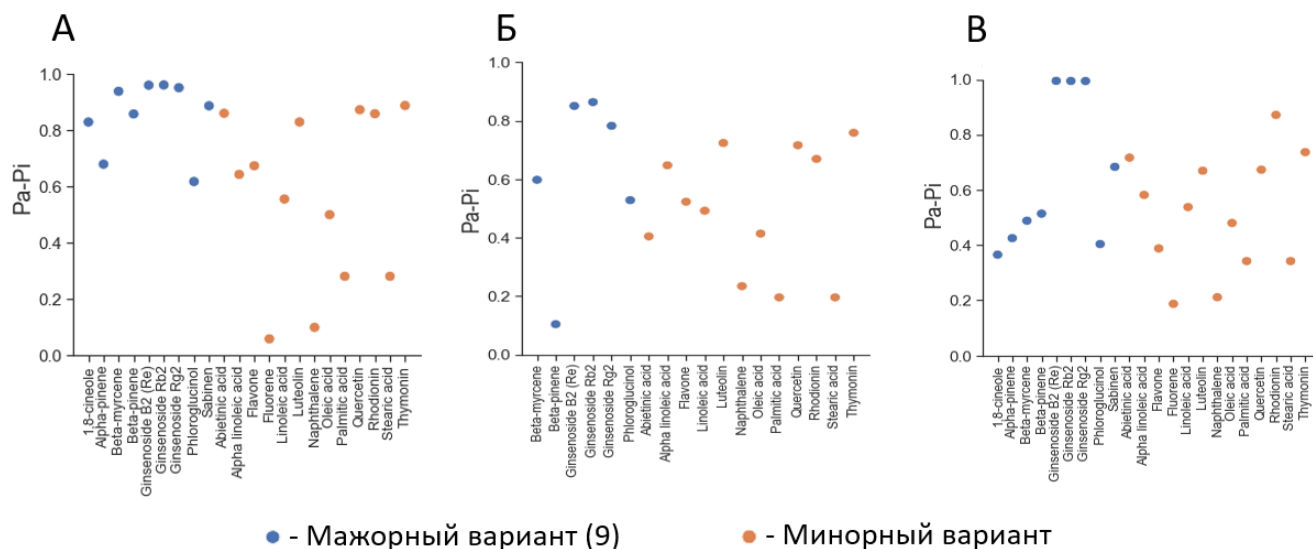


Рисунок 31 – Распределение соединений с наивысшими вероятностными оценками проявления активности между двумя вариантами фармацевтической композиции. А - агонист апоптоза, Б - стимулятор каспазы-3, В - ингибитор транскрипционного фактора NF-κB

Как видно из рисунка 31, соединения, для которых получены наивысшие оценки Pa-Pi ($Pa-Pi > 0,9$), преимущественно представлены среди 9 компонентов мажорного набора ФЛА. Значения полученных оценок, превышающих порог 0,5, приведены в таблице 15.

Таблица 15 – Наивысшие значения оценок Pa-Pi, полученные для механизмов Apoptosis agonist, Caspase 3 stimulant и Transcription factor NF kappa B inhibitor

Соединение	Механизм	Значения Pa-Pi
1,8-cineole	Apoptosis agonist	0,83
Abietinic acid	Apoptosis agonist	0,86
Alpha linoleic acid	Apoptosis agonist	0,64
Alpha-pinene	Apoptosis agonist	0,68

Соединение	Механизм	Значения Ра-Рi
Beta-myrcene	Apoptosis agonist	0,94
Beta-pinene	Apoptosis agonist	0,86
Flavone	Apoptosis agonist	0,67
Fluorene	Apoptosis agonist	0,06
Ginsenoside B2 (Re)	Apoptosis agonist	0,96
Ginsenoside Rb2	Apoptosis agonist	0,96
Ginsenoside Rg2	Apoptosis agonist	0,95
Linoleic acid	Apoptosis agonist	0,55
Luteolin	Apoptosis agonist	0,83
Phloroglucinol	Apoptosis agonist	0,62
Quercetin	Apoptosis agonist	0,87
Rhodionin	Apoptosis agonist	0,86
Sabinen	Apoptosis agonist	0,89
Thymonin	Apoptosis agonist	0,89
Alpha linoleic acid	Caspase 3 stimulant	0,65
Beta-myrcene	Caspase 3 stimulant	0,60
Flavone	Caspase 3 stimulant	0,52
Ginsenoside B2 (Re)	Caspase 3 stimulant	0,85
Ginsenoside Rb2	Caspase 3 stimulant	0,86
Ginsenoside Rg2	Caspase 3 stimulant	0,78
Luteolin	Caspase 3 stimulant	0,72
Phloroglucinol	Caspase 3 stimulant	0,53
Quercetin	Caspase 3 stimulant	0,72
Rhodionin	Caspase 3 stimulant	0,67
Thymonin	Caspase 3 stimulant	0,76
Abietinic acid	Transcription factor NF kappa B inhibitor	0,73

Соединение	Механизм	Значения Ра-Рi
Alpha linoleic acid	Transcription factor NF kappa B inhibitor	0,58
Beta-pinene	Transcription factor NF kappa B inhibitor	0,51
Flavone	Transcription factor NF kappa B inhibitor	0,39
Ginsenoside B2 (Re)	Transcription factor NF kappa B inhibitor	0,99
Ginsenoside Rb2	Transcription factor NF kappa B inhibitor	0,99
Ginsenoside Rg2	Transcription factor NF kappa B inhibitor	0,99
Linoleic acid	Transcription factor NF kappa B inhibitor	0,54
Luteolin	Transcription factor NF kappa B inhibitor	0,67
Quercetin	Transcription factor NF kappa B inhibitor	0,67
Rhodonin	Transcription factor NF kappa B inhibitor	0,87
Sabinen	Transcription factor NF kappa B inhibitor	0,68
Thymonin	Transcription factor NF kappa B inhibitor	0,74

По результатам анализа результатов прогноза PASS с применением компьютерной программы PharmaExpert, описанные механизмы ассоциированы с противоопухолевым действием в отношении рака мочевого пузыря, поэтому данное направление исследований было выбрано приоритетным. Экспериментальное тестирование было выполнено на клеточной линии рака мочевого пузыря человека RT112.

В результате *in vitro* экспериментов продемонстрирована более высокая цитотоксичность мажорной фармкомпозиции ФЛА в отношении клеток рака мочевого пузыря человека линии RT-112. Также в экспериментах *in vitro* двумя методами — двойным окрашиванием аннексином V и йодидом пропидия, а также детектированием в клетках активной каспазы-3 – подтверждена гибель опухолевых клеток рака мочевого пузыря по механизму апоптоза. Полученные экспериментальные данные *in vitro* соответствуют результатам компьютерного прогноза *in silico*, и указывают на преимущества мажорного состава фармкомпозиции ФЛА [186].

3.6.2 Оценка спектров биологической активности соединений для фитокомпонентов *Rubus chamaemorus* L.

На базе Кафедры фармакогнозии ФГБОУ ВО «Санкт-Петербургский государственный химико-фармацевтический университет» был выполнен анализ фитохимического состава листьев морошки обыкновенной (*Rubus chamaemorus* L.)² [177].

Результаты компьютерного прогноза для идентифицированных соединений позволили установить перспективные направления дальнейших исследований. В частности, установлено, что для соединений под номерами **1** (4-О- α -L-арабинофуранозилэллаговая кислота), **3** (кверцетин-3-О- β -D-глюкуронид), **4** (кверцетин-3-О- β -D-2''-галлоилглюкуронид) и **5** (кемпферол-3-О- β -D-глюкуронид) (структурные формулы приведены на рисунке 11 в разделе «Материалы и методы») при пороге $R_a > 0,7$ прогнозируется антитромботический эффект. Для соединения под номером **2** ((-)-эпикатехин), антитромботический эффект прогнозируется при $R_a > 0,3$. Значения R_a и P_i , рассчитанные для соединений *Rubus chamaemorus* L., приведены в таблице 16.

В результате исследования действия индивидуальных компонентов *Rubus chamaemorus* L. на систему гемостаза *in vitro* с использованием крови здоровых доноров (мужчин, от 18 до 34 лет) были впервые установлены их антикоагуляционные и антиагрегантные свойства. Показано, что соединения **1** и **3** проявляются антиагрегантную активность, близкую к таковой для ацетилсалициловой кислоты. Соединения **1**, **3** и **5** полностью подавляли активацию тромбоцитов (в изученных концентрациях) и влияли на уровень экспрессии CD62 в то время, как соединения **2** и **4** на уровень экспрессии CD62 не влияли [188].

²Работы выполнены в сотрудничестве с кафедрой фармакогнозии ФГБОУ ВО «Санкт-Петербургский государственный химико-фармацевтический университет» (к.ф.н. А.К. Уэйли, к.б.н. В.Г. Лужанин)

Таблица 16 – Результаты прогноза антитромботического эффекта для фитокомпонентов *Rubus chamaemorus* L

№	Pa	Pi
1	0,757	0,005
2	0,306	0,082
3	0,749	0,005
4	0,749	0,005
5	0,774	0,004

3.6.3 Исследование влияния мажорных компонентов *Cichorium intybus* L. на активность цитохрома P450 и глутатионтрансферазы

Предметом исследования³ были выбраны два химических соединения, выделенные из травы цикория обыкновенного – цикориевая и хлорогеновая кислоты.

В рамках данной работы исследовался гепатопротекторный эффект надземной части травы цикория обыкновенного, а также влияние на систему биотрансформации-детоксикации (CYP450 и глутатионтрансферазы (GST)) [189]. Важно заметить, что гепатопротекторный эффект цикориевой и хлорогеновой кислот был изучен ранее и, согласно литературным данным, обусловлен их антиоксидантной активностью [190, 191]. Анализ результатов компьютерного прогноза PASS показал, что гепатопротекторный эффект исследуемых химических соединений прогнозируется со значениями Pa-Pi равными 0,35 и 0,45 для цикориевой и хлорогеновой кислот, соответственно. Кроме того, среди механизмов действия, ассоциированных с гепатопротекторным эффектом обнаружен механизм антиоксидантного действия – «ловушка свободных радикалов» (хлорогеновая кислота – 0,81, цикориевая кислота 0,68). Таким образом, результаты прогноза гепатопротекторного действия для цикориевой и хлорогеновой кислот при пороге $Pa > Pi$ согласуются с известными данными.

³ Работа выполнена совместно с Центром доклинических исследований ФГБНУ «Всероссийский НИИ лекарственных и ароматических растений» (н.с., к.б.н. И.А. Лупанова).

Результаты прогноза видов биологической активности, ассоциированных с активацией ферментов системы биотрансформации-детоксикации приведены в таблице 17.

Таблица 17 – Оценки Pa-Pi рассчитанные для видов биологической активности, ассоциированных с активацией ферментов системы биотрансформации-детоксикации

Вид биологической активности, прогнозируемой PASS	Цикориевая кислота	Хлорогеновая кислота
MGST1 expression enhancer	0,73	0,48
CYP26B1 expression enhancer	0,68	0,57
CYP4A11 expression enhancer	0,63	0,09
CYP2C9 expression enhancer	0,62	0,67
CYP1A1 expression enhancer	0,60	0,57
CYP17A1 expression enhancer	0,60	0,45
CYP3A7 expression enhancer	0,60	0,32
CYP3A4 expression enhancer	0,54	0,48
CYP1B1 expression enhancer	0,48	0,21
CYP1A inducer	0,47	0,16
CYP2B6 inducer	0,46	0,44
CYP2E1 inducer	0,43	0,26
CYP11A1 expression enhancer	0,41	0,30
CYP2E1 expression enhancer	0,41	0,26
CYP2B6 expression enhancer	0,37	0,38
CYP19A1 expression enhancer	0,37	0,31
CYP2A6 expression enhancer	0,35	0,30
CYP3A5 expression enhancer	0,33	0,36
CYP24A1 expression enhancer	0,33	0,28
CYP1A1 inducer	0,29	0,09
CYP2C9 inducer	0,25	0,28
CYP1A2 expression enhancer	0,25	0,04
CYP11B2 expression enhancer	0,24	0,12
CYP21A2 expression enhancer	0,21	0,18
CYP26A1 expression enhancer	0,14	0,11
CYP11B1 expression enhancer	0,12	0,02
CYP3A4 inducer	0,10	0,28
CYP3A inducer	0,09	0,28

Вид биологической активности, прогнозируемой PASS	Цикориевая кислота	Хлорогеновая кислота
CYP3A5 inducer	0,06	0,06

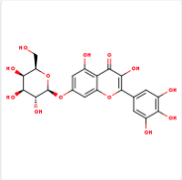
3.6.4 БД SistemатX

С целью расширения возможностей практического использования фармакологического потенциала растений Бразилии, по просьбе бразильских коллег, нами был прогноз *in silico* спектров биологической активности для природных соединений из БД SistemатX. Результаты прогноза биологической активности 8 940 структур химических соединений были включены в эту базу данных [121]. Пример представления результатов прогноза PASS в БД SistemатX приведен на рисунке 32.

SISTEMATX About Team Contact Help How to Cite Login

3,5-dihydroxy-7-[(2S,3R,4S,5R,6R)-3,4,5-trihydroxy-6-(hydroxymethyl)oxan-2-yl]oxy}-2-(3,4,5-trihydroxyphenyl)-4H-chromen-4-one

← Back Amplify



NMR 1H Data
NMR 1H by SPINUS WEB
NMR 13C Data
PASS Data
PLATO Data
PubChem Reference
ChEMBL Reference
ChemSpider Reference
MolpredictX Reference

PASS DATA data for compound 15059

Number of MNA descriptors: Overall - 44, New - 0

Activity	Prediction
Monophenol monooxygenase inhibitor	99%
Cardioprotectant	99%
Neuroprotector	99%
Antimutagenic	99%
Carbonic anhydrase XIII inhibitor	98%
Membrane integrity agonist	98%
APOB expression inhibitor	98%
Anticarcinogenic	97%
Beta-N-acetylhexosaminidase inhibitor	97%
Free radical scavenger	97%
Estrogen-related receptor alpha agonist	96%
Chemopreventive	96%
Iodide peroxidase inhibitor	95%
Transferase inhibitor	95%
Hemostatic	95%
Carbonyl reductase [NADPH] inhibitor	94%
IRS1 expression inhibitor	94%
CDP-glycerol glycerophosphotransferase inhibitor	93%
Hypoglycemic	93%

Рисунок 32 – Пример результатов прогноза PASS в базе данных SistemатX

3.7 Компьютерный анализ фармакологического потенциала фармакопейных растений

Ведущей мотивацией для разработки ИВП является содействие широкому кругу научных сотрудников, работающих в области фармацевтической химии и других смежных специальностей, в выборе наиболее перспективных путей исследования фитокомпонентов (или смесей фитокомпонентов) лекарственных растений. В настоящем разделе нами рассматривается возможность использования компьютерных оценок для анализа фармакологического потенциала фитокомпонентов 25, наиболее охарактеризованных с точки зрения фитохимического состава по данным Phyto4Health фармакопейных растений России. Перечень растений приведен в разделе 3.2.

По данным Phyto4Health, отобранные растения, суммарно содержат в своем фитохимическом составе 1820 уникальных химических соединений. Число фармакологических эффектов, прогнозируемых для данных соединений при пороге $P_a > P_i$, приведены на рисунке 33.

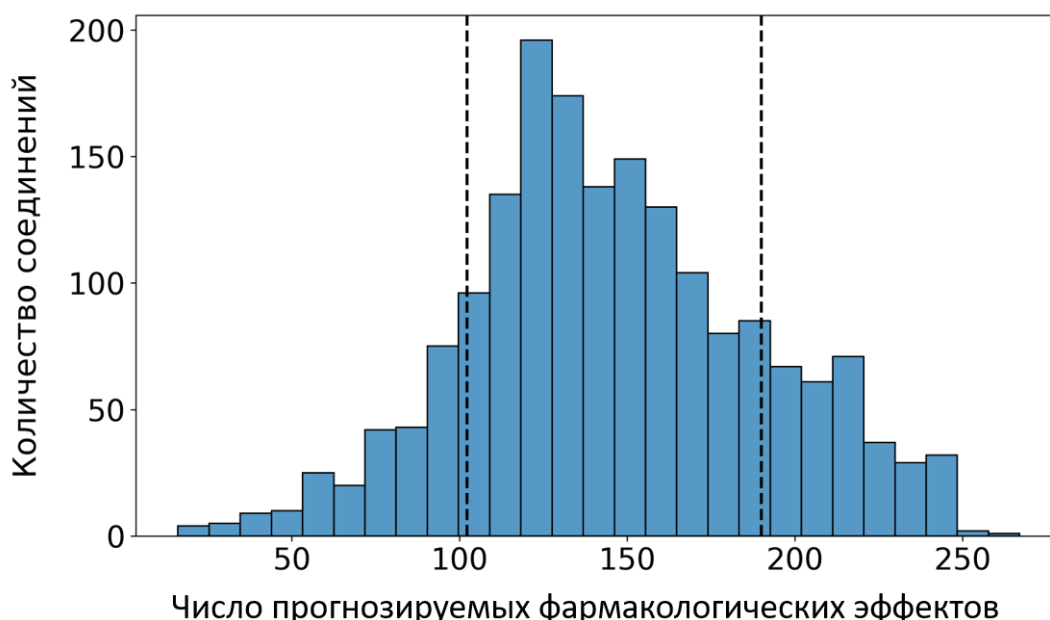


Рисунок 33 – Распределение количества прогнозируемых фармакологических эффектов для фитокомпонентов 25 отобранных растений

Из рисунка 33 видно, что для большинства химических соединений (1224, 68%) прогнозируется от 103 до 190 фармакологических эффектов при пороге $P_a > P_i$. Для 14% (261) химических соединений прогнозируется от 16 до 103 фармакологических эффектов, а для 18% (335) химических соединений прогнозируется от 190 до 267 фармакологических эффектов. Кроме того, важно заметить, что положительные результаты компьютерного прогноза охватывают 352 из 358 фармакологических эффектов, доступных для прогноза компьютерного программой PASS Refined 2022 (см. таблицу 7). Полученные данные косвенно указывают на высокое химическое разнообразие рассматриваемого набора структур, что указывает на многочисленные возможности их исследования в качестве «кандидатов» терапевтических средств для лечения различных нозологий.

Рассматриваемые химические структуры используются в составе фармацевтических композиций, разработанных на основе отобранных фармакопейных растений. Поэтому дальнейший анализ будет проводится с учётом качественного фитохимического состава 25 отобранных растений. Число фармакологических эффектов, прогнозируемых для отдельных соединений с учётом содержания этих соединений в отдельных частях растений приведено на рисунке 34.

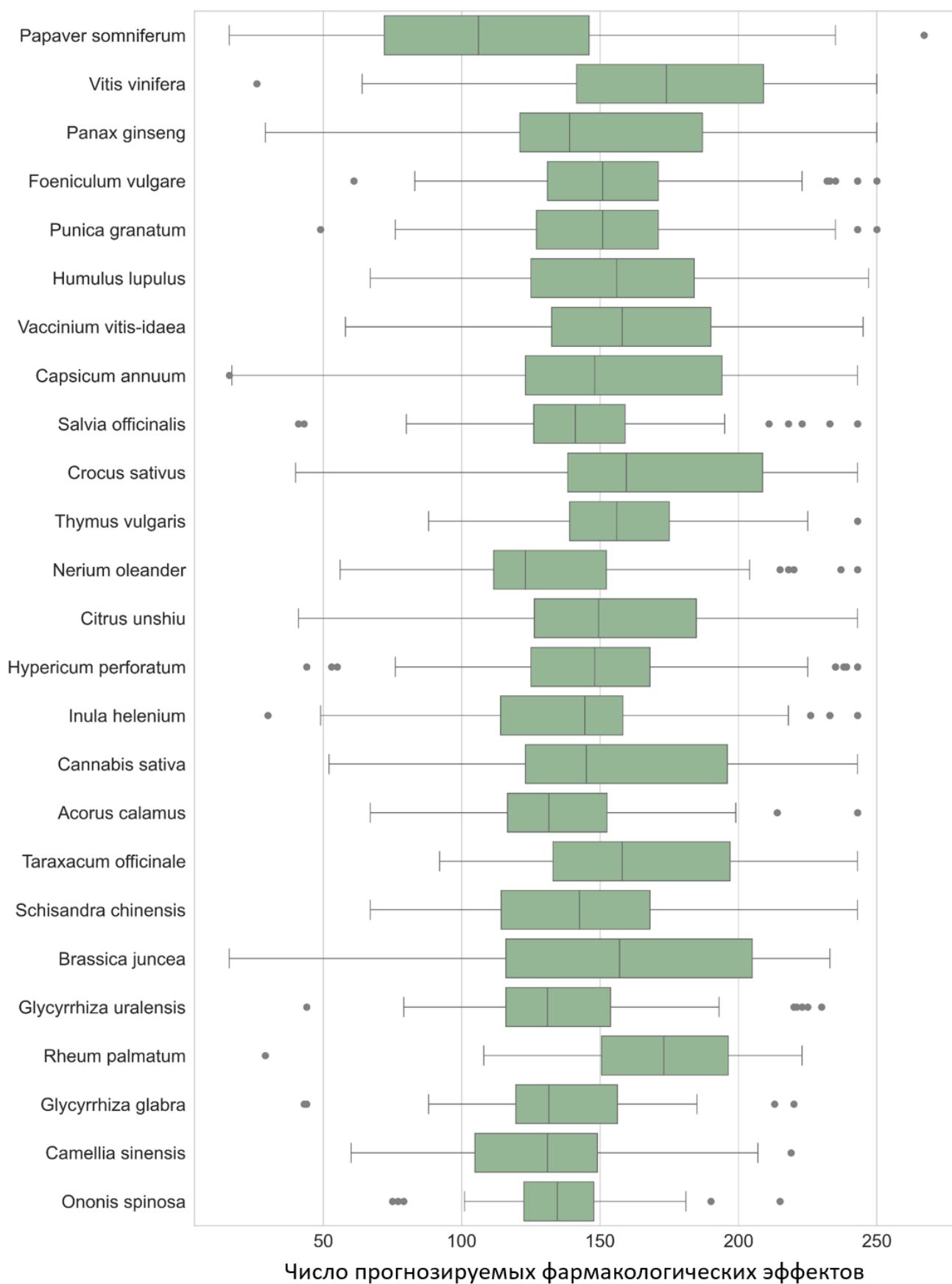


Рисунок 34 – Количество фармакологических эффектов, прогнозируемых при пороге $P_a > P_i$ для фитокомпонентов 25 растений

Как видно, из рисунка 34, медиана количества прогнозируемых фармакологических эффектов для фитокомпонентов рассматриваемых растений не превышает 174. Соединения, характеризующие минимальным количеством прогнозируемых фармакологических эффектов, представлены в фитохимическом составе 3-х растений: Горчица сарептская (лат. *Brassica juncea*), мак снотворный (лат. *Papaver somniferum*), перец стручковый (лат. *Capsicum annuum*). Структурные формулы этих соединений приведены на рисунке 35.

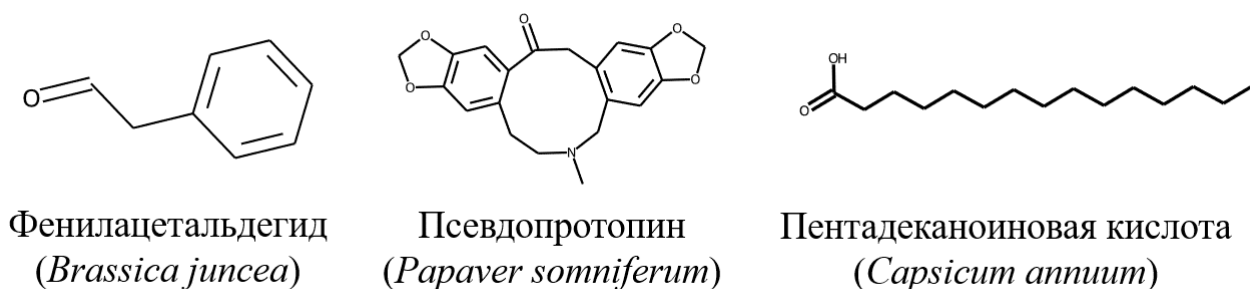


Рисунок 35 – Структурные формулы фитокомпонентов, характеризующихся наименьшим количеством прогнозируемых фармакологических эффектов

Для того чтобы оценить степень изученности фармакологического потенциала отдельных лекарственных растений, нами был выполнен автоматизированный поиск текстов научных публикаций в информационной системе PubMed. Запрос в информационную систему был сформирован следующим образом: «(название фармакологического эффекта) AND (название растения)». Важно заметить, что такой подход имеет ряд недостатков, в числе которых отсутствие учёта синонимов названий фармакологических эффектов и названий растений. Однако, он позволяет в автоматическом режиме сформировать большой набор релевантных публикаций.

В результате выполнения ранее описанной процедуры, удалось сформировать набор из 111276 релевантных публикаций, представленных в информационной системе PubMed. Соотношение прогнозируемых и известных из литературы фармакологических эффектов, описанных в публикациях, приведено на рисунке 36.

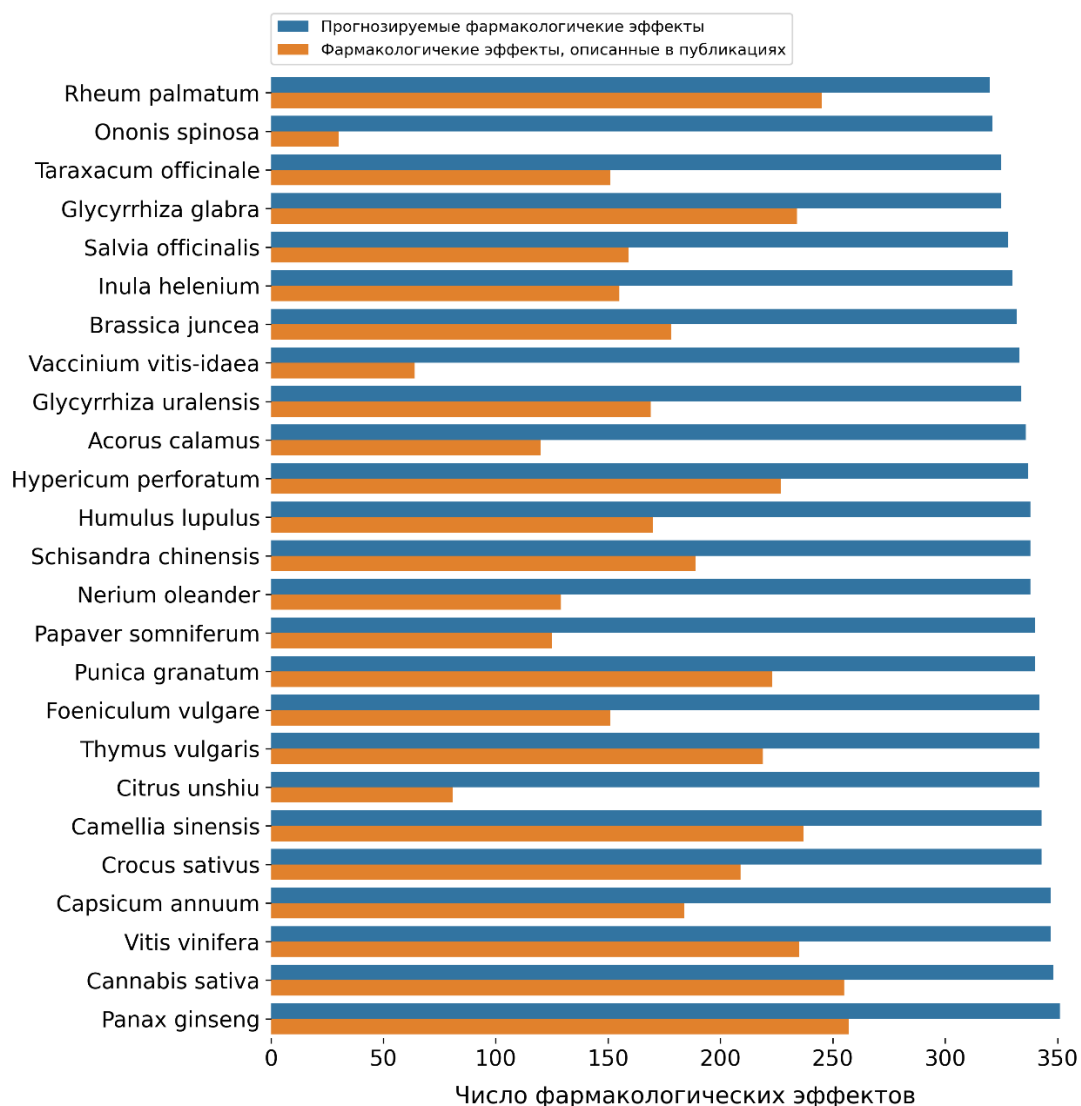


Рисунок 36 – Соотношение изученных и прогнозируемых фармакологических эффектов 25 фармакопейных растений России

Как видно на рисунке 36, среди 25 видов фармакопейных растений России наибольшим количеством известных фармакологических эффектов характеризуются женьшень обыкновенный (лат. *Panax ginseng*), конопля посевная (лат. *Cannabis sativa*), ревень пальчатый (лат. *Rheum palmatum*), камелия китайская (лат. *Camellia sinensis*), виноград культурный (лат. *Vitis vinifera*), солодка голая (лат. *Glycyrrhiza glabra*). Высокая степень изученности возможных фармакологических эффектов этих растений не вызывает удивления, поскольку, они имеют длительный опыт традиционного применения, а методы подготовки и стандартизации лекарственного сырья представлены в фармакопеях различных стран. Кроме того, эти растения широко применяются в составе пищевых добавок. В свою очередь,

отдельный интерес вызывают растения, для которых идентифицировано наименьшее количество известных фармакологических эффектов. Как видно из рисунка 36, к таким растениям относятся: стальник колючий (лат. *ononis spinosa*), брусника (лат. *vaccinium vitis-idaea*) и мандарин уншиу (лат. *citrus unshiu*). Сравнительно низкое количество исследованных фармакологических эффектов можно объяснить сложностями культивирования данных растений и последующей наработки растительного сырья. Кроме того, важно заметить, что сравнительно низкую изученность фармакологических эффектов мандарина уншиу можно объяснить тем, что данное растение применяется преимущественно в парфюмерно-косметической промышленности в виде эфирных масел и сывороток. Качественная характеристика изученности фармакологических эффектов для отдельных растений приведена на рисунках 37 и 38.

Как видно из рисунков 37 и 38, значительная часть фармакологических эффектов, прогнозируемых для фитокомпонентов ОЛЛР, остается не исследованной. В частности, среди представленных фармакологических эффектов обнаруживаются те, которые не были исследованы для отобранных растений. Общее количество таких эффектов составило 40. В их числе фармакологические эффекты для лечения микоплазмоза, глаукомы, антигипотензивный эффект, лечения эмфиземы, фармакологические эффекты для лечения симптомов болезни Паркинсона и другие.

Далее, для более детального анализа для каждого растения, мы отобрали 50 фармакологических эффектов, которые прогнозируются для наибольшего количества их фитокомпонентов при пороге $P_a > 0.5$. Соответствующие данные приведены в таблице 18.



Рисунок 37 – Тепловая карта сопоставления результатов компьютерного прогноза с опубликованными экспериментальными данными по биологической активности экстрактов ОЛРР

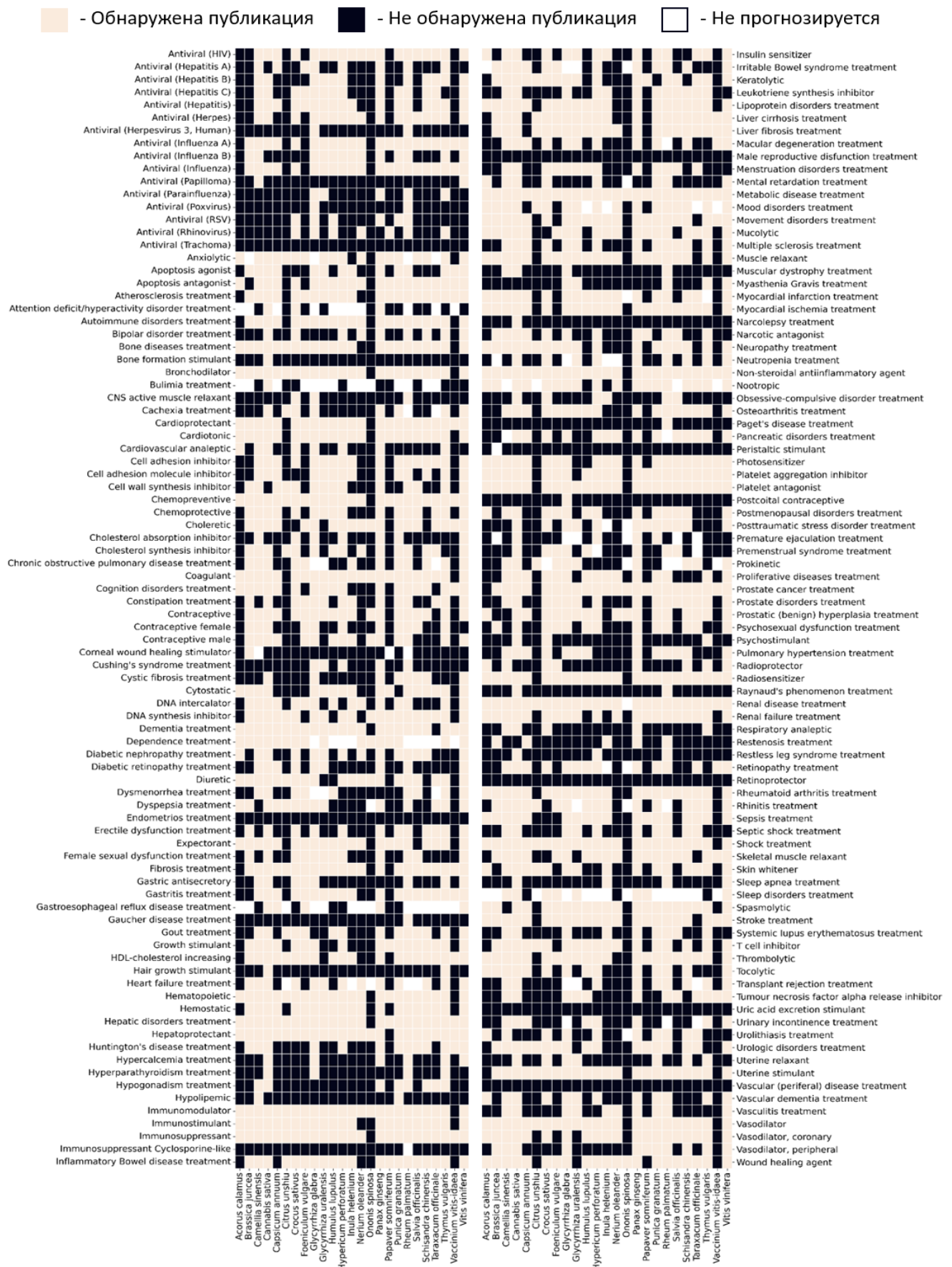


Рисунок 38 – Тепловая карта сопоставления результатов компьютерного прогноза с опубликованными экспериментальными данными по биологической активности экстрактов ОЛРР

Таблица 18 – Наиболее перспективные направления исследований фитокомпонентов ОЛЛР по результатам компьютерного прогноза

ОЛРР	Фармакологический эффект ($P_a > 0.5$)	Количество соединений
<i>Panax ginseng</i>	Antihypercholesterolemic	164
<i>Glycyrrhiza uralensis</i>	Antihypercholesterolemic	137
<i>Panax ginseng</i>	Antipsoriatic	124
<i>Glycyrrhiza uralensis</i>	Proliferative diseases treatment	110
<i>Humulus lupulus</i>	Antihypercholesterolemic	105
<i>Glycyrrhiza uralensis</i>	Antineoplastic (brain cancer)	103
<i>Panax ginseng</i>	Antileukemic	96
<i>Vitis vinifera</i>	Cardiovascular analeptic	92
<i>Schisandra chinensis</i>	Antihypercholesterolemic	88
<i>Punica granatum</i>	Antihypercholesterolemic	87
<i>Vitis vinifera</i>	Antihypoxic	86
<i>Panax ginseng</i>	Hypolipemic	84
<i>Humulus lupulus</i>	Cardiovascular analeptic	83
<i>Thymus vulgaris</i>	Antihypercholesterolemic	80
<i>Foeniculum vulgare</i>	Respiratory analeptic	79
<i>Humulus lupulus</i>	Skin whitener	79
<i>Humulus lupulus</i>	Antipruritic, allergic	78
<i>Foeniculum vulgare</i>	Antipruritic, allergic	77
<i>Capsicum annuum</i>	Cardiovascular analeptic	76
<i>Glycyrrhiza uralensis</i>	Respiratory analeptic	75
<i>Capsicum annuum</i>	Antihypercholesterolemic	74
<i>Schisandra chinensis</i>	Cardiovascular analeptic	74
<i>Taraxacum officinale</i>	Antihypercholesterolemic	74
<i>Foeniculum vulgare</i>	Antihypercholesterolemic	74
<i>Vitis vinifera</i>	Respiratory analeptic	73
<i>Humulus lupulus</i>	Radioprotector	73
<i>Foeniculum vulgare</i>	Cardiovascular analeptic	71
<i>Schisandra chinensis</i>	Antipsoriatic	70
<i>Capsicum annuum</i>	Antipsoriatic	70
<i>Citrus unshiu</i>	Antipsoriatic	70
<i>Citrus unshiu</i>	Respiratory analeptic	70
<i>Foeniculum vulgare</i>	Antipsoriatic	70
<i>Thymus vulgaris</i>	Antipsoriatic	69
<i>Citrus unshiu</i>	Analeptic	69
<i>Humulus lupulus</i>	Respiratory analeptic	67
<i>Citrus unshiu</i>	Cardiovascular analeptic	67
<i>Foeniculum vulgare</i>	Chemoprotective	66
<i>Humulus lupulus</i>	Antihypoxic	66
<i>Citrus unshiu</i>	Antihypercholesterolemic	66

ОЛРР	Фармакологический эффект ($P_a > 0.5$)	Количество соединений
<i>Hypericum perforatum</i>	Antihypercholesterolemic	64
<i>Capsicum annuum</i>	Vasodilator, peripheral	63
<i>Citrus unshiu</i>	Antithrombotic	61
<i>Punica granatum</i>	Cardiovascular analeptic	61
<i>Rheum palmatum</i>	Antihypercholesterolemic	61
<i>Citrus unshiu</i>	Chemoprotective	61
<i>Hypericum perforatum</i>	Antipruritic, allergic	60
<i>Taraxacum officinale</i>	Radioprotector	60
<i>Citrus unshiu</i>	Antipruritic, allergic	59
<i>Capsicum annuum</i>	Radioprotector	59
<i>Glycyrrhiza uralensis</i>	Radioprotector	59

Как видно из приведенных в таблице 18 данных, в число наиболее перспективных направлений исследования ОЛРР относятся 13 фармакологических эффектов, которые потенциально могут быть исследованы для 16 ОЛРР.

Приведенные на рисунках 37 и 38 сведения и данные из таблицы 18 указывают, что фармакологический потенциал ОЛРР остается в достаточной степени не изученным, а представленные в Государственной фармакопее Российской Федерации растения характеризуются высоким разнообразием возможных путей их исследования.

ЗАКЛЮЧЕНИЕ

В последние годы произошло стремительное развитие методов машинного обучения, позволяющих выполнять прогноз различных свойств химических соединений *in silico*. Отчасти это связано с агрегацией большого объема данных в различных БД, предоставляющих сведения о структуре и свойствах ПС, которые могут быть использованы для построения математических моделей. Однако применимость этих БД для анализа фармакологического потенциала лекарственных растений России ограничена несколькими факторами. В их числе высокое видовое разнообразие произрастающих растений и как следствие разнообразие структур фитокомпонентов растений. Кроме того, большинство существующих источников не содержат исчерпывающей информации о фитохимическом составе отдельных частей растений. Приведенные недостатки ограничивают применение существующих ресурсов для анализа фитохимического состава ОЛРР. Поэтому нами была создана информационно-вычислительная платформа Phyto4Health, обеспечивающая доступ к информации о фитохимическом составе отдельных частей растений России, представленных в 14-м издании Государственной Фармакопеи.

Платформа Phyto4Health предоставляет информацию о лекарственных растениях, структуре их фитокомпонентов, результатах оценки биологической активности *in vitro*, а также значениях молекулярных свойств для 3 128 различных ПС. Вычислительный компонент платформы позволяет пользователю получить на основе оценок *in silico* сведения о вероятностных профилях фармакологических эффектов, механизмов действия, нежелательных побочных эффектах, а также цитотоксическом эффекте в отношении нормальных и опухолевых клеточных линий.

При сравнении нашего массива химических структур с БД фитокомпонентов растений других географических регионов установлено, что собранный нами набор ПС охватывает отдельную область химического пространства, что позволяет существенно обогатить ранее доступную информацию, необходимую при проведении вычислительных экспериментов.

В результате ретроспективной валидации компьютерного прогноза установлено, что оценки, полученные *in silico*, хорошо согласуются с реальными данными лабораторных экспериментов.

В лабораторных экспериментах показана эффективность использования созданных нами «инструментов» для проведения виртуального скрининга и анализа фармакологического потенциала комплексных смесей фитокомпонентов (экстрактов). На примере препарата ФЛА было выявлено новое направление его исследований в качестве препарата для терапии рака мочевого пузыря. Для индивидуальных компонентов *Rubus chamaemorus* L. впервые был предсказан антитромботический эффект, который был подтвержден данными *in vitro* экспериментов. Другим примером являются оксикоричные кислоты, представленные в фитохимическом составе *Cichorium intybus* L. Для двух мажорных фитокомпонентов был выполнен прогноз их влияния на CYP450 и GST, полученные результаты также были подтверждены в рамках лабораторных экспериментов.

Таким образом, нами впервые создана оригинальная свободно доступная в сети Интернет информационно-вычислительная платформа Phyto4Health, предоставляющая сведения о структуре и биологической активности фитокомпонентов официальных лекарственных растений России, а также содержащая вычислительные инструменты для прогнозирования профилей фармакологических эффектов, механизмов действия, нежелательных побочных эффектов и цитотоксического действия в отношении опухолевых и нормальных клеточных линий. Использование Phyto4Health позволяет выявить наиболее перспективные направления практического использования фармакологического потенциала доступного на территории России растительного сырья.

ВЫВОДЫ

1. Собран массив информации, содержащий сведения о фитохимическом составе 233 растений, включенных в Государственную фармакопею 14-го издания, и используемых в качестве растительного сырья на территории РФ. Массив содержит 3 128 структурных формул уникальных фитокомпонентов, охарактеризованных с точки зрения их молекулярных свойств, содержания в отдельных частях растений, и результатах исследования биологической активности *in vitro*.
2. Реализована информационно-вычислительная платформа Phyto4Health, предоставляющая пользователям возможность получить известные сведения о фитохимическом составе официальных растений России, а также свойствах их отдельных фитокомпонентов. Вычислительные инструменты платформы позволяют осуществить поиск структурных аналогов, а также выполнить прогноз вероятных видов биологической активности фитокомпонентов.
3. Сопоставление представленных в БД фитокомпонентов с фитохимическим составом лекарственных растений других стран показало, что содержащаяся в нашей БД информация во многих случаях является уникальной и отражает широкое разнообразие структурного и фармакологического потенциала лекарственных растений России.
4. На примере фармацевтических композиций (мажорный и минорный варианты ФЛА), двух фитокомпонентов *Cichorium intybus* L. и пяти фитокомпонентов *R. Chamaemorus* продемонстрирована возможность использования накопленного массива данных и компьютерного прогноза для анализа фармакологического потенциала и выявления новых путей исследования. Полученные результаты компьютерного прогноза подтверждены в экспериментах *in vitro*.

СПИСОК ОПУБЛИКОВАННЫХ РАБОТ ПО ТЕМЕ ДИССЕРТАЦИИ

Статьи в рецензируемых научных журналах

1. **Ионов Н.С.**, Барышникова М.А., Бочаров Е.В., Погодин П.В., Лагунин А.А., Филимонов Д.А., Карпова Р.В., Косоруков В.С., Стилиди И.С., Матвеев В.Б., Бочарова О.А., Поройков В.В. Возможности оценок *in silico* для разработки фармкомпозиции фитоадаптоген, цитотоксичной для клеток рака мочевого пузыря // Биомедицинская химия. – 2021. – Т. 67. – №. 3. – С. 278-288.
2. Уэйли А.К., Понкротова А.О., Орлова А.А., Серебряков Е.Б., Смирнов С.Н., Прокш П., **Ионов Н.С.**, Поройков В.В., Лужанин В.Г. Фитохимический анализ вторичных метаболитов полифенольной природы в листьях морошки обыкновенной (*Rubus Chamaemorus* L.) // Химико-фармацевтический журнал. – 2021. – Т. 55. – №. 3. – С. 22-27.
3. Costa R., Lucena L., Silva L., Zocolo G., Herrera-Acevedo C., Scotti L., Da Costa F., **Ionov N.**, Poroikov V., Muratov E., Scotti M. SistemX: Web portal of natural products // Journal of Chemical Information and Modeling. – 2021. – Т. 61. – №. 6. – С. 2516-2522. (Q1)
4. Лупанова, И. А., Мизина, П. Г., **Ионов, Н. С.**, Поройков, В. В., Хлебников, А. И., Мартынич, И. А. Влияние цикориевой и хлорогеновой кислот из *Cichorium intybus* L. на активность цитохрома P450 и глутатионтрансферазы // Биофармацевтический Журнал. – 2022. – Т. 14. – №. 5. – С. 8-18.
5. **Ionov N.**, Druzhilovskiy, D., Filimonov, D., Poroikov, V. Phyto4Health: Database of Phytocomponents from Russian Pharmacopoeia Plants. // Journal of Chemical Information and Modeling. – 2023. – Т. 63. – №. 7. – С. 1847-1851. (Q1)
6. Bocharova O.A., **Ionov N.S.**, Kazeev I.V., Shevchenko V.E., Bocharov E.V., Karpova R.V., Sheychenko O.P., Aksenov A.A., Chulkova S.V., Kucheryanu V.G., Revishchin A.V., Pavlova G.V., Kosorukov V.S., Filimonov D.A., Lagunin A.A., Matveev V.B., Pyatigorskaya N.V., Stilidi I.S., Poroikov V.V. // Molecular Informatics. – 2023. – Т. 42. – №. 1. – С. 2200176. (Q2)

Свидетельство на государственную регистрацию БД

1. **Ионов Н.С.**, Филимонов Д.А., Дружиловский Д.С., Поройков В.В. Свидетельство о государственной регистрации базы данных Phyto4Health № 2023622658. Москва: Федеральная служба по интеллектуальной собственности, 14.07.2023 г.

Работы, опубликованные в сборниках материалов научных конференций

1. **Ионов Н.С.**, Барышникова М.А., Бочаров Е.В., Погодин П.В., Лагунин А.А., Филимонов Д.А., Карпова Р.В., Бочарова О.А., Поройков В.В. Возможности использования оценок *in silico* для экспериментальной разработки состава фармкомпозиции, цитотоксичной в отношении опухолевых клеток // XXVII Симпозиум «Биоинформатика и компьютерное конструирование лекарств». Москва, Россия, 2021. С. 57.
2. **Ionov N.**, Druzhilovskiy D., Filimonov D., Lagunin A., Poroikov V. Phyto4health - secondary metabolites database of Russian officinal medicinal plants // XXVIII Symposium «Bioinformatics and computer-aided drug discovery». Москва, Россия, 2022. P. 112.
3. **Ионов Н.С.**, Дружиловский Д.С., Поройков В.В. Phyto4Health – База данных вторичных метаболитов фармакопейных растений России // Международной научной конференции «От биохимии растений к биохимии человека». Москва, Россия, 2022.
4. **Ионов Н.С.** Application of *in silico* methods for assessment of pharmacological potential of russian medicinal plants // «Персонализированная медицина: технологии профилактики и ранней диагностики заболеваний» Орёл, Россия, 2023
5. **Ионов Н.С.**, Дружиловский Д.С., Филимонов Д.А., Поройков В.В. Phyto4health - база данных фитокомпонентов фармакопейных лекарственных растений России // Научно-практической конференции «Гаммермановские чтения. Пермь, Россия, 2023. С. 137-140.

БЛАГОДАРНОСТИ

Автор выражает искреннюю признательность:

- научному руководителю доктору биологических наук, члену-корреспонденту РАН Владимиру Васильевичу Поройкову за постановку интересной и актуальной задачи, а также руководство и сопровождение на всех этапах выполнения диссертационной работы.

- доктору биологических наук, профессору Ольге Алексеевной Бочаровой (НМИЦ онкологии им. Н.Н. Блохина), за комплексные работы экспериментальных и клинических исследований противоопухолевой и геропротекторной активности инновационных препаратов Мультифитоадаптоген и Фитоладаптоген, использованных в данной работе.

- кандидату биологических наук Ирине Александровне Лупановой (ФГБНУ ВИЛАР) за выполнение лабораторного исследования надземной части *Cichorium intybus* L.

- кандидату фармацевтических наук Андрею Кеннету Уэйли и кандидату биологических наук. Владимиру Геннадьевичу Лужанину (ФГБОУ ВО СПХФУ Минздрава России) за выполнение лабораторных исследований, связанных с оценкой антитромботического эффекта *Rubus chamaemorus* L.

- сотрудникам лаборатории структурно-функционального конструирования лекарств кандидату физико-математических наук Дмитрию Алексеевичу Филимонову, кандидату биологических наук Дмитрию Сергеевичу Дружиловскому, кандидату биологических наук Анастасии Владимировне Рудик, кандидату биологических наук Сергею Михайловичу Иванову, доктору биологических наук, профессору Алексею Александровичу Лагунину за содействие в разработке информационно вычислительной платформы Phyto4Health.

- сотрудникам ИБМХ, участвовавшим в межлабораторном семинаре, за продуктивное обсуждение диссертационной работы.

ФИНАНСИРОВАНИЕ РАБОТЫ

Работа выполнена в рамках Программы фундаментальных научных исследований в Российской Федерации на долгосрочный период (2021 - 2030 годы) (№ 122030100170-5).

СПИСОК ЛИТЕРАТУРЫ

1. Chen Y. et al. Characterization of the Chemical Space of Known and Readily Obtainable Natural Products // J. Chem. Inf. Model. 2018. Vol. 58, № 8. P. 1518–1532.
2. Cragg G.M. Paclitaxel (Taxol®): A success story with valuable lessons for natural product drug discovery and development // Med. Res. Rev. 1998. Vol. 18, № 5. P. 315–331.
3. Newman D.J. Developing natural product drugs: Supply problems and how they have been overcome // Pharmacol. Ther. Elsevier Inc., 2016. Vol. 162, P. 1–9.
4. Yuan H., Ma Q., Ye L., Piao G. et al. The traditional medicine and modern medicine from natural products // Molecules. 2016. Vol. 21, № 5., P. 559
5. Atanasov A.G. et al. Natural products in drug discovery: advances and opportunities // Nat. Rev. Drug Discov. 2021. Vol. 20, № 3. P. 200–216.
6. Eberhardt L., Kumar K., Waldmann H. Exploring and Exploiting Biologically Relevant Chemical Space // Curr. Drug Targets. 2011. Vol. 12, № 11. P. 1531–1546.
7. Newman D.J. Natural products and drug discovery // Natl. Sci. Rev. 2022. Vol. 9, № 11. P. nwac206.
8. Newman D.J., Cragg G.M. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019 // J. Nat. Prod. 2020. Vol. 83, № 3. P. 770–803.
9. Feher M., Schmidt J.M. Property Distributions: Differences Between Drugs, Natural Products, and Molecules from Combinatorial Chemistry. // ChemInform. 2003. Vol. 34, № 17. P. 218–227.
10. Cragg G.M., Newman D.J. Biodiversity: A continuing source of novel drug leads // Pure Appl. Chem. 2005. T. 77, № 1. C. 7–24.
11. Aronson J.K., Green A.R. Me-too pharmaceutical products: History, definitions, examples, and relevance to drug shortages and essential medicines lists // Br. J. Clin. Pharmacol. 2020. Vol. 86, № 11. P. 2114–2122.
12. Lanthier M.L., Kerr K.W., Miller K.L. An Analysis of Follow-On Development in New Drug Classes, January 1986–June 2018 // Clin. Pharmacol. Ther. 2019. Vol. 106, № 5. P. 1125–1132.

13. Leeson P.D., Springthorpe B. The influence of drug-like concepts on decision-making in medicinal chemistry // *Nat. Rev. Drug Discov.* 2007. Vol. 6, № 11. P. 881–890.
14. Wright G.D. Unlocking the potential of natural products in drug discovery // *Microb. Biotechnol.* 2019. Vol. 12, № 1. P. 55–57.
15. Маджидов Т.И. и др. // Введение в хемоинформатику. Казань. 2013. т. 1.
16. Baurin N. et al. Drug-like annotation and duplicate analysis of a 23-supplier chemical database totalling 2.7 million compounds // *J. Chem. Inf. Comput. Sci.* ACS Publications, 2004. Vol. 44, № 2. P. 643–651.
17. Sirois S. et al. Assessment of chemical libraries for their druggability // *Comput. Biol. Chem.* 2005. Vol. 29, № 1. P. 55–67.
18. Tanimoto T.T. IBM internal report // Nov. 1957. Vol. 17. P. 1957.
19. Dice L.R. Measures of the Amount of Ecologic Association Between Species // *Ecology.* 1945. Vol. 26, № 3. P. 297–302.
20. Ochiai A. Zoogeographical Studies on the Soleoid Fishes Found in Japan and its Neighbouring Regions-III // *Nippon Suisan Gakkaishi.* 1957. Vol. 22. P. 522–525.
21. Rosén J. et al. Novel Chemical Space Exploration via Natural Products // *J. Med. Chem.* 2009. Vol. 52, № 7. P. 1953–1962.
22. van Deursen R., Blum L. C., Reymond J. L. Visualisation of the chemical space of fragments, lead-like and drug-like molecules in PubChem // *Journal of computer-aided molecular design.* 2011. Vol. 25. P. 649-662.
23. Krier M., Bret G., Rognan D. Assessing the scaffold diversity of screening libraries // *J. Chem. Inf. Model.* 2006. Vol. 46, № 2. P. 512–524.
24. Grabowski K., Baringhaus K.H., Schneider G. Scaffold diversity of natural products: Inspiration for combinatorial library design // *Nat. Prod. Rep.* 2008. Vol. 25, № 5. P. 892–904.
25. Miyao T. et al. Chemography of natural product space // *Planta Med.* 2015. Vol. 81, № 6. P. 429–435.
26. Probst D., Reymond J.L. Visualization of very large high-dimensional data sets as minimum spanning trees // *J. Cheminform.* Springer International Publishing, 2020. Vol. 12, № 1. P. 1–13.

27. Ibezim E.C. et al. Computer-aided linear modeling employing QSAR for drug discovery // *Sci. Res. Essays*. 2009. Vol. 4, № 13. P. 1559–1564.
28. Singh S.B., Culberson J.C. Chemical space and the difference between natural products and synthetics // *Natural product chemistry for drug discovery*. Royal Society of Chemistry: Cambridge, UK, 2010. P. 28–43
29. Lipinski C.A. Lead- and drug-like compounds: The rule-of-five revolution // *Drug Discov. Today Technol.* 2004. Vol. 1, № 4. P. 337–341.
30. Shen J. et al. Virtual Screening on Natural Products for Discovering Active Compounds and Target Information // *Curr. Med. Chem.* Bentham Science Publishers, 2005. Vol. 10, № 21. P. 2327–2342.
31. He M. et al. Traditional Chinese medicine database and application on the web // *J. Chem. Inf. Comput. Sci.* 2001. Vol. 41, № 2. P. 273–277.
32. Qiao X. et al. A 3D structure database of components from Chinese traditional medicinal herbs // *J. Chem. Inf. Comput. Sci.* 2002. Vol. 42, № 3. P. 481–489.
33. Wishart D.S. et al. DrugBank 5.0: A major update to the DrugBank database for 2018 // *Nucleic Acids Res.* Oxford University Press, 2018. Vol. 46, № D1. P. D1074–D1082.
34. Gu J. et al. Use of Natural Products as Chemical Library for Drug Discovery and Network Pharmacology // *PLoS One*. 2013. Vol. 8, № 4. P. 1–10.
35. Chen H. et al. A comparative analysis of the molecular topologies for drugs, clinical candidates, natural products, human metabolites and general bioactive compounds // *Medchemcomm.* 2012. Vol. 3, № 3. P. 312–321.
36. El-Elimat T. et al. Chemical Diversity of Metabolites from Fungi, Cyanobacteria, and Plants Relative to FDA-Approved Anticancer Agents // *ACS Med. Chem. Lett.* American Chemical Society, 2012. Vol. 3, № 8. P. 645–649.
37. Xue R. et al. TCMID: Traditional Chinese medicine integrative database for herb molecular mechanism analysis // *Nucleic Acids Res.* 2013. Vol. 41, № D1. P. 1089–1095.
38. Kang H. et al. HIM-herbal ingredients in-vivo metabolism database // *J. Cheminform.* 2013. Vol. 5, № 5. P. 1–6.
39. Ye H. et al. HIT: Linking herbal active ingredients to targets // *Nucleic Acids Res.* 2011. Vol. 39, № SUPPL. 1. P. 1055–1059.

40. Ntie-Kang F. et al. AfroDb: a select highly potent and diverse natural product library from African medicinal plants. // PLoS One. 2013. Vol. 8, № 10. P 1–15.
41. Choi H. et al. NPCARE: Database of natural products and fractional extracts for cancer regulation // J. Cheminform. Springer International Publishing, 2017. Vol. 9, № 1. P. 1–9.
42. CHEMnetBASE [Электронный ресурс]. URL: <https://dnp.chemnetbase.com/chemical/ChemicalSearch.xhtml?dswid=-4417> (дата обращения: 14.08.2023).
43. Adamu R.M. et al. Virtual discovery of a hetero-cyclic compound from the Universal Natural Product Database (UNPD36) as a potential inhibitor of interleukin-33: Molecular docking and dynamic simulations // J. Biomol. Struct. Dyn. Taylor & Francis, 2022. Vol. 40, № 19. P. 8696–8705.
44. Ambinter [Электронный ресурс]. URL: <https://www.ambinter.com/> (дата обращения: 14.08.2023).
45. InterBioScreen Ltd. | Compound Libraries [Электронный ресурс]. URL: <https://www.ibscreen.com/> (дата обращения: 14.08.2023).
46. Ziemert N., Alanjary M., Weber T. The evolution of genome mining in microbes-a review // Natural Product Reports. Royal Society of Chemistry, 2016. Vol. 33, № 8. P. 988–1005.
47. Lachance H. et al. Charting, navigating, and populating natural product chemical space for drug discovery // Journal of Medicinal Chemistry. 2012. Vol. 55, № 13. P. 5989–6001.
48. Pilon A.C. et al. NuBBEDB: an updated database to uncover chemical and biological information from Brazilian biodiversity // Sci. Reports 2017 71. Nature Publishing Group, 2017. Vol. 7, № 1. P. 1–12.
49. Pathania S., Singh P.K. Analyzing FDA-approved drugs for compliance of pharmacokinetic principles: should there be a critical screening parameter in drug designing protocols? // Expert Opin. Drug Metab. Toxicol. Taylor & Francis, 2021. Vol. 17, № 4. P. 351–354.

50. Benet L.Z. et al. BDDCS, the Rule of 5 and drugability // *Adv. Drug Deliv. Rev.* Elsevier, 2016. Vol. 101. P. 89–98.
51. Veber D.F. et al. Molecular properties that influence the oral bioavailability of drug candidates // *J. Med. Chem. American Chemical Society*, 2002. Vol. 45, № 12. P. 2615–2623.
52. Stone S. et al. Cheminformatic analysis of natural product-based drugs and chemical probes // *Nat. Prod. Rep.* 2022. Vol. 39, № 1. P. 20–32.
53. Paclitaxel: Uses, Interactions, Mechanism of Action | DrugBank Online [Электронный ресурс]. URL: <https://go.drugbank.com/drugs/DB01229> (дата обращения: 14.08.2023).
54. Wani M.C. et al. Plant antitumor agents. VI. Isolation and structure of taxol, a novel antileukemic and antitumor agent from *Taxus brevifolia* // *J. Am. Chem. Soc.* 1971. Vol. 93, № 9. P. 2325–2327.
55. Cragg G.M. et al. The Taxol Supply Crisis. New NCI Policies for Handling the Large-Scale Production of Novel Natural Product Anticancer and Anti-HIV Agents // *J. Nat. Prod.* 1993. Vol. 56, № 10. P. 1657–1668.
56. Malik S. et al. Production of the anticancer drug taxol in *Taxus baccata* suspension cultures: A review // *Process Biochem.* 2011. Vol. 46, № 1. P. 23–34.
57. Fett-Neto A.G. et al. Improved Growth and Taxol Yield in Developing Calli of *Taxus cuspidata* by Medium Composition Modification // *Bio/Technology*. 1993. Vol. 11, № 6. P. 731–734.
58. Guo1 B.H. et al. Taxol synthesis // *African J. Biotechnol.* 2006. Vol. 5, № 1. P. 15–20.
59. Brower V. Back to Nature: Extinction of Medicinal Plants Threatens Drug Discovery // *JNCI J. Natl. Cancer Inst.* 2008. Vol. 100, № 12. P. 838–839.
60. Atanasov A.G. et al. Discovery and resupply of pharmacologically active plant-derived natural products: A review // *Biotechnol. Adv.* 2015. Vol. 33, № 8. P. 1582–1614.
61. Henrich C.J., Beutler J.A. Matching the power of high throughput screening to the chemical diversity of natural products // *Nat. Prod. Rep. The Royal Society of Chemistry*, 2013. Vol. 30, № 10. P. 1284–1298.

62. Gyllenhaal C. et al. Ethnobotanical approach versus random approach in the search for new bioactive compounds: Support of a hypothesis // *Pharm. Biol.* Taylor & Francis, 2012. Vol. 50, № 1. P. 30–41.
63. Spjut R.W. Limitations of a random screen: Search for new anticancer drugs in higher plants // *Econ. Bot.* 1985. Vol. 39, № 3. P. 266–288.
64. David B., Wolfender J.-L., Dias D.A. The pharmaceutical industry and natural products: historical status and new trends // *Phytochem. Rev.* 2015. Vol. 14, № 2. P. 299–315.
65. Kong D.X., Li X.J., Zhang H.Y. Where is the hope for drug discovery? Let history tell the future // *Drug Discovery Today*. 2009. Vol. 14, № 3–4.
66. Labanca F., Ovesna J., Milella L. *Papaver somniferum* L. taxonomy, uses and new insight in poppy alkaloid pathways // *Phytochemistry reviews*. 2018. Vol. 17. P. 853-871.
67. Rosenman R.H. et al. Observations on the clinical use of visammin (khellin) // *J. Am. Med. Assoc.* 1950. Vol. 143, № 2. P. 160–165.
68. Perla V., Jayanty S.S. Biguanide related compounds in traditional antidiabetic functional foods // *Food Chem.* 2013. Vol. 138, № 2–3. P. 1574-1580
69. Thompson A.J., Lochner M., Lummis S.C.R. The antimalarial drugs quinine, chloroquine and mefloquine are antagonists at 5-HT₃ receptors // *Br. J. Pharmacol.* Wiley Online Library, 2007. Vol. 151, № 5. P. 666–677.
70. Briskin D.P. Medicinal plants and phytomedicines. Linking plant biochemistry and physiology to human health // *Plant Physiology*. 2000. Vol. 124, № 2. P. 507-514
71. Rohwer J.G. Flowering Plants. Dicotyledons Vol. II: Magnoliid, Hamamelid and Caryophyllid Families // *The Families and Genera of Vascular Plants*. 1993. P. 366-391
72. Barbosa W.L.R. et al. Selecting medicinal plants for development of phytomedicine and use in primary health care // *Bioact. Compd. phytomedicine*. In-techOpen, 2012. P. 3–24.
73. Takenaka T. Classical vs reverse pharmacology in drug discovery // *BJU International*, Supplement. 2001. Vol. 88, № 2.
74. Li J.W.H., Vederas J.C. Drug discovery and natural products: End of an era or an endless frontier? // *Science*. 2009. Vol. 325, № 5937.

75. Lee J.A. et al. Modern phenotypic drug discovery is a viable, neoclassic pharma strategy // J. Med. Chem. ACS Publications, 2012. Vol. 55, № 10. P. 4527–4538.
76. Butterweck V., Nahrstedt A. What is the best strategy for preclinical testing of botanicals? A critical perspective // Planta Medica. 2012. Vol. 78, № 8.
77. Rishton G.M. Reactive compounds and in vitro false positives in HTS // Drug Discovery Today. 1997. Vol. 2, № 9.
78. Wang B. et al. The screening toolbox of bioactive substances from natural products: A review // Fitoterapia. 2011. Vol. 82, № 8. P. 1141-1151
79. Teicher B.A. In vivo/Ex vivo and in situ assays used in cancer research: A brief review // Toxicologic Pathology. 2009. Vol. 37, № 1. P. 114-122
80. Sliwoski G. et al. Computational methods in drug discovery // Pharmacol. Rev. ASPET, 2014. Vol. 66, № 1. P. 334–395.
81. Ferreira L.G. et al. Molecular docking and structure-based drug design strategies // Molecules. 2015. Vol. 20, № 7. P. 13384-13421.
82. Hansch C. et al. Correlation of biological activity of phenoxyacetic acids with Hammett substituent constants and partition coefficients // Nature, 1962. Vol. 194, №. 4824. P. 178-180.
83. Filimonov D.A. et al. Computer-aided prediction of biological activity spectra for chemical compounds: opportunities and limitations // Biomed. Chem. Res. Methods. 2018. Vol. 1, № 1. P. e00004-e00004.
84. Golender V. Вычислительные методы конструирования лекарств. 1978.
85. Chen H. et al. Beyond the scope of free-wilson analysis: Building interpretable QSAR models with machine learning algorithms // J. Chem. Inf. Model. 2013. Vol. 53, № 6. P. 1324-1336
86. Chen B. et al. Development of quantitative structure activity relationship (QSAR) model for disinfection byproduct (DBP) research: A review of methods and resources // Journal of Hazardous Materials. 2015. Vol. 299. P. 260-279
87. Lagunin A.A. et al. Chemo- and bioinformatics resources for in silico drug discovery from medicinal plants beyond their traditional use: A critical review // Natural Product Reports. 2014. Vol. 31, № 11. P. 1585-1611

88. OECD principles for the validation, for regulatory purposes, of (quantitative) structure-activity relationships models. // *Biotechnology*
89. Roy K., Das R. A Review on Principles, Theory and Practices of 2D-QSAR // *Curr. Drug Metab.* 2014. Vol. 15, № 4. P. 346-379
90. Verma J., Khedkar V., Coutinho E. 3D-QSAR in Drug Design - A Review // *Curr. Top. Med. Chem.* 2010. Vol. 10, № 1. P. 95-115
91. Муатс А. и др. QSAR-анализ аффинитета ряда экистероидов на основе 2.5 D-симплексного представления молекулярной структуры // *Вісник Одеського національного університету. Хімія.* 2016. № 21, Вип. 1. С. 80–91.
92. Sorokina M., Steinbeck C. Review on natural products databases: where to find data in 2020 // *J. Cheminform.* Springer, 2020. Vol. 12, № 1. P. 20.
93. Vivek-Ananth R.P. et al. MeFSAT: a curated natural product database specific to secondary metabolites of medicinal fungi // *RSC Adv.* 2021. Vol. 11, № 5. P. 2596-2607.
94. Naghizadeh A. et al. UNaProd: A Universal Natural Product Database for Materia Medica of Iranian Traditional Medicine // *Evidence-based Complement. Altern. Med.* 2020. Vol. 2020.
95. Xu T. et al. NPDB: a natural product database with relational data between natural products and biological sources. 2020.
96. Wishart D.S. et al. NP-MRD: the natural products magnetic resonance database // *Nucleic Acids Res.* Oxford University Press, 2022. Vol. 50, № D1. P D665--D677.
97. Medina-Franco J.L. Towards a unified Latin American Natural Products Database: LANaPD // *Futur. Sci. OA.* 2020. Vol. 6, № 8. P. FSO468
98. Madariaga-Mazón A. et al. DiaNat-DB: A molecular database of antidiabetic compounds from medicinal plants // *RSC Adv.* 2021. Vol. 11, № 9. P. 5172-5178
99. Barazorda-Ccahuana H.L. et al. PeruNPDB: the Peruvian Natural Products Database for in silico drug screening // *Sci. Rep.* 2023. Vol. 13, № 1. P. 7577
100. Tan X. et al. ACNPD: The Database for Elucidating the Relationships Between Natural Products, Compounds, Molecular Mechanisms, and Cancer Types // *Front. Pharmacol.* 2021. Vol. 12. P. 746067

101. Diallo B.N. et al. SANCDB: an update on South African natural compounds and their readily available analogs // J. Cheminform. 2021. Vol. 13, № 1. P. 37.
102. Mendez D. et al. ChEMBL: Towards direct deposition of bioassay data // Nucleic Acids Res. 2019. Vol. 47, № D1. P. D930-D940.
103. Hastings J. et al. The ChEBI reference database and ontology for biologically relevant chemistry: Enhancements for 2013 // Nucleic Acids Res. 2013. Vol. 41, № D1. P. D456-D463
104. Sorokina M. et al. COCONUT online: collection of open natural products database // J. Cheminform. BioMed Central, 2021. Vol. 13, № 1. P. 1–13.
105. Rutz A. et al. The LOTUS initiative for open knowledge management in natural products research // Elife. eLife Sciences Publications Limited, 2022. Vol. 11. P. e70780.
106. Gallo K. et al. SuperNatural 3.0—a database of natural products and natural product-based derivatives // Nucleic Acids Res. Oxford Academic, 2023. Vol. 51, № D1. P. D654–D659.
107. Bioregistry - Universal Natural Products Database [Электронный ресурс]. URL: <https://bioregistry.io/registry/unpd> (дата обращения: 17.08.2023).
108. MassBank of North America [Электронный ресурс]. URL: <https://mona.fiehnlab.ucdavis.edu/> (дата обращения: 17.08.2023).
109. MassBank | MassBank Europe Mass Spectral DataBase [Электронный ресурс]. URL: <https://massbank.normandata.eu//MassBank/> (дата обращения: 17.08.2023).
110. Horai H. et al. MassBank: a public repository for sharing mass spectral data for life sciences // Journal of mass spectrometry. 2010. Vol. 45. №. 7. P. 703-714.
111. Guijas C. et al. METLIN: A Technology Platform for Identifying Knowns and Unknowns // Anal. Chem. American Chemical Society, 2018. Vol. 90, № 5. P. 3156–3164.
112. Wishart D.S. et al. HMDB: The human metabolome database // Nucleic Acids Res. 2007. Vol. 35, № SUPPL. 1. P. D521-D526.
113. Jewison T. et al. YMDB: The yeast metabolome database // Nucleic Acids Res. 2012. T. 40, № D1. P. D815-D820.

114. Wang M. et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking // Nat. Biotechnol. 2016 348. Nature Publishing Group, 2016. Vol. 34, № 8. P. 828–837.
115. Steinbeck C., Krause S., Kuhn S. NMRShiftDB - Constructing a Free Chemical Information System with Open-Source Components // J. Chem. Inf. Comput. Sci. 2003. Vol. 43, № 6. P. 1733-1739
116. Fishedick J.T. et al. NMR spectroscopic search module for Spektraris, an online resource for plant natural product identification - Taxane diterpenoids from *Taxus x media* cell suspension cultures as a case study // Phytochemistry. 2015. Vol. 113. P. 87-95.
117. López-Pérez J.L. et al. NAPROC-13: a database for the dereplication of natural product mixtures in bioassay-guided protocols // Bioinformatics. Oxford Academic, 2007. Vol. 23, № 23. P. 3256–3257.
118. Lagunin A.A. et al. Computer evaluation of hidden potential of phytochemicals of medicinal plants of the traditional Indian ayurvedic medicine // Biomed. Khim. 2015. Vol. 61, № 2. P. 286—297.
119. Pílon-Jiménez B.A. et al. BIOFACQUIM: A Mexican Compound Database of Natural Products // Biomolecules. 2019. Vol. 9, № 1. P. 31.
120. Olmedo D.A. et al. Cheminformatic characterization of natural products from Panama // Mol. Divers. 2017. Vol. 21, № 4. P. 779–789.
121. Costa R.P.O. et al. The Sistemax Web Portal of Natural Products: An Update // J. Chem. Inf. Model. American Chemical Society, 2021. Vol. 61, № 6. P. 2516–2522.
122. Günthardt B.F. et al. Comprehensive Toxic Plants-Phytotoxins Database and Its Application in Assessing Aquatic Micropollution Potential // J. Agric. Food Chem. American Chemical Society, 2018. Vol. 66, № 29. P. 7577–7588.
123. Ntie-Kang F. et al. NANPDB: A Resource for Natural Products from Northern African Sources // J. Nat. Prod. American Chemical Society, 2017. Vol. 80, № 7. P. 2067–2076.
124. Ntie-Kang F. et al. CamMedNP: Building the Cameroonian 3D structural natural products database for virtual screening // BMC Complement. Altern. Med. 2013. Vol. 13, № 1. P. 88.

125. Ntie-Kang F. et al. ConMedNP: a natural product library from Central African medicinal plants for drug discovery // RSC Adv. The Royal Society of Chemistry, 2013. Vol. 4, № 1. P. 409–419.
126. Derese S. et al. Mitishamba database: a web based in silico database of natural products from Kenya plants. University of Nairobi, 2015.
127. Mohanraj K. et al. IMPPAT: A curated database of Indian Medicinal Plants, Phytochemistry and Therapeutics // Sci. Rep. Nature Publishing Group, 2018. Vol. 8, № 1. P. 1–17.
128. Nguyen-Vo T.H. et al. VIETHERB: A Database for Vietnamese Herbal Species // J. Chem. Inf. Model. American Chemical Society, 2019. Vol. 59, № 1. P. 1–9.
129. Klementz D. et al. StreptomeDB 2.0—an extended resource of natural products produced by streptomycetes // Nucleic Acids Res. Oxford Academic, 2016. Vol. 44, № D1. P. D509–D514.
130. Nupur L.N.U. et al. ProCarDB: A database of bacterial carotenoids // BMC Microbiol. BioMed Central Ltd., 2016. Vol. 16, № 1. P. 1–8.
131. Huang W. et al. PAMDB: A comprehensive Pseudomonas aeruginosa metabolome database // Nucleic Acids Res. Oxford Academic, 2018. Vol. 46, № D1. P. D575–D580.
132. Shikov A.N. et al. Medicinal plants from the 14th edition of the Russian Pharmacopoeia, recent updates // Journal of Ethnopharmacology. 2021. Vol. 268. P. 113685
133. Golovkin B.N. Biologicheski aktivnye veshchestva rastitel'nogo proiskhozhdeniiâ: «Nauka», 2001.
134. Фармакопее. Издание XIV. Том IV. Биологические лекарственные препараты, Растительные средства и препараты, Гомеопатические фармацевтические субстанции [Электронный ресурс]. URL: <https://docs.rucml.ru/feml/pharma/v14/vol4/> (дата обращения: 28.08.2023).
135. Home — The Plant List [Электронный ресурс]. URL: <http://www.theplantlist.org/> (дата обращения: 28.08.2023).
136. Federhen S. The NCBI Taxonomy database // Nucleic Acids Res. 2012. Vol. 40, № D1. P. D136-D143

137. Академия Google [Электронный ресурс]. URL: <https://scholar.google.ru/schhp?hl=ru> (дата обращения: 28.08.2023).
138. PubMed [Электронный ресурс]. URL: <https://pubmed.ncbi.nlm.nih.gov/> (дата обращения: 28.08.2023).
139. Weininger D., Weininger A., Weininger J.L. SMILES. 2. Algorithm for Generation of Unique SMILES Notation // J. Chem. Inf. Comput. Sci. American Chemical Society, 1989. Vol. 29, № 2. P. 97–101.
140. Heller S. et al. InChI - The worldwide chemical structure identifier standard // J. Cheminform. BioMed Central, 2013. Vol. 5, № 1. P. 1–9.
141. Pletnev I. et al. InChIKey collision resistance: An experimental testing // J. Cheminform. BioMed Central, 2012. Vol. 4, № 12. P. 1–9.
142. Dalby A. et al. Description of Several Chemical Structure File Formats Used by Computer Programs Developed at Molecular Design Limited // J. Chem. Inf. Comput. Sci. American Chemical Society, 1992. Vol. 32, № 3. P. 244–255.
143. PubChem [Электронный ресурс]. URL: <https://pubchem.ncbi.nlm.nih.gov/> (дата обращения: 29.08.2023).
144. Xiong G. et al. ADMETlab 2.0: an integrated online platform for accurate and comprehensive predictions of ADMET properties // Nucleic Acids Res. Oxford Academic, 2021. Vol. 49, № W1. P. W5–W14.
145. Protti Í.F. et al. Do Drug-likeness Rules Apply to Oral Prodrugs? // ChemMedChem. John Wiley & Sons, Ltd, 2021. Vol. 16, № 9. P. 1446–1456.
146. Kim S. et al. An update on PUG-REST: RESTful interface for programmatic access to PubChem // Nucleic Acids Res. Nucleic Acids Res, 2018. Vol. 46, № W1. P. W563–W570.
147. Welcome to Python.org [Электронный ресурс]. URL: <https://www.python.org/> (дата обращения: 26.09.2023).
148. pandas - Python Data Analysis Library [Электронный ресурс]. URL: <https://pandas.pydata.org/> (дата обращения: 26.09.2023).
149. Requests: HTTP for Humans™ — Requests 2.31.0 documentation [Электронный ресурс]. URL: <https://requests.readthedocs.io/en/latest/> (дата обращения: 26.09.2023).

150. MySQL : MySQL Connector/Python Developer Guide [Электронный ресурс]. URL: <https://dev.mysql.com/doc/connector-python/en/> (дата обращения: 26.09.2023).
151. Pogodin P. V et al. PASS Targets: Ligand-based multi-target computational system based on a public data and naïve Bayes approach // SAR and QSAR in Environmental Research. 2015. Vol. 26. №. 10. P. 783-793.
152. Ivanov S.M. et al. ADVERPred–Web Service for Prediction of Adverse Effects of Drugs // J. Chem. Inf. Model. American Chemical Society, 2018. Vol. 58, № 1. P. 8–11.
153. Lagunin A.A. 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 // International Journal of Molecular Sciences. 2023. Vol. 24, № 2. P. 1689.
154. MySQL [Электронный ресурс]. URL: <https://www.mysql.com/> (дата обращения: 03.10.2023).
155. HTML | MDN [Электронный ресурс]. URL: <https://developer.mozilla.org/ru/docs/Web/HTML> (дата обращения: 03.10.2023).
156. Руководство по CSS - CSS | MDN [Электронный ресурс]. URL: <https://developer.mozilla.org/ru/docs/Web/CSS/Reference> (дата обращения: 03.10.2023).
157. Современный учебник JavaScript [Электронный ресурс]. URL: <https://learn.javascript.ru/> (дата обращения: 03.10.2023).
158. PHP: Hypertext Preprocessor [Электронный ресурс]. URL: <https://www.php.net/> (дата обращения: 03.10.2023).
159. DataTables. Javascript table library [Электронный ресурс]. URL: <https://datatables.net/> (дата обращения: 16.05.2024).
160. Filimonov D. et al., Chemical Similarity Assessment through Multilevel Neighborhoods of Atoms: Definition and Comparison with the Other Descriptors // J. Chem. Inf. Comput. Sci. American Chemical Society, 1999. T. 39, № 4. С. 666–670.
161. Filimonov D.A. et al. QNA-based «Star Track» QSAR approach // SAR QSAR Environ. Res. 2009. Vol. 20, № 7–8. P. 679–709.

162. Mauri A. et al. Mixtures, metabolites, ionic liquids: a new measure to evaluate similarity between complex chemical systems // J. Cheminform. BioMed Central Ltd., 2016. Vol. 8, № 1. P. 1–3.
163. Druzhilovskiy D.S. et al. Computational Approaches to Identify a Hidden Pharmacological Potential in Large Chemical Libraries // Supercomput. Front. Innov. 2020. Vol. 7, № 3. P. 57–76.
164. Morgan H. L. The generation of a unique machine description for chemical structures-a technique developed at chemical abstracts service // Journal of chemical documentation, 1965. Vol. 5, №. 2. P. 107-113.
165. Rogers D., Hahn M. Extended-connectivity fingerprints // Journal of chemical information and modeling, 2010. Vol. 50, №. 5. P. 742-754.
166. Getting Started with the RDKit in Python [Электронный ресурс]. URL: <https://www.rdkit.org/docs/GettingStartedInPython.html> (дата обращения: 06.09.2024).
167. Becht E. et al. Dimensionality reduction for visualizing single-cell data using UMAP // Nature biotechnology, 2019. Vol. 37, №. 1. P. 38-44.
168. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction [Электронный ресурс]. URL: <https://umap-learn.readthedocs.io/en/latest/> (дата обращения: 06.09.2024).
169. Sander J. et al. Density-based clustering in spatial databases: The algorithm gdbscan and its applications // Data mining and knowledge discovery, 1998. Vol. 2. P. 169-194.
170. DBSCAN [Электронный ресурс]. URL: <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.DBSCAN.html> (дата обращения: 06.09.2024).
171. Ahmed M., Seraj R., Islam S. M. S. The k-means algorithm: A comprehensive survey and performance evaluation // Electronics, 2020. Vol. 9. №. 8. P. 1295.
172. Rousseeuw P. J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis // Journal of computational and applied mathematics, 1987. Vol. 20. P. 53-65.

173. silhouette_score [Электронный ресурс]. URL: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.silhouette_score.html (дата обращения: 06.09.2024).
174. KMeans [Электронный ресурс]. URL: <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html#kmeans> (дата обращения: 06.09.2024).
175. Zhang X. et al. In silico modeling on ADME properties of natural products: Classification models for blood-brain barrier permeability, its application to traditional Chinese medicine and in vitro experimental validation // J. Mol. Graph. Model. Elsevier, 2017. Vol. 75. P. 347–354.
176. Mosmann T. Rapid colorimetric assay for cellular growth and survival: Application to proliferation and cytotoxicity assays // J. Immunol. Methods. Elsevier, 1983. Vol. 65, № 1–2. P. 55–63.
177. Lowry O.H. et al. Protein measurement with the folin phenol reagent // J. Biol. Chem. 1951. Vol. 193. P. 265–275.
178. Omura T., Sato R. The Carbon Monoxide-binding Pigment of Liver Microsomes: II. Solubilization, purification, and properties // J. Biol. Chem. Elsevier, 1964. Vol. 239, № 7. P. 2379–2385.
179. Bateman A. UniProt: a worldwide hub of protein knowledge // Nucleic Acids Res. Oxford Academic, 2019. Vol. 47, № D1. P. D506–D515.
180. Ionov N. et al. Phyto4Health: Database of Phytocomponents from Russian Pharmacopoeia Plants // J. Chem. Inf. Model. American Chemical Society, 2023. Vol. 63, № 7. P. 1847–1851.
181. Marvin [Электронный ресурс]. URL: <https://chemaxon.com/marvin> (дата обращения: 16.05.2024).
182. Zhang H.W., Song Y.C., Tan R.X. Biology and chemistry of endophytes // Nat. Prod. Rep. The Royal Society of Chemistry, 2006. Vol. 23, № 5. P. 753–771.
183. Wani Z.A. et al. Plant-endophyte symbiosis, an ecological perspective // Appl. Microbiol. Biotechnol. Springer Verlag, 2015. Vol. 99, № 7. P. 2955–2965.

184. Xu H.Q. et al. EMNPD: a comprehensive endophytic microorganism natural products database for prompt the discovery of new bioactive substances // J. Cheminform. BioMed Central Ltd, 2023. Vol. 15. № 1. P. 1–10.
185. Martin Y. C., Kofron J. L., Traphagen L. M. Do structurally similar molecules have similar biological activity? // Journal of medicinal chemistry, 2002. Vol. 45. №. 19. P. 4350-4358.
186. Ionov N.S. et al. Possibilities of *in silico* estimations for the development of pharmaceutical composition phytoladaptogene cytotoxic for bladder cancer cells // Biomeditsinskaya Khimiya. Institute of biomedical chemistry, Moscow, Russia, 2021. Vol. 67, № 3. P. 278–288.
187. Whaley A.K. et al. Phytochemical Analysis of Polyphenol Secondary Metabolites in Cloudberry (*Rubus Chamaemorus* L.) Leaves // Pharm. Chem. J. Springer, 2021. Vol. 55, № 3. P. 253–258.
188. Лужанин В.Г. и др. Влияние выделенных из *Rubus chamaemorus* метаболитов на систему гемостаза в условиях *in vitro* // Разработка и регистрация лекарственных средств. 2024. Т. 13, № 1. С. 149–158.
189. Лупанова И. и др. Влияние цикориевой и хлорогеновой кислот из *Cichorium intybus* L. на активность цитохрома P450 и глутатионтрансферазы // Биофармацевтический журнал. 2022. Т. 14. №. 5. С. 8-18.
190. Yun N., Kang J.W., Lee S.M. Protective effects of chlorogenic acid against ischemia/reperfusion injury in rat liver: molecular evidence of its antioxidant and anti-inflammatory properties // J. Nutr. Biochem. Elsevier, 2012. Vol. 23, № 10. P. 1249–1255.
191. Zhu D. et al. Cichoric acid regulates the hepatic glucose homeostasis via AMPK pathway and activates the antioxidant response in high glucose-induced hepatocyte injury // RSC Adv. Royal Society of Chemistry, 2017. Vol 7, № 3. P. 1363–1375.