

ОТЗЫВ

официального оппонента, доктора биологических наук, заведующего лабораторией информационных технологий в фармакологии и компьютерного моделирования лекарств Научного центра инновационных лекарственных средств с опытно-промышленным производством Федерального государственного бюджетного образовательного учреждения высшего образования «Волгоградский государственный медицинский университет» Министерства здравоохранения Российской Федерации Васильева Павла Михайловича на диссертацию Ионова Никиты Сергеевича на тему «Разработка информационно-вычислительной платформы для оценки фармакологического потенциала фитокомпонентов лекарственных растений», представленной к защите на соискание ученой степени кандидата биологических наук по специальности 1.5.8. – Математическая биология, биоинформатика

Актуальность проблемы

В настоящее время существенно возрос интерес к созданию новых лекарственных препаратов на основе соединений природного происхождения. Исторически большинство новых лекарств разрабатывалось из природных продуктов и из полученных из них соединений. Примерами могут служить хинин (1820 г.), аспирин (1897 г.), пенициллин (1928 г.) и др.

По данным из различных литературных источников, до 80% когда-либо используемых лекарств были природными веществами или их аналогами, а за последние 40 лет доля новых лекарств, полученных на основе природных соединений, составляла в разные годы от 50% до 70%; при этом доля препаратов категории «first-in-class», созданных на основе соединений природного происхождения, составляла примерно 80%.

Природные соединения характеризуются существенно меньшей ксенобиотичностью, в сравнении с соединениями, полученными преимущественно синтетическим путем, что объясняется наличием в организме человека рецепторных и ферментных систем, обеспечивающих адекватные фармакологические реакции и эффективный метаболизм

поступающих извне компонентов естественного происхождения. По этой причине для лекарственных препаратов на основе природных соединений характерно более мягкое, часто пролонгированное действие и они обычно не вызывают выраженных побочных эффектов.

Поэтому неудивительно, что во всем мире ведется интенсивная работа по созданию баз данных по структуре и свойствам природных соединений – уже известно более 60 таких ресурсов, отражающих химический состав и биологическую активность компонентов растений разных географических регионов.

Однако до сих пор не существовало специализированной базы данных, включающей полную и верифицированную информацию о структуре и биологической активности всех известных соединений, содержащихся в лекарственных растениях, разрешенных к применению в Российской Федерации.

Таким образом, создание оригинальной базы данных по структуре и активности соединений, присутствующих в официальных лекарственных растениях Российской Федерации, является весьма актуальной задачей.

Как известно, полная стоимость разработки нового лекарственного препарата очень высока и может достигать 2 млрд. долларов. Больше половины этой суммы составляют доклинические исследования, что обуславливает необходимость оптимизации данного этапа. Важную роль в этом играют доэкспериментальные методы поиска новых лекарственных веществ. Методы *in silico* позволяют существенно сократить временные, материальные и финансовые затраты на экспериментальные исследования за счет предварительного отсева мало активных структур, оптимизируя таким образом направленный поиск высоко активных веществ. Максимально быстро и эффективно это можно выполнять при использовании соответствующих профильных Интернет-систем.

Таким образом, разработка нового информационно-вычислительного Web-ресурса для оценки фармакологической активности химических соединений, содержащихся в лекарственных растениях, также является весьма актуальной задачей.

Диссертация Ионова Н.С. посвящена разработке информационно-

вычислительной платформы для оценки фармакологического потенциала фитокомпонентов лекарственных растений на основе созданной оригинальной базы данных по структуре и активности соединений, присутствующих в официальных лекарственных растениях Российской Федерации, что, в контексте вышеизложенного, обуславливает высокую актуальность проведенной работы.

Обоснованность и достоверность результатов исследования

Работу отличает корректно поставленная цель, грамотно сформулированные задачи, адекватно подобранные методы исследования.

Проведенное исследование характеризуется масштабностью и тщательностью подготовки исходного материала, выполненной с использованием официальных источников и широко известных и надежных баз данных: 1) из 14-го издания Государственной фармакопеи Российской Федерации получено 233 видовых названия официальных лекарственных растений России; 2) с помощью веб-ресурса The Plant List и БД NCBI Тахопому выявлено более 1000 наименований указанных растений и после анализа полученного списка определены 533 наиболее известных их синонима; 3) с использованием тематических публикаций и информационных систем Google Scholar и PubMed найдены 3128 уникальных фитокомпонентов, входящих в состав 233 официальных лекарственных растений РФ, а также 9484 значения по их содержанию в отдельных частях растений; 4) из репозиториев химических структур PubChem, ChEMBL и с помощью веб-сервиса ADMETlab получен массив из 31280 значений молекулярных характеристик указанных фитокомпонентов; 5) из БД PubChem и ChEMBL получены сведения о 13688 взаимодействиях этих фитокомпонентов с 802 биомолекулами человека, с указанием количественных значений активности. Подготовленные и структурированные массивы данных послужили информационным наполнением создаваемой реляционной БД Phyto4Health. При анализе структурного разнообразия БД Phyto4Health была использована информация из шести БД о структурах более чем 26 тыс. природных соединений из разных географических регионов мира. В ретроспективной валидации использовали 2732 структуры природных соединений из БД LOTUS.

Следует отметить высокий уровень математической подготовки автора. Он самостоятельно разработан логическую схему формируемой М:М реляционной базы данных. Используемые методы прогноза, такие, как метод Байеса, анализ по сходству с применением метрик Танимото и Тодескини, анализ на выбросы, методы кластерного анализа и снижения размерности содержательно излагаются в терминах математической статистики, многомерного анализа и теории множеств.

Отдельно необходимо отметить высокий профессиональный уровень автора как программиста, умелое использование современных языков и методов программирования и специализированных программ. Для препроцессинга и последующего формирования готовой базы данных Phyto4Health автором использована известная и хорошо зарекомендовавшая себя реляционная СУБД MySQL. Веб-интерфейс информационно-вычислительной платформы Phyto4Health разработан автором с использованием языка программирования JavaScript, с применением формального языка декорирования веб-страниц CSS на основе языка гипертекстовой разметки HTML. Серверная часть платформы разработана с использованием языка программирования PHP. Самостоятельно разработан ряд процедур на языке Python, включающих применение специализированных библиотек requests, pandas, mysql-connector-python, UMAP, scikit-learn. Автор квалифицированно применяет для доказательных исследований такие хорошо зарекомендовавшие себя программы, как PASS, PharmaExpertm, MarvinSketch. При этом современная версия системы PASS имплементирована в платформу Phyto4Health.

Необходимо особо отметить, что на основе созданной верифицированной базы данных по химическим структурам и свойствам уникальных фитокомпонентов официальных лекарственных растений Российской Федерации впервые создан оригинальный свободно доступный специализированный информационно-вычислительный Web-сервис Phyto4Health.

Специально следует подчеркнуть, что работа выполнена в рамках программы фундаментальных научных исследований в Российской Федерации на долгосрочный период (2021-2030 гг.) № 122030100170-5.

Таким образом, результаты диссертационного исследования неоднократно проходили экспертизу высокопрофессиональных специалистов указанной программы.

Большой объем данных, полученных из надежных источников, их глубокий анализ, тщательная предварительная обработка и последующая профессиональная структуризация, аргументированное построение алгоритмов и модулей разрабатываемой системы, корректная вычислительная обработка с помощью профильных программ весьма объемных данных, демонстрация высокой прогнозной эффективности новой информационно-вычислительной Web-платформы – все это позволяет квалифицировать результаты диссертационной работы как обоснованные и достоверные, а основные положения, выносимые на защиту, выводы и практические рекомендации, как вполне обоснованные.

Научная новизна и практическая ценность диссертации

Исследование носит фундаментально-прикладной характер. Бесспорны научная новизна и практическая значимость полученных результатов.

Впервые проведено обобщение известных данных о структуре и биологической активности фитокомпонентов официальных лекарственных растений России. Впервые создана реляционная база данных, включающая информацию о 3128 уникальных соединениях, входящих в состав 233 лекарственных растений РФ, 9484 значения о содержании указанных соединений в отдельных частях растений и сведения о 13688 взаимодействиях этих соединений с 802 биомешинами человека.

Выполнено сопоставление структурного разнообразия сформированной базы данных с иными базами данных природных соединений из разных географических регионов мира. Показано, что содержащаяся в ней информация является достаточно уникальной и отражает широкое химическое разнообразие фитокомпонентов официальных лекарственных растений России.

Выполнена ретроспективная и проспективная валидация вычислительной компоненты создаваемой системы. Показана высокая точность и эффективность прогноза различных видов фармакологической активности соединений природного происхождения.

Как итог, впервые создана новая информационно-вычислительная платформа Phyto4Health, которая позволяет получать сведения о химической структуре, молекулярных свойствах, известных и прогнозных видах биологической и фармакологической активности соединений официальных лекарственных растений России и выполнять поиск их структурных аналогов. Следует подчеркнуть, что разработанная платформа является свободно доступным Интернет-ресурсом и может быть использована в поиске аналогов природных соединений с различными видами лекарственной активности.

Таким образом, диссертационная работа Ионова Н.С. имеет высокую степень научной новизны, а полученные при ее выполнении результаты характеризуются высокой практической значимостью.

Теоретическая и научно-практическая значимость

В диссертационном исследовании разработана новая методология обработки информации о соединениях природного растительного происхождения, существенно расширяющая и уточняющая первичные исходные данные. В рамках этой методологии выполняется интеллектуальная структуризация обрабатываемых данных, что обеспечивает необходимые условия для создания верифицированной реляционной базы данных с широкими поисковыми функциями и делает возможным создание на основе такой БД мощного Web-сервиса. Эта методология может быть применена для создания аналогичных инструментов, ориентированных на обработку данных по другим природным соединениям, например, животного или микробиологического происхождения.

Разработан новый подход к анализу структурного разнообразия массивов природных соединений, выполнено сравнение созданной БД соединений официальных лекарственных растений России с шестью БД природных соединений других географических регионов мира. Средний уровень структурного сходства составил 0.64, что свидетельствует о высокой степени оригинальности созданной БД.

Проведена вычислительная ретроспективная и экспериментальная проспективная валидации созданной информационно-вычислительной платформы, результаты которых полностью согласуются в компьютерном

прогнозом.

Свободно доступный Web-сервис Phyto4Health, созданный как результат выполненного диссертационного исследования, позволяет выполнять виртуальный скрининг и направленный поиск новых природных соединений и их производных с высокой лекарственной активностью.

Особую ценность данной работе придает то, что она выполнена комплексно и логически последовательно: 1) сформирован большой массив данных по структуре и активности фитокомпонентов официальных растений РФ; 2) выполнен детальный анализ, предварительная обработка и последующая структуризация этих данных; 3) сформирован математический аппарат и построены алгоритмы обработки указанной информации; 4) сформирована оригинальная верифицированная реляционная база данных; 5) на этой основе создана специальная свободно доступная уникальная информационно-вычислительная платформа Phyto4Health; 6) показана высокая прогностическая способность и эффективность созданного Web-сервиса.

Следует отдельно подчеркнуть, что в результате проделанной работы автором создана новая информационно-вычислительной Web-платформа, которая является реальным рабочим инструментом для поиска новых лекарственных веществ на основе природных соединений.

Работа выполнена в рамках Программы фундаментальных научных исследований в Российской Федерации на долгосрочный период (2021-2030 гг.), полученные диссертантом результаты опубликованы в авторитетных международных научных журналах.

Таким образом, диссертационное исследование Ионова Н.С. имеет высокую теоретическую и практическую значимость.

Соответствие диссертации паспорту специальности

Поставленные цели, задачи, область исследования, предмет исследования, примененные современные и обоснованные методы, полученные выводы и рекомендации являются весьма актуальными для биоинформатики. Считаю, что представленная диссертации полностью соответствует паспорту специальности 1.5.8. Математическая биология, биоинформатика по следующим направлениям.

п.4. Математическое и компьютерное моделирование биологического действия ксенобиотиков. Компьютерное конструирование лекарств. Анализ взаимосвязей «структура-активность». Компьютерная фармакология и токсикология.

п.12. Разработка и применение новых вычислительных алгоритмов для анализа экспериментальных данных в биологии и медицине.

п.14. Математические модели, численные методы, алгоритмы и программные средства применительно к процессам получения, накопления, обработки и систематизации биологических и медицинских данных и знаний.

п.16. Разработка и применение методов машинного обучения и искусственного интеллекта для анализа и прогнозирования свойств биологических объектов на основе анализа больших биомедицинских данных.

Полнота освещения результатов диссертации в печати

По теме исследования опубликовано 11 работ, из них шесть статей в журналах, рекомендованных ВАК Минобрнауки РФ и индексируемых в базах данных Scopus и Web of Science, в том числе две в журналах первой четверти Q1 и одна в журнале второй четверти Q2. Основные положения диссертации докладывались на 5 международных и всероссийских конференциях. Получено свидетельство о государственной регистрации базы данных Phyto4Health.

Структура и оформление работы

Диссертация оформлена в классическом стиле, в соответствии с существующими требованиями, изложена на 141 странице, включает 38 рисунков и 18 таблиц. Работа состоит из Списка сокращений; Введения; главы 1 «Обзор литературы», включающей пять разделов; главы 2 «Материалы и методы», включающей восемь разделов; главы 3 «Результаты и обсуждение», включающей семь разделов самостоятельно полученных данных; Заключение; Выводов; Списка работ, опубликованных по теме диссертации; Благодарностей; Финансирования работы; Списка литературы.

Список литературы включает в себя 191 источник и охватывает период в 54 года с 1971 по 2024 гг. Необходимо специально отметить значительный объем и временную широту литературного поиска.

В первой главе проведен обзор и анализ литературных данных по теме диссертации.

В первом разделе главы 1 описано значение природных соединений для разработки новых лекарственных веществ. По результатам анализа литературных данных сформулировано определение низкомолекулярных химических соединений природного происхождения. Проведено сопоставление долей разрешенных к применению лекарственных препаратов синтетического происхождения и созданных на основе природных соединений. Отмечается, что наибольшее число разрешенных к применению лекарственных препаратов “first-in-class” (77%) разработано на основе веществ природного происхождения. Сделан вывод о том, что природные соединения являются перспективным источником разнообразных химических структур с богатым лекарственным потенциалом, которые могут стать эффективными препаратами для лечения различных социально-значимых заболеваний.

Во втором разделе главы 1 выполнен обзор литературных данных по анализу химического разнообразия пространств природных соединений. Отмечено, что природные соединения охватывают более широкое химическое пространство, чем библиотеки синтетических соединений и одобренные FDA лекарства, а также имеют намного более сложную организацию химической структуры.

В третьем разделе главы 1 изложены различные аспекты разработки лекарственных веществ на основе природных соединений – как облегчающие, так и затрудняющие этот процесс. Проанализированы публикации, описывающие различные подходы к данной проблеме и определяемые: 1) источниками соединений; 2) скринингом библиотек экстрактов; 3) знаниями традиционной медицины; 4) экологическими данными; 5) хемотаксономикой; 6) типами фармакологического тестирования; 7) методами компьютерного виртуального скрининга. Современные методы компьютерного поиска изложены более детально, отмечены такие подходы, как молекулярный докинг, анализ молекулярного подобия, построение зависимостей «структура – активность».

В четвертом разделе главы 1 представлен обзор существующих баз

данных по структуре и активности соединений природного происхождения. Описаны семь крупных репозиториях общим объемом более 1 млн. 157 тыс. структур; базы данных для дубликации природных соединений, в том числе семь БД, аннотированных масс-спектрами и три БД, аннотированных спектрами ЯМР; 17 баз данных, характеризующих химическое разнообразие отдельных географических регионов общим объемом более 69 тыс. структур; ряд БД, характеризующие отдельные таксономические группы.

В пятом разделе главы 1 приведено заключение по литературному обзору, в котором отмечено, что территория России охватывает несколько климатических зон. Это обуславливает высокое биоразнообразие эндемичных растений и разнообразие синтезируемых ими фармакологических веществ. Данные об этих растениях и соединениях доступны только в печатной форме. Подчеркивается, что создание базы данных и разработка на её основе информационно-вычислительной платформы, предоставляющей пользователям сведения о структурах фитокомпонентов лекарственных растений России и информацию об их биологической активности, является актуальной задачей.

Во второй главе подробно описаны материалы и методы проведенных исследований. Следует отметить, что примененные диссертантом методики сбора и обработки данных валидны и соответствуют задачам исследования.

В первом разделе главы 2 описана процедура извлечения данных. Изложена методика формирования набора общепринятых видовых названий растений на основе Государственной Фармакопеи Российской Федерации 14-го издания, с уточнением синонимов с помощью веб-ресурса The Plant List и таксономических характеристик с помощью базы данных NCBI Taxonomy. Изложена методика формирования набора структурных формул фитокомпонентов (в четырех форматах), выполненного с использованием информационных систем Google Scholar и PubMed. Описана методика формирования набора значений десяти молекулярных характеристик фитокомпонентов, выполненного с использованием репозитория химических структур PubChem и ChEMBL, а также с использованием веб-сервиса ADMETlab. Описана методика извлечения информации о взаимодействии химических соединений с молекулярными мишенями

человека, для чего использовались БД PubChem и ChEMBL.

Во втором разделе главы 2 описаны методы работы с базами данных. Детализирована процедура получения информации из базы данных PubChem, которая выполнялась с применением специализированного программного интерфейса PUG REST. Приведен URL-шаблон загрузки данных для соединения из PubChem с помощью этого программного средства. Указано, что генерацию URL-адресов, а также дальнейшую обработку полученных данных проводили с использованием языка программирования Python и его специализированных библиотек. Детализирована процедура получения информации из базы данных ChEMBL, которая выполнялась с использованием СУБД MySQL, языка программирования Python и его специализированных библиотек.

В третьем разделе главы 2 излагается математический аппарат программы PASS и состав четырех баз знаний, которые используются при прогнозе спектра биологических активностей на информационно-вычислительной платформе Phyto4Health. Описано использование программы PharmaExpert в вычислительных экспериментах, проводимых в рамках диссертационного исследования.

В четвертом разделе главы 2 указано, что хранение и обработка собранных данных выполнялась с использованием СУБД MySQL.

В пятом разделе главы 2 указано, что разработка веб-интерфейса информационно-вычислительной платформы Phyto4Health была выполнена с использованием языка гипертекстовой разметки HTML, формального языка декорирования веб-страниц CSS и языка программирования JavaScript, а серверная часть – с использованием языка программирования PHP.

В шестом разделе главы 2 описаны поисковые инструменты. Описан поиск по текстовым запросам, который реализован с использованием DataTables – jQuery-плагина для создания динамических HTML-таблиц. Детально описаны алгоритмы поиска по сходству химических структур: 1) расчёт коэффициента Танимото между структурами, представленными в формате MNA-дескрипторов; 2) расчёт коэффициента Танимото по Тодескини между структурами, представленными в формате QNA-дескрипторов. Приводятся математические формулы расчета всех указанных

метрик.

В седьмом разделе главы 2 описаны методы анализа структурного сходства массивов природных соединений с использованием следующих алгоритмов. Алгоритм DBSCAN основан на плотности пространственной кластеризации для приложений с шумами. Для описания структур использовались молекулярные «отпечатки пальцев» с расширенной связностью ECFP, сходство рассчитывалось по Танимото. Кластерный анализ выполнялся методом k-средних с предварительным снижением размерности пространства оценок сходства с применением алгоритма равномерной аппроксимации и проекции UMAP.

В восьмом разделе главы 2 описаны методы ретроспективной и проспективной валидации информационно-вычислительной платформы Phyto4Health. Для этого использовалась современная версия программы PASS, имплементированная в данный Web-ресурс.

В качестве массива структур природного происхождения для ретроспективной валидации использовалась БД LOTUS. В ней было найдено 2732 соединения, для которых в БД PubChem имелись результаты экспериментального тестирования 83 видов биологической активности.

Проспективная валидация выполнена на основе современной версии системы PharmaExpert для соединений из созданной БД Phyto4Health. Осуществлялся прогноз: 1) цитотоксического действия в отношении клеток рака мочевого пузыря двух композиций препарата «Фитоладаптоген» (мажорной из 22 соединений и минорной из 13 соединений); 2) антиагрегантного и антикоагулянтного действия пяти фитокомпонентов морошки обыкновенной; 3) гепатопротекторного эффекта, ингибирования цитохрома P450, ингибирования глутатионтрансферазы для двух соединений, выделенных из травы цикория обыкновенного.

В третьей главе представлены результаты проведенных исследований и их обсуждение.

В первом разделе главы 3 приведена логическая структура создаваемой М:М реляционной базы данных, которая содержит 13 таблиц, объединённых между собой посредством внешних ключей. Изложено содержание каждой таблицы и ее полей.

Во втором разделе главы 3 описаны результаты извлечения и агрегации данных. Для 233 видовых названий официальных лекарственных растений России из 14-го издания Государственной фармакопеи Российской Федерации найдено 533 наиболее часто используемых синонимов. Выявлено 3128 уникальных фитокомпонентов, входящих в состав указанных растений, найдены структуры этих соединений, а также 9484 значений о содержании данных веществ в отдельных частях растений. Дополнительно найдены сведения о 13688 взаимодействиях с 802 биомолекулами человека, с указанием уровня активности IC_{50} , K_i , EC_{50} . В результате создана база данных платформы Phyto4Health.

В третьем разделе главы 3 показаны результаты создания доступного в сети Интернет веб-интерфейса разрабатываемой информационно-вычислительной платформы. Приведены общий интерфейс и интерфейсы поиска по текстовым запросам и по структурному сходству, примеры результатов выполнения поисковых запросов и результатов компьютерного прогноза фармакологических эффектов.

В четвертом разделе главы 3 приводятся и обсуждаются результаты структурного сопоставления массива соединений созданной базы данных Phyto4Health с данными о фитокомпонентах традиционной медицины других стран. Рассмотрены шесть таких БД общим объемом 26011 природных соединений. Средний взвешенный коэффициент сходства Танимото на основе MNA-дескрипторов по всем этим БД, с учетом их представительности, составил 0.64. При этом значения коэффициента сходства изменялись от 0.57 (BIOFACQUIM, Мексика) до 0.71 (NANPDB, Северная Африка).

Оценка сходства с использованием подхода ECFP/UMAP и кластерного анализа показала, что соединения БД Phyto4Health охватывают отдельную область химического пространства.

Сделан вывод о том, что сформированный набор данных в БД Phyto4Health охватывает отдельную область химического пространства по сравнению с соединениями других БД.

В пятом разделе главы 3 изложены и обсуждены результаты ретроспективной валидации информационно-вычислительной платформы.

Установлено, что при прогнозе 83 видов активности природных соединений из БД LOTUS среднее значение чувствительности составило 0,8, специфичности 0,6, а среднее значение сбалансированной точности прогноза составляет 0,7. Сделан вывод о том, что результаты компьютерного прогноза согласуются с данными, полученными в ходе *in vitro* экспериментов по определению биологической активности природных соединений.

В шестом разделе главы 3 изложены и обсуждены результаты проспективной валидации информационно-вычислительной платформы.

Прогноз для двух композиций препарата «Фитоладаптоген» (мажорной из 22 соединений и минорной из 13 соединений) спектра таргетных видов активности, связанных с противоопухолевым действием, показал, что наибольшее число соединений с наиболее релевантными оценками активностей ($P_a - P_i > 0.9$) обнаружено в мажорной композиции. Сделан вывод о том, что по результатам компьютерного прогноза мажорная композиция должна быть более активна. Последующие эксперименты *in vitro* показали более высокую цитотоксичность мажорной фармкомпозиции препарата «Фитоладаптоген» в отношении клеток рака мочевого пузыря человека линии RT-112. Сделан вывод о том, что полученные экспериментальные данные *in vitro* соответствуют результатам компьютерного прогноза *in silico*, и указывают на преимущества мажорного состава изучаемой фармкомпозиции.

Прогноз антиагрегантного и антикоагулянтного действия пяти фитокомпонентов морошки обыкновенной показал, что у трех из них следует ожидать проявления антитромботического действия. Остальные два соединения должны быть малоактивными. Результаты последующих экспериментов *in vitro* с использованием крови здоровых доноров-мужчин полностью совпали с результатами компьютерного прогноза.

Прогноз для двух соединений, выделенных из травы цикория обыкновенного (цикориевая и хлорогеновая кислоты) спектра таргетных активностей, релевантных гепатопротекторному действию, ингибированию цитохрома P450 и ингибированию глутатионтрансферазы показал, что расчетные оценки активности этих двух веществ согласуются с известными экспериментальными данными.

В седьмом разделе главы 3 приведены и обсуждаются результаты компьютерного прогноза фармакологического потенциала компонентов 25 фармакопейных растений России, наиболее охарактеризованных с точки зрения фитохимического состава по данным Phyto4Health. Эти растения, суммарно содержат в своем составе 1820 уникальных химических соединений, для большинства из которых по прогнозу можно ожидать проявления от 103 до 190 фармакологических активностей. Делается вывод о том, что полученные данные косвенно указывают на высокое химическое разнообразие рассматриваемого набора структур, а это указывает на многочисленные возможности их исследования в качестве «кандидатов» на лекарственные препараты для лечения различных патологий. При этом наиболее перспективными следует считать 16 соединений, для которых наиболее целесообразно исследовать 13 фармакологических активностей.

Заключение и выводы завершают диссертацию и включают практические рекомендации, которые логично вытекают из результатов работы, отражают поставленные задачи, аргументированы и имеют научно-практическую значимость.

Соответствие содержания автореферата основным положениям и выводам диссертации

Автореферат адекватно отражает основное содержание диссертационного исследования, полностью соответствует разделам, положениям и выводам диссертационной работы и оформлен в соответствии с предъявляемыми требованиями.

Вопросы и замечания

Принципиальных замечаний и возражений по диссертационной работе нет, однако в процессе ознакомления с ней возникли следующие вопросы, требующие дополнительного уточнения, но не затрагивающие существа работы.

1. В системе PASS применяемая для прогноза QSAR-база сформирована преимущественно с использованием синтетических соединений. В диссертации же она используется для прогноза активности природных веществ. В разделе 2.8 указано, что «... использование моделей машинного обучения, разработанных на основе обучающих выборок,

содержащих синтетические соединения с целью оценки различных свойств молекул природного происхождения, может оказаться малоэффективным». Поясните, можно ли все-таки использовать обучающие выборки из синтетических соединений для прогноза биологической активности природных соединений?

2. Почему для определения оптимального числа кластеров и их компактности не используется оценка по F-критерию Фишера, традиционно применяемая в методе k-средних?

3. В разделе 3.4 указано: «В результате проведенного нами расчета установлено, что на основе плотности данные разделяются на 23 кластера». Но далее из описания результатов видно, что для анализа используется 28 кластеров. Еще ниже указано, что «... соединения в БД Phyto4Health представлены только в восьми из 29 кластеров». Поясните, какое все-таки было найдено оптимальное число кластеров – 23, 28 или 29?

4. Какой статистический показатель использовался для определения оптимального числа кластеров, каково его граничное значение и какая величина была рассчитана для оптимального числа кластеров?

5. Какова статистическая достоверность p выбранного оптимального числа кластеров?

Заключение

Диссертация Ионова Никиты Сергеевича на тему «Разработка информационно-вычислительной платформы для оценки фармакологического потенциала фитокомпонентов лекарственных растений» является законченной научно-квалификационной работой, в которой содержится решение актуальной задачи биоинформатики по созданию нового уникального информационно-вычислительного Web-ресурса на основе разработанной диссертантом методологии обработки информации о соединениях природного происхождения, расширяющей первичные исходные данные и предусматривающей их интеллектуальную структуризацию, что позволяет проводить комплексные исследования по поиску *in silico* природных соединений и их производных с высокой фармакологической активностью.

С учетом актуальности, высокого методического уровня, научной

новизны, теоретической и практической значимости исследования считаю, что представленная работа соответствует требованиям п. 9 «Положения о присуждении ученых степеней», утвержденного постановлением Правительства РФ № 842 от 24.09.2013 (в редакции постановления Правительства РФ № 101 от 26.01.2023), предъявляемым к диссертациям на соискание ученой степени кандидата наук, а ее автор, Ионов Никита Сергеевич, заслуживает присуждения искомой степени кандидата биологических наук по специальности 1.5.8. – Математическая биология, биоинформатика.

Официальный оппонент:

Заведующий лабораторией

информационных технологий в фармакологии

и компьютерного моделирования лекарств

НЦИЛС ФГБОУ ВО «Волгоградский

государственный медицинский университет»

Минздрава России,

доктор биологических наук



Васильев Павел Михайлович

15. 11. 2024г.

ФГБОУ ВО «Волгоградский государственный медицинский университет» Минздрава России

Адрес: 400131, г. Волгоград, пл. Павших борцов, д. 1

Тел: +7 (8442) 38-50-05

E-mail: post@volgmed.ru

<https://www.volgmed.ru/>

