МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

ИМЕНИ М.В. ЛОМОНОСОВА

На правах рукописи

Сулимов Алексей Владимирович

МОЛЕКУЛЯРНЫЙ ДОКИНГ: МАТЕМАТИЧЕСКИЕ МОДЕЛИ, СУПЕРКОМПЬЮТЕРНАЯ РЕАЛИЗАЦИЯ И ИХ ПРИМЕНЕНИЕ ДЛЯ РАЗРАБОТКИ ЛЕКАРСТВ

Специальность 1.1.10 – «Биомеханика и биоинженерия»

АВТОРЕФЕРАТ

диссертации на соискание ученой степени

кандидата физико-математических наук

Москва – 2022

Работа выполнена в лаборатории вычислительных систем и прикладных технологий программирования Научно-исследовательского вычислительного центра МГУ имени М.В. Ломоносова.

Научный руководитель **Тыртышников Евгений Евгеньевич**
Доктор физико-математических наук, академик РАН

Официальные оппоненты **Поройков Владимир Васильевич**
Доктор биологических наук, член-корреспондент РАН, профессор, ФГБНУ «Научно-исследовательский институт биомедицинской химии имени В.Н. Ореховича», главный научный сотрудник, заведующий отделом биоинформатики и лабораторией структурно-функционального конструирования лекарств

Шайтан Константин Вольдемарович
Доктор физико-математических наук, профессор,
МГУ имени М.В. Ломоносова, Биологический факультет,
Кафедра биоинженерии, профессор

Быстров Владимир Сергеевич
Доктор физико-математических наук,
Институт математических проблем биологии РАН – филиал ФГУ «Федеральный исследовательский центр Институт прикладной математики им. М.В. Келдыша РАН», Отдел перспективных информационных технологий, Группа компьютерного моделирования молекулярных наноструктур и биосистем, главный научный сотрудник, руководитель группы

Защита диссертации состоится 09 июня 2022 г. в 17:30 на заседании диссертационного совета МГУ.011.1 (МГУ.01.04) Московского государственного университета имени М.В. Ломоносова по адресу: 119991 Москва, Ленинские горы, д. 1, стр. 2, Физический факультет МГУ, ауд. 5-68. E-mail: info@physics.msu.ru

С диссертацией можно ознакомиться в отделе диссертаций научной библиотеки МГУ имени М.В. Ломоносова (Ломоносовский пр-т, д. 27) и на сайте ИАС «ИСТИНА»: https://istina.msu.ru/dissertations/453727234/

Автореферат разослан « ___ » _____ 20___ г.

Ученый секретарь диссертационного совета МГУ.011.1 (МГУ.01.04)

кандидат технических наук А.Э. Сидорова

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность работы

В настоящее время рациональная разработка лекарств проводится во многих случаях в соответствии со следующей парадигмой: биологически активная молекула должна связаться в определенном месте биологической макромолекулы и изменить течение заболевания. Наиболее часто такой макромолекулой является белок, играющий ключевую роль в развитии заболевания, и биологически активная молекула должна связаться с активным центром этого белка. Эти белки часто называют терапевтическими мишенями или просто белками-мишенями.

Такие разработки проводятся в основном методами экспериментального тестирования многих тысяч, а иногда и миллионов соединений, прежде чем будет открыто новое лекарство [1]. В настоящее время затраты на создание нового лекарства могут составлять более 2 миллиардов долларов США и занимать время до 15 лет [2]. С каждым годом к новым лекарствам усиливаются требования по токсичности, и разработать в качестве лекарства новое химическое соединение становится всё труднее. Поэтому задача повышения эффективности рациональной разработки лекарств с каждым годом становится всё более и более актуальной. Побочные эффекты, связанные с приемом лекарства могут проявиться на поздних стадиях разработки или даже после получения разрешения на его широкое применение. Поэтому этот процесс должен продолжаться непрерывно подобно непрерывно работающему конвейеру: если лидерное соединение становится бесперспективным из-за выявленных побочных эффектов на каком-либо этапе разработки, вслед за ним должны быть на подходе и другие, уже прошедшие предыдущие этапы соединения-кандидаты. Данный конвейер должен работать до тех пор, пока новое лекарство не пройдет самую широкую апробацию.

Одним из способов существенного повышения эффективности такого конвейера становится применение методов молекулярного моделирования. Для

поиска соединений, молекулы которых избирательно связываются с активными центрами заданных белков-мишеней, широко применяется молекулярный докинг. Если 20 лет назад успехи разработок новых лекарств с применением программ докинга были скорее случайными, то в настоящее время известно уже довольно много публикаций, в которых при поиске лекарственных соединений регулярно использовался докинг [А3]. Несмотря на наличие нескольких десятков программ докинга [3], точность предсказания энергии связывания молекулы-лиганда предполагаемого биологически активного соединения с белком-мишенью ещё далека от того уровня, когда на основе расчетов можно оптимизировать структуру молекулы, различая лиганды с низкой активностью от лигандов со средней активностью, а последние – от лигандов с высокой активностью по воздействию на заданный белок.

Чтобы такие предсказания стали надежными необходимо существенно повысить точность программ докинга. Поэтому разработка новых программ докинга, в том числе и суперкомпьютерных программ докинга, которые позволяют сосредоточить большие вычислительные мощности на моделировании взаимодействия одного лиганда с белком-мишенью, а значит выполнять такое моделирование с минимальными модельными упрощениями и с более высокой точностью, представляется чрезвычайно актуальной. Каждая разрабатываемая программа докинга уникальна, благодаря особенной комбинации заложенных в ней физических, химических и математических моделей, сделанных упрощений и аппроксимаций, используемых алгоритмов и их программной реализации. Учитывая это, разработка отечественных программ докинга, и говоря шире, целой технологии докинга, включающей и подготовку для докинга лигандов, и создание атомистических моделей белков-мишеней, организация докинга многих тысяч лигандов на суперкомпьютерах, сбор и анализ полученных результатов докинга, является актуальной.

С другой стороны, важную роль играет и использование больших вычислительных ресурсов для перебора с помощью докинга многих тысяч лигандов, чтобы выявить наиболее перспективные кандидаты для

экспериментального тестирования. Актуальность этого подтверждается и чрезвычайно широким использованием программ докинга с началом пандемии COVID-19. За 2 прошедших года появились десятки работ, посвященных поиску с помощью докинга соединений, молекулы которых избирательно связываются с белками-мишенями коронавируса SARS-CoV-2, вызвавшего пандемию. Такие соединения могут стать основой противовирусных препаратов, подавляющих репликацию коронавируса в организме.

Научная новизна

Разработана методика проведения суперкомпьютерного виртуального скрининга многих тысяч низкомолекулярных соединений для выявления кандидатов в ингибиторы актуальных терапевтических белков-мишеней с использованием суперкомпьютерных мощностей. Поиск ингибиторов осуществляется с использованием оригинальной программы докинга SOL [A15,A17,P3] и квантово-химической программы MOPAC [4].

Разработана оригинальная атомистическая модель главной протеазы коронавируса SARS-CoV-2, вызвавшего пандемию COVID-19, и на ее основе с применением методики суперкомпьютерного виртуального скрининга найдены новые соединения, подавляющие репликацию коронавируса в культуре клеток. Эти соединения могут стать основой нового класса противовирусных лекарств прямого действия на коронавирус SARS-CoV-2 [A1,A2,A3].

Впервые разработаны программы докинга, осуществляющие поиск глобального минимума системы белок-лиганд методом глобальной оптимизации, основанным на разложении многомерных массивов в тензорные поезда [P1,A10,A11,A14]. При этом программные комплексы SOL-T [A11] и SOL-P [A10] не используют при докинге сетку заранее рассчитанных потенциалов взаимодействия пробных атомов лигандов с белком. С помощью программного комплекса SOL-P впервые проведен успешный докинг олигопептидов имеющих до 25 внутренних вращательных степеней свободы включительно [A7].

Степень разработанности темы исследования

В отечественной литературе 15 лет назад, когда начинались исследования, не было публикаций, посвященных разработке отечественных программ докинга и их применению на начальном этапе разработки новых лекарств, а о зарубежных программах только начинали появляться первые публикации. Даже и сейчас, среди нескольких десятков известных в мире программ докинга [А3] нет доступных отечественных программ. Каждая из этих программ имеет свои уникальные черты организации работы, алгоритма, положенных в основу модельных представлений и аппроксимаций, методов расчета энергии взаимодействия атомов белка и лиганда, методов учета растворителя – воды, а чаще его отсутствия или крайне упрощенного учета.

Большинство программ докинга демонстрирует сравнительно неплохую точность позиционирования лиганда в активном центре белка-мишени, но точность оценки свободной энергии связывания лиганда с белком, определяющей надежность предсказания активности данной молекулы-лиганда, представляется неудовлетворительной [А3]. Большинство существующих программ докинга создано в рамках парадигмы: быстрее и еще быстрее, так что докинг одного лиганда проводится многими программами за несколько минут, и даже быстрее, на ноутбуке. До сих пор для докинга используются в основном ноутбуки или обычные персональные компьютеры.

Изучение публикаций о применении суперкомпьютеров для докинга показало, что во всех обнаруженных на эту тему публикациях суперкомпьютерные вычислительные мощности применяются для докинга многих миллионов лигандов с помощью обычных (непараллельных) программ докинга. Применение методов квантовой химии для решения задачи докинга началось сравнительно недавно [А3]. Главным образом методы квантовой химии используются для вычисления функции оценки энергии связывания белок-лиганд, для лучших положений лиганда в активном центре белка-мишени, полученными обычными «классическими» программами докинга.

Цель и задачи диссертационного исследования

Разработать математические методы, алгоритмы и программы докинга, провести их тестирование и применить их для реальных разработок ингибиторов заданных белков-мишеней. Исследовать причины, ограничивающие точность докинга, и разработать параллельные суперкомпьютерные программы докинга, свободные от большинства ограничений, присущих существующим программам докинга. Исследовать возможность применения методов квантовой химии для разработки ингибиторов заданных белков-мишеней. Для достижения поставленной цели были сформулированы следующие задачи:

1. Разработка методики суперкомпьютерного виртуального скрининга многих тысяч низкомолекулярных соединений для выявления кандидатов в ингибиторы терапевтических белков-мишеней с помощью программы докинга SOL и программы квантовой химии MOPAC с использованием суперкомпьютерных вычислительных ресурсов;
2. Разработка вспомогательных программ для подготовки молекул для докинга и последующего вычисления энтальпии связывания, а также для проведения массированных суперкомпьютерных расчетов и анализа полученных результатов;
3. Применение разработанной методики суперкомпьютерного виртуального скрининга для поиска ингибиторов главной протеазы коронавируса SARS-CoV-2, вызвавшего пандемию COVID-19;
4. Разработка суперкомпьютерной программы прямого докинга, не использующей сетку заранее рассчитанных потенциалов взаимодействия пробных атомов лигандов с белком;
5. Применение разработанной суперкомпьютерной программы прямого докинга для проведения успешного докинга гибких лигандов, имеющих более 15 внутренних вращательных степеней свободы;

Объект и предмет исследований

Объектом данного исследования являются связывание молекул низкомолекулярных соединений (лигандов) с активными центрами терапевтических белков-мишеней, отвечающих за развитие тех или иных заболеваний, молекулярные модели белков-мишеней, базы данных и библиотеки низкомолекулярных органических соединений.

Предметом исследований является компьютерное моделирование связывания лигандов с белками-мишенями, суперкомпьютерные расчеты, молекулярный докинг, разработка методов, алгоритмов и программ докинга, его математические модели, суперкомпьютерная реализация и применение докинга для разработки лекарств.

Методология и методы диссертационного исследования

Методология проведенного исследования опирается на использование физических и химических методов и подходов к описанию межмолекулярного взаимодействия белков-мишеней и молекул-лигандов, на классические силовые поля, описывающие межмолекулярные и внутримолекулярные межатомные взаимодействия, на полуэмпирические квантово-химические методы, на континуальные модели растворителя, а также на современные методы глобальной и локальной оптимизации систем с большим числом степеней свободы и методы многопроцессорных суперкомпьютерных вычислений.

Проведенные исследования позволили выработать методологию подготовки лигандов и белков для докинга, методологию суперкомпьютерного докинга многих тысяч молекул-лигандов и выявление среди них молекул кандидатов в ингибиторы заданных белков-мишеней.

Теоретическая значимость работы

Теоретическая значимость работы заключается в разработке математических методов, алгоритмов и их программной реализации для позиционирования (докинга) молекул-лигандов в активном центре белка-мишени и оценке свободной энергии связывания белок-лиганд.

Практическая значимость работы

Практическая значимость работы заключается в том, что с помощью разработанной методики суперкомпьютерного виртуального скрининга баз данных органических низкомолекулярных соединения были открыты новые классы антикоагулянтов на основе низкомолекулярных ингибиторов факторов свертываемости крови Xa, XIa, а также ингибиторов главной протеазы вызвавшего пандемию COVID-19 коронавируса SARS-CoV-2.

Положения, выносимые на защиту

1. Разработка методики суперкомпьютерного виртуального скрининга низкомолекулярных соединений с использованием программы докинга SOL и программы квантовой химии MOPAC для выявления кандидатов в ингибиторы терапевтических белков-мишеней. Создание скриптов – вспомогательных программных модулей, по автоматической подготовке молекулярных структур для запуска докинга и квантово-химических расчетов, а также для проведения массированных суперкомпьютерных расчетов и анализа полученных результатов.
2. Разработка атомистической модели главной протеазы коронавируса SARS-CoV-2, вызвавшего пандемию COVID-19, и с помощью разработанной методики суперкомпьютерного виртуального скрининга выявление в базах данных низкомолекулярных органических соединений новых кандидатов в ингибиторы главной протеазы.
3. Разработка программного комплекса SOL-P, находящего глобальный минимум системы белок-лиганд в силовом поле MMFF94 с использованием метода глобальной оптимизации, основанного на разложении многомерных массивов в тензорные поезда.
4. Применение программного комплекса SOL-P для успешного докинга олигопептидов, имеющих до 25 внутренних вращательных степеней свободы.

Степень достоверности

Достоверность полученных результатов следует из корректности постановки задачи, опирающейся на парадигму докинга, согласно которой лиганд связывается с белком-мишенью вблизи глобального минимума системы белок-лиганд. Проблема докинга сводится к поиску глобального минимума на энергетической поверхности белок-лиганд. Высокая степень достоверности результатов определяется использованием проверенных методов глобальной оптимизации, использованием известных силовых полей, методов квантовой химии и континуальных моделей растворителя. Достоверность полученных результатов подтверждается так же хорошим совпадением найденных при докинге положений лигандов с экспериментальными положениями этих лигандов для широкого набора тестовых комплексов.

Личный вклад

Автор диссертации лично разработал методику суперкомпьютерного виртуального скрининга для поиска среди множества низкомолекулярных органических соединений кандидатов в ингибиторы исследуемых белков-мишеней. В работах [A5,A8,A9,A12,A13], опубликованных в соавторстве, основополагающий вклад по применению методики суперкомпьютерного виртуального скрининга принадлежит соискателю. Автор диссертации разработал и реализовал вспомогательные программы по автоматизации подготовки входных файлов программы докинга SOL и программы квантовой-химии MOPAC, которые необходимы для проведения виртуального скрининга. Автор диссертации лично разработал и реализовал вспомогательные скрипты для проведения массированных расчетов на суперкомпьютерах МГУ Ломоносов-1 и Ломоносов-2. Автор диссертации анализировал полученные результаты и корректировал применяемую методику виртуального скрининга с целью улучшения качества выявления ингибиторов для исследуемых белков-мишеней: факторов свертывания крови Xa, XIa и урокиназы (uPA).

Автор диссертации является одним из разработчиков основного инструмента методики суперкомпьютерного виртуального скрининга – программы докинга SOL [P3]. В работе [A17], опубликованной в соавторстве, соискатель осуществлял тестирование разработанной программы докинга SOL, готовил тестовые наборы данных для проверки, сравнивал найденные при докинге положения лигандов с экспериментальными положениями соответствующих лигандов в активных центрах белков-мишеней. В работе [A16], опубликованной в соавторстве, соискатель адаптировал использование программы докинга SOL для многопроцессорных вычислений. В работе [A15], опубликованной в соавторстве, основополагающий вклад по проведению расчетов и анализу результатов программы SOL в рамках соревнований по докингу CSAR2011-2012 принадлежит соискателю.

В работах [A1–A4], опубликованных в соавторстве, вклад автора является определяющим. Автор диссертации лично разработал атомистическую модель главной протеазы коронавируса SARS-CoV-2, применил разработанную методику суперкомпьютерного виртуального скрининга для поиска кандидатов в ингибиторы главной протеазы коронавируса SARS-CoV-2, а также лично анализировал результаты и отбирал наиболее перспективные соединения для экспериментальной проверки.

Автор диссертации является одним из разработчиков новых программ докинга TTDock, SOL-T и SOL-P, в работе которых используется метод глобальной оптимизации на основе разложения многомерных массивов в тензорные поезда. При этом программные комплексы SOL-T и SOL-P не используют при докинге сетку заранее рассчитанных потенциалов взаимодействия пробных атомов лигандов с белком. В работах [A10–A11,A14], опубликованных в соавторстве, соискатель внес основополагающий вклад в разработку архитектуры программ, в создание тестовых наборов комплексов, в проведение тестовых расчетов и в анализ полученных результатов. В работах [A6,A7], опубликованных в соавторстве, соискатель внес основополагающий

вклад в подготовку соединений, имеющих до 25 внутренних вращательных степеней свободы и в проведение докинга программным комплексом SOL-P.

Публикации

Всего опубликовано 77 статей. По теме диссертации опубликовано всего 17 статей, из них 17 статей в научных изданиях, индексируемых в базах данных Web of Science, Scopus, RSCI и РИНЦ. Кроме того, опубликовано 2 монографии, а также получены свидетельства о государственной регистрации 4 программ для ЭВМ.

Апробация результатов

Основные результаты диссертации были представлены и обсуждены на 23 международных и всероссийских конференциях (статьи в сборниках – 2, тезисы – 10, доклады – 23).

Структура и объем диссертации

Диссертационная работа включает: введение, четыре главы, заключение, результаты и выводы, список цитируемой литературы. Работа изложена на 148 страницах и содержит 11 рисунков. Библиография включает 202 источника.

СОДЕРЖАНИЕ РАБОТЫ

В первой главе рассмотрена парадигма рациональной разработки лекарств, основанная на знании молекулярных механизмов развития заболеваний. Согласно этой парадигме, молекула лекарства должна избирательно связаться с активным центром белка, играющего ключевую роль в развитии заболевания, и изменить течение болезни. Рассмотрены методы молекулярного моделирования, применяемые при разработке лекарств, и наиболее подробно из них рассмотрен молекулярный докинг. Отмечено, что пандемия COVID-19 высветила чрезвычайно высокую востребованность программ докинга на начальной стадии разработки лекарств. Программы докинга находят положения молекул-лигандов в активных центрах белков-мишеней и дают оценку свободной энергии связывания белок-лиганд.

Чем выше эта энергия, тем с большей вероятностью данный лиганд свяжется с белком и ингибирует его работу при экспериментальном тестировании. Кроме того, ингибиторы с большой энергией связывания могут стать основой более эффективных и менее токсичных лекарств, поскольку меньшая концентрация ингибитора будет приводить к желаемому терапевтическому эффекту. Приведен обзор популярных сейчас программ докинга. В обзоре основное внимание уделено алгоритмам докинга, алгоритмам глобальной оптимизации, методам расчета энергии в этих программах, методам учета растворителя, поскольку связывание белок-лиганд проходит в водном растворе. Уделено внимание и упрощениям, которые сделаны в различных программах докинга для ускорения расчетов, поскольку это влияет на точность докинга. Отдельный параграф посвящен рассмотрению применения суперкомпьютеров для докинга. Отмечено, что каждая программа докинга имеет свои характерные черты, связанные с алгоритмом докинга, с использованным силовым полем для вычисления энергии молекулярной системы, с учетом или не учетом растворителя в той или иной модели, набором упрощений, чаще всего направленных на ускорение расчетов за счет их точности. При этом у большинства программ докинга точность позиционирования лигандов удовлетворительна, но надежность прогнозирования величины свободной энергии связывания белок-лиганд остается неудовлетворительной, несмотря на успехи применения докинга для разработки лекарств.

Во второй главе описывается методика (Рисунок 1) виртуального суперкомпьютерного скрининга больших баз данных низкомолекулярных соединений с целью нахождения ингибиторов заданных белков-мишеней. Основными инструментами методики являются оригинальная программа докинга SOL [A15] и квантово-химическая программа MOPAC [4]. На первом этапе методики осуществляется разработка трехмерной молекулярной модели белка-мишени и её проверка путем успешного позиционирования программой докинга SOL закристаллизованного вместе с белком-мишенью нативного лиганда. На втором этапе методики осуществляется подготовка для докинга

трехмерных структур лигандов низкомолекулярных соединений исследуемых баз данных. На третьем этапе с помощью программы SOL проводится докинг всех подготовленных трехмерных конформаций лигандов в разработанную модель белка-мишени. По окончании данного этапа проводится отбор лучших лигандов на основании функции оценки программы SOL свободной энергии связывания в силовом поле MMFF94 [5]. Для каждого отобранного лиганда берется только лучшая его конформация. На четвертом заключительном этапе для отобранных конформаций программой MOPAC проводится расчет энтальпии связывания белка с лигандом. Расчет проводится квантово-химическим методом PM7 [6] с учетом растворителя в модели COSMO [7].

Рисунок 1 – Методика суперкомпьютерного виртуального скрининга.

Описана реализация методики виртуального скрининга на суперкомпьютерных системах с использованием вспомогательных программ, позволяющих максимально автоматизировать все этапы. По результатам проведения методики суперкомпьютерного виртуального скрининга происходит отбор кандидатов в ингибиторы исследуемого терапевтического белка-мишени. Отбор осуществляется на основании оценки свободной энергии связывания и на основании рассчитанной величины энтальпии связывания методом квантовой химии, а также на основании визуального анализа найденных при докинге положений лигандов в активном центре белка-мишени.

В третьей главе диссертации дано описание разработанной атомистической модели 6W63 главной протеазы коронавируса SARS-CoV-2, вызвавшего пандемию COVID-19, а также тестирование этой модели. Подробно описано проведение суперкомпьютерного виртуального скрининга в построенную модель главной протеазы SARS-CoV-2 для трех различных наборов соединений: базы данных существующих лекарственных соединений ZINC 15 [8], базы данных соединений кафедры органической химии Воронежского государственного университета [9] и коммерческой базы данных Maybridge Screening Collection [10]. По результатам проведения виртуального скрининга отобраны для экспериментальной проверки кандидаты в ингибиторы.

В Таблице 1 приведены результаты экспериментальной проверки отобранных соединений базы данных Воронежского государственного университета, представлена активность *in vitro* ингибирования рекомбинантной главной протеазы коронавируса SARS–CoV–2, а также результаты подавление репликации коронавируса в культуре клеток *in vivo*. Как видим, в результате экспериментального тестирования несколько молекул показали активность по ингибированию главной протеазы SARS–CoV–2 в микромолярном диапазоне IC_{50}, и несколько из них подавляли репликацию SARS–CoV–2 в культуре клеток на уровне микромолярных значений EC_{50} и имели низкую цитотоксичность.

Таблица 1 – Результаты ингибирования коронавируса SARS-CoV-2.

Лиганд	IC_{50}, μM	ΔG_{bind}, ккал/моль	ΔH_{bind}, ккал/моль	EC_{50}, μM	SI
2425	1.77 ±0.01	–7.1	–54.4	–	–
2900	4.05 ±0.02	–6.7	–44.6	–	–
6276	8.59 ±0.03	–6.6	–44.0	–	–
15689	6.25 ±0.03	–6.9	–55.5	18.19 ±0.51	11.9
16826	0.20 ±0.01	–6.6	–47.8	–	–
18302	41.62 ±0.12	–6.9	–50.8	4.77 ±0.33	45.1
214225	41.54 ±0.15	–6.8	–46.1	–	–
223350	24.50 ±0.12	–7.1	–49.5	6.53 ±0.42	27.3

В главе 4 диссертации описана разработка нового суперкомпьютерного программного комплекса прямого докинга SOL-P. Комплекс состоит из программ подготовки, анализа и поиска глобального минимума на сложной многомерной энергетической поверхности системы белок-лиганд (Рисунок 2). Работа программного комплекса SOL-P состоит из трех основных этапов. На этапе I осуществляется подготовка к докингу входных файлов, и расчетным модулем TT-Dock проводится поиск спектра низкоэнергетических минимумов комплекса белок-лиганд в рамках силового поля MMFF94 в вакууме. На этапе II осуществляется фильтрация одинаковых локальных минимумов. На этапе III проводится тщательная локальная оптимизация энергии каждого уникального минимума градиентным методом L-BFGS. Модуль поиска низкоэнергетических минимумов TT-Dock использует TT-алгоритм глобальной оптимизации основанный на разложении многомерных массивов в тензорные поезда и новом методе TT-CROSS [11]. Модуль TT-Dock не используют сетку заранее рассчитанных потенциалов взаимодействия пробных атомов лигандов с белком. Благодаря этому поиск глобального минимума осуществляется не только при варьировании положения и ориентации лиганда как целого, а также его торсионных углов, но и при варьировании декартовых координат заданных

атомов белка. В качестве области докинга модуль TT-Dock использует куб с ребром 10 Å, центрированный в геометрическом центре закристаллизованного нативного положения лиганда. Выделенные подвижные атомы белка изменяют декартовые координаты в пределах небольших кубов с длинной ребра 1 Å. В результате работы программного комплекса SOL-P определяется глобальный минимум и сохраняется целый спектр уникальных низкоэнергетических локальных минимумов комплекса белок-лиганд. Спектр сохраненных минимумов используется для анализа работы программного комплекса SOL-P.

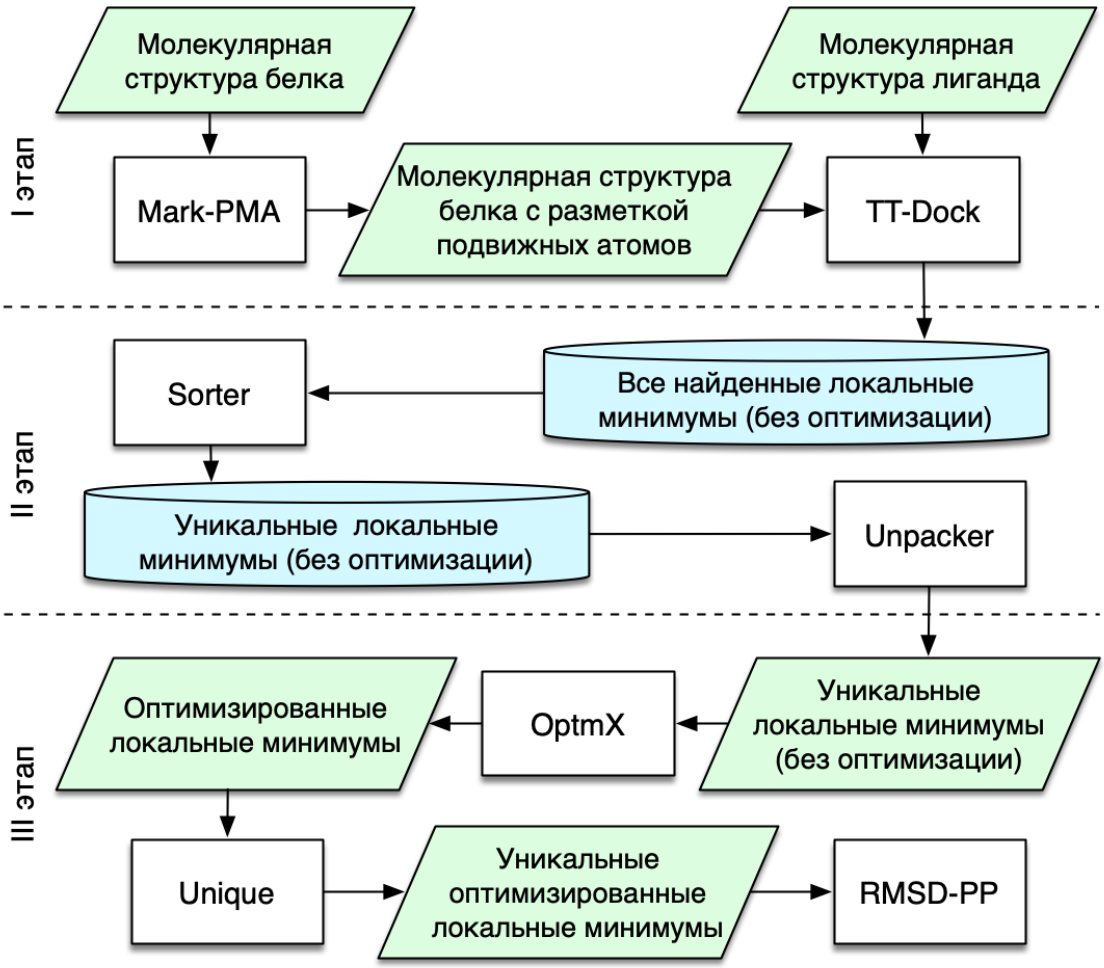

Рисунок 2 – Блок-схема программного комплекса SOL-P.

При исследовании параметров глобальной оптимизации расчетного модуля TT-Dock было показано успешное выполнение докинга нативнного лиганда комплекса 3CEN, имеющего 7 внутренних вращательных степеней свободы, и включением подвижности для 48 атомов белка. То есть в

пространстве поиска со 157 измерениями было показано нахождение глобального минимума, которому соответствует поза лиганда вблизи закристаллизованного с белком положения нативного лиганда.

Далее было проведено исследование возможностей программного комплекса SOL-P при докинге олигопептидов – молекул, состоящих из аминокислотных остатков, и имеющих большое количество внутренних вращательных степеней свободы. Из Protein Data Bank было выбрано 9 комплексов с хорошим разрешением (1.3 – 2.1 Å) и с закристаллизованными олигопептидами, состоящими из разного числа различных аминокислотных остатков. Выбранные комплексы содержали олигопептиды, которые имели от 10 до 25 торсионов. Для сравнения качества докинга сначала был проведен докинг с помощью классической программы докинга SOL, а потом уже проведен докинг с помощью программного комплекса SOL-P. Программа SOL была выбрана потому, что в ней энергия молекулярной системы вычисляется в том же силовом поле MMFF94, как и в программном комплексе SOL-P, но поиск глобального минимума осуществляется с помощью генетического алгоритма. Точность докинга определялась путем вычисления среднеквадратичного отклонения атомов лиганда, находящегося в самом выгодном по энергии положении, от соответствующих атомов лиганда, закристаллизованного вместе с белком в структуре PDB. Успешным в данном исследовании считался докинг при RMSD менее 2.3 Å (обычно используется критерий 2 Å). Результаты докинга приведены в Таблице 2.

Докинг программой SOL с использованием стандартных параметров генетического алгоритма (Std) прошел успешно только для одного комплекса 6DQQ с небольшим олигопептидом (10 торсионов, 43 атома). Докинг при использовании повышенных параметров генетического алгоритма (High) прошел успешно для четырех комплексов 6DQQ, 1OLA, 2OLB и 6DQU. При этом населенность первого кластера при 50 запусках алгоритма была более 2 поз только при докинге на повышенных параметрах для комплекса 6DQQ, что указывает на ненадежность полученных результатов для других комплексов.

Таблица 2 – Качество докинга олигопептидов программами SOL и SOL-P.

Кол-во торсион. лиганда	PDB ID	Кол-во атомов лиганда	Кол-во атомов белка	SOL Std RMSD, Å	SOL High RMSD, Å	SOL-P RMSD, Å
10	6DQQ	43	8216	1.08	0.83	1.24
12	6M9C	60	5166	9.04	5.01	4.91
15	1B9J	66	8242	8.56	3.41	1.22
18	1OLA	62	8007	9.94	1.39	1.28
18	6DTH	82	8098	10.13	8.91	2.12
21	2D5W	82	9619	10.73	9.18	12.01
22	6DTG	81	8226	10.01	27.43	1.04
23	2OLB	69	8241	9.56	1.43	1.37
25	6DQU	95	8129	8.49	2.11	2.28

Далее, был проведен докинг тех же олигопептидов с помощью нового программного комплекса SOL-P без учета подвижности атомов белка. Для семи тестовых комплексов кроме двух (6M9C и 2D5W) докинг оказался успешным. Как видим, новый программный комплекс SOL-P справляется с докингом олигопептидов гораздо лучше, чем программа SOL. В работе [12] отмечается, что часто для докинга олигопептидов используются априорные соображения, позволяющие сильно ограничить область докинга. В данном исследовании не использовались какие-либо априорные соображения о предпочтительных конформациях олигопептида и упрощений при расчете энергии, и при этом была получена высокая точность позиционирования олигопептидов. Таким образом, показано, что разработанный программный комплекс прямого докинга SOL-P для 7 тестовых комплексов из 9 успешно выполняет докинг гибких олигопептидов, состоящих из 3-6 аминокислотных остатков и содержащих 10-25 внутренних вращательных степеней свободы и до 95 атомов лиганда. Это удалось сделать за счет применения в SOL-P TT-алгоритма докинга на основе нового алгоритма глобальной оптимизации, связанного с разложением многомерных тензоров в тензорные поезда.

В Заключении представлены основные результаты проведенных исследований и выводы, а также отмечено, что разработанная методика суперкомпьютерного виртуального скрининга больших баз данных лигандов показала свою эффективность на начальной стадии разработки новых ингибиторов заданных белков-мишеней, играющей важную роль в непрерывном конвейере рациональной разработки лекарств. Созданием суперкомпьютерного программного комплекса прямого докинга SOL-P, использующего новый алгоритм глобальной оптимизации, сделан важный шаг с целью повышения точности докинга путем применения больших вычислительных ресурсов и многопроцессорных вычислений.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ И ВЫВОДЫ

1. Разработана методика суперкомпьютерного виртуального скрининга баз данных низкомолекулярных соединений с использованием программы докинга SOL и программы квантовой химии MOPAC для выявления кандидатов в ингибиторы терапевтических белков-мишеней.
2. Разработана атомистическая модель главной протеазы коронавируса SARS-CoV-2, вызвавшего пандемию COVID-19, и с помощью разработанной методики суперкомпьютерного виртуального скрининга выявлены новые кандидаты в ингибиторы главной протеазы, среди которых оказались соединения, подавляющие репликацию этого коронавируса в культуре клеток.
3. Разработан суперкомпьютерный программный комплекс SOL-P, находящий глобальный минимум системы белок-лиганд в силовом поле MMFF94 с использованием метода глобальной оптимизации, основанного на разложении многомерных массивов в тензорные поезда.
4. С помощью программного комплекса SOL-P впервые успешно выполнен докинг олигопептидов, имеющих до 25 внутренних вращательных степеней свободы включительно.

СПИСОК ЦИТИРУЕМОЙ ЛИТЕРАТУРЫ

1. Phatak S.S., Stephan C.C., Cavasotto C.N. High-throughput and in silico screenings in drug discovery // Expert Opin. Drug Discov. Taylor & Francis, 2009. Vol. 4, № 9. P. 947–959.
2. DiMasi J.A., Grabowski H.G., Hansen R.W. Innovation in the pharmaceutical industry: New estimates of R&D costs // J. Health Econ. 2016. Vol. 47. P. 20–33.
3. Yuriev E., Holien J., Ramsland P.A. Improvements, trends, and new ideas in molecular docking: 2012-2013 in review // J. Mol. Recognit. 2015/03/27. 2015. Vol. 28, № 10. P. 581–604.
4. Stewart J.J.P. Stewart Computational Chemistry. MOPAC2016 [Electronic resource]. Colorado Springs, CO, USA, 2016. URL: http://openmopac.net/MOPAC2016.html (accessed: 08.04.2022).
5. Halgren T.A. Merck molecular force field // J. Comput. Chem. 1996. Vol. 17, № 5–6. P. 490–641.
6. Stewart J.J.P. Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters // J. Mol. Model. 11/29. 2013. Vol. 19, № 1. P. 1–32.
7. Klamt A., Schüürmann G. COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient // J. Chem. Soc., Perkin Trans. 2. The Royal Society of Chemistry, 1993. № 5. P. 799–805.
8. Sterling T., Irwin J.J. ZINC 15 – ligand discovery for everyone // J. Chem. Inf. Model. American Chemical Society, 2015. Vol. 55, № 11. P. 2324–2337.
9. The Department of Organic Chemistry of Voronezh State University [Electronic resource]. URL: http://www.vsu.ru/english/depts/chem.html.
10. Maybridge Screening Libraries [Electronic resource]. URL: https://www.fishersci.com/us/en/brands/I9C8LZ5R/products.html.
11. Oseledets I., Tyrtyshnikov E. TT-cross approximation for multidimensional arrays // Linear Algebra Appl. 2010. Vol. 432, № 1. P. 70–88.
12. Ciemny M., Kurcinski M., Kamel K., Kolinski A., Alam N., Schueler-Furman O., Kmiecik S. Protein-peptide docking: opportunities and challenges. // Drug Discov. Today. England, 2018. Vol. 23, № 8. P. 1530–1537.

ПУБЛИКАЦИИ АВТОРА ПО ТЕМЕ ДИССЕРТАЦИИ в журналах, индексируемых в базах данных Web of Science, Scopus, RSCI и РИНЦ

A1. Application of docking and quantum chemistry to the search for inhibitors of SARS-CoV-2 main protease / A. Tashchilova, **A. Sulimov**, I. Ilin, D. Kutov, V. Sulimov // Russian Supercomputing Days, RuSCDays 2021 : Communications in Computer and Information Science / eds. V. Voevodin, S. Sobolev. – Cham: Springer International Publishing, 2021. – P. 17–28. – DOI 10.1007/978-3-030-92864-3_2. Импакт-фактор по SJR: 0.16.

A2. Разработка противовирусных лекарств на основе ингибиторов главной протеазы SARS-CoV-2 / **А.В. Сулимов**, Х.С. Шихалиев, О.В. Пьянков, Д.Н. Щербаков, В.Ю. Чиркова, И.С. Ильин, Д.К. Кутов, А.С. Тащилова, М.Ю. Крысин, Д.В. Крыльский, Н.В. Столповская, Е.А. Волосникова, С.В. Беленькая, В.Б. Сулимов // Биомедицинская химия. – 2021. – Vol. 67, № 3. – P. 259–267. – DOI 10.18097/PBMC20216703259. Импакт-фактор по SJR: 0.192.

A3. Docking paradigm in Drug Design / V.B. Sulimov, D.C. Kutov, A.S. Taschilova, I.S. Ilin, E.E. Tyrtyshnikov, **A.V. Sulimov** // Current Topics in Medicinal Chemistry. – 2021. – Т. 21, № 6. – С. 507–546. – DOI 10.2174/1568026620666201207095626. Импакт-фактор по JCR Science Edition: 3.442.

A4. In search of non-covalent inhibitors of SARS-CoV-2 main protease: Computer aided drug design using docking and quantum chemistry / **A.V. Sulimov**, D.C. Kutov, A.S. Taschilova, I.S. Ilin, N.V. Stolpovskaya, K.S. Shikhaliev, V.B. Sulimov // Supercomputing Frontiers and Innovations. – 2020. – Т. 7, № 3. – С. 41–56. – DOI 10.14529/jsfi200305. Импакт-фактор по SJR: 0.238.

A5. Synthesis, docking, and in vitro anticoagulant activity assay of hybrid derivatives of pyrrolo[3,2,1-ij]quinolin-2(1H)-one as new inhibitors of factor Xa and factor XIa / N. Novichikhina, I. Ilin, A. Tashchilova, **A. Sulimov**, D. Kutov, I. Ledenyova, M. Krysin, K. Shikhaliev, A. Gantseva, E. Gantseva, N. Podoplelova, V. Sulimov // Molecules. – 2020. – Т. 25, № 8. – С. 1889. – DOI 10.3390/molecules25081889. Импакт-фактор по JCR Science Edition: 3.01.

A6. Docking of oligopeptides / **A.V. Sulimov**, D.K. Kutov, I.S. Ilin, K.S. Shikhaliev, D.A. Zheltkov, E.E. Tyrtyshnikov, V.B. Sulimov // Russian Chemical Bulletin. – 2019. – Т. 68, № 9. – С. 1780–1786. – DOI 10.1007/s11172-019-2624-y. Импакт-фактор по JCR Science Edition: 1.014.

A7. Supercomputer docking with a large number of degrees of freedom / **A.V. Sulimov**, D.C. Kutov, I.S. Ilin, D.A. Zheltkov, E.E. Tyrtyshnikov, V.B. Sulimov // SAR and QSAR in Environmental Research. – 2019. – Т. 30, № 10. – С. 733–749. – DOI 10.1080/1062936X.2019.1659412. Импакт-фактор по JCR Science Edition: 2.287.

A8. New factor Xa inhibitors based on 1,2,3,4-tetrahydroquinoline developed by molecular modelling / I.S. Ilin, E.N. Lipets, **A.V. Sulimov**, D.C. Kutov, K.S. Shikhaliev, A.Y. Potapov, M.Y. Krysin, F.I. Zubkov, L.V. Sapronova, F.I. Ataullakhanov, V.B. Sulimov // Journal of Molecular Graphics and Modelling. – 2019. – Т. 89 – С. 215–224. – DOI 10.1016/j.jmgm.2019.03.017. Импакт-фактор по JCR Science Edition: 1.885.

A9. Современные методы разработки новых лекарственных средств, влияющих на систему гемостаза / **А.В. Сулимов**, Д.К. Кутов, А.С. Тащилова, И.С. Ильин, Н.А. Подоплелова, М.А. Пантелеев, И.В. Леденева, Х.С. Шихалиев, В.Б. Сулимов // Вопросы гематологии/онкологии и иммунопатологии в педиатрии. – 2019. – Vol. 18, № 4. – P. 136–152. – DOI 10.24287/1726-1708-2019-18-4-136-152. Импакт-фактор по CiteScore: 0.04.

A10. Tensor train global optimization: Application to docking in the configuration space with a large number of dimensions / **A.V. Sulimov**, D.A. Zheltkov, I.V. Oferkin, D.C. Kutov, E.V. Katkova, E.E. Tyrtyshnikov, V.B. Sulimov // 3rd Russian Supercomputing Days, RuSCDays 2017 : Communications in Computer and Information Science / ред. V.V. Voevodin, S.I. Sobolev. – Cham: Springer, 2017. – Т. 793. – С. 151–167. – DOI 10.1007/978-3-319-71255-0_12. Импакт-фактор по SJR: 0.17.

A11. Evaluation of the docking algorithm based on tensor train global optimization / I.V. Oferkin, D.A. Zheltkov, E.E. Tyrtyshnikov, **A.V. Sulimov**, D.C. Kutov, V.B. Sulimov // Bulletin of the South Ural State University, Ser. Mathematical

Modelling, Programming & Computer Software. – 2015. – Т. 8, № 4. – С. 83–99. – DOI 10.14529/mmp150407. Импакт-фактор по SJR: 0.177.

A12. Application of molecular modeling to development of new factor Xa inhibitors / V.B. Sulimov, I.V. Gribkova, M.P. Kochugaeva, E.V. Katkova, **A.V. Sulimov**, D.C. Kutov, K.S. Shikhaliev, S.M. Medvedeva, M.Y. Krysin, E.I. Sinauridze, F.I. Ataullakhanov // BioMed Research International. – 2015. – Т. 2015, № Article ID 120802, 15 pages. – DOI 10.1155/2015/120802. Импакт-фактор по JCR Science Edition: 2.134.

A13. Application of molecular modeling to urokinase inhibitors development / V.B. Sulimov, E.V. Katkova, I.V. Oferkin, **A.V. Sulimov**, A.N. Romanov, A.I. Roschin, I.B. Beloglazova, O.S. Plekhanova, V.A. Tkachuk, V.A. Sadovnichiy // BioMed Research International. – 2014. – Т. 2014 – С. 625176. – DOI 10.1155/2014/625176. Импакт-фактор по JCR Science Edition: 1.579.

A14. TTDock: метод докинга на основе тензорных поездов / Д.А. Желтков, И.В. Офёркин, Е.В. Каткова, **А.В. Сулимов**, В.Б. Сулимов, Е.Е. Тыртышников // Вычислительные методы и программирование. – 2013. – Vol. 14 – P. 279–291. Импакт-фактор по РИНЦ: 0.614.

A15. Application of the docking program SOL for CSAR benchmark / **A.V. Sulimov**, D.C. Kutov, I.V. Oferkin, E.V. Katkova, V.B. Sulimov // Journal of Chemical Information and Modeling. – 2013. – Т. 53, № 8. – С. 1946–1956. – DOI 10.1021/ci400094h. Импакт-фактор по JCR Science Edition: 4.068.

A16. Реализация поддержки параллельных вычислений в программах докинга SOLGRID и SOL / И.В. Оферкин, **А.В. Сулимов**, О.А. Кондакова, В.Б. Сулимов // Вычислительные методы и программирование. – 2011. – Vol. 12, № 2. – P. 9–23. Импакт-фактор по РИНЦ: 0.614.

A17. Компьютерный дизайн лекарственных средств: программа докинга SOL / А.Н. Романов, О.А. Кондакова, Ф.В. Григорьев, **А.В. Сулимов**, С.В. Лущекина, Я.Б. Мартынов, В.Б. Сулимов // Вычислительные методы и программирование. – 2008. – Vol. 9, № 2. – P. 213–233. Импакт-фактор по РИНЦ: 0.614.

ПУБЛИКАЦИИ АВТОРА ПО ТЕМЕ ДИССЕРТАЦИИ в монографиях

M1. Сулимов, В.Б. Докинг: Молекулярное моделирование для разработки лекарств. Докинг: молекулярное моделирование для разработки / В.Б. Сулимов, **А.В. Сулимов**. – Москва: ИИнтелл, 2017. – 348 р. – ISBN 978-5-98956-025-7.

M2. Молекулярное моделирование для разработки новых лекарств на основе ингибиторов урокиназы / В.А. Садовничий, В.Б. Сулимов, Е.В. Каткова, А.Н. Романов, **А.В. Сулимов**, И.В. Оферкин, Д.В. Стамбольский, И.Б. Белоглазова, В.А. Ткачук // Постгеномные исследования и технологии под ред. члена-корреспондента РАН С.Д.Варфоломеева. – Изд. МГУ Москва, 2011. – Р. 103–140. – ISBN 978-5-317-03968-4.

СВИДЕТЕЛЬСТВА АВТОРА О РЕГИСТРАЦИИ ПРОГРАММ ДЛЯ ЭВМ ПО ТЕМЕ ДИССЕРТАЦИИ

P1. Желтков Д.А., Оферкин И.В., **Сулимов А.В.**, Сулимов В.Б., Тыртышников Е.Е. TTDock // Свидетельство о государственной регистрации программы для ЭВМ № 2013615623. Правообладатель ФГБУН ИВМ РАН. Зарегистрировано в Реестре программ для ЭВМ 17.06.2013

P2. Сулимов В.Б., **Сулимов А.В.** Keenbase // Свидетельство о государственной регистрации программы для ЭВМ № 2007612082. Зарегистрировано в Реестре программ для ЭВМ 22.05.2007

P3. Романов А.Н., Григорьев Ф.В., Кондакова О.А., Сулимов В.Б., **Сулимов А.В.** SOL // Свидетельство о государственной регистрации программы для ЭВМ № 2006611167. Зарегистрировано в Реестре программ для ЭВМ 14.02.2006

P4. Григорьев Ф.В., Кондакова О.А., Лущекина С.В., Романов А.Н., Сулимов В.Б., **Сулимов А.В.** FARS // Свидетельство о государственной регистрации программы для ЭВМ № 2006611168. Зарегистрировано в Реестре программ для ЭВМ 31.03.2006