

Каткова Екатерина Владимировна

ПРИМЕНЕНИЕ МЕТОДОВ МОЛЕКУЛЯРНОГО МОДЕЛИРОВАНИЯ ДЛЯ
РАЗРАБОТКИ НОВЫХ ЛЕКАРСТВ

Специальность 03.01.02 – «Биофизика», 03.01.08 – «Биоинженерия»

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата физико-математических наук

Работа выполнена на кафедре биофизики физического факультета Московского государственного университета имени М.В. Ломоносова.

Научные руководители: доктор биологических наук, профессор
Атауллаханов Фазоил Иноятович
доктор физико-математических наук
Сулимов Владимир Борисович

Официальные оппоненты: **Поройков Владимир Васильевич**, доктор биологических наук, профессор, заведующий лабораторией Структурно-функционального конструирования лекарств, Федеральное государственное бюджетное учреждение Научно-исследовательский институт биомедицинской химии имени В.Н. Ореховича Российской академии наук
Мазо Михаил Абрамович, кандидат физико-математических наук, старший научный сотрудник, Федеральное государственное бюджетное учреждение науки Институт химической физики им. Н.Н. Семенова Российской академии наук

Ведущая организация: Федеральное государственное бюджетное учреждение науки
Институт вычислительной математики Российской академии наук

Защита состоится «18» сентября 2014 года в 15-30 на заседании диссертационного совета Д501.002.11 при Московском государственном университете имени М.В. Ломоносова по адресу: 119991, ГСП-1, Москва, Ленинские горы, д.1, стр. 2, физический факультет МГУ, аудитория 5-19.

С диссертацией можно ознакомиться в библиотеке Московского государственного университета имени М.В.Ломоносова

Автореферат разослан « » _____ 2014 года.

Ученый секретарь
диссертационного совета Д 501.002.11,
кандидат технических наук

Сидорова Алла Эдуардовна

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Для решения задач, связанных с рациональной разработкой новых лекарственных средств, в настоящее время все чаще прибегают к методам молекулярного моделирования, поскольку все стадии разработки нового лекарственного вещества, как правило, требуют значительных финансовых затрат и занимают порядка 10 лет. На первом этапе разработки нового лекарственного соединения необходимо предсказать структуры молекул-ингибиторов, которые обычно представляют собой низкомолекулярные органические соединений и которые избирательно связываются с активными центрами белков, обуславливающих болезнь, блокируя их активность. Для этих целей в настоящее время все чаще используются методы рационального компьютерного *in silico* дизайна, которые позволяют уменьшить как временные, так и материальные затраты за счет сокращения числа соединений, которые необходимо синтезировать и экспериментально тестировать в процессе разработки нового лекарства.

В настоящее время существует множество методов для моделирования взаимодействия лекарственных соединений с соответствующими мишенями и вычисления энергии связывания ингибитора и белка. Сущность вычислительного аспекта проблемы заключается в необходимости вычислять энергию взаимодействия лиганд-мишень с высокой точностью (1-2 ккал/моль или 0.05-0.1 эВ), и с высокой скоростью, позволяющей оперативно проводить расчеты, по крайней мере, десятков тысяч соединений.

Одним из ключевых методов в данной работе является докинг – поиск положения лиганда (молекулы-ингибитора или кандидата в ингибиторы) в активном центре белка-мишени, соответствующего минимальному значению энергии системы белок-лиганд. Преимущество докинга состоит в том, что, отличаясь от эмпирических методов QSAR достаточно высоким качеством прогнозирования, он также позволяет производить расчеты нескольких тысяч соединений в день, что недоступно пока методам молекулярной динамики.

Однако наряду с повышением скорости вычислений для современных методов компьютерного моделирования стоит также проблема повышения точности вычислений. Для этих целей часто используются методы постпроцессинга, в которых производится более точная оценка энергий связывания для выбранных в результате

скрининга молекул. Эти методы обычно включают в себя локальную оптимизацию молекул (то есть поиск локального минимума в окрестности найденного глобального минимума), которая может происходить, например, с использованием силовых полей, а также более аккуратный учет составляющих свободной энергии связывания лиганда и белка (например, более аккуратный учет растворителя, энтропийного вклада и т.д.). На этом этапе также могут быть применены, например, более точные, и одновременно более ресурсоемкие, методы квантовой химии.

В процессе работы были выбраны два белка: сериновая протеаза урокиназа (uPA) и серин/треониновая протеинкиназа PAK1. Каждый из рассмотренных белков представлял интерес в качестве мишени для новых ингибиторов, способных в дальнейшем стать основой для лекарственных средств, в том числе и противоопухолевых.

Основное внимание в диссертационной работе уделено новой реализации методики докинга и постпроцессинга, её тестированию, использованию полуэмпирических квантово-химических методов для уточнения энергии связывания белок-лиганд, а также применению этих методов для разработки новых ингибиторов урокиназы и протеинкиназы PAK1. Возможность использования высокопроизводительных вычислений также рассмотрена в данной работе. Во всех случаях расчеты выполнялись на суперкомпьютерных ресурсах МГУ им. М.В. Ломоносова «Чебышев» и «Ломоносов».

Цели и задачи работы.

Целью работы является повышение эффективности рациональной разработки лекарств с помощью методов молекулярного моделирования (молекулярного докинга, методов локальной оптимизации (постпроцессинга) лиганда в силовом поле белка, методов квантовой химии).

В соответствии с поставленной целью решались следующие **задачи**:

1. Разработка и использование методов молекулярного моделирования для решения задачи поиска новых ингибиторов урокиназы и протеинкиназы PAK1.
2. Построение молекулярных моделей активных центров белков урокиназы и PAK1.
3. Исследование влияние параметров генетического алгоритма на качество докинга (в рамках программы докинга SOL).

4. Тестирование методов молекулярного моделирования применительно к выбранным белкам-мишеням.

5. Сравнение различных методов расчета энергий связывания в силовых полях (докинг и постпроцессинг) и с помощью методов квантовой химии, а также оценка энергий сольватации, полученных разными методами.

6. Поиск ингибиторов урокиназы и PAK1 в базах данных готовых соединений с помощью процедуры виртуального скрининга.

7. Дизайн новых структур ингибиторов урокиназы.

Научная новизна результатов:

1. Исследована возможность применения совокупности методов докинга и полуэмпирической квантовой химии (в рамках недавно разработанного метода PM7, реализованного в программе MORAC и применяемого для поиска новых ингибиторов впервые) для расчетов энергий связывания белок-лиганд. Такая методика демонстрирует высокую эффективность при поиске новых ингибиторов.

2. С помощью молекулярного моделирование впервые определено расположение некаталитического активного центра протеинкиназы PAK1, связывающего соединение ВРІРР (5-(3-бромифенил)-5,11-дигидро-1,3-диметил-1*H*-индено[2',1':5,6]пиридо[2,3-*d*]пиримидин-2,4,6(3*H*)-трион).

3. Найдены новые ингибиторы белков урокиназы и протеинкиназы PAK1.

Научная и практическая значимость данной работы заключается в том, что на основе полученных данных выдвинуты предположения о структуре некаталитического активного центра белка PAK1, а также разработаны новые ингибиторы белков урокиназа и PAK1. Для разработки новых ингибиторов был впервые применен полуэмпирический квантово-химический метод PM7 из программного комплекса MORAC, в котором учтены как водородные связи, так и дисперсионные межмолекулярные взаимодействия, что позволяет увеличить точность расчетов энергии связывания и эффективность разработок новых ингибиторов и новых лекарств на их основе.

Результаты данной работы могут быть применены, в частности, для прогнозирования структур новых ингибиторов урокиназы, PAK1 и других белков-мишеней, а уже найденные в данной работе соединения могут быть использованы для

проведения дальнейших доклинических и клинических испытаний с целью введения новых лекарственных средств в клиническую практику.

Положения, выносимые на защиту.

1. При найденных в работе оптимальных параметрах генетического алгоритма, качество позиционирования лигандов программой SOL достигает желаемого высокого уровня: 70-80% лигандов после докинга имеют среднеквадратичное отклонение от нативного положения менее 3 Å.
2. Программа докинга SOL и программа постпроцессинга DISCORE показывают высокое качество предсказания ингибирующей активности лигандов, не уступающее многим зарубежным аналогам по результатам международного конкурса CSAR.
3. Применение полуэмпирического метода PM7 из программного пакета MOPAC продемонстрировало хорошие результаты для задач расчета энтальпии связывания белка и лиганда, и показывает лучшую корреляцию с экспериментом, нежели аналогичные расчеты в силовом поле MMFF94. На этом основании предложен комбинированный метод молекулярного моделирования, включающий в себя докинг (программа SOL) в совокупности с постпроцессингом в рамках квантово-химического полуэмпирического метода PM7 (программный пакет MOPAC). Такой комбинированный метод дает возможность эффективно предсказывать ингибирующую активность низкомолекулярных лигандов для заданных белков-мишеней.
4. Найденные на основе методов молекулярного моделирования новые соединения, ингибирующие урокиназу и PAK1, в экспериментах *in vitro*, демонстрируют микромолярную (ингибиторы урокиназы) и субмикромолярную (ингибиторы PAK1) активность.

Личный вклад диссертанта заключается в сборе и анализе литературных данных, разработке путей решения поставленных задач, отработке и тестировании методов докинга и постпроцессинга применительно к выбранным мишеням, проведении скрининга баз данных готовых соединений, анализе результатов и выдвижении предположений о возможной структуре новых ингибиторов белков урокиназы и PAK1, проведении вычислений полуэмпирическими методами квантовой химии в программном комплексе MOPAC, интерпретации результатов,

подготовке публикаций и докладов по теме диссертационной работы. Все результаты, составляющие основное содержание диссертации, получены автором самостоятельно.

Апробация работы и публикации. Материалы диссертации были представлены на международной конференции "Ломоносов" (Москва, 2009), на научных конференциях «Ломоносовские чтения» (Москва, 2011, 2013 гг.), на Московском международном конгрессе «Биотехнология: состояние и перспективы развития» (Москва, 2011), на XIX Российском национальном конгрессе "Человек и лекарство" (симпозиум «Биоинформатика и компьютерное конструирование лекарств», Москва, 2012), на IV съезде фармакологов России «Инновации в современной фармакологии» (Казань, 2012), на 2-ом Русско-Французском семинаре по хемо- и биоинформатике (2nd French-Russian Workshop on Chemoinformatics and Bioinformatics, Казань, 2012), на 224-ой национальной конференции Американского химического общества (244th ACS National Meeting, Филадельфия, Пенсильвания, 2012).

Основные результаты диссертационной работы опубликованы в 15 печатных изданиях, в том числе в 5 статьях в изданиях из перечня ведущих рецензируемых научных журналов и изданий ВАК РФ, в 8 тезисах докладов на конференциях и в 2 сборниках трудов.

Структура диссертации. Диссертация состоит из введения, 5 глав, заключения и списка цитируемой литературы из 154 наименований. Работа изложена на 155 страницах машинописного текста и включает 21 рисунок и 26 таблиц.

СОДЕРЖАНИЕ РАБОТЫ

Первая глава содержит описание современных методов рациональной разработки лекарственных средств, в частности, докинга. Рассмотрены способы позиционирования лигандов в активном центре белка, способы представления белка при моделировании активного центра, а также методы расчета энергий связывания белок-лиганд. Представлены сведения о роли в организме и о структурах белков урокиназы и протеинкиназы PAK1.

Во **второй главе** подробно рассматриваются методы молекулярного моделирования, которые использовались в данном исследовании для разработки новых ингибиторов. К ним относятся сеточный и прямой докинг, а также

постпроцессинг в рамках силового поля MMFF94 и в рамках методов полуэмпирической квантовой химии.

В параграфе 2.1 рассматривается реализация сеточного докинга в программе SOL. Описаны приближения, используемые в программе и позволяющие ускорить время расчетов, такие как: представление белка в виде жесткой сетки потенциалов, объединение некоторых типов атомов, отсутствие локальной оптимизации положений атомов лиганда в процессе докинга. Энергия десольватации также вычисляется при помощи упрощенной Обобщенной модели Борна таким образом, чтобы нелокальную задачу электростатического экранирования свести к локальным потенциалам на сетке.

В результате функция оценки энергии взаимодействия белка и лиганда (скоринг-функция) программы SOL выглядит следующим образом:

$$\Delta E = \alpha E_{VdW} + \beta E_{ES} + \beta E_{GB-lig-prot} + \gamma E_{GB-prot-prot} + \delta E_{GB-lig-lig} + \mu N_{tors}, \quad (1)$$

где E_{VdW} , E_{ES} – энергия ван-дер-ваальсового и электростатического взаимодействия между протеином и лигандом, описываемая в рамках модели силового поля MMFF94, N_{tors} – число торсионных степеней свободы лиганда, α , β , γ , δ , μ – коэффициенты, $E_{GB-lig-prot}$, $E_{GB-prot-prot}$, $E_{GB-lig-lig}$ – части энергии взаимодействия с растворителем в рамках упрощенного Обобщенного Борновского приближения.

Также приведено описание генетического алгоритма, с помощью которого осуществляется поиск глобального минимума энергии связывания, и указаны параметры, которые влияют на его работу.

В параграфе 2.2. приведено описание постпроцессинга в программе DISCORE, задачей которой является более точное вычисление энергии взаимодействия белка и лиганда за счет отказа от упрощений программы SOL, приведенных выше. В рамках программы DISCORE проводится локальная оптимизация лиганда в силовом поле MMFF94, при которой могут оставаться подвижными как атомы лиганда, так и атомы белка. Энергия взаимодействия белок-лиганд вычисляется по формуле:

$$\Delta G_{bind} = k_1 \Delta G_{Coulomb} + k_2 \Delta G_{VdW} + k_3 \Delta G_{pol} + k_4 \Delta G_{np} + k_5 \Delta G_{LS} + k_6 \cdot 0.33 N_{TORS} + k_7 \Delta G_{rot-tr} \quad (2)$$

где $\Delta G_{Coulomb}$ – энергия прямого кулоновского взаимодействия зарядов лиганда и протеина, ΔG_{VdW} – энергия прямого ван-дер-ваальсового взаимодействия, ΔG_{pol} и ΔG_{np} – полярный и неполярный вклады в энергию десольватации, соответственно, ΔG_{LS} – энергия внутренних напряжений лиганда, $k_6 \cdot 0.33 N_{TORS}$ – энтропийная составляющая, обусловленная замораживанием внутренних торсионных степеней свободы лиганда

при его связывании с мишенью, N_{TORS} – число внутренних торсионных степеней свободы, ΔG_{rot-tr} – свободная энергия, обусловленная замораживанием при связывании степеней свободы, отвечающих за вращение и трансляцию лиганда как целого (это энтропийный вклад), k_i ($i = 1, \dots, 7$) – безразмерные подгоночные коэффициенты. Полярный вклад в энергию десольватации в рамках программы DISCORE может быть рассчитан с помощью одной из трех континуальных моделей растворителя (PCM, COSMO или SGB).

В параграфе 2.3 рассмотрены нюансы полуэмпирических квантово-химических расчетов в программном комплексе MOPAC. Рассмотрен программный модуль MOZYME, который позволяет рассчитывать системы, содержащие до 15000 атомов (такие как белки) благодаря методу локализованных молекулярных орбиталей LMO (LocalizedMolecular Orbital). Этот метод исключает из вычислений локализованные молекулярные орбитали, находящиеся друг от друга на большом расстоянии, что позволяет вычислять не все двухэлектронные двуцентровые и одноэлектронные двуцентровые интегралы. Помимо этого не так давно в программном комплексе MOPAC был реализован метод параметризации PM7, в который включен учет межмолекулярных взаимодействий благодаря добавлению поправок на образование водородных и галогенных связей, а также на дисперсионные взаимодействия.

В параграфе 2.3 рассматривается реализация прямого докинга на примере программы FLM, в которой не используется сетка заранее рассчитанных потенциалов. В отличие от сеточного докинга, прямой докинг позволяет учитывать подвижность не только атомов лиганда, но и атомов белка, а также проводить локальную оптимизацию атомов в процессе докинга. Помимо этого программа FLM осуществляет поиск не только одного глобального минимума, но находит множество локальных минимумов энергии взаимодействия белок-лиганд и свободная энергия молекулярной системы рассчитывается по всем этим минимумам в мультигармоническом приближении: строится аппроксимация потенциальной энергии полиномом второй степени в каждом локальном минимуме, то есть система представляется в виде гармонического осциллятора. Затем вычисляются частоты собственных колебаний гармонического осциллятора и статистическая сумма, по которой вычисляется свободная энергия молекулярной системы.

В параграфах 2.4 и 2.5 описаны экспериментальные методики измерения активности соединений в отношении урокиназы и РАК1 *in vitro*.

Третья глава посвящена тестированию описанных во второй главе методов молекулярного моделирования на нативных комплексах урокиназа-лиганд, а также применению этих методов для дальнейшей разработки новых ингибиторов урокиназы.

Среди исследованных параметров генетического алгоритма (к ним относятся: число независимых запусков программы, число поколений, размер популяции, размер пула родителей и способ выбора особей в пул родителей, количество особей элиты, способ кроссинговера, а также параметры, характеризующие кроссинговер и мутации) были выбраны те, которые оказывают наибольшее влияние на время расчетов и качество докинга. Их оптимальные значения для программы SOL приведены в таблице 1.

Таблица 1. Оптимальные значения параметров генетического алгоритма.

Число независимых запусков	50
Размер популяции	300000
Число поколений	1000

Далее, в этой главе на основе данных базы Protein Data Bank построены молекулярные модели урокиназы для 45 структур комплексов урокиназа-лиганд. Для этих комплексов проведена оценка качества позиционирования программой SOL и определено, что 80 % лигандов имеют среднеквадратичное отклонение от нативного положения (RMSD) < 3Å. Также в подавляющем большинстве случаев в первый кластер (кластер с минимальной энергией) попадает довольно большое число лигандов, что говорит о том, что глобальный минимум энергии связывания белок-лиганд с большой вероятностью найден.

Для расширенного набора комплексов урокиназа-лиганд, в который вошли не только нативные комплексы, но и комплексы урокиназы с соединениями, заведомо не являющимися ее ингибиторами, проведена проверка качества оценки энергии связывания программой SOL. Количество ложно положительных и ложно отрицательных результатов после докинга 89 лигандов программой SOL, а также коэффициенты корреляции рассчитанных энергий с энергиями связывания, полученными из экспериментальных данных, приведены в таблице 2.

Таблица 2. Коэффициент корреляции рассчитанных значений энергий с экспериментальными энергиями и число ложно положительных и ложно отрицательных результатов при докинге программой SOL для 89 комплексов урокиназа-лиганд.

Коэффициент корреляции (для 89 комплексов)	0.45	
Коэффициент корреляции (для 45 нативных комплексов)	0.35	
Число ложноположительных	28	Всего 33
Число ложноотрицательных	5	

Помимо этого была проведена оценка способности программы SOL выбирать активные соединения среди множества неактивных с помощью построения кривой обогащения (зависимость количеств активных лигандов среди первых N выбранных лигандов от доли выбранных N лигандов среди всех лигандов) и вычисления коэффициента обогащения (площади под кривой обогащения). Чем ближе коэффициент обогащения к единице, тем выше качество оценочной функции. Кривая обогащения для 1888 неактивных в отношении урокиназы лигандов из базы данных Национального Института Рака, США, NCI Diversity и 7 активных ингибиторов урокиназы приведена на рисунке 1.

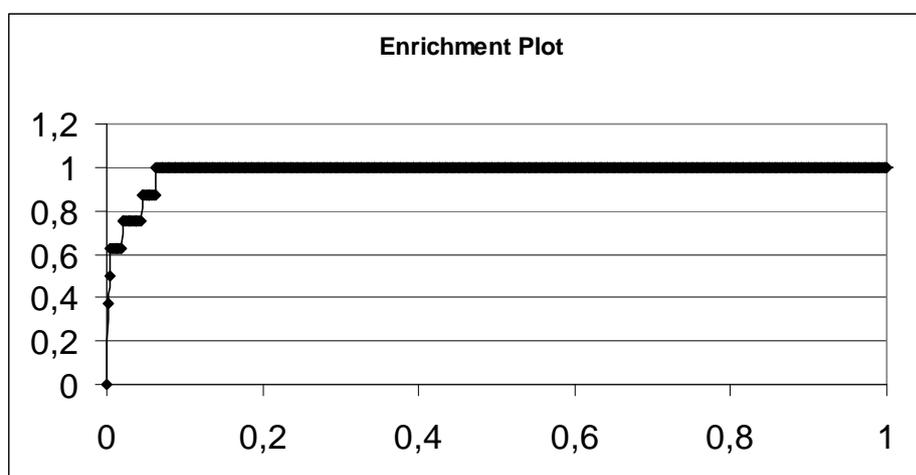


Рис. 1. Кривая обогащения, построенная для программы докинга SOL на основе 7 активных в отношении урокиназы и 1888 неактивных в отношении урокиназы лигандов. Коэффициент обогащения для этой кривой = **0.98**.

Значение коэффициента обогащения равно 0.98 и считается достаточно высоким, чтобы утверждать, что программа успешно находит активные соединения во множестве неактивных.

Чтобы сравнить результаты сеточного докинга и прямого докинга также были проведены вычисления с помощью программы FLM. С точки зрения качества докинга она показала неплохие результаты: для 43 нативных комплексов урокиназа-лиганд из 45 удалось найти такое «задоченное» положение лиганда, которое имеет низкую энергию и отличается от нативного с $RMSD < 2 \text{ \AA}$. Коэффициент корреляции при этом составил 0.44. Однако поиск глобального минимума энергии в этом случае потребовал большого времени вычислений - на порядок больше, чем программа SOL (более 20000 процессоро-часов), что делает невозможным на данном этапе применение прямого докинга для задач, где требуется докинг большого числа лигандов.

Докинг программой SOL соединений с известными активностями в отношении урокиназы показал способность программы отделять активные соединения от неактивных (с некоторой долей ложных результатов). Поэтому следующим шагом стал поиск активных соединений в базах данных готовых соединений (в которые вошли базы данных ZINC, ACB Blocks, Vitas-M Lab) с помощью программы SOL. По результатам анализа докинга было заказано 44 соединения, 14 из которых показали слабую ингибирующую активность. Чтобы находить соединения, обладающие большей ингибирующей активностью, был предпринят следующий шаг – включение постпроцессинга в процедуру молекулярного моделирования.

Процедура постпроцессинга в рамках силового поля MMFF94 была выполнена с помощью программы DISCORE. Предварительно было проведено сравнение континуальных моделей растворителя для расчета энергий гидратации как малых молекул, так и белков, с результатами экспериментальных данных и молекулярной динамики.

Обучение программы DISCORE было проведено на известных ингибиторах урокиназы. И для обучающего и для тестового набора наблюдалось улучшение корреляции рассчитанных значений энергий с экспериментальными, а также уменьшение числа ложно положительных результатов (таблица 3).

Также процедура постпроцессинга была выполнена в рамках полуэмпирических квантово-химических расчетов: в этом случае локальная оптимизация атомов комплекса белок-лиганд проводилось с помощью программного комплекса MORAC с использованием метода параметризации PM7 в результате чего

рассчитывались энтальпии образования комплекса белок-лиганд. Тестирование программы на 10 нативных комплексах урокиназа-лиганд показало, во-первых, важность учета растворителя при таких расчетах, а также необходимость учета подвижности атомов белка. При этом оптимальным кажется включение в оптимизацию не всех атомов белка, а только ближайших атомов белка, расположенных не далее, чем на расстоянии 2 \AA от одного из атомов лиганда.

Таблица 3. Сравнение коэффициентов корреляции и числа ложно положительных и ложно отрицательных результатов для 89 комплексов для программы докинга SOL (89 комплексов) и программы докинга DISCORE, обученной на наборе из 50 урокиназных комплексов.

	Коэффициент корреляции	Число ложно положительных результатов	Число ложно отрицательных результатов
Докинг SOL (все комплексы)	0.45	28	5
Постпроцессинг Discore (обучающий набор)	0.58	8	3
Постпроцессинг Discore (тестовый набор)	0.52	9	4
Постпроцессинг Discore (все комплексы)	0.55	17	7

Помимо поиска новых соединений - кандидатов в ингибиторы урокиназы - в базах данных готовых соединений также был выполнен дизайн новых оригинальных структур потенциальных ингибиторов урокиназы. В результате синтеза и проверки предложенных структур *in vitro* из 18 синтезированных соединений 8 проявили ингибирующую активность в отношении урокиназы. Для 6 активных соединений IC_{50} составляет 200 \mu M и более. Среди синтезированных и проверенных в эксперименте соединений можно выделить два наиболее активных ингибитора урокиназы: 1,2-бензотиазол-3-ил-гуанидин ($IC_{50} = 33 \text{ \mu M}$) и его 5-бром производная ($IC_{50} = 4\text{-}20 \text{ \mu M}$) (рисунок 2).

По полученным экспериментальным результатам для каждого метода вычисления энергии связывания белок-лиганд были рассчитаны коэффициенты корреляции с экспериментальными данными. В случае программы SOL он составил - (-0.2), при использовании программы DISCORE – 0.12, при докинге с помощью FLM

– 0.10, и при вычислении энтальпий связывания в МОРАС – 0.52. Таким образом, корреляция с экспериментальными данными наблюдалась только в случае квантово-химических расчетов.

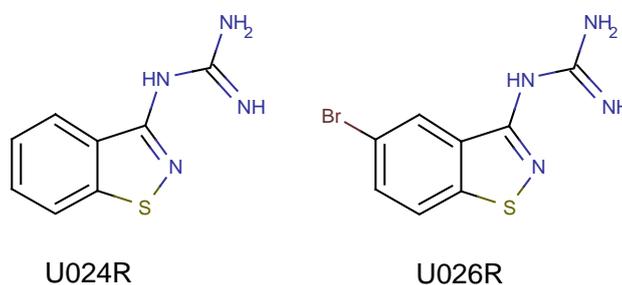


Рис. 2. Структуры соединений U024R и U026R.

Четвертая глава посвящена применению методов компьютерного моделирования для поиска новых ингибиторов протеинкиназы РАК1. Поскольку положение некаталитического активного центра белка РАК1 точно не известно, первым этапом работы стало определение этого активного центра с помощью докинга ингибитора ВР1РР в разные области белка. В результате был выбран куб докинга k11 с координатами в структуре 1F3M - (X, Y, Z; длина ребра)=(12.0, 6.0, 47.0; 32) (рисунок 3). При этом куб докинга включает в себя четыре области связывания (кармана), в которые ложатся различные изомеры и таутомеры соединения ВР1РР.

Поиск новых ингибиторов РАК1 проводился в базе данных Национального Института Рака, США, NCI Diversity, содержащей около 1888 интересных для данного исследования соединений. В результате докинга программой SOL 41 соединение было отобрано для экспериментальной проверки. Среди них 17 соединений показали активность в отношении РАК1 в экспериментах *in vitro* и оказались как ингибиторами, так и активаторами РАК1.

Постпроцессинг был проведен в программном комплексе МОРАС: оптимизировалось положение лиганда и ближайших атомов белка, находящихся на расстоянии до 2 Å от лиганда, с учетом растворителя на последнем шаге. При этом корреляции рассчитанных в МОРАС энтальпий с энергиями, полученными из экспериментальных данных, хотя и оказались невысокими (что может быть следствием неконкурентного механизма действия данных соединений), но возросли относительно корреляций результатов программы SOL и составили соответственно 0.17 для рН=8 и 0.21 для рН=7.

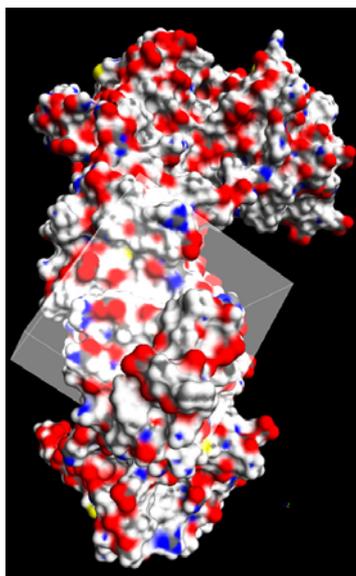


Рис. 3. Структура белка РАК1 и пространство куба докинга k11.

В ходе докинга соединения из NCI занимали те же области (карманы связывания) I, II, III и IV, что и конформеры и изомеры соединения ВРІРР. Для каждого из этих карманов были выявлены особенности и определены аминокислотные остатки, играющие роль при взаимодействии лиганда и белка.

Карман I имеет довольно узкую геометрию и требует очень точного соответствия заходящего в него соединения этой геометрии. Основной вклад во взаимодействие с лигандом вносят положительно заряженный Lys135 (цепь А), а также Pro428 и Met424 (цепь С), причем известно, что две последних аминокислотных остатка принимают участие в связывании областей субстрата в каталитическом домене и включены в процесс распознавания субстрата. Также во взаимодействие дополнительно могут быть включены аминокислотные остатки Val425, Gly426, Trp430, и Pro469 каталитической цепи С.

Карман II полностью открыт, и соединениям зайти в него гораздо проще, чем в карман I. Также, это единственный карман связывания, в котором все «задоченные» соединения стимулировали киназную активность. Среди аминокислот белка, вносящих значительный вклад во взаимодействие с лигандом, можно выделить Asn138 (цепь А) и Thr136 (цепь А), а также His86 (цепь В). Кроме того, геометрия данного кармана определяется взаимодействиями с Val87 и Ala91 цепи В.

В случае кармана III лишь несколько из «задоченных» лигандов оказались способны активировать РАК1, причем все эти лиганды обладали множественными отрицательными зарядами, что делает их слабыми кандидатами на роль

лекарственных средств. В этом кармане можно выделить следующие аминокислоты, ответственные за взаимодействие с лигандом: His83 и Gln102 (цепь А), или в некоторых случаях Gly98 (цепь В), эффективно связываются с отрицательными функциональными группами лигандов. Ala104 и Met99 цепи В, а также Ile85 и Pro100 цепи А определяют общую геометрию данного кармана.

Исследование области IV показало, что она, вероятно, не несет явной функциональной нагрузки.

По результатам данного анализа можно заключить, что карман II представляется наиболее актуальным для поиска высоко аффинных лигандов, которые смогут активировать или ингибировать РАК-1, а также определить аминокислоты белка РАК1, которые вероятнее всего участвуют во взаимодействии с лигандами и определяют специфичность связывания в карманах I, II и III: His83, His86, Gly98, Gln102, Lys135, Thr136, Asn138, Met424, Pro428.

Найденные соединения не представляют собой конкурентные ингибиторы каталитического центра РАК, а значения скоринг-функций (ΔG) соединений хотя и коррелируют с их способностью влиять на активность РАК1 в эксперименте, но напрямую не связаны с вызываемым ими функциональным эффектом. Это означает, что соединения, возможно, связываются, не оказывая функциональных влияний напрямую, а для оказания функционального эффекта нужны определенные более или менее специфические взаимодействия.

Результатам участия программы докинга SOL и программы постпроцессинга DISCORE в международном конкурсе CSAR 2011 Benchmark Exercise посвящена **пятая глава**. Целью такого участия была возможность независимой оценки качества работы программ и сравнения их с мировыми аналогами. Всего, по данным организаторов, в конкурсе принимало участие 17 групп. Участникам конкурса предоставлялись структуры белков и лигандов, и задачей являлось правильно позиционировать эти лиганды в активном центре белка и ранжировать их по энергии связывания.

Для оценки качества предсказания активности лигандов организаторами конкурса использовалось значение величины AUC (area under the curve) - площади под ROC-кривой, характеризующей количество ложно положительных и ложно отрицательных результатов и таким образом оценивающей правдоподобие

предсказанной ингибирующей способности лигандов (наилучшее предсказание дает AUC равный 1, наихудшее – 0,5). Еще одним критерием оценки докинга стал коэффициент корреляции рассчитанных значений скоринг-функций с экспериментальными энергиями, который в данном случае вычислялся только по связывающимся с белком лигандам.

Программа SOL для некоторых мишеней показывала высокое качество предсказания. Например, для белка LpxC AUC = 0.95, для urokinase AUC = 0.97, однако для некоторых белков, она не смогла отделить «плохие» лиганды от хороших (для белка Chk1 AUC = 0.51). На рисунках 4-6 приведены значения AUC для первого и второго наборов для разных групп, участвующих в конкурсе в виде диаграмм. Из них видно, что программы SOL и DISCORE для белков Urokinase и LpxC показали очень хорошие результаты по сравнению с другими группами.

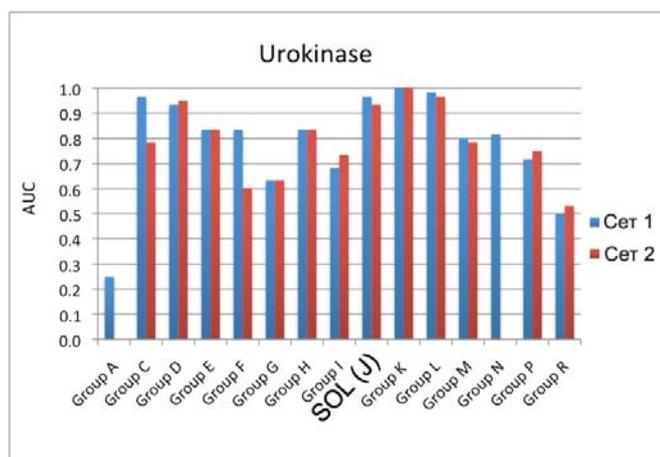


Рис. 4. На диаграмме столбцами обозначены значения AUC в случае первого (первый столбец) и второго (второй столбец) сетов для разных групп для белка Urokinase. Группе J соответствуют значения AUC для SOL и DISCORE.

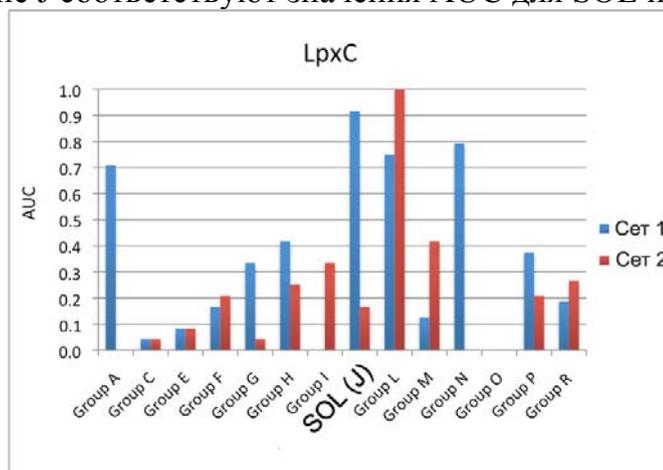


Рис. 5. На диаграмме столбцами обозначены значения AUC в случае первого (первый столбец) и второго (второй столбец) сетов для разных групп для белка LpxC. Группе J соответствуют значения AUC для SOL и DISCORE.

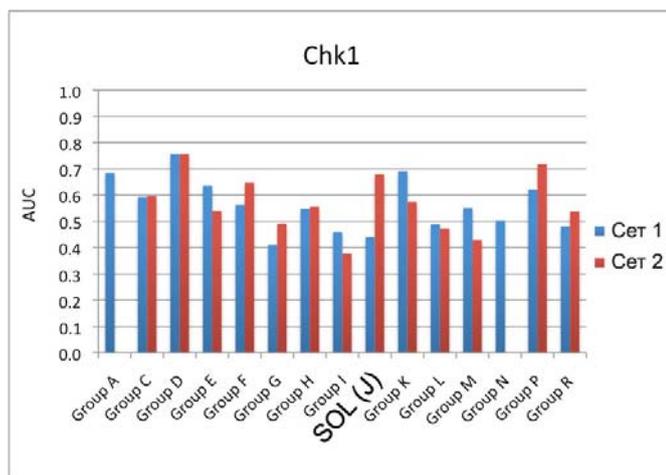


Рис. 6. На диаграмме столбцами обозначены значения AUC в случае первого (первый столбец) и второго (второй столбец) сетов для разных групп для белка Chk1. Группе J соответствуют значения AUC для SOL и DISCORE.

По предоставленной организаторами информации о среднеквадратичном отклонении «задоченных» положений, присланных участниками, от нативных положений активных лигандов в белках было определено качество позиционирования лигандов. Программа докинга SOL продемонстрировала хорошее качество позиционирования лигандов в активные центры тестируемых белков для большинства случаев: 56 лигандов из 91 было позиционировано со среднеквадратичным отклонением от нативного положения (RMSD) $< 2 \text{ \AA}$ (62%), 65 лигандов – $< 3 \text{ \AA}$ (71%). По результатам конкурса программа SOL заняла одно из первых мест в тестировании CSAR 2011.

РЕЗУЛЬТАТЫ И ВЫВОДЫ:

1. На примере белков урокиназы и протеинкиназы PAK1 проведена разработка, тестирование и использование методов молекулярного моделирования для решения задачи поиска новых ингибиторов. В исследование включены такие методы молекулярного моделирования как сеточный и прямой докинг, а также постпроцессинг, как с помощью силового поля, так и в рамках методов квантовой химии.

2. В рамках тестирования программы докинга SOL проведено исследование влияние параметров генетического алгоритма на качество докинга. К параметрам, наиболее сильно влияющим как на качество докинга, так и на время счета относятся: число независимых запусков программы, число поколений и размер популяции. При

найденных в работе оптимальных параметрах генетического алгоритма, качество позиционирования лигандов программой SOL достигает желаемого высокого уровня: 70-80% лигандов после докинга имеют среднеквадратичное отклонение от нативного положения менее 3 Å. Также выбраны менее жесткие параметры генетического алгоритма, более подходящие для скрининга многих лигандов.

3. Проведено сравнение сеточного (программа SOL) и прямого (программа FLM) докинга. Прямой докинг требует гораздо большего времени вычислений, однако позволяет более полно исследовать энергетическую поверхность и находить спектр низкоэнергетических локальных минимумов. Показано, что программа прямого докинга FLM находит среди низкоэнергетических локальных минимумов и нативное положение лиганда почти для всех исследованных в данной работе комплексов, и в большинстве случаев это положение обладает минимальной энергией. Коэффициент корреляции рассчитанных энергий связывания белок-лиганд с экспериментальными данными в случае прямого докинга выше, чем в случае сеточного докинга, несмотря на пренебрежение учетом растворителя и отсутствие подгоночных коэффициентов в функции энергии связывания.

4. Проведено обучение и тестирование программы постпроцессинга DISCORE, которая осуществляет локальную оптимизацию лиганда в найденном после докинга глобальном минимуме энергии комплекса белок-лиганд и расчет энергии связывания с использованием подгоночных коэффициентов. Также проведено сравнение различных континуальных моделей растворителя, реализованных в программе DISCORE. Показано, что обучение программы DISCORE на комплексах урокиназа-лиганд позволяет увеличить корреляцию с экспериментальными данными для тестового набора комплексов урокиназа-лиганд, однако применительно к другим комплексам (комплексы PAK1-лиганд, а также комплексы, с которыми проводилась работа в рамках конкурса CSAR) DISCORE не всегда позволяет улучшить результаты докинга.

5. Проведено сравнение программ докинга SOL и постпроцессинга DISCORE с аналогичными зарубежными программами в рамках международного конкурса CSAR. Программа докинга SOL и программа постпроцессинга DISCORE показали высокое качество предсказания ингибирующей активности лигандов, не уступающее многим зарубежным аналогам.

6. Проведен анализ программного комплекса МОРАС (в рамках метода параметризации РМ7) применительно к задаче поиска новых ингибиторов. Выбраны оптимальные условия работы программного комплекса, показано влияние учета растворителя на результаты расчета энтальпий связывания, а также влияние учета подвижности атомов белка. Соединение методов докинга (поиска глобального минимума энергии взаимодействия) и квантовой химии в качестве постпроцессинга (локальная оптимизация найденного в результате докинга положения и оценка энтальпии связывания белка и лиганда) дает более хорошие результаты по сравнению с использованием докинга и постпроцессинга в рамках только силового поля MMFF94.

7. Проведено построение молекулярных моделей активных центров белков урокиназы и РАК1. Для белка РАК1 положение некаталитического активного центра изначально не было известно и было определено путем докинга известного ингибитора РАК1 в различные области белка.

8. Проведен поиск ингибиторов урокиназы и РАК1 в базах данных готовых соединений с помощью процедуры виртуального скрининга и анализ экспериментальной проверки данных соединений. Для белка РАК1 это позволило найти 17 активных соединений: как ингибиторов, так и активаторов РАК1, в том числе и обладающих субмикромольной активностью.

9. На основе анализа докинга баз данных готовых соединений проведен дизайн новых ингибиторов урокиназы методами молекулярного моделирования. Из 18 синтезированных соединений 9 показали ингибирующую активность в отношении урокиназы в эксперименте *in vitro*, в том числе два соединения обладали микромольной активностью.

СПИСОК ПУБЛИКАЦИЙ ПО ТЕМЕ ДИССЕРТАЦИИ:

Работы, опубликованные соискателем в перечне ведущих рецензируемых научных журналов и изданий, рекомендованных ВАК:

1. **Каткова Е.В.** Исследования влияния параметров генетического алгоритма на эффективность докинга с помощью программы SOL // Вычислительные методы и программирование. – 2012. – Т. 13. С. 536-550.

2. Желтков Д.А., Офёркин И.В., **Каткова Е.В.**, Сулимов А.В., Сулимов В.Б., Тыртышников Е.Е. TTDock: метод докинга на основе тензорных поездов // Вычислительные методы и программирование. – 2013. – Т. 14. – С. 279-291.

3. Sulimov A.V., Kutov D.C., Oferkin I.V., **Katkova E.V.**, Sulimov V.B. Application of the docking program SOL for CSAR Benchmark // J. Chem. Inf. Model. – 2013. – Vol. 53, № 8. – P. 1946-1956.

4. **Каткова Е.В.**, Оферкин И.В., Сулимов В.Б. Применение квантово-химического полуэмпирического метода PM7 для разработки новых ингибиторов урокиназы // Вычислительные методы и программирование. – 2014. – Т.15. – С. 258-271.

5. Sulimov V.B., **Katkova E.V.**, Oferkin I.V., Sulimov A.V., Romanov A.N., Roschin A.I., Beloglazova I.B., Plekhanova O.S., Tkachuk V.A., Sadovnichiy V.A. Application of molecular modeling to urokinase inhibitors development // Biomed. Res. Int. – 2014. – Article ID 625176. – 15 p. – URL: <http://www.hindawi.com/journals/bmri/2014/625176/>

Работы, опубликованные соискателем в сборниках трудов:

6. **Каткова Е.В.**, Жабин С.Н., Сулимов В.Б. Проведение квантовомеханических расчетов геометрии олигопептидов с помощью программы Pritoda, входящей комплекс для моделирования наноразмерных атомно-молекулярных структур // Сборник Трудов научно-исследовательского центра фотоники и оптоинформатики Под ред. И.П. Гурова и С.А. Козлова, Санкт-Петербург. 2009. С. 362-372.

7. Садовничий В.А., Сулимов В.Б., **Каткова Е.В.**, Романов А.Н., Сулимов А.В., Оферкин И.В., Стамбольский Д.В., Белоглазова И.Б., Ткачук В.А. Молекулярное моделирование для разработки новых лекарств на основе ингибиторов урокиназы // Монография, сборник: «Постгеномные исследования и технологии» под ред. члена-корреспондента РАН С.Д.Варфоломеева, Москва. Изд. МГУ Москва. 2011. С. 103-140.

Работы, опубликованные соискателем в виде тезисов конференций:

8. **Каткова Е.В.**, Жабин С.Н. Компьютерная разработка синтетических вакцин: докинг с ограничениями // Тезисы доклада, Международная научная конференция студентов, аспирантов и молодых учёных «Ломоносов-2009» (Москва, 13-18 апреля 2009 г.), с.34.

9. В.А.Садовничий, В.Б.Сулимов, **Е.В.Каткова**, А.Н.Романов, А.В.Сулимов, И.В.Оферкин, К.М.Федулов, А.И.Роцин, С.М.Никитин, И.В.Тайдаков, И.Б.Белоглазова, Д.В.Стамбольский, В.А.Ткачук Разработка новых лекарств на основе ингибиторов урокиназы // Тезисы доклада, Научная конференция «Ломоносовские чтения» - 2011. Секция вычислительной математики и кибернетики (Москва, 14-23 ноября 2011 г.).

10. **Каткова Е.В.**, Augilar V. (США), Onufriev A. (США), Масленников Е.Д., Сулимов В.Б. Сравнение качества и эффективности различных моделей растворителя // Тезисы доклада, Научная конференция «Ломоносовские чтения» - 2013. Секция вычислительной математики и кибернетики (Москва, 16-24 апреля 2013 г.).

11. **Е. Katkova** Docking and postprocessing: application to design of urokinase inhibitors // Theses, 2nd French-Russian Workshop on Chemoinformatics and Bioinformatics (Kazan, September 17th 2012), P. 16.

12. Садовничий В.А., Сулимов В.Б., **Каткова Е.В.**, Романов А.Н., Сулимов А.В., Оферкин И.В., Белоглазова И.Б., Стамбольский Д.В., Ткачук В.А. Применение молекулярного моделирования для разработки ингибиторов урокиназы // Тезисы доклада, Московский международный конгресс «Биотехнология: состояние и перспективы развития» (Москва, 21-25 марта 2011 г.), Материалы конгресса, часть 2, С. 252-253.

13. **Каткова Е.В.**, Оферкин И.В., Сулимов А.В., Сулимов В.Б. Применение полуэмпирических квантово-химических методов для разработки ингибиторов урокиназы // Тезисы доклада, IV съезд фармакологов России «Инновации в современной фармакологии» (Казань, 18-21 сентября 2012 г.), С.83.

14. Sulimov V.B., **Katkova E.V.**, Oferkin I.V., Kochugaeva M.P. PCM, COSMO, and Surface GB Implementation in Program DISOLV: Development and Validation // Theses, 244th ACS National Meeting (Philadelphia, Pennsylvania, August 19–23, 2012) oral presentation COMP 441.

15. Sulimov A.V., Kutov D.K., **Katkova E.V.**, Sulimov V.B. In Application of the Docking Program SOL for CSAR Benchmark // Theses, 244th ACS National Meeting (Philadelphia, Pennsylvania, August 19–23, 2012) oral presentation COMP 48.