

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
УЧРЕЖДЕНИЕ НАУКИ ИНСТИТУТ ОРГАНИЧЕСКОЙ ХИМИИ
ИМ. Н.Д. ЗЕЛИНСКОГО РОССИЙСКОЙ АКАДЕМИИ НАУК**

на правах рукописи



**СВИТАНЬКО
ИГОРЬ ВАЛЕНТИНОВИЧ**

**МОДЕЛИРОВАНИЕ В НАПРАВЛЕННОМ СИНТЕЗЕ
ВЕЩЕСТВ С ЗАДАНЫМИ СВОЙСТВАМИ**

02.00.03 — Органическая химия

ДИССЕРТАЦИЯ
на соискание ученой степени
доктора химических наук

Москва, 2018

Оглавление

1. Введение. Общие закономерности соответствия структура– свойство	6
2. Обзор литературы. Моделирование как первая стадия направленного синтеза соединений с заданными свойствами	10
2.1. Идеология моделирования активных веществ для направленного синтеза	10
2.2. Информационные технологии	12
2.3. Подходы к моделированию веществ с заданными свойствами, их ограничения и упрощения.....	15
2.3.1. QSAR (Quantitative Structure–Activity Relationship) / KCCC (количественные соотношения структура–свойство).....	15
2.3.2. Практическое моделирование QSAR.....	41
2.3.3. QSAR исследования душистых веществ.....	51
2.4. Молекулярный докинг	57
2.4.1. Моделирование биологически активных соединений, основанное на структурных параметрах лиганда и мишени	57
2.4.2. Программа Lead Finder.....	62
2.5. Объекты применения классического молекулярного докинга	71
2.5.1. Ингибиторы поли(ADP–рибозо)полимеразы (ПАРП1).....	71
2.5.2. Ингибиторы Syk-киназ.....	75
2.6. Преодоление принципиальных ограничений классического докинга	85
2.6.1. FEP (Free Energy Perturbation) / ВСЭ (Возмущение свободной энергии).....	86
2.6.2. Механистическая (динамическая) составляющая молекулярного моделирования.....	90
2.7. Моделирование стереохимического синтеза (переходных состояний, конечных продуктов).....	93

2.7.1. Стереохимия реакции Дильса–Альдера. Термодинамический/ кинетический контроль	93
2.8. Квантово-химические методы молекулярного моделирования	95
2.8.1. Некаталитические трансформации	99
2.8.2. Каталитические трансформации	100
2.8.3. Моделирование переходного состояния внутримолекулярной реакции Дильса–Альдера	102
3. Обсуждение результатов	108
3.1. Этапы оптимизации моделирования направленного синтеза	108
3.2. Классический QSAR, расчет дискретных взаимодействий атомов и функциональных групп	112
3.2.1. 3D-QSAR, описание поверхности взаимодействия	114
3.2.2. Молекулярная электростатическая поверхность (МЭП)	114
3.2.3. Формализация проблемы QSAR	120
3.2.4. Математическое решение проблемы QSAR	121
3.3. 3D-QSAR исследование амбрового запаха в ряду норлабдановых соединений	122
3.4. Моделирование мускусного запаха	130
3.5. Определение связи структура бицикломочевин (гликольбурилов) – нейротропная активность	134
3.5.1. 3D-QSAR моделирование в направленном синтезе биологически активных замещенных тиогликольбурилов с учетом конформационной лабильности лиганда	142
3.6. Докинг, дискретные взаимодействия	146
3.6.1. Методология и последовательность действий	146
3.6.2. Ингибиторы ПАРП1, определенные на базе классического (дискретного) докинга	150
3.6.3. Ингибиторы Суk-киназы, классический (дискретный) докинг	155

3.7. Докинг, интегральные взаимодействия. Метод возмущения свободной энергии (ВСЭ)	165
3.7.1. Моделирование протеинкиназ методом ВСЭ	165
3.7.2. Моделирование с учетом конформаций. Конформационная фокусировка	183
3.7.3. Конформационное равновесие лигандов Суk-киназы	185
3.7.4. Применение квантовохимических подходов в моделировании ингибиторов киназ	191
3.8. Моделирование катализа	195
3.8.1. Моделирование асимметрического синтеза с участием индуктора	195
3.8.2. Дальнейшее продвижение методов динамического моделирования в катализе. Применение ВСЭ и методов квантовой механики в моделировании гомогенного катализа	199
3.8.3. Применение модельных принципов расчета переходных состояний к структуре переходного состояния ферменто-катализируемой внутримолекулярной реакции Дильса–Альдера	214
3.8.4. Подходы к моделированию гетерогенного катализа	224
4. Экспериментальная часть. Описания использованных в практическом моделировании типовых методик расчетного эксперимента; схемы вспомогательных синтезов	227
4.1. Математические методы, примененные в расчетах методами 3D-QSAR	227
4.1.1. Основные этапы построения молекулярных представлений для последующего QSAR-анализа	227
4.1.2. Определения и постановка задачи расчетов методами QSAR	228
4.1.3. Дескрипторное описание соединений обучающей выборки	230
4.2. Программы моделирования (докинг и ВСЭ)	233
4.2.1. Методики моделирования, использованные в данной	

работе	233
4.3. Синтетические методики и схемы	243
4.3.1. Синтез веществ с амбровым запахом электрохимическим окислением склареола	243
4.3.2. Синтез циклоалкиловых эфиров и амбролигнана электрокаталитическим присоединением спиртов к двойной связи	244
4.3.3. Моделирование асимметрического синтеза методом ВСЭ, реакция Дильса–Альдера.....	246
4.3.4. Вспомогательные синтезы промежуточных соединений и соединений сравнения в синтезе ингибиторов Suk- киназы	246
5. Заключение	252
6. Сокращения и условные обозначения	255
7. Список терминов	257
8. Список литературы	259

1. Введение.

Общие закономерности соответствия структура–свойство

С философской точки зрения, большинство научных соответствий – это причинно-следственные отношения между базовыми структурными особенностями и вытекающими из них статическими или динамическими индивидуальными либо обобщенными характеристиками, переход от простого к сложному.

В математике это соответствие базовый постулат (лемма) – теорема или построение более высокого уровня; в физике – зависимость природы и проявления физического процесса от базовых физических законов, постулатов и элементарной структуры вещества; в теоретической медицине особое внимание «уделено безусловной необходимости выяснения молекулярных механизмов функций, мишеней действия физиологически активных веществ и обязательного соотнесения данных моделирования с теми же процессами в условиях *in vivo* в целостном организме» [1]; наконец, социология «изучает общие принципы воспроизводства (функционирования) и изменения основных (простых и сложных) форм социальных взаимодействий, в том числе общество как целостную систему социальных взаимодействий, на основе широкого привлечения эмпирических данных, фактов реальной жизни, выделяя *повторяющееся, устойчивое* в этих взаимодействиях в различных сферах общественной жизни» [2].

Все перечисленные закономерности можно объединить термином «корреляция повторяющихся элементов структуры со свойством макросистемы, включающей эти элементы». При этом каждый конкретный случай выявления такой зависимости выглядит тем более достоверно, чем больше он подтвержден количественными корреляциями (оцифрован), а совокупность (критическая масса) достоверных количественных соотношений позволяет формулировать общую для данной научной тематики теорию «количественных соотношений структура–свойство» (Quantitative Structure–Activity Relationship, QSAR).

Химия по мере своего развития все менее воспринимается как наука описательная и все более включается в процесс всеобщего применения принципов соответствия структура–свойство. Это выражается в растущей популярности

направленного синтеза соединений с заданными свойствами, где структура соединения прогнозируется на основе предварительных статистических и/или физических расчетов.

Что мы видели в органическом синтезе целевых соединений 10–15 лет назад? Фундаментальную проблему: практически полное отсутствие математического или физического моделирования структур для направленного синтеза веществ с заранее заданными свойствами. Планирование эксперимента, основанное на интуиции синтетика, т.е. богатом практическом опыте и/или хорошем знании литературы.

Несмотря на значительный прогресс в области математической химии в последние годы, эта проблема всё еще актуальна, вследствие распространенности сугубо эмпирическое планирование химического эксперимента в области медицинской химии, органического катализа.

По статистике, только одно из пяти тысяч синтезированных соединений, обладающих биологической активностью (потенциальных лекарственных средств), доходит до испытаний *in vivo*, и еще на порядок меньше доходят до выпуска коммерческого лекарственного препарата. Полный цикл выведения препарата на рынок длится от 10 до 15 лет, и затраты в этом цикле могут быть весьма значительными (Рисунок 1) [3].

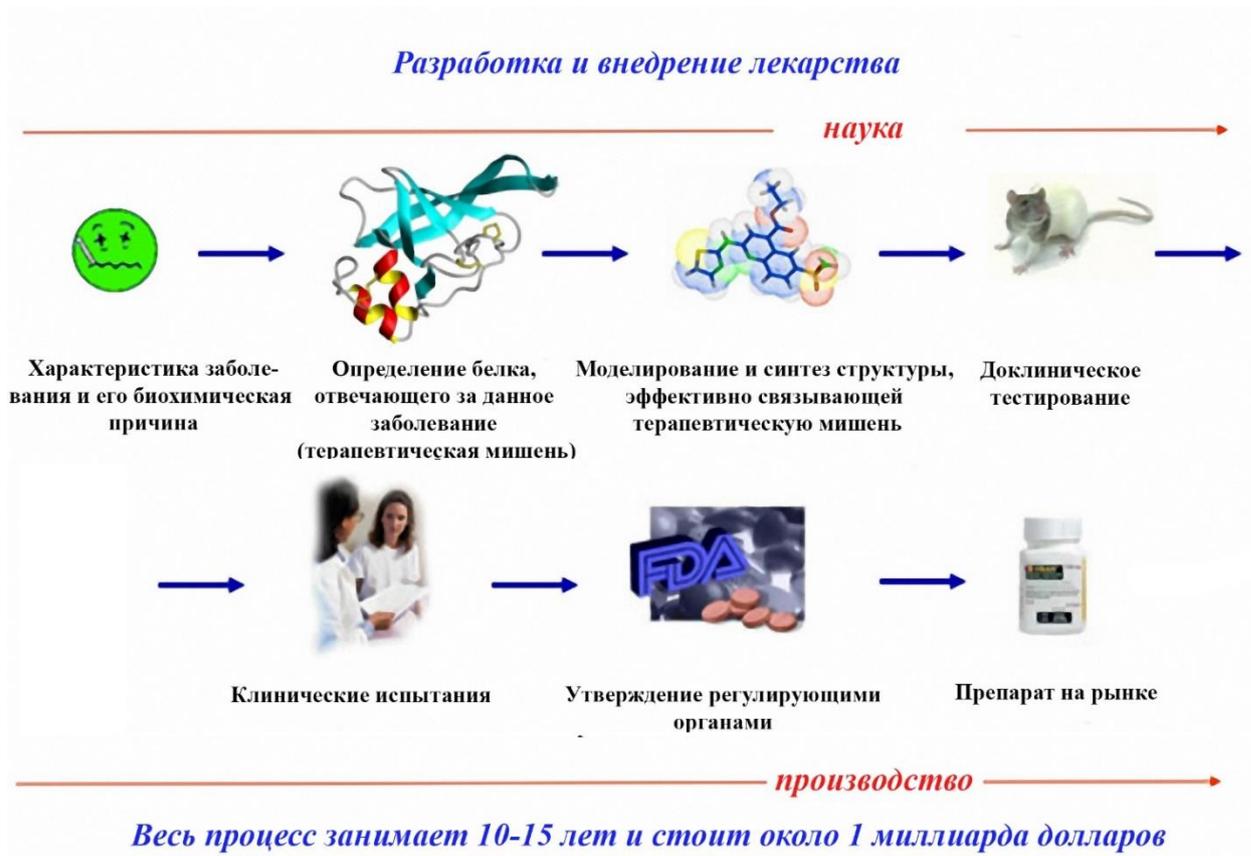
Как следствие:

- перерасход средств и ресурсов на проведение «слепого эксперимента» – синтеза множества структур, потенциально («интуиция синтетика») обладающих нужной активностью;
- кадровая проблема: слабая вовлеченность молодых специалистов в планирование работы по синтезу – по причине малого опыта научных исследований, пока слабой «интуиции синтетика» или недостаточного владения литературой;
- доведение до предпромышленных (предклинических) испытаний заведомо недостаточного количества практически важных веществ, длительное время самих испытаний.

Конечная цель расчетов в области прогнозирования активности, моделирования соединений с заданной активностью для направленного синтеза – получение и выпуск промышленных форм.

Для примера приведем картину полного процесса внедрения биологически активного вещества (Рисунок 1).

Рисунок 1. Процедура разработки и внедрения лекарственного препарата



При очевидном прогрессе в понимании абсолютной необходимости *направленного* синтеза веществ с заранее заданными свойствами большая часть органического синтеза (медицинского, каталитического, новых материалов) в настоящее время, тем не менее, происходит без предварительного моделирования целевых соединений. Молекулярный дизайн соединений, обладающих требуемым свойством (третья стадия процесса на Рисунке 1), существенно ускоряет и удешевляет процесс направленного синтеза и является *первой стадией химического эксперимента*. В данной работе показано, что это синтетическое планирование, в зависимости от применяемого метода моделирования, может относиться не только и не столько к медицинской химии, но и к органическому синтезу и катализу в целом.

Целью настоящей работы является разработка принципиально новых подходов к направленному синтезу биологически активных соединений и продуктов

каталитических превращений с заранее заданными свойствами на основе компьютерного молекулярного дизайна. Для достижения поставленной цели необходимо было решить следующие основные задачи:

1. Разработать более точные по сравнению с существующими методы молекулярного моделирования целевых продуктов, промежуточных соединений и механизмов реакций, и применить их в направленном органическом синтезе.
2. Предварительно, до какого-либо эксперимента, определить строение необходимого продукта и оценить вид и степень его активности.
3. Экспериментально подтвердить предсказанные методологии направленного синтеза и экспериментально оценить прогнозируемые свойства синтезированных целевых структур на наличие заданного вида активности.

Кроме изменения базовой схемы направленного синтеза, в работе использованы новые, усовершенствованные алгоритмы моделирования потенциально активного вещества: методы описания и сравнения молекулярной электростатической поверхности; классический докинг расширен конформационной фильтрацией, учетом растворителя; динамической составляющей, позволяющей верифицировать механизм реакции образования целевого продукта; осуществлен переход к квантово-химическим расчетам в определении энергетики переходных состояний, что впервые позволило распространить моделирование биологически активных соединений на каталитический синтез и найти подходы к моделированию гетерогенного катализа.

Направленным синтезом получены вещества с заданной активностью, проявившие в валидирующем эксперименте целевые свойства. Этим показана эффективность разработанных расчетных методов в связке структура–активность продуктов органического синтеза.

По материалам диссертации опубликованы 24 статьи в журналах, рекомендованных ВАК, 4 главы в книгах и получено 1 авторское свидетельство на изобретение. Результаты работы были представлены автором в виде устных и приглашенных докладов и тезисов на 1 российской и 9 международных научных конференциях.

2. Обзор литературы. Моделирование как первая стадия направленного синтеза соединений с заданными свойствами

2.1. Идеология моделирования активных веществ для направленного синтеза

Важнейшей фундаментальной проблемой в области органического синтеза является поиск подходов к рациональному планированию работ, который, в принципе, может быть достигнут посредством кардинального изменения логики синтеза веществ с заранее заданными свойствами, на основе информационных технологий.

Принципиально новые подходы к синтезу практически важных веществ с заранее заданными свойствами могут быть основаны на предварительном компьютерном молекулярном дизайне, с последующим синтезом предсказанных структур, обязательным испытанием их специфических свойств и разработкой оптимальных методологий, обеспечивающих их внедрение в практику. Таким образом, решение проблемы предварительного прогнозирования синтезируемых практически важных соединений с заранее заданными свойствами, а именно, развитие уже существующих теоретических подходов к моделированию и их применение к направленному синтезу в медицинской химии и био-, гомогенном и гетерогенном катализе является чрезвычайно *актуальной* задачей.

Базовым принципом является неразрывная связь предварительного моделирования структуры для синтеза соединений с заданными свойствами и последующего экспериментального подтверждения. Такой тандем позволяет экономить временные и материальные ресурсы, избегая проведения заведомо бесперспективных трудоемких экспериментов. При этом существенную роль играет надежность используемых теоретических подходов и их соответствие поставленной задаче. Данная работа направлена как на развитие уже существующих теоретических подходов, так и на их применение к различным проблемам медицинской химии и био-, гомогенного и гетерогенного катализа – областей, где теоретическое моделирование используется либо обязано использоваться наиболее широко по вышеизложенным причинам.

Для решения вопроса предварительного моделирования активных веществ в медицинской и каталитической химии необходимо проанализировать текущее состояние дел в:

- моделировании биологически активных соединений, где исследуется природа процесса координации перспективного биологически активного соединения и мишени, соответствующей определенному биологическому свойству, т.е. де факто природа биокатализа;
- моделировании гомогенного и гетерогенного катализа, где с помощью расчетов и уникальных методов спектральных исследований можно определить механизм элементарного каталитического акта и состав продуктов каталитического синтеза;

т.е. областей, где особенно важно достижение конечного результата – синтеза веществ с практически важными характеристиками за меньшее время с меньшими усилиями и расходами.

Имеется большое количество работ по молекулярному дизайну соединений в направленном синтезе, использующих как данные о самих соединениях, так и данных о мишенях, с которыми они взаимодействуют, и применяющие самые разнообразные способы подбора дескрипторов – структурных, электрических и других факторов, определяющих свойство вещества, а также способы их сравнения [4 – 9].

Обязательным при этом является эксперимент, в котором подтверждается корректность модельных исследований – измерение степени и направления активности полученного в направленном синтезе вещества. Без такой завершающей валидации математических построений не обходится ни одна современная работа в области молекулярного моделирования.

Объединение объектов исследований – биологически активных соединений и катализаторов – в одной работе оправданно по причине существенного сходства закономерностей соответствия природы и механизма первой (и часто – скорость-определяющей) стадии каталитического действия – координации лиганда и мишени. При этом лиганд может быть потенциальным лекарственным препаратом, а может и исходным веществом в каталитическом синтезе; а мишенью служит как терапевтическая мишень, так и катализатор – как гомогенный, так и гетерогенный. Важно подобрать методы расчетов, не зависящие от природы реагирующих веществ, а зависящие только от их физических параметров.

Важно отметить, что при разработке как современных лекарственных средств, так и катализаторов химических процессов, помимо их эффективности, решающую

роль также имеет их селективность, то есть относительное сродство к целевому белку/объектам катализа, по сравнению с родственными нецелевыми белками/не вступающими в целевую реакцию катализа веществами.

2.2. Информационные технологии

Глобальное разделение методов поиска потенциально активных структур для последующего направленного синтеза этих соединений состоит в различных подходах к моделированию их активности.

Вариант 1. В распоряжении исследователя *нет никаких экспериментальных данных о механизме* (или хотя бы о структуре обоих составляющих – активного вещества и мишени) *проявления искомой активности*. В этом случае исследователь не может сформировать физическую картину связывания активной структуры и мишени, взаимодействие с которой индуцирует эту активность. Ничего не остается, как использовать для построения модели связывания косвенные данные.

В таком подходе экспериментально определенные значения активности непосредственно связаны со структурной информацией [5,10,11]. Структура молекулы математически формализована, и необходимая информация об особенностях молекулярной структуры кодируется в формате, пригодном для моделирования. Отправной точкой в этом подходе является молекулярная топология.

Набор структур в качестве исходных данных содержит информацию, которая позволяет выявить особенности структуры всего пула данных. Для каждого вида активности аспекты структуры различаются в разных списках – количество колец, химические типы колец, общие вариации скелета, молекулярный размер, расположение гетероатомов, водородные связи и другие донорно-акцепторные взаимодействия, порядок связей и т.д. Тем не менее, вопрос, на который моделирование пытается ответить, заключается в следующем: «Какая комбинация вариации структурных признаков соответствует вариациям в активности? Легко ли увидеть важное изменение структуры при визуализации структуры или оно более тонкое и неявное? Какие методы можно надежно использовать для выбора значимых из множества вариаций структурных элементов?»

Эти вопросы определяют задачу моделирования.

3D-информация о молекулярной структуре также совершенно необходима

для моделирования топологических дескрипторов, используемых для построения модели. Во многих работах показано, что модели, основанные на таком подходе, являются предсказательными и характеристичными для данной активности. Таким образом, синтез новых кандидатов (и не обязательно лекарственных) может быть направлен на достижение желаемой цели.

Необходимо отметить также *общепринятый методический постулат в терминологии соответствия структура–свойство и в методах прогнозирования активности*.

Методы, основанные на формализации структурной информации молекул (без учета структуры мишени), в литературе принято называть QSAR, QSPR (structure–activity, structure–property); однако исследователи, работающие в области докинга и связанных с ним алгоритмов (структура мишени известна), практически никогда не применяют данные аббревиатуры в своих работах (так, по их мнению, они абстрагируются от неточностей первого подхода).

Это *ошибочное и искусственно разделяющее эту область науки мировоззрение*. Базисный постулат QSAR (или, в узком понимании, CADD – Computer-aided drug design): активность является функцией структуры. Взаимосвязь может быть линейной или нелинейной, определена параметрическим или непараметрическим методом, от этого она не делится на «правильную» и «неправильную». Задача QSAR сформулирована как разработка методов, позволяющих находить существенные взаимосвязи между структурой и активностью, а именно, функциональные связи. Полученная в результате применения метода модель представляет собой количественную взаимосвязь, конечной целью которой является предсказательная способность и интерпретируемость.

В данной работе, во избежание путаницы в терминах, будет применяться та же (ошибочная, на наш взгляд) разделенная методология, однако *аббревиатура QSAR, по определению выше, это общий принцип связи структура–свойство, и, независимо от метода определения и расчета степени такой связи, охватывает все работы в области количественных соотношений структура–свойство, какими бы методами исследований ни пользовались ученые*. Ключевые опубликованные работы [6,7,12], как и наши исследования (см. Обсуждение результатов), объединяют терминологию, принципы и объекты применения обоих подходов и показывают, что

методы определения соответствия структура–свойство дополняют друг друга, будучи специфическими для каждого случая применения.

Косвенные данные о взаимодействии активной структуры и мишени мы можем получить из экспериментальных результатов взаимодействия известных активных структур с этой мишенью. Вычленив общие черты таких известных веществ («обучающая выборка»), можно статистически распространить найденную картину на новые вещества, с неизвестной активностью, и определить, обладает ли это новое вещество данной активностью. Таким образом, метод сводится к математическому манипулированию физическими данными об объектах, обладающих активностью, и, в общем случае, не касается физических данных мишени.

Вариант 2. *Известны экспериментальные данные о структурах взаимодействующих субстратов* (как активного вещества, так и мишени). В этом случае, пользуясь математическим и химическим инструментарием, исследователь может смоделировать физическую картину взаимодействия субстратов и, рассчитав энергию такого взаимодействия, визуализировать ее и рассчитать степень проявления искомой активности испытываемой структурой. Подходы, основанные на знании структуры взаимодействующих молекул особенно эффективны, когда к информации о структуре добавляется знание о структуре активного центра рецептора, ответственного за определенную активность. При этом активность подразумевается не только биологическая, а и каталитическая, поскольку «рецептор» (мишень) – любое химического соединения, координирующееся с молекулами лиганда, например, комплексообразование катализатора с исходными соединениями катализируемой реакции. Информацию о структуре компонентов комплекса дают спектральные исследования или прямопоказывающие методы (например, оптическая спектроскопия).

Таким образом, независимо от вариантов молекулярного моделирования в направленном синтезе веществ с заданной активностью (определение «биологической» сознательно опускаем, оно сужает область описываемой работы), такое моделирование решает следующие задачи:

- определение структурных параметров, соответствующих определенному свойству, т.е. решение задачи QSAR в базовом варианте;
- моделирование взаимодействия лиганда и рецептора;

- определение механизма биохимических превращений;
- прогнозирование активности гипотетических, еще не синтезированных молекул;
- расчет направлений биотрансформации природных соединений.

2.3. Подходы к моделированию веществ с заданными свойствами, их ограничения и упрощения

Знание химической структуры потенциально активного вещества и мишени еще не гарантирует точной модели «сайта» – точек связывания взаимодействующих объектов. Это связано с множественностью вариантов координации ключ–замок, а также со следующими существенными приближениями и упрощениями модели.

2.3.1. QSAR (Quantitative Structure–Activity Relationship) / KCCC (количественные соотношения структура–свойство)

Как классический QSAR (структура мишени неизвестна, моделирование основано на известных свойствах обучающей выборки), так и классический докинг (мишень известна, рассчитываются взаимодействия любого вида атомов и/или функциональных групп) оперируют дискретными взаимодействиями атомов, функциональных групп, точечных диполей; и скалярными 2D- и 3D-срезами векторного электростатического/магнитного полей – по ковалентному, Ван-дер-Ваальсовому радиусу, другие построения пространственного графа, описываемого трехмерными фигурами замкнутых и незамкнутых контуров.

Суммарная энергия связывания рассчитывается как арифметическая сумма дискретных взаимодействий. Статические координированные конгломераты (комплексы катализатора или биологической мишени с исходными веществами) рассчитываются схожим образом.

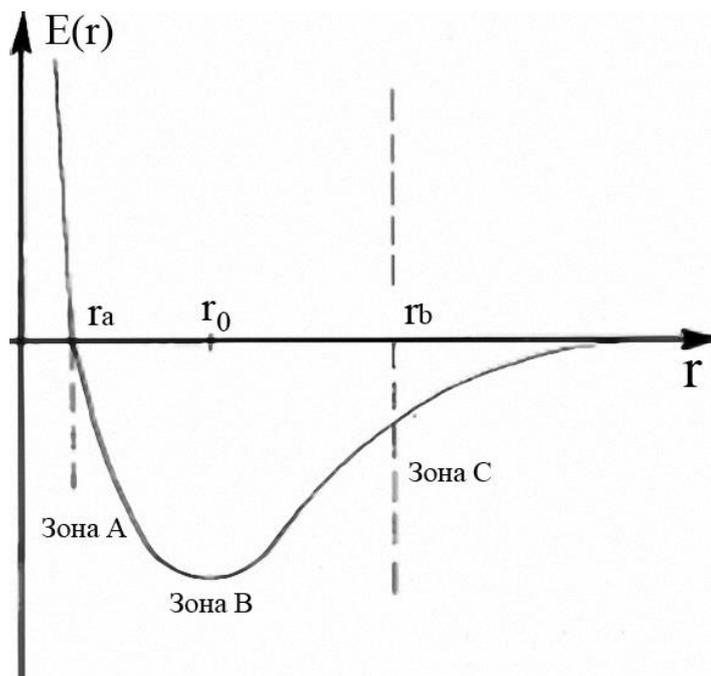
Главным во всех исследованиях структура–свойство является вопрос «как?» связываются молекулы лиганда (лекарства, реагенты каталитического процесса или душистые вещества) с мишенью (рецептором, катализатором и т.д.). Второй вопрос: на каком расстоянии, и какие физические параметры молекулы при межмолекулярных взаимодействиях являются определяющими?

Очевидная классификация межмолекулярных взаимодействий лиганд–рецептор представлена в Таблице 1 [13].

Таблица 1. Классификация межмолекулярных взаимодействий в зависимости от расстояния между молекулами

Малые расстояния (Зона А на Рисунке 2)	Средние расстояния (Зона В на Рисунке 2)	Дальние расстояния (Зона С на Рисунке 2)
Зона химического взаимодействия (Интегралы перекрывания)	1. Электростатика 2. Взаимная поляризация 3. Отталкивание–притяжение 4. Поляризация 5. Энтропия 6. Дисперсия	1. Электростатика 2. Поляризация 3. Энтропия

Рисунок 2. Межмолекулярный потенциал. Зона А – малые расстояния, зона В – средние, зона С – дальние



Определение зон взаимодействия предложено в работе [14]:

Зона I: в этом пространстве взаимодействие лиганд–рецептор происходит через межмолекулярные силы.

Зона II: охватывает первую зону и определяется как пространство, в котором действуют только ионные силы. Зона непосредственной фильтрации молекул лиганда по признакам комплементарности рецептору. Процесс определения комплементарности связан с комплементарностью молекулярных электростатических

потенциалов лиганда и рецептора, иначе стерические взаимодействия невозможны в принципе.

Зона III: нет влияния рецептора, здесь происходит броуновское движение молекул, с выходом на границу зоны II.

Величиной, характеризующей взаимодействие препарата и рецептора, является константа равновесия, которая обычно измеряется *in vitro*. Эта константа прямо зависит от многих факторов, но в первую очередь – от энергии первичного физического взаимодействия лиганда и рецептора [15]. Т.е. еще в 1964 г. был выдвинут тезис, что без электростатической комплементарности не может быть никакого стерического или химического взаимодействия молекул.

Все статические дискретные взаимодействия – это также существенное упрощение механизма активного процесса. Необходимо рассматривать процесс координации молекула–база в динамике, и учитывать при этом несвязывающие взаимодействия.

Такое взаимодействие возможно, только если рассматривать динамику изменения полной энергии задействованных в процессе молекул, ионов и прочих субстратов.

Координация молекул происходит не в безвоздушном пространстве; если это био- или гомогенные каталитические процессы, то они всегда происходят в присутствии молекул растворителя. Гетерогенных моно- и бимолекулярных процессов, естественно, также не бывает.

Т.е. как дескрипторы в подходе 1, так и докинг в подходе 2 должны рассматриваться в многопараметрической системе, учитывающей как минимум молекулы растворителя, а также молекулы других веществ, участвующих в скорость-определяющей стадии процесса.

Таким образом, реальную картину взаимодействия координирующихся веществ может дать только моделирование динамики всего процесса, включая переходное состояние, в интегральном виде, с учетом изменений полной энергии системы в процессе взаимодействия, т.е. с возможностью, насколько позволяют ресурсы, включить в рассмотрение и реагирующие субстраты, и растворитель, и влияние внешних факторов.

Рассмотрим оба подхода подробно.

Подход 1 (классический QSAR).

Экспериментальные (литературные) данные об активности веществ на 99% основаны на результатах измерений эффективности взаимодействия, которую характеризуют полученные в реакции продукты, а также экспериментальные результаты определения биологической или каталитической активности. Соответственно, в отсутствие сведений о механизме проявления активности, обучающая выборка для косвенного моделирования будет основываться не на механизме взаимодействия, а на его результатах – т.е. априори не будет точной, поскольку один и тот же результат может быть достигнут при разных типах и механизмах взаимодействий, а статистические данные – основываться на различных подходах к конструированию дескрипторов.

Следствие незнания механизмов взаимодействий: невозможно сказать точно, какая из конформаций или какой из стереоизомеров активной структуры проявляет оную активность. Остается либо выбирать для исследований вещества с жестким каркасом, либо учитывать в обучающей выборке все возможные конформеры (стереоизомеры).

Первая работа по расчетным методам в QSAR [16], как и все последующие, основаны на том, что априори считается, что активные структуры в скорость-определяющей стадии взаимодействуют с мишенью без химической реакции; а скорость-определяющая стадия взаимодействия – это связывающая координация активной молекулы и мишени. При участии химического взаимодействия молекулы и мишени подход 1, в соответствии с многочисленными литературными данными, дает неточный результат прогнозирования активности.

Какие общие черты, определяющие активность (дескрипторы) мы можем использовать в классическом QSAR?

2.3.1.1. Молекулярные дескрипторы

Химические дескрипторы лежат в основе моделирования QSAR, отражающих уровни и качество представления структуры соединения – от молекулярной формулы (так называемой 1D), матрицы связности атомов в двумерной структуре (2D)

до трехмерной, конформационно-зависимой (3D) и более высокой, с учетом взаимной ориентации и временной динамики молекул (4D и выше) [17]. Всеобъемлющий анализ молекулярных дескрипторов по методологии выбора и применения по состоянию на момент наиболее активного использования приведен в обзоре Раевского [18], а перечисление, вместе с их сферами влияния на активность, математические зависимости, примеры и ссылки, сделано в 2009 г. [19].

2D-дескрипторы для моделирования QSAR (2D-прогнозирование) – двумерное представление молекулы, как правило, называют топологическим представлением, и определяет связь атомов в молекуле с точки зрения наличия и характера химических связей.

Такое двумерное представление позволяет определить так называемые молекулярные 2D-дескрипторы. Основные преимущества этих параметров QSAR заключаются в том, что они

- содержат простую и полезную информацию о молекулярной структуре,
- не зависят от конформационной составляющей, и, следовательно,
- не зависят от возможной оптимизации геометрии структуры.

2D-дескрипторы неоднозначно характеризуют молекулярную топологию, и, как следствие, не могут использоваться при реконструкции молекулы в химических взаимодействиях. Для более точного описания молекулы можно использовать определенные упорядоченные последовательности 2D дескрипторов, например, отдельно по трехмерным осям – но это уже ближе к 3D-описанию.

Молекулярный граф – это топологическое представление химических соединений; обычно обозначается как $G = (V, E)$, где V – множество вершин графа, описывающее молекулу, и E – множество элементов, представляющих собой бинарные отношения между парами вершин, которые называются ребрами (связями между атомами). Молекулярный граф без всех атомов водорода называют H-обедненным молекулярным графом, молекулярный граф с атомами водорода называется H-заполненным (или просто молекулярный граф).

Граф-теоретические матрицы являются наиболее распространенным математическим инструментом формализации структурной информации, предоставляемой молекулярными графами. Граф-теоретической матрицей могут быть одновременно

вершины матрицы, если обе строки и столбцы относятся к вершинам (атомам), а также матричные элементы (пары вершин), или, например, кромки матриц, если обе строки и столбцы относятся к ребрам графа.

Существенны также элементы информации о парах вершин, таких как их связности (матрица смежности), топологических расстояний (матрица расстояний), с учетом суммы весов атомов вдоль соединительных путей (взвешенных матриц). Диагональные элементы матрицы могут быть равны нулю или формализовать химическую информацию о вершинах (дополненные матрицы). Кроме наиболее распространенных атомных свойств, часто встречаются локальные инварианты вершин, которые являются численными данными, полученными из молекулярной топологии и используемыми для характеристики свойств атомов молекул (например, степень вершин, сумма расстояния вершин, эксцентricность атомов).

Инварианты графа – это математические величины, полученные из графического представления молекулы и представляющие граф-теоретические свойства, которые сохраняются изоморфизмом, т. е. свойствами с идентичными значениями для изоморфных графов. Инвариант графа может быть уникальным полиномом, числовой последовательностью или единственным числовым индексом, полученным при операциях с граф-теоретическими матрицами. Предложено несколько алгоритмов, связанных с информацией о молекулярном графе, и применено к различным молекулярным матрицам и различным весовым схемам, что привело к появлению нескольких новых классов связанных с графиком инвариантов [9,20].

Индексы, полученные на основе молекулярного графа, обычно называются топологическими. Это численные индексы молекулярной топологии, математически выведенные прямым и однозначным образом из структурного графа молекулы. Они могут отражать структурные особенности молекулы – форму, симметрию, ветвление скелета или его цикличность, информацию о типе атома и мультиплетности связей в молекуле. На самом деле, топологические индексы, как правило, делятся на две категории: топоструктурного и топохимического характера.

Топоструктурные индексы отражают информацию о связности и расстояниях между атомами в молекуле; топохимические индексы отражают количественную информацию о пространственном расположении, а также физико-химические свойства атомов, такие как их химическая идентичность и состояние гибридизации.

Топологические индексы, основанные на связности атомов и расстояниях (рассчитанными на основе формальной матрицы) между ними, отличаются от геометрических дескрипторов, которые основаны на геометрии межатомных расстояний.

Еще один важный класс графов представлен так называемыми автокорреляционными индексами. Они были впервые введены в 1980 году [21], чтобы визуализировать матрицу связности атомов в зависимости от их пространственного расположения; алгоритм анализа автокорреляционных дескрипторов сформулирован Consonni и Todeschini [22]. Приемлемые автокорреляционные дескрипторы формулируются из положения о том, что атомы молекул являются множеством дискретных пространственных точек с свойством этих атомов в качестве функции, значение которой оценивается в этих точках. Дескриптор автокорреляции выражает корреляцию числовых значений.

Требования к атрибутам топологических индексов сформулированы в [23]:

- 1) имеют структурную интерпретацию,
- 2) имеют хорошую корреляцию, по крайней мере, с одним свойством,
- 3) различаются изомеры,
- 4) могут быть применены к структуре с неизвестным свойством,
- 5) желательно быть независимыми и
- 6) простыми,
- 7) не основанными на экспериментальных свойствах,
- 8) не иметь связей с другими дескрипторами,
- 9) могут иметь качественную природу,
- 10) используют общепринятые конструктивные решения,
- 11) имеют правильные геометрические и/или физические параметры,
- 12) обладают вариативностью при вариативности базовой структуры.

Большинство топологических дескрипторов имеют вышеуказанные характеристики, поэтому они успешно применены при описании структурного сходства/несходства молекул и при моделировании QSAR/QSPR.

3D-дескрипторы для моделирования QSAR (3D-QSAR)

В общем случае, определение QSAR постулирует, что биологическая селективность целевого объекта и той матрицы, куда этот объект координируется, определяется специфическими взаимодействиями, например, водородными связями мишени с лигандом. Однако, большинство работ сходится на том, что селективность связывания лигандов с мишенями определяется главным образом нековалентными эффектами полей, генерируемых в околопространственной близости этих лигандов. Таким образом, системное сэмплирование, или, другими словами, картина таких полей активных структур должны быть определяющими в молекулярных дескрипторах, на базе которых проводится прогнозирование этой активности. Этот постулат на основе здравого смысла и физической картины мира определил создание 3D-QSAR, в его оригинальном и по-прежнему преобладающем виде Comparative Molecular Field Analysis (CoMFA) [24].

Наиболее важной проблемой, связанной с внедрением CoMFA, был конфликт между тысячами молекулярных дескрипторов, необходимых для описания поля лиганда, и совсем небольшим набором биологических действий, что приводит к разбросу предсказательной способности в зависимости от малейших корреляций в виде дескриптора или методе его цифрового описания.

Впервые методология 3D-QSAR была сформулирована в [25], что было следующим шагом по сравнению с предшественником – моделью кристаллической решетки [26]. Его развитие вылилось в появление CoMFA [27]. Успех CoMFA в создании прогнозных моделей полностью опирался на новый статистический подход, Partial Least Squares of Latent Variables (Наименьшие квадраты латентных переменных) (PLS) – чисто математический подход, примененный к химии профессором Сванте Волдом из Университета Умео, Швеция [28].

Революционным было то, что мы можем извлечь полезные корреляции из ситуаций, когда было больше неизвестных параметров, чем точных. Анализ линейной регрессии защищает пользователя от случайных корреляций, когда используется множество переменных, и в дело вступает закон больших чисел. Исправляются кросс-корреляции между переменными, и исключается возможность корреляции модели систематическим учетом каких-либо параметров структур обучающей выборки [29].

В одной из первоначальных работ Крамера [30] рассматривались основные дескрипторы, полученные в результате изучения данных о физических свойствах большого набора справочных данных химических веществ. Выяснилось, что только два фактора отвечают за разброс параметров в модели: выбор системы отсчета (абсолютные координаты или относительные) и точность описания структуры. С первым легко бороться, привязав начало координат к любой точке молекулы (атому, точке на энергетической поверхности), со вторым – только мощностью используемого расчетного оборудования.

2.3.1.2. Прогнозирование активности

Для того, чтобы определить приоритеты синтеза, необходимо оценить прочность связывания того, что будет синтезировано, и рецептора (мишени). К сожалению, даже для систем, в которых структура комплекса лиганда с рецептором/ферментом точно определена экспериментально, прогнозирование активности по-прежнему является сложным и, среди прочего, должно учитывать ограничения существующей методологии, оценку изменений энтропии, влияние мультиполей как электростатической, так и иной природы, а также поляризуемости.

Хотя интерполяция на основе экспериментальных данных может быть эффективной, как в случае с CoMFA и другими прогностическими моделями, а методы моделирования, которые дорисовывают связанное соединение с известной активностью в представляющее интерес патентно чистую структуру, часто дают эффективные прогнозы, точное предсказание активности «с нуля» все еще нереально.

Шкала относительного вклада энтропии и энтальпии в энергию связывания лиганда и мишени для различных рассчитанных либо экспериментально определенных комплексов, представлена в [31,32]. Параметры энтальпии связывания, а также изменения площади поверхности, числа вращающихся связей, зафиксированных при связывании, и других параметров, связанных с энтропией связывания, обычно рассчитываются молекулярной механикой. В основных компонентах модели преобладают два параметра, ΔH и ΔS . Конечно, предварительный скрининг критической массы веществ при помощи молекулярной механики с целью отброса структур с низкой энергией связывания, это существенный вспомогательный ход.

Проблема метода CoMFA, как и любого статистического метода, связана с процессом согласования обучающей выборки соединений и лигандов, а также выбор конформации и ориентации каждого лиганда, которые обязаны быть учтены в модели QSAR. На практике эта задача выливается в «размножение» более высоких статистических критериев, т.е. исследованием биологической активности от конформационной зависимости лигандов обучающей выборки. Конформеров же может быть множество.

А, поскольку в статистическом методе, которым является QSAR, нет возможности визуализировать и минимизировать энергию комплекса лиганд–мишень, перебирая конформации, то и потребности в вычислительных мощностях возрастают для соответственно.

Однако, следует иметь в виду, что любая статистически приемлемая модель QSAR, полученная в результате изменения структуры обучающей выборки (и, тем самым, выравнивания предсказательной способности системы), представляет собой действительную альтернативу перебору конформаций; и что изучение любых различий между такими моделями может улучшить понимание активности. Естественно, что стоимость такого «тестирования по площадям» существенна, но не идет ни в какое сравнение с «тестированием по площадям» путем массированного синтеза всего подряд – что, по мнению синтетика, «должно проявить требуемую активность».

В наших исследованиях мы применяли различные методы формирования дескрипторов, сравнивая результаты по каждому из методов, с использованием элементов машинного обучения, и вычленили из них только совпадающие результаты, что показало свою эффективность при расчетах ПАРП1.

Требования моделирования 3D-QSAR при оперировании 3D дескрипторами и PLS методологией учитываются частично или полностью в каждом программном пакете. Так, CoMSIA [33] (производство фирмы Tripos), оперирует расширенными (по сравнению с CoMFA) возможностями расчета стерических и электростатических взаимодействий, водородной связи и гидрофобных эффектов.

Использование фармакофорных эффектов в расчетах 3D-QSAR может ограничить число активных конформаций, сгенерированных обычными системами (GRID [34], являющийся составной частью пакета программ фирмы Molecular Discovery

[35,36], ПО Cresset [37] использует поле лиганда для координирования его положения в связке с мишенью. ПО Schrodinger содержит модуль PHASE для расчетов 3D-QSAR [38].

Два подхода уже учитывают комплементарные биологические области сайта связывания для прогнозирования лигандов на основе моделей 3D QSAR: программы COMBINE [39] и AFMoC [40]; есть еще более новаторские методологии 3D QSAR, HASSLE [41] and MSTD [42]. Наконец, существуют системы, выполняющие очевидную функцию (почему-то никем до них не реализованную) – создание "псевдо-рецептора" на основе анализа лигандов. Это делает последний вариант программы COMPASS 3D – QMOD [43].

Однако, ни один из этих методов не может обеспечить:

- конформационную фильтрацию лиганда в координации его с рецептором-мишенью;
- динамические трансформации лиганда в процессе координации его в комплекс с мишенью;
- учет влияния статического и динамического окружения комплекса, а именно внешние поля, влияние растворителя, транс-взаимодействия со стороны неактивных компонентов молекул лиганда и/или рецептора.

А это вполне может быть критичным при селективной *in vivo* активности. При рассмотрении и разработке любой методологии всегда необходимо учитывать фундаментальную основу любого исследования QSAR: единственной возможной причиной любого различия в биологической активности двух структур является их структурная разница – независимо от того, насколько сложными могут быть физико-химические и биологические взаимодействия, которые связывают эту структурную разницу с наблюдаемым биологическим эффектом. И основа для 3D-QSAR состоит в том, что эта разница в структуре лиганда лучше всего, прежде всего и изначально, выражена в полях лиганда. Эти соображения обеспечивают фильтрацию параметров сравнения структур в методологии 3D QSAR, поскольку позволяют безболезненно объединить структурно эквивалентные группы при расчетах.

Исследователи используют топологические индексы [44], 2D-проекции структур [45] или 3D-модели с дискретно описанной поверхностью [4]. Это все – варианты структурного моделирования ключ-замок.

Естественно, что для моделирования предпочтительна визуализация взаимодействия лиганд–мишень.

Для 3D визуализации применяются следующие программные комплексы:

Название	URL
Avogadro	http://avogadro.cc
BALLView	http://www.ball-project.org/ballview
gMol	https://github.com/tjod/gMol/wiki
Jamberoo	https://sourceforge.net/projects/jbonzer
LPMV	https://sourceforge.net/projects/lpmolecularviewer
Luscus	https://sourceforge.net/projects/luscus
Molecular Rift	https://github.com/Magnusnorrby/MolecularRift
OpenStructure	http://www.openstructure.org
PLIP	https://github.com/ssalentin/plip
PyMOL	https://sourceforge.net/projects/pymol
RasTop	https://sourceforge.net/projects/rastop
OpenRasMol	https://sourceforge.net/projects/openrasmol
SPADE	http://sites.google.com/view/spade
QuteMol	http://qutemol.sourceforge.net

Avogadro – 3D молекулярный визуализатор и редактор с плагином, написанным на языках Python и C++, который оптимизирует химическую связь при редактировании молекулы в режиме реального времени.

BALLView обеспечивает 3D-визуализацию в рамках системы *BALL cheminformatics*.

gMol обеспечивает базовую 3D-визуализацию молекулярной информации для системы *Open Babel*.

Jamberoo – написанный на Java 3D молекулярный визуализатор и редактор.

LP Molecular Viewer – ActiveX/ATL контроль для ввода интерактивной 3D визуализации молекулярных данных в продукты Microsoft.

Luscus – 3D визуализатор и редактор, разработанный с прицелом на информацию об электронной структуре.

Molecular Rift интегрируется с приборами виртуальной реальности Oculus Rift для обеспечения визуализации трехмерных молекулярных данных.

OpenStructure – вычислительная система структурной биологии, которая позволяет манипулировать структурной информацией; включает в себя интерактивную оболочку Python.

PLIP (Protein–Ligand Interaction Profiler) разработано как веб-приложение; анализирует и визуализирует 3D-взаимодействие белка и лиганда.

PyMOL – 3D визуализатор, включающий полный интерфейс Python для разработки скриптов и плагинов.

RasTop и *OpenRasMol* – базовая визуализация для системы RasMol.

SPADE (Structural Proteomics Application Development Environment) – графический интерфейс Python для структурной информации.

QuteMol обеспечивает высококачественную привлекательную визуализацию трехмерных молекулярных данных.

Общая схема представления дескрипторов свойств со стороны различных групп исследователей можно продемонстрировать иллюстрацией из лекции Гастайгера (Рисунок 3) [46].

Из Рисунка 3 видно, что подход к моделированию у математиков сводится к системе сходящихся функций, физиков – к определению объема и свойств физического пространства, занимаемого молекулой, химиков – какие свойства следуют из физико-химических параметров модели. При таком положении дел результаты (структурные модели для направленного синтеза) также будут различны у представителей различных наук для одних и тех же целевых веществ или систем.

Де-факто – несовместимо, реально – подходы могут дополнять друг друга в комплексном молекулярном моделировании целевой молекулы. Одной из основных целей данной работы и является попытка соединить противоречивые результаты различных подходов в моделировании веществ для направленного синтеза.

Рисунок 3. Разложение дескрипторов QSAR по трем естественнонаучным осям



Различия в подходах к моделированию, в частности, является причиной недоверия химиков к классическому QSAR как к методу моделирования структур с заранее заданной активностью [47].

Эти противоречия не мешают QSAR быть основным методом предсказания активности в тех случаях, когда в моделировании невозможно определить и использовать структуру мишени и структуру комплекса лиганд–мишень. Ренессанс QSAR в 2012– 2016 гг. отражен в работах [48,49].

Для расчета дескрипторов также существуют средства автоматизации [50]:

QSAR/ADMET моделирование

Название	URL
4D-FAP	http://www.ra.cs.uni-tuebingen.de/software/4DFAP
BlueDesc	http://www.ra.cs.uni-tuebingen.de/software/bluedesc
MolSig	http://molsig.sourceforge.net
PaDEL-descriptor	http://www.yapcwsoft.com/dd/padeldescriptor

Название	URL
<i>TMACC</i>	http://comp.chem.nottingham.ac.uk/download/tmacc

4D Flexible Atom–Pair Kernel (4D FAP) рассчитывает ‘4D’–сходство на основе сравнения молекулярных графов, соответствующих каждой конформации, которая должна быть посчитана в рамках QSAR.

BlueDesc – консольный расчет дескрипторов, конвертирующий MDL SD формат в ARFF и ***LIBSVM*** с использованием ***CDK*** и ***JOELib2*** в машинном обучении. Базовый набор – 174 дескриптора из обеих библиотек.

MolSig учитывает дескрипторы на основе молекулярных графов, содержащих стереохимическую информацию.

PaDEL-descriptor – расчет молекулярных дескрипторов и уникальных характеристик. Вычисляет 1875 дескрипторов (1444 1D, 2D дескрипторов и 431 3D дескрипторов) и 12 типов уникальных характеристик.

Topological maximum cross correlation descriptors (топологические дескрипторы максимальной кросс-корреляции) (***TMACC***) генерирует легко интерпретируемые 2D автокорреляционные дескрипторы, хорошо подходит для QSAR молекулярного моделирования

Средства автоматизации существуют и для построения моделей [50]:

Название	URL
<i>AZOrange</i>	https://github.com/AZCompTox/AZOrange
<i>camb</i>	https://github.com/cambDI
<i>eTOXlab</i>	https://github.com/manuelpastor/eTOXlab
<i>Open3DGRID</i>	http://open3dgrid.sourceforge.net
<i>Open3DQSAR</i>	http://open3dqsar.sourceforge.net
<i>QSAR-tools</i>	https://github.com/dkoes/qsar-tools

AZOrange – пакет машинного обучения, который поддерживает построение модели QSAR в полном цикле – от вычисления дескриптора до автоматизированного построения модели и ее проверки. Точности моделирования улучшена за счет использования нескольких высокопроизводительных алгоритмов машинного

обучения для эффективного выбора конкретного набора данных для статистической обработки.

Chemistry aware model builder (camb) – пакет для генерирования количественных моделей. Его возможности включают в себя расчет дескрипторов (в том числе 905 двумерных и 14 дескрипторов типа «отпечатков пальцев» для малых молекул, 13 дескрипторов последовательности цельных белков и 8 типов дескрипторов аминокислот), формирование моделей, моделирование комплекса и его визуализацию.

eTOXLab – портативная система моделирования, разворачиваемая как виртуальная машина.

Open3DGrid и ***Open3DQSAR*** все-в-одном пакеты моделирования 3D QSAR. *Open3DGrid* генерирует molecular interaction fields (MIFs) (поле молекулы) в разнообразных форматах, а *Open3DQSAR* строит прогнозные модели на базе этих полей. Расчеты визуализуются программой PyMOL в режиме реального времени.

QSAR-tools – набор скриптов на языке Python, использующих RDKit для построения линейных QSAR моделей на базе 2D химических данных.

Построенные модели можно автоматически применить к неизвестному соединению, и получить результат прогнозирования:

Название	URL
<i>SMARTCyp</i>	http://www.farma.ku.dk/smartcyp
<i>Toxtree</i>	http://toxtree.sourceforge.net
<i>UG-RNN</i>	http://cdb.ics.uci.edu/cgi-bin/tools/AquaSolWeb.py

SMARTCyp – QSAR модель, прогнозирующая сайты связывания P450-цитохромного связывания лекарственных молекул непосредственных анализом 2D (n/t/ диполь–диполь) энергетических взаимодействий.

Toxtree – приложение на JAVA для оценки «токсической опасности» молекул, используя модули прогноза токсичности, таких как токсичность при дыхании, попадании на кожу или в глаз, ковалентное связывание белка и ДНК, промотирование цитохромом P450 метаболизм лекарства и т.д.

UG-RNN/AquaSol – нейронная сеть, построенная на неориентированном молекулярном графе и предсказывающая растворимость.

Системы визуализации полученной информации:

Название	URL
<i>CheS-Mapper</i>	http://ches-mapper.org
<i>DataWarrior</i>	http://www.openmolecules.org/datawarrior
<i>DecoyFinder</i>	http://urvnutrigenomica-ctns.github.io/DecoyFinder
<i>Scaffold Hunter</i>	http://scaffoldhunter.sf.net
<i>Synergy Maps</i>	https://github.com/richlewis42/synergy-maps
<i>VIDEAN</i>	https://github.com/jimenamartinez/VIDEAN
<i>WCSE</i>	http://www.cheminfo.org/wikipedia
<i>WebChemViewer</i>	http://sourceforge.net/projects/webchemviewer

CheS-Mapper (chemical space mapper) . 3D вьюер химических баз данных. Подстраивает набор данных к выбранному пользователем 3D-пространству.

DataWarrior – инструмент визуализации и анализа химических данных с богатым набором доступных методов вычислений, метрик подобия, возможностей моделирования и представлений полученных результатов.

DecoyFinder предоставляет графический интерфейс для формирования обучающей выборки по заданной активности из большой библиотеки.

Scaffold Hunter предоставляет графический интерфейс Java для визуализации связи между соединениями в наборе данных.

Synergy Maps визуализирует синергетическую активность комбинаций препаратов.

VIDEAN (visual and interactive descriptor analysis) – визуальный инструмент для итеративного выбора подмножества дескрипторов, подходящих для прогнозирования целевого свойства с помощью статистических методов.

WCSE (Wikipedia chemical structure explorer) работает как веб-приложение и предоставляет 2D-интерфейс для визуализации и поиска 2D-молекул.

WebChemViewer онлайн система для просмотра и работы со списками соединений и связанных с ними данных.

2.3.1.3. Электростатические (кулоновские) взаимодействия

Координация структур дальнего порядка происходит на основе электростатического взаимодействия (можно называть его диполь–дипольным, Ван-дер-Ваальсовым, сольволитическим и пр., но суть от этого не меняется – первично взаимодействие электрических полей, имеющих векторную природу, и наводящимися одним или несколькими точечными источниками – например, атомами). В случае, если молекулы не комплементарны с точки зрения электрического диполь–дипольного взаимодействия (причем рассматривать молекулу как комплект диполей нужно в жестком каркасе), то до координации и химического связывания может и не дойти, даже при полной комплементарности взаимодействующих структур. Из чего следует, что в зонах взаимодействия I и II (см. выше) рассмотрение непосредственного атом–атомного взаимодействия имеет вторичную природу, первична электростатика, как постулируется еще в работе 1964 г. [14].

При сближении трехмерных объектов в первую очередь проявляется электростатическое интегральное взаимодействие полей. Электростатическое поле имеет векторную природу; т.е. для визуализации взаимодействующих объектов их необходимо дискретизировать – более наглядно представить в виде скалярных 3D-поверхностей, описывающих молекулу на условно фиксированном расстоянии от центров ядер. Это может быть ковалентный или Ван-дер-Ваальсовый радиус либо другая поверхность, описывающая молекулу или молекулярный конгломерат.

Предположим, это поверхность, полученная прокаткой шарика радиусом атома водорода по ковалентной поверхности молекулы [51]. Каждая точка такой поверхности характеризуется двумя параметрами – координатой и зарядом, рассчитанным как показано выше.

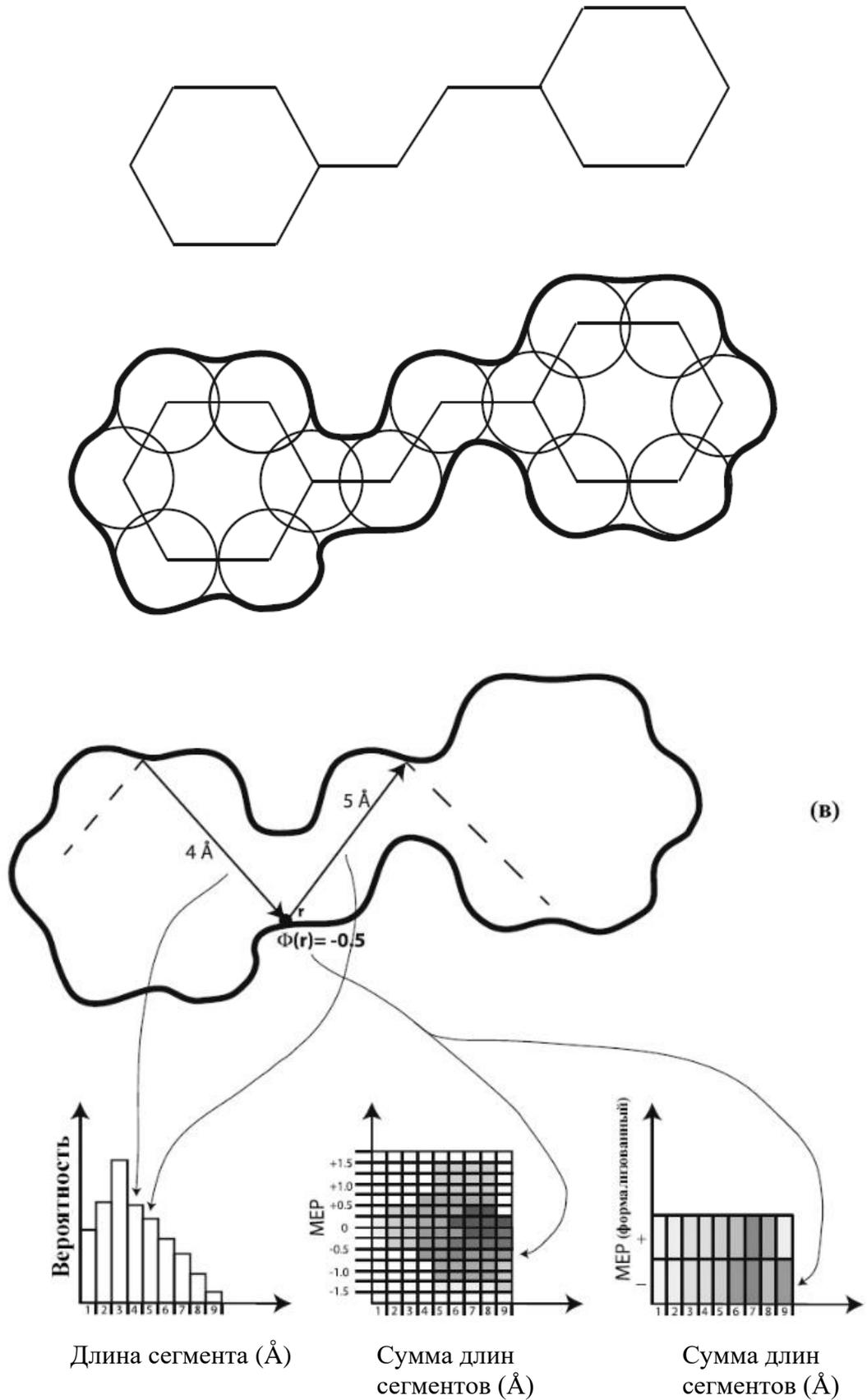
Возникает резонный вопрос: насколько существенна электростатическая составляющая, или, другими словами, интегральное взаимодействие векторных электрических полей молекул (либо молекулы и подложки (рецептора)) при определении энергии связывания в случаях, когда таковое связывание является скоростью-определяющей стадией процесса, а в исследуемом процессе нет разрыва–образования химических связей (т.е. исследуемый процесс – элементарная стадия координации взаимодействующих субстратов)?

Электростатическое взаимодействие происходит на существенно большем расстоянии, чем образование водородной (а уж тем более ковалентной) связи. Сольватационное взаимодействие (сольватное поле) относится сюда же, поскольку в большинстве случаев имеет либо электростатическую природу (дальний порядок), либо стерическую (ближний порядок). И, если объекты обладают несовместимыми по зарядовому расположению поверхностями, ни о каком химическом взаимодействии говорить не приходится. И наоборот, комплементарными являются координирующиеся объекты, имеющие схему расположения зарядовых областей с противоположными зарядами на молекулярной поверхности. Исходя из вышесказанного, теория «ключ–замок», по нашему мнению, имеет электростатическую основу, а электростатическое поле, окружающее молекулу, является важным параметром взаимодействия ее с окружением.

Как формировать это поле, описано в работе [52]. Оно создается с помощью окружности, прокатываемой на расстоянии атомных радиусов. Диаметр окружности обычно выбирается диаметром атома водорода.

Если работать с внутренним объемом данной поверхности, т.е. ввести в нее точечный источник света, то трассировка луча образует, в общем случае, два последовательных сегмента с отражением в точке g . Входящий сегмент имеет длину 4Å , выходящий из отражения сегмент имеет длину 5Å . Молекулярный электростатический потенциал в точке отражения (рассчитывается как функция зарядов атомов) равен $-0,5D$, где D – размерность дескриптора. Дескриптор "1D" аккумулирует распределение длин лучевых сегментов, аппроксимируемых гистограммой. «2D»– дескрипторы аккумулируют совместную вероятность наблюдения суммы длин сегментов, центрированных на отражении, в сочетании с потенциалом, измеренным в точке отражения. Для дескриптора, описывающего поверхностный потенциал, числовое значение потенциала используется непосредственно для идентификации ячейки гистограммы, для «формализованного» дескриптора во внимание принимается только знак потенциала (Рисунок 4).

Рисунок 4. Построение 2D-потенциала (сегментное построение),
переход к 3D-поверхности



2.3.1.4. Расчет молекулярной электростатической поверхности (МЭП)

Дискретный электростатический потенциал является точечным свойством (заряд), которое легко вычислить в любой точке пространства (координаты) из молекулярной волновой функции. Как уже сказано, на электростатические интерактивные взаимодействия приходится основная (часто единственная) часть дальнего взаимодействия молекул (лиганда и мишени).

Эти электростатические взаимодействия могут быть выражены в терминах мультипольных координат. Т.е. атом-центрированные поля (единичные) суммируются по стандартным правилам (закон аддитивности Гюйгенса) [53].

Модель суммирования единичных полей, генерированных отдельными атомами, обеспечивает интуитивно понятный вид распределения заряда в молекуле, который можно сравнить с более строгими методами.

- Несколько базовых методов моделирования точечной зарядовой поверхности неэмпирическими методами:
- использование регулярной точечного поверхностного молекулярного графа [54],
- атом-центрированные зарядовые конфигурации [55]. Здесь используется «алгоритм молекулярной поверхности Коннолли» для определения зарядов точек на нескольких концентрических оболочках, окружающих молекулу;
- процедура CHELP [56]. Процедура CHELP выбирает 14 точек для каждой оболочки, т.е. допускает большую дискретизацию описания оболочки. При этом соотношение радиусов оболочек всегда постоянно.

Все процедуры используют приблизительно 200–400 точек в каждой оболочке, с интервалами точек 0,8–1,0 Å. Алгоритм исключает точки внутри молекулярной поверхности из рассмотрения. Каждый из этих методов используется со своими волновыми функциями, и воспроизводит молекулярные свойства, такие как дипольный момент, достаточно хорошо. Появились также программы типа PDM88, определяющие молекулярный электростатический потенциал как функцию связь-центрированных взаимодействий [57].

Однако, многими исследованиями было установлено, что проблемы конформационного вращения делают CHELP и подобные методы непригодными для

использования в конформационном анализе.

Разнообразие разработанных методов расчета «только по лигандам» для различных молекулярных систем было просуммировано в классической работе Кубиньи [58]. Применимость метода, как уже было отмечено, прямо зависит от имеющейся информации о взаимодействии биологически активных веществ с рецептором. Если известна хотя бы количественная экспериментальная оценка биологического эффекта обучающей выборки лигандов, то применимы различные процедуры QSAR [59,60], включая такие 3D-методы, как CoMFA [27].

Если к тому же известна структура рецептора (белка), можно использовать методы докинга [61,62].

Для систем, о которых неизвестна конфигурация мишени, остается применить только метод молекулярного сходства. Основная идея такого подхода заключается в том, что две или более молекул обладают одинаковой биологической активностью, если они имеют сходные химические или физико-химические характеристики. Общий подход заключается в наложении активных молекул для нахождения общих структурных субъединиц, которые могут быть важны для биологического эффекта. При этом возникает ряд проблем – например, то, что значения молекулярного электростатического потенциала Ван-дер-Ваальсовых взаимодействий или водородных связей близки к нулю для такой процедуры. Одна из проблем касается выбора соответствующих химических элементов. Широкий спектр структурных и топологических дескрипторов, а также электронные свойства, такие как плотность электронов, атомные заряды и электростатические потенциалы, могут быть рассчитаны и применены в рамках теории электростатического взаимодействия.

Впервые метод молекулярного подобия, как и QSAR в общем плане, был применен к душистым веществам, первыми из которых были вещества с запахом сандалового дерева. Учитывались как стереохимические особенности соединений (благо в основном это конформационно нелабильные структуры), так и поверхностный электростатический потенциал. Полученные структурные модели, естественно, синтезировались, чем подтверждались теоретические подходы и расчеты (Таблица 2).

Последовательно далее изучались амбровый запах (также слаболабильные структуры), затем уже мускусный и другие.

Таблица 2. Примеры обучающей выборки по сандаловому запаху

№	Соединение обучающей выборки	Наличие и отнесение запаха	Лит.
1	(+)- <i>трет</i> -Бутилбицикло[4.4.0]декан-3-ол	чистый запах сандалового дерева	[63,64]
2	(<i>Z</i>)-(-)- β -Сандалол	типичный сандал, примесь древесного	[65]
3	Дигидро- β -Сандалол	сильный сандал	[66]
4	Деметилированный дигидро- β -Сандалол	типичный сандал	[66]
5	Деметилированный β -Сандалол	типичный сандал	[66]
6	(<i>Z</i>)- β -Сандалаль	сладкий, ср. сандал	
7	<i>экзо</i> -Изокамфанилциклогексан-3'-ол (акс)	похожий на сандал	[67][68]
8	<i>экзо</i> -Изокамфанилциклогексан-5'-ол (акс)	похожий на сандал	[67,67]
9	(-)-Мадрол	однозначно сандал, животный	[69,70]
10	α -Норсандаленон	древесный, сандал, напоминает иононы	[71,72]
11	β -Норсандаленон	древесный, зеленый сандал	[71,72]
12	Осирол®	типичный сандал	[73,74]
13	1-Метил-2- <i>цис</i> -метилциклопропил-5'-гекс-3-ил- <i>цис</i> -циклопропилметанол	сандал, сливочный, теплый, сильный	[75]
14	(+)-(<i>Z</i>)- α -Сандалол	древесный, кедровый, мягкий сандал	[66,71,75]
15	<i>экзо</i> -Изокамфанилциклогексан-2'-ол (акс)	без запаха	[68,76,77]
16	<i>экзо</i> -Изокамфанилциклогексан-2'-ол (эkv)	без запаха	[68,76,77]
17	<i>экзо</i> -Изокамфанилциклогексан-3'-ол (эkv)	без запаха	[68,76,77]
18	<i>экзо</i> -Изокамфанилциклогексан-4'-ол (акс)	напоминает сандал	[68,76,77]
19	<i>экзо</i> -Изокамфанилциклогексан-4'-ол (эkv)	без запаха	[68,76,77]
20	<i>экзо</i> -Изокамфанилциклогексан-5'-ол (эkv)	без запаха	[68,76,77]
21	<i>экзо</i> -Изокамфанилциклогексан-6'-ол (акс)	без запаха	[68,76,77]
22	<i>экзо</i> -Изокамфанилциклогексан-3'-ол (эkv)	без запаха	[68,76,77]
23	1-Метил-2- <i>транс</i> -метилциклопропил-5'-гекс-3-ил- <i>цис</i> -циклопропилметанол	запах лактона	[75]
24	(<i>Z</i>)-(-)-Окса- β -сандалол	без запаха	[64]
25	Тетрагидро- β -сандалол	без запаха	[70]
26	(<i>Z</i>)-(-)-Кето- β -сандалол	древесный, кедровый	[78]

Масло Восточно-Индийского сандалового дерева, *Santalum album*, обладает очень приятным, сладко-древесным и молочно-ореховым ароматом с превосходными свойствами закрепителя (фиксатора более легких ароматов). В начале 80-х годов Брунке и Найпавер [79] изучали корреляции структура-запах для аромата сандалового дерева, открывая особенности молекулярной структуры, характерные для аромата сандалового дерева. Работы с сандаловым запахом продолжились позднее [80], но общая модель для всех соединений, обладающих данным свойством, так и не была найдена. И только когда был применен метод наложения молекул на основе молекулярных распределений электростатического потенциала [81], удалось получить обобщенную модель свойства.

Существуют другие успешные примеры использования электростатических взаимодействий для изучения взаимосвязи между молекулярными характеристиками и биологической активностью [82].

Основная идея вычисления подобия молекулярных полей (включая электростатические) для различных молекул – поместить их в трехмерную «сетку», состоящую из точек, имеющих два параметра – координату и значение электростатического потенциала в каждой точке. Точки прилегают к молекуле, и поэтому используются как хорошее приближение молекулярной формы соответствующей структуры, т.е. представляют собой карту электростатического потенциала на молекулярной поверхности соединения. Эти точки поверхности затем используются для расчета стерического и электростатического подобия между двумя или более структурами путем суперпозиции структур, а затем путем сопоставления расстояний, причем выбираются те точки сетки, которые лежат в одном и том же месте в трехмерном пространстве с одним и тем же началом координат. Количество совпадающих точек сетки определяет степень стерического сходства между соединениями. Следующим шагом является оценка того, насколько электростатические потенциалы различаются в соответствующих точках сетки. Таким образом, с помощью этого алгоритма можно обнаружить области высокого и/или низкого уровня, а также электростатическое сходство.

Совпадение молекулярных поверхностей является важной задачей для оценки сходства между двумя структурами. Чем больше точек совпадения, тем более точны

результаты, полученные в результате расчетов подобия. Существует множество различных методов подгонки двух структур. Подход, который использовали в [81], отличается расширенным алгоритмом уплотнения, который предпочтителен, когда различные структуры имеют большие структурные различия, как это происходит в случае молекул запаха сандалового дерева (или, например, классический пример – различие молекул веществ, обладающих миндальным запахом). Этот метод основан на работе [83], где вместо суперпозиции парных атомов (ребер графа), сравниваются их стерические и электростатические поля.

Другое расширение, предложенное Машек и соавт. [84], учитывает подобные типы атомов при сравнении, в дополнение к стерическим и электростатическим полям. В процессе наложения молекул уплотнительный потенциал рассчитывается для образования комплекса с различных углов. Он состоит в юстировке полной энергии и определяется [85] как электростатический поверхностный потенциал. Для определения молекулярной электростатической поверхности (МЭП) введено понятие энергии взаимодействия между молекулой и протоном, расположенным на определенном расстоянии, для каждой точки МЭП.

Мы в своей работе (см. «Обсуждение результатов») использовали это определение в построении МЭП душистых веществ и потенциально биологически активных нейробиологических веществ.

С помощью информации о заряде точки на молекулярной поверхности можно обнаружить гидрофобные либо гидрофильные взаимодействия, поскольку они представляют собой электростатические взаимоотношения между двумя структурами.

Возможность учитывать конформационную гибкость соединений по методу Монте-Карло (Monte Carlo simulated annealing (MCSA)) очень полезна. Здесь нет необходимости определять ни пары атомов, ни относительную ориентацию совмещаемых молекул, что было бы довольно трудно делать со структурами, существенно отличающимися по объему; это приводит к огромному количеству конформаций и разнообразию вводимых энергетических параметров. Вероятность P использования одной конформации определяется следующим уравнением:

$$P(\Delta E) = e^{-\left(\frac{\Delta E}{kT}\right)}$$

k – константа Больцмана, ΔE – разность энергии между двумя итерациями и T – абсолютная температура.

Из чего следует, что количество использованных конформаций зависит от

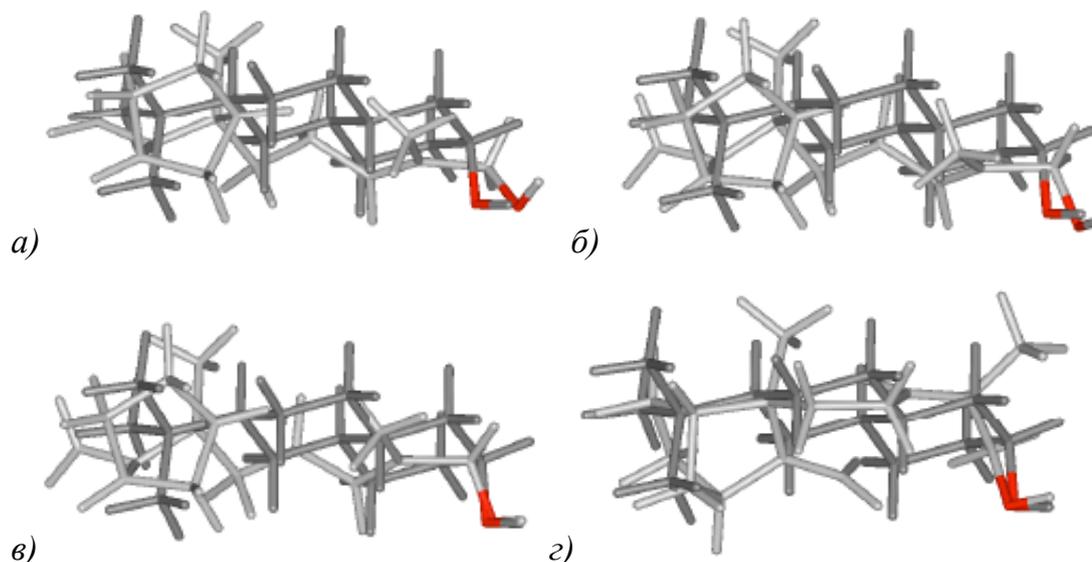
температуры: чем выше температура, тем больше приходится их учитывать, и наоборот. Учет конформационной гибкости гораздо более эффективен, чем использование методов на основе пар атомов.

2.3.1.5. Построение поверхности молекулы.

Пример использования МЭП для прогнозирования сандалового запаха приведен в работе [86]. Использованы два метода генерации сферы:

- Ван-дер-Ваальсова координатная поверхность, образованная атомной структурой, причем полная поверхность является суперпозицией отдельных сфер, образованных атомами, в соответствии с алгоритмом [87]. Ключевые точки поверхности располагаются на равном расстоянии друг от друга.
- Поверхность потенциальной энергии (по сути, МЭП) использует рассчитанные заряды, являющиеся компиляцией атомных потенциалов. Сгенерированная поверхность обладает двумя характеристиками – зарядом и координатой каждой дискретной точки. В отличие от п.1, эта поверхность отражает такие внутренние электростатические взаимоотношения, как диполь–дипольные взаимодействия, водородные связи, двойные связи, поскольку они влияют на электростатическую природу каждого атома. Конечно, такие параметры необходимо рассчитывать при помощи квантовой механики [88]. Примененные вместе, эти два метода дают следующую картину совмещения двух молекул (Рисунок 5).

Рисунок 5. Визуализация совмещения двух молекул (серая и темно-серая) при фиксации *a)* O–O, *б)* C–O, *в)* O–H, и *г)* C–O–H связей



Метод позволил предсказать свойства несуществующих, смоделированных для синтеза структур, дал возможность приблизиться к пониманию механизма действия душистого вещества и комплементарного рецептора, а также к решению вопроса – в какой конформации и с какой стереохимией душистое вещество обладает искомым свойством.

Работа по сандаловому запаху рассмотрена так подробно, поскольку запах сандалового дерева обладают вещества самого разного строения, что затрудняет поиск общих структурных элементов, отвечающих за основу запаха, Сравнение МЭП дает больше информации о нахождении активных рецепторов. Поэтому в цитируемой работе применено сочетание различных методов сравнения как молекулярной формы, так и электронных свойств соединений. Да, конечно, приходится для ограничения вычислительных потребностей рассматривать выборку соединений с ограниченной конформационной подвижностью. Обучающая выборка из 25 различных соединений с запахом сандалового дерева, плюс обязательное количество неактивных соединений, показали, что применение к ним двух методов дало практически полное соответствие рассчитанной модели. Смоделированные соединения были синтезированы, что валидировало расчетные методы. И, самое главное, можно сделать базовый вывод, что для определения более детальных дескрипторов и расширения возможностей QSAR необходимы квантовохимические расчеты, глубокий регрессионный анализ и нейросетевые технологии. Мы в своих исследованиях («Обсуждение результатов») руководствовались именно этими принципами.

2.3.2. Практическое моделирование QSAR

Отсутствие информации по структуре мишени, на которую координируется активное вещество, приводит к алгоритму моделирования (Схема 1).

Рассмотрим схему детально, с учетом многочисленных опубликованных описаний подобных схем.

Обучающая выборка. Поскольку методы обработки дескрипторов в рамках классического QSAR – сугубо математические, то подбор структур обучающей выборки, основанной на физико-химических параметрах соединений, а также выбор оных дескрипторов во многом определяет результат исследования [8].

Схема 1



В случае биологически активных соединений эта выборка должна отвечать следующим критериям, общим для QSAR: а) вещества не должны вступать в химические взаимодействия с мишенями; б) в выборке должна быть критическая масса структур для расчета; в) должны быть количественные данные об активности веществ в выборке; причем часть выборки является обучающей, а часть – контрольной, валидирующей последующие расчеты. И там, и там должны присутствовать как активные, так и неактивные соединения со схожими структурами (последнее – для улучшения распознавания дескрипторов).

Дескрипторы. Как изложено выше, базовым принципом данной работы было использование различных имплементаций электростатического поля, генерируемого одной или несколькими молекулами. Соответственно, выбранный дескриптор для данного обзора – вторичные реализации действия электростатического поля молекулы, с учетом влияния (электростатического или электродинамического) сторонних объектов в системе лиганд–мишень. Другие дескрипторы, коих множество, и количество их с годами только увеличивается, по изложенным причинам мы рассматривать не будем.

Электростатическое поле генерируется атомами, составляющими молекулу, и, в соответствии с законом Фарадея, электрический потенциал $\varphi_{\text{общ.}}$ в любой удаленной точке равен $c_{\text{общ.}}\varphi_{\text{общ.}} = c_1q_1 + c_2q_2 + \dots + c_kq_k$, где $q_1 \dots q_k$ = известные заряды в фиксированных точках (например, заряды атомов в молекуле), $c_1 \dots c_{\text{общ.}}$ – «вклады» этих точечных зарядов, обратно пропорциональные квадрату расстояния, зависящие также от экранирования, магнитных и прочих факторов. Именно так отрисовываются скалярные «поверхности потенциальной энергии», в нашей интерпретации – молекулярные электростатические поверхности (МЭП) различных химических визуализаторов. С термином МЭП также связаны параметры различных физико-химических процессов, что приводит к некоторым недопониманиям между химиками, но мы здесь тем не менее воспользуемся этой аббревиатурой, которую сами же и ввели для этой цели в 1994 г. [51].

Базовым дескриптором, определяющим свойства молекул, будем считать взаимодействия электрических полей молекул. Эти поля имеют векторную природу, и для расчета дискретных взаимодействий (других пока в QSAR не используется) необходимо дискретизовать МЭП лиганда. Очевидно, это должно быть на расстоянии максимального действия (ковалентный или Ван-дер-Ваальсов радиус), и быть легко считаемой скалярной величиной, производной векторного поля.

Такие скалярные описания 3D-поверхности могут быть:

- Проекция на пространственные оси.
- Описывающие молекулу пространственные примитивы.
- Точки, расположенные на фиксированном либо динамически варьирующемся расстоянии от ядер атомов,
- 2D-фигуры достаточно малого размера [51], облегающими данную поверхность.

Молекулярная электростатическая поверхность (МЭП) как основной фактор влияния на связывание молекул широко использовался в дальнейшем для прогнозирования запаха.

Формализуя QSAR-подход, можно сказать, что здесь происходит построение *квази-модели* – структуры (не обязательно атомно-молекулярной), отражающей в максимальной степени характеристичные свойства всех активных молекул обучающей выборки. К этой структуре и «прикладываются» молекулы, для которых требуется определить априорную активность.

Нижняя часть схемы, в соответствии с изложенными выше принципами построения любого QSAR-исследования, является обязательной.

2.3.2.1. Теории запахов и их место в расчетах QSAR.

Как ни парадоксально, но до сих пор наиболее точным применением классического и 3D-QSAR оказалось применение его к классификации запахов. Причин этому две:

Профессионал-аналитик различает тысячи оттенков и может с достаточной точностью классифицировать вклад каждого запахового компонента в суммарный аромат; при этом подавляющее большинство аналитиков сходятся в оценке одоранта. В этом – существенное отличие от практического определения биологической или каталитической активности, где иерархический вид активности, вид и степень ее проявления, подборка обучающей выборки на этих принципах сопряжены с аппаратными и, как минимум, *in vitro* исследованиями.

Классический QSAR, как уже было сказано, имеет ограничение: координирующиеся структуры не должны вступать в химическую реакцию. Душистые вещества, по определению, отвечает этому требованию. Биологически или каталитически активные вещества – совсем не обязательно, и когда химическая реакция вмешивается в расчет координации лиганд–мишень, этот расчет является фикцией.

Всего было предложено около 50 различных теорий запахов, в которых принята попытка найти разумное объяснение возникновению ощущения запаха. Но они до сих пор остаются теориями, поскольку не получили достаточного экспериментального подтверждения [89].

2.3.2.2. Выбор дескрипторов для QSAR одорантов.

Влияние различных функциональных групп одорантов на генерируемый запах и поиски оптимальных структур для различных классов душистых веществ начались более 70 лет назад, и не прекращаются до сих пор. Как и попытки классификации и кластеризации одорантов и биологических мишеней различных запахов.

Все это накладывается на весьма сложный механизм человеческого обоняния: предполагается, душистые вещества воздействуют на рецепторы, связанные с G-белком, выстилающим носовой эпителий. Но основное различие между обонятельными рецепторами и другими рецепторами, связанными с G-белком, заключается в

том, что обонятельное восприятие происходит из-за комбинаторного кодирования нескольких сотен обонятельных рецепторов [90].

Современное понимание механизмов обнаружения и фиксации запахов и того, как мозг реконструирует эти воздействия в "карту запахов", сформулировано Ричардом Акселем в Колумбийском университете (Нью-Йорк, штат Нью-Йорк, США) и Линдой Бак в исследовательском центре рака Фреда Хатчинсона в Сиэтле (штат Вашингтон, США), которые получили Нобелевскую премию 2004 года по физиологии и медицине "за свои открытия запаховых рецепторов и схему обонятельной системы". Принципиальная схема человеческой системы фиксации и передачи запахов [91] приведена на Рисунке 6.

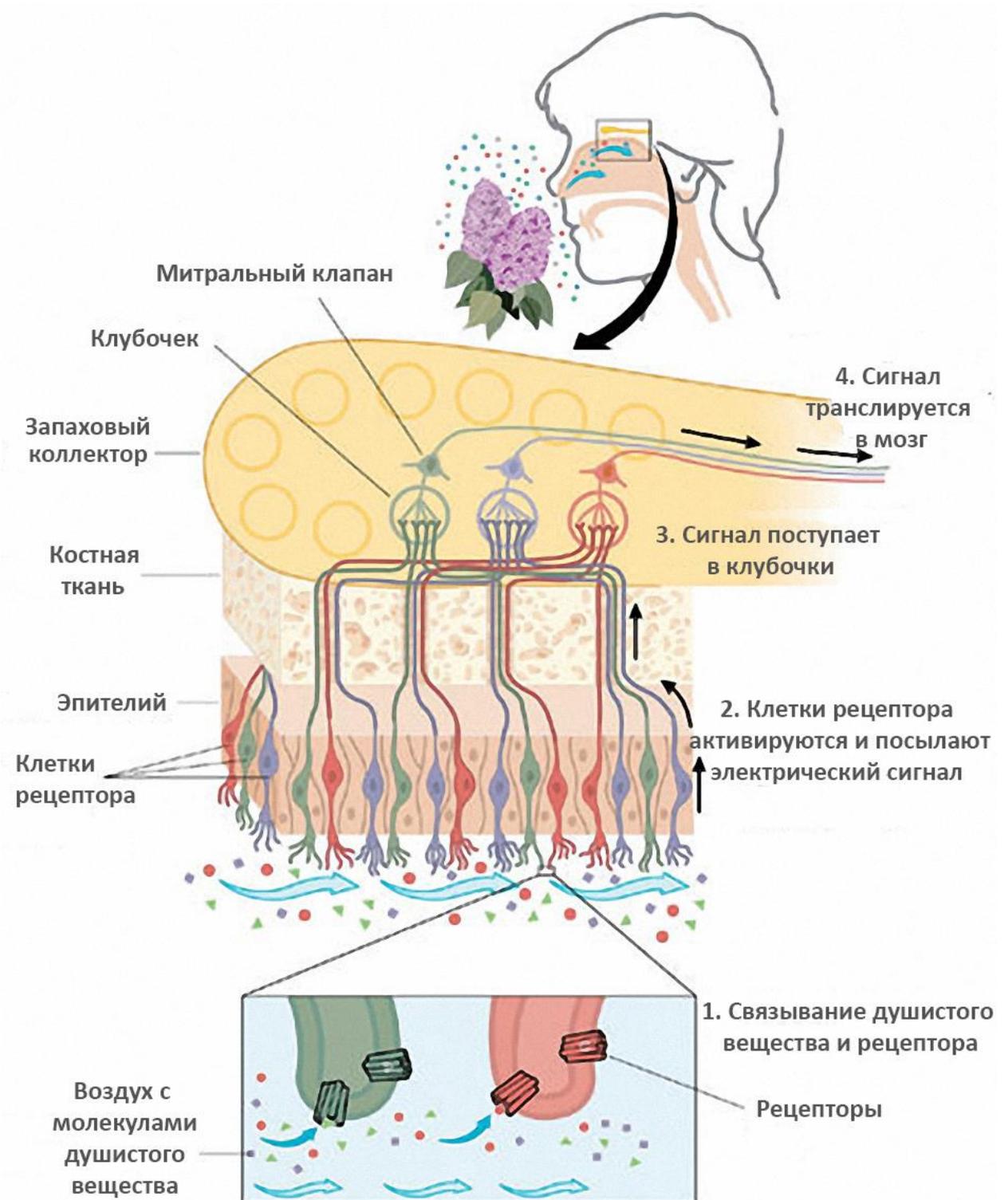
Рецепторы запаха (РЗ) локализуются на поверхности сенсорных нейронов, которые занимают пространство в верхней части назального эпителия. Постулируется, что каждая обонятельная клетка рецептора соответствует только одному запаху. При активации сигналы от рецепторных клеток передаются в клубочках – в обонятельные луковицы. Рецепторные клетки одного и того же типа случайным образом распределяются в слизистой оболочке носа, но сходятся в одном и том же гломеруле (коллекторе). В гломеруле нервные окончания рецептора возбуждают митральные клетки, которые передают сигнал в центральную нервную систему.

РЗ на поверхности сенсорных нейронов показывают специфичную структуру и чувствительность, зависящую от природы одоранта. Мышь, например, имеет примерно 1200 разных РЗ, в то время как люди – менее чем 400. Первичные данные, полученные от РЗ, комбинируются и используются для кодирования индивидуального составного запаха.

Каждый РЗ либо связывается с несколькими одорантами, и, соответственно, каждый одорант обнаруживается определенной комбинацией РЗ. Затем мозг переводит этот специфический комбинаторный "код рецептора" в отчетливый комплексный запах.

Сигналы, поступающие от нейронов в обонятельной луковице (коллекторе), поступают к частично перекрывающимся кластерам нейронов в обонятельном эпителии. Это взаимодействие формирует сенсорную карту, которая радикально отличается от первоначальной, потому что обонятельные нейроны эпителия могут получать сигналы от десятков различных РЗ.

Рисунок 6. Принципиальная схема человеческой системы фиксации и распознавания запахов



Это повышает вероятность того, что один из «собирающих» нейронов может интегрировать сигналы от разных РЗ, которые обнаруживают тот же вид запаха, и,

таким образом, обеспечивают «начальный этап в реконструкции образа запаха от его особенностей деконструкции» [92].

Следуя этой схеме, нейроны в обонятельном коллекторе могут выступать в качестве "случайных детекторов", которые проявляют активность только при воздействии на них определенного сочетания душистых веществ. Показано в этой связи, что существуют нейроны, неактивные при воздействии отдельных душистых веществ, но проявляющие нейроактивность при воздействии бинарных смесей тех же душистых веществ [91].

Одной из самых больших проблем в теории запахов является корреляция РЗ с лигандами. Определение вида и энергии взаимодействия РЗ–лиганд покажет, как различные РЗ генерируют различные восприятия запаха, однако эти исследования биологической природы запаха представляют собой весьма сложный эксперимент. До сих пор достоверно известно лишь очень небольшое число РЗ млекопитающих, четко связанных с определенными одорантами.

Если загадочные вопросы и нерешенные проблемы делают науку захватывающим, то теория запахов – это целиком такое поле. До сих пор нет понимания, как 350–400 доказанных рецепторов запахов способны учитывать идентифицировать тысячи различных запахов с их оттенками. Доминирующие теории обоняния – известная как теория одотопов постулирует, что форма молекулы признается специфическим рецептором, который каким-то образом определяет, как она пахнет. Т.е. вместо жесткого механизма ключ–замок рецепторы связываются только с небольшой частью одоранта так, что разные молекулы могут активировать одни те же рецепторы различным образом. Однако, несмотря на свое ведущее состояние, теория формы не может объяснить почему, например, различные молекулы могут иметь идентичный запах и почему подобные молекулы могут иметь сильно отличающийся запах [93].

Базовый вопрос для всех теорий запаха – стадия 1, т.е. механизм воздействия (связывается или не связывается, если не связывается, то каким образом воздействует?) душистого вещества и рецептора.

При просмотре достаточного количества литературы о воздействии вещества на организм с генерацией ощущения запаха, можно объединить сотни разных точек зрения в три коллекции (назовем их базовыми теориями запаха):

- 1) волновые воздействия на рецепторы (колебательная теория);
- 2) контактные воздействия ключ–замок;
- 3) контактные воздействия «ключ без замка».

Аргументация каждой точки зрения приводится вполне разумная, поэтому, не будучи биологом, нельзя четко остановиться на какой-либо из них. А биологи так и не пришли к единому мнению.

2.3.2.3. *Колебательная теория запахов*

Среди различных альтернативных теорий запаха наиболее противоречивой является так называемая «колебательная теория обоняния». Впервые представленная в 1930-х годах, она утверждает, что запах молекулы определяется внутримолекулярными колебаниями, а не формой молекулы. Колебательная теория запаха полагает зависимость запаха от колебательных частот в инфракрасном спектре. Все началось с работы Дайсона [94], в которой постулировалось, что колебательные процессы химической связи между характеристичными атомами являются запах–определяющими. Механизм объясняется как удаленная передача запаха (влияние на одорантные рецепторы) через модулирование частоты колебаний химических связей одорантов.

В 1996 году Лука Турин, биофизик, работавший в Университетском колледже Лондона, Великобритания, осовременил базовые положения теории, предположив, что передача молекулярных колебаний для активации запахового рецептора может происходить через туннелирование электронов [95]. Электроны в рецепторах одорантов туннелируются между различными энергетическими состояниями, если частота колебаний лиганда соответствует разности энергий этих состояний, тем самым активируя рецептор. Различные одоранты имеют уникальные колебательные характеристики, и различные частоты колебаний будут фиксироваться различными рецепторами, что объясняет, почему подобные молекулы пахнут по-разному, а разные по структуре молекулы могут пахнуть одинаково.

Идея Турина не получала признания до 2003 года, но дискуссия возобновилась после того, как научный писатель Чандлер Берр написал книгу о теории запахов, Турине, его «еретической» теории и игнорировании ее научным сообществом [96], и перешла в принципиальную дискуссию о необходимости валидации

любых теоретических построений в научной практике [97].

Колебательная теория запаха подтверждается исследованиями, показывающими, что насекомые, люди и другие животные могут отличить две версии одной и той же молекулы одоранта – нормальную и дейтерированную. Дейтерированные и нормальные молекулы одоранта имеют одинаковую форму и характеристики поверхности, и все же люди и другие животные могут почувствовать разницу.

Общая теория молекулярного распознавания обонятельной (и не только обонятельной) системы утверждает, что рецепторы признают структуру или форму запаха (вкуса, других воздействий) [94] молекулы. Замена водорода одорантов в эксперименте различения запахов *Drosophila melanogaster* показала, что мухи различают изотопный состав. Мухи не только дифференцируют изотопические одоранты, но и выборочно реагируют на одно из соединений. Эти наблюдения несовместимы с моделью ключ–замок обоняния, и поддерживают существование молекулярного вибрационно-чувствительного обонятельного компонента [98].

Франко и др. [99]. считали, что дейтерированные одоранты генерируют специфические C–D колебания, которые, в свою очередь, возбуждают специфический «модифицированный запах», не зависящий от «структуры и химических свойств молекул одоранта». Однако, человек не способен различать изотопы по запаху: он не чувствует различий между нормальным и дейтерированным ацетофеноном [100].

Общий вывод: не найдено существенных подтверждений, что функции обоняния, как «другие спектральные чувства, зрение и слух» [94], имеют колебательную природу. Биологические системы, даже если они генерируют инфракрасный или ультрафиолетовый спектр излучения, не могут осуществлять функции туннелирования электронов [99].

Гипотетическая связь между запахом и спектральными особенностями не подтверждается следующим:

- во-первых, запахи могут сильно отличаться, даже когда инфракрасные спектры очень похожи. Например, инфракрасный спектр циветона (макроциклического кетона с запахом мускуса) практически идентичен его кислой форме (без запаха) [101]. Различия запаха в оптических изомерах с идентичными инфракрасными спектрами также не поддерживают теорию вибрации.

- во-вторых, запахи могут быть очень похожи, даже когда спектры сильно отличаются.

Таким образом, неудачные попытки установить количественную связь между запахом и спектрами, скорее, работают против колебательной теории запахов.

Однако, альтернативная «теория формы», или механизм «ключ–замок» также неточны. Химия запаха включает множество факторов, помимо формы или химического строения. Например, этанол и этантиол имеют подобные формы, но химически (и запах тоже) отличаются, как роза и сунс. Использование колебательной теории в обнаружении изотопов по различию запаха биохимически бессмысленно. Нюх собаки имеет химический смысл, а не спектральный. QSAR во взаимосвязи между запахом и химическим строением проявляется, когда учитывается не только форма, но и другие понятия, такие как молекулярные вес, геометрия функциональных групп, полярность, кислотность, основность и стерические взаимодействия. У людей или у мух одоранты связывают и активируют рецепторы за счет сочетания химических свойств, и это ничем не отличается от других видов рецепторных молекулярных сигналов в биологии [102].

2.3.2.4. Ключ–замок

Дискретная, базирующаяся на фиксированных мишенях и фиксированных функциональных особенностях лигандов (активных структур) теория, первая и самая многочисленная по количеству аргументов. Базовые мишени дискретных запахов найдены у животных и человека, базовые структуры и функциональные группы, отвечающие за различные запахи, зафиксированы и классифицированы, (например, [86,103,104]).

Серьезные исследования по душистым веществам проводились практически всеми QSAR-апологетами. Пик пришелся на середину 90-х, когда работы в области QSAR привели к логичному интересу к стереохимически-ориентированным молекулярным дескрипторам [105,106]. Дело в том, что практически все душистые вещества, обладающие оптической активностью, показывают, что более сильный запах имеет какой-то один стереоизомер.

В работе [107] проведено генерирование хиральных молекулярных дескрипторов для QSAR-моделирования двух выборок душистых веществ со сложной

стереохимией. Первая выборка состояла из производных камфоры с известной хиральностью оптически активных центров. Для расчетов было формализовано положение пяти функциональных групп относительно хирального центра. Модель QSAR была сформирована с использованием множественной линейной регрессии (MLR) и валидирована, как это общепринято, внешней контрольной выборкой известных душистых веществ.

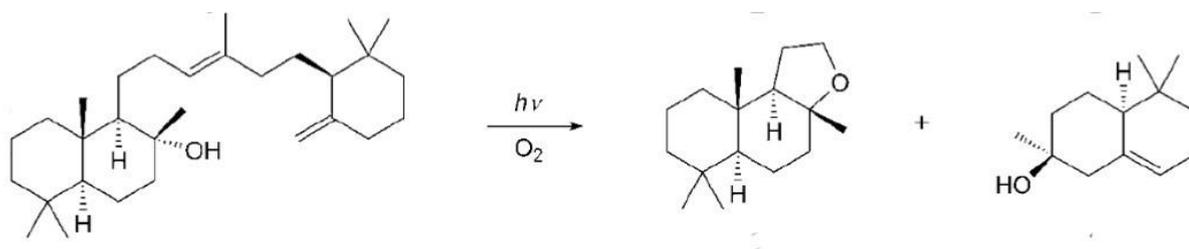
2.3.3. QSAR исследования душистых веществ

Амбровый запах в исследованиях QSAR используется потому, что он имеет стойкость, наиболее выраженные оттенки, не забывает чувствительность запаховых рецепторов человека, как, например, мускусный, – и человеческий нос может отличить эти оттенки с очень малым разбросом.

Амбра, или «морское золото» – комки с землистым сладким запахом – начиная с X века находили на берегу океана и использовали в спиртовом растворе для тонкой парфюмерии. Обладает стабильным «вечерним» ароматом и отличными фиксирующими (для легколетучих душистых веществ) свойствами.

Амбра – это продукт метаболизма кашалотов (*Physeter macrocephalus L.*). Согласно самой популярной теории [108], это патологическое уплотнение, возникающее из-за пищевых травм. Содержит тритерпеновый спирт *амбреин*, соединение без запаха, фотолитически расщепляющийся до душистых веществ с амбровым запахом (Схема 2).

Схема 2



Амбреин

Ледерер [109] первым заметил, что чем дольше кусочки амбры находятся в воде, тем тоньше запах. Коммерческое использование амбровых одорантов стимулировало стремительное развитие химии искусственных веществ с этим. Синтез новых одорантов был поддержан развитием SAR (Structure–Activity Relationship).

Первые исследования SAR проводились Ола и сотрудниками и привели к «амбровому треугольнику» [110]. Согласно этому правилу, соединения с запахом амбры должны иметь скелет *транс*-декалина с осевыми заместителями в положениях 1, 2 и 4 (типичным примером являются соединения **24** и **25**, Таблица 3). Одна из этих функциональных групп должна была включать атом кислорода. Однако «амбровый треугольник» постулировался исключительно для группы производных *транс*-декалина и не объясняло отсутствие запаха амбры для других молекул, которые соответствуют структурным требованиям «амбрового треугольника» (Таблица 3, соединения **30**, **60**, **67** и др.); или, например, наличие запаха амбры для молекул с *цис*-декалиновым скелетом (Таблица 3, соединения **54**, **77**, **78** и др.) или вовсе со скелетом другого структурного типа (Таблица 3, соединения **1**, **81** и др.) [111 – 113].

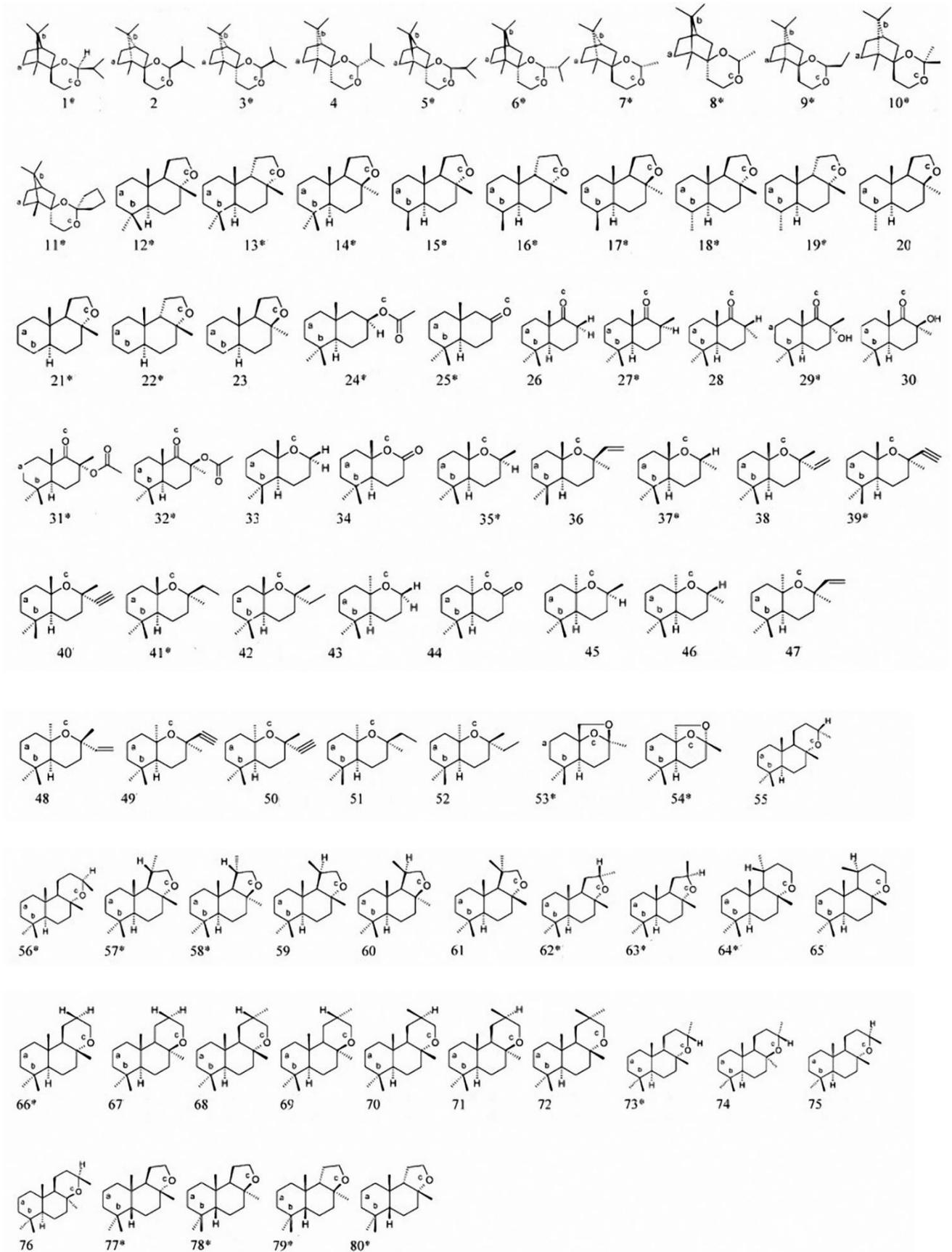
В дальнейших исследованиях были сформулированы структурные требования к нескольким химическим группам амбрового запаха.

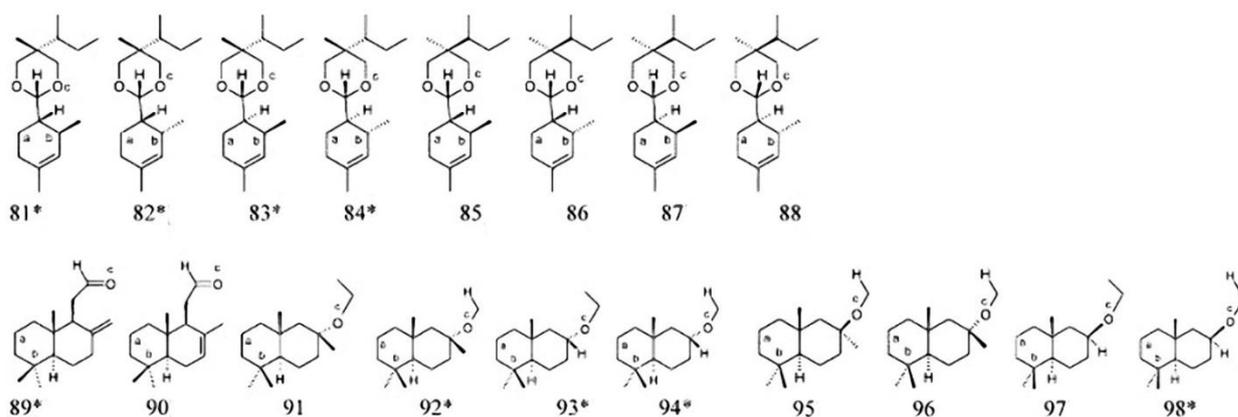
П.Ф.Влад и соавт. установили, что в декалиновых системах амбровый треугольник, образованный атомом кислорода и двумя атомами водорода, должен стабилизировать НСМО амбровых соединений, из чего следовало, что НСМО принимают участие в «орбитальном контролируемом переходе электронного заряда» между активными молекулами с амбровым запахом и рецептором запаха [114].

Теория о наличии определенных структурных фрагментов, отвечающих за запах амбры, получила дальнейшее развитие у Димогло [115] и Горбачева и Росситера [113], применявших электронно-топологический подход [116,117]. Димогло определил два характерных структурных фрагмента: первый включал определенные атомы углерода и атом кислорода, связанный с вторичным или третичным углеродом; второй состоял из двух метильных групп, прикрепленных к четвертичному атому углерода, с одинаковой стереохимией.

Эти два фрагмента объясняют исчезновение запаха амбры, когда пятичленное кольцо **63** заменяется шестичленным в соединении **65**. Однако, это не объясняет присутствие запаха в соединении **64**. Следовательно, корреляция между *стерической доступностью* (СД) функциональной группы (гидроксил, эфир, эфир) и амбровым запахом [118,119] должна быть скоррелирована СД атома кислорода [114]. Для соединений с более сложными структурами помимо СД атома кислорода необходимо учитывать СД определенной метильной группы [114].

Таблица 3. Набор веществ с амбровым запахом,^a
исследованных методом CoMFA [28]





^a Соединения, обозначенные *, имеют запах амбры. Атомы, задействованные в расчетах CoMFA, маркированы буквами.

В «дорасчетные» времена присутствие или отсутствие запаха амбры успешно коррелировало только с СД функциональной эфирной группы в бициклических производных *Амброкса* [117]. В последующих исследованиях была предпринята попытка найти дополнительные структурные фрагменты в сочетании с известными параметрами СД для соединений определенных структурных типов (не только с амбровым запахом). А найденные закономерности превратить в модельные структуры и синтезировать их. Далее – уже модельные структуры валидировать при помощи компьютерной симуляции.

В работе [120] синтезированы шесть новых стереоизомерных производных камфоры, обладающих амбровым запахом, нарисованных на основе QSAR-гипотез [121], и рассчитанных при помощи программы молекулярной симуляции CATALYST [122].

Эти QSAR-гипотезы включали одну функцию, ориентированную на акцепторные свойства водородной связи [123] и четыре – на основе гидрофобных параметров соединений. Гипотеза подтверждена успешно.

Установлено, что реакция рецепторов запаха чувствительна и к стереохимии душистых веществ. Запах стереомеров может отличаться от присутствия до отсутствия, разным уровнем интенсивности или оттенком. Стереохимия – это важнейший фактор, который влияет не только на запах [124,125], а также на свойства (включая физико-химические) множества классов биологически активных соединений, таких как аминокислоты, углеводы, различные фармацевтические препараты [126 – 131]

Таким образом, существует важная задача в медицинской, биоорганической и теоретической химии – разработки прогнозных моделей QSAR стереоизомеров [131].

Классическая работа по QSAR-исследованиям амбрового запаха [132] суммирует вышеперечисленные методы в применении к амбровому запаху. Комбинаторный подход QSAR был применен именно к перечисленным (Таблица 3) 98 соединениям с амбровым запахом (классический набор, которым до них и после пользовались многие исследователи, включая нашу группу) со сложной стереохимией для определения относительной интенсивности запаха стереоизомеров одной структуры. Набор хиральных дескрипторов сгенерирован коммерчески доступными методами MOE (Molecular Operating Environment, Среда Молекулярного Моделирования) [133,134], CoMFA [27], CoMMA [135], Dragon, VolSurf [136] и MolconnZ [137] с использованием хиральных топологических дескрипторов [138].

Каждый дескриптор в канонической последовательности операций рассматривается независимо от остальных, и в каждом расчете используется максимальное количество методов QSAR. Таким образом нивелируются ошибки и приближения отдельных методов.

Авторы назвали метод *Combinatorial quantitative structure–activity relationships* (*Combi–QSAR*, *Комбинаторный QSAR*). Дескрипторы CoMMA использовались в сочетании с дескрипторами MOE, а дескрипторы MolconnZ – в сочетании с хиральными дескрипторами. Каждый дескриптор был проверен на соответствие четырём разными способами, включая подобные параметры ближайших k дескрипторов (k -NN), *Support Vector Machines* (*SVM*, *наложение векторных параметров*), статистическое дерево решений, и статистическое наложение 28 различных моделей QSAR.

Различные репрезентативные обучающие и контрольные выборки были подвергнуты двойной фильтрации, чтобы избежать переобучения и обеспечения надежной предсказательной силой. Использовались два метода валидации: рандомизация целевого свойства (в данном случае интенсивность запаха), также известная как Y -тест, и оценка точности прогнозирования с использованием контрольной выборки. Классификация k NN в сочетании с дескрипторами CoMFA дала наилучший прогностический результат QSAR среди все полученных моделей свойства, с коэффициентом прогнозирования 0,7, что является наивысшей величиной для QSAR того

времени, и ненамного ниже, чем результаты прогнозирования биологической активности, полученные методами докинга (с известными структурами как молекул лигандов, так и мишени).

Хиральные дескрипторы. Авторы работы [139] использовали парные хиральные дескрипторы (ХД), рассчитанные из химических графов амбровых веществ. Такие дескрипторы не требуют информации о 3D молекулярной структуре, за исключением конфигурации непосредственного окружения хиральных атомов. Это непосредственное окружение определяет частичный заряд хирального атома или группы. На этом строится модель максимального проявления искомого запаха.

Достаточно неожиданный (на первый взгляд) дескриптор определения вида одорантов выбран в работе [140] – *индекс удерживания одоранта на хроматографической колонке*. QSAR в этом случае звучит как QSRR (Quantitative Structure–Retention Relationships). Но неожиданный только на первый взгляд, поскольку этот индекс зависит, по мнению авторов [141], от: χ (индекс молекулярной связности), $n_{\text{донг}}$ (количество доноров водородной связи), MW (molecular weight, молекулярной массы), DPSA1 (зарядовая разница между частично положительно и отрицательно заряженными участками молекулярной поверхности), FPSA1 (распределение преимущественно положительно заряженных участков), qh_{max} (максимумы положительного заряда на водороде), μ (дипольный момент), i_{pc} (информационная составляющая матриц связности), и $4\chi_4$ (индекс молекулярной связности для четырех кластеров). Последний параметр связан с конкретным экспериментом – 1341 душистых веществ разделяли на четырех неподвижных фазах с разной полярностью: OV-101, DB-5, OV-17 и Carbowax 20M.

Используя стандартные методы QSAR, авторы [37] получили степень прогнозирования активности для обучающей выборки в 0,6–0,8 для различных запахов.

Возвращаясь к различным подходам к теории запахов – колебательной и ключ–замок, необходимо упомянуть и промежуточные точки зрения, основанные и на геометрии молекулы душистого вещества, и на его волновой активности. В этих теориях упор делается на то, что в момент взаимодействия душистого вещества и запахового рецептора (комплементарность определяется стереохимическими параметрами молекулы) специфичность одоранта распознается как набор волнового

воздействия, усваиваемого рецептором и передаваемого в виде электрического сигнала в центральную нервную систему в соответствии со схемой на Рисунке 5.

2.4. Молекулярный докинг

2.4.1. Моделирование биологически активных соединений, основанное на структурных параметрах лиганда и мишени

Подход 2. Возможность точного расчета в кристаллохимии сделало моделирование лигандов для определения их совместимости с молекулярной поверхностью белка более реальным процессом. Конечно, первые полученные данные по моделированию соединений (комплементарность с гемоглобином и пр.) были основаны на химической интуиции, физических моделях (в том числе шаростержневых моделях Драйдинга) и очень примитивных системах молекулярного моделирования.

Если говорить о биологически активных соединениях, т.е. наиболее развитой области молекулярного моделирования, то на сегодняшний день в этой области доминирует концепция, согласно которой причиной каждого заболевания является несбалансированная (повышенная или пониженная) активность тех или иных белков, характерных для данного заболевания, и называемых, соответственно, его терапевтическими мишенями. Считается также, что направленное действие химических соединений на соответствующие мишени (приводящее к ингибированию, или активации белка) должно приводить к положительному терапевтическому эффекту.

Соответственно, основа молекулярного моделирования направленного лекарственного действия – определение комплементарности лекарства и биологической мишени человеческого организма. К сегодняшнему дню в распоряжении исследователей имеются 3D-структуры основных мишеней, актуальным является моделирование и расчет энергии связывания предполагаемых лекарственных кандидатов с белками–мишенями известной структуры. В рамках данной концепции задача направленного синтеза состоит в том, чтобы синтезировать только те соединения, которые имеют потенциальное сродство к белку – мишени определенной болезни.

В большинстве случаев речь идет об образовании нековалентного комплекса белок–лиганд, который влияет на активность белка, т.е. необходимо предсказать

структуру и прочность такого комплекса, расположение лиганда на поверхности потенциальной энергии белка при образовании комплекса, энергетический эффект комплексообразования.

Расчет молекулы и/или комплекса молекула–рецептор, типы силовых полей:

- Для малых молекул (MM2/MM3/MM4, Tinker, Momec, Cosmos, GAFF и др.)
- Для биологических макромолекул (AMBER, GROMOS, CHARMM, OPLS, ECEPP, CVFF/CFF, др.)
- Универсальные (Tripos, MMFF, UFF)
- «Крупнозернистые» (MARTINI)

Системы, обеспечивающие расчет прочности такого комплекса и предсказание ориентации связей между молекулами с помощью дискретных (т.е. связанных с суммированием взаимодействий отдельных атомов и функциональных групп) вычислений [142]:

- AutoDock (<http://autodock.scripps.edu>)
- FlexX (<http://www.biosolveit.de/FlexX/>)
- Dock (<http://dock.compbio.ucsf.edu>)
- Surfex (<http://www.biopharmics.com>, <http://www.tripos.com>)
- Fred (<http://www.eyesopen.com/products/applications/fred.html>)
- Gold (http://www.ccdc.cam.ac.uk/products/life_sciences/gold/)
- PLANTS (<http://www.tcd.uni-konstanz.de/research/plants.php>)
- 3DPL (<http://www.chemnavigator.com/cnc/products/3dpl.asp>)
- Lead Finder (<http://www.moltech.ru>)
- Molegro Virtual Docker (<http://www.molegro.com>)
- ICM Pro (http://www.molsoft.com/icm_pro.html)
- Ligand fit, Libdock and CDocker (<http://accelrys.com/services/training/life-science/StructureBasedDesignDescription.html>)
- DockSearch (<http://www.ibmc.msk.ru>)
- eHiTS (<http://www.simbiosys.ca/ehits/index.html>)
- Glide (<http://www.schrodinger.com/productpage/14/5/>)
- DockingShop (http://vis.lbl.gov/~scrivelli/Public/silvia_page/DockingShop.html)
- HADDOCK (<http://www.nmr.chem.uu.nl/haddock/>)

Основным инструментом нашей работы в области классического докинга был разработанный в 2009 г. в нашей лаборатории (ИОХ РАН) программный комплекс Lead Finder.

Макроскопические физико-химические или биохимические свойства белков определяются комплексом их конформационного и ионизационного состояний, а не каким-либо одним экземпляром структуры белка. Из этого следует, что разработка методов подсчета молекулярной энергии не менее важна для моделирования структуры и свойств белка, чем разработка методов отбора проб для оценки различных состояний комплексных структур.

Моделирование ионизационных свойств остатков белка были предметом интенсивных исследований в течение нескольких десятилетий. Сложность такой задачи моделирования обусловлена необходимостью прогнозирования микроскопических ионизационных свойств конкретного аминокислотного остатка со всеми другими окружающими аминокислотами в некоторых фиксированных ионизационных состояниях, и усреднения этих результатов по всем возможным ионизационным состояниям окружающих.

Электростатические взаимодействия традиционно рассматривались как основные причины сдвигов ионизационных свойств аминокислотных остатков [143]. Учет электростатических взаимодействий в расчете ионизационных свойств белков был впервые предложен в работах Варшеля, который ввел микро-[144] и полумикро-[145 – 147] модели. Эти модели позволили физически оценить свободные энергии ионизируемых групп белка; однако, они требовали значительной компьютерной поддержки – особенно микромодели, которые могли считать и белок, и растворитель на атомной модели. По этой причине был разработан ряд упрощенных макроскопических подходов для неявной обработки участия растворителя, которые включали обобщенную модель Пуассона–Больцмана [148 – 150], модель Борна [151] и экранированные кулоновские взаимодействия [152].

В дополнение к физическим моделям, описывающим электростатические эффекты в белках, для расчета ионизационных свойств белков был разработан ряд методов для учета конформационной лабильности белка. Гибкая молекула белка сначала рассчитывается микро- и полумикрометодами, затем уже – различными методами расчета полной свободной энергии [146,153]. Для оценки свободных энергий

заряженных остатков белка применяется молекулярная механика, поскольку методы более высокого уровня требуют огромных вычислительных мощностей [154 – 156].

Из-за высокой вычислительной стоимости расчета обширной выборки конформационных вариаций белков был разработан ряд менее требовательных, хотя и менее точных приближений, которые рассматривали дискретный набор локально/глобально оптимизированной боковой цепи конформеров (вместо всего белка) [157,158] и расчетов методом Монте-Карло конформационного пространства и состояний ионизации белков [159,160]. Проблема выборки была недавно рассмотрена в элегантном, но ресурсоемком способе с использованием теории усреднения поля [161], который неявно учитывал различные конформационные и ионизационные состояния боковых цепей [162].

Таким образом, обширная явная выборка пространства конформационного и ионизационного состояния белка остается актуальной проблемой для практических расчетов свойств ионизации белка.

В основе метода, примененного в программном пакете Lead Finder, была разработка общего, широко применимого теоретического метода прогнозирования макроскопических свойств макромолекулярной системы путем оценки функций состояния отдельных молекулярных фрагментов системы. Основная идея вполне очевидна. Предположим, что существует система, состоящая из N узлов, взаимодействующих друг с другом, и каждый узел имеет M состояний. Общее количество состояний системы M^N . Однако, число значений, которые необходимо вычислить для вычисления функций разбиения, может быть меньше M^N в зависимости от топологии взаимодействий между частицами. Развитие этих соображений привело авторов к выводу об эквивалентности нахождения функции разбиения и концепции верификации, т.е. граф-теоретического построения, где состояние каждого узла системы определяется состояниями смежных узлов [163].

Предполагая, что белковая структура может быть аппроксимирована как система независимых взаимодействующих узлов, таких как аминокислоты, конформационная лабильность которых ограничена вращением вокруг определенной центральной оси, авторами использован метод верификации общего состояния комплекса как суммы состояний отдельных его частей.

2.4.1.1. Алгоритм TSAR

Алгоритм вычисления разбиения боковых функциональных групп белка был назван TSAR (Thermodynamic Sampling of Amino acid Residues, Термодинамическое Сэмплирование Аминокислотных Фрагментов). Алгоритм TSAR применен для расчета кривых титрования белковых аминокислот. Для выполнения энергетических расчетов использовали упрощенную версию полуэмпирического функционала молекулярной механики из программного обеспечения Lead Finder для молекулярного докинга [163].

В этом алгоритме для упрощения вычислений опущено рассмотрение влияния дальних электростатических взаимодействий, а также влияние растворителя, что снижает его ценность для расчетов докинга. Наконец, введен набор процедур для упрощения результирующего белкового графа путем устранения ребер низкой значимости и уменьшения числа состояний, которые узлы графа могут принимать, пока результирующий граф не будет достаточно простым, чтобы быть рассчитанным за реальное время.

Точность разработанного алгоритма подтверждена на подмножестве из 486 экспериментально определенных значений рКа белковых остатков, половина из которых использовалась в качестве тренировочного набора для параметризации функции скоринга, а остальная часть состояла из тестового набора. Коэффициент корреляции для расчетных и экспериментальных значений рКа составлял в среднем 0,80, от 0,95 (для тирозина) до 0,61 (для лизина). Rmsd (root mean square deviation, чем меньше его значение, тем лучше структуры совмещены в пространстве) для всех значений рКа был 0,59. Из этого следует, что прогнозная способность алгоритма TSAR успешно продемонстрирована на предсказании значений рКа, что открывает возможности для применения TSAR в других областях белкового моделирования, которые требуют расчета функций разбиения и усреднения в ансамблях отдельных белковых состояний (т.е. расчета по частям). Эти области могут включать в себя решение таких важных проблем, как докинг в полностью гибкие белки, моделирование структур мутантных белков и вычисление термической стабильности белков.

Заслуживает особого упоминания, что в рамках алгоритма TSAR возможен и

более обобщенный подход, где узел графа, представляющий положение в аминокислотной цепи белка, фактически может содержать более одной аминокислоты. В таком случае становится возможным моделирование свойств библиотек мутантных белков. В «Обсуждении результатов» рассмотрены такие расширения применения данного алгоритма.

2.4.2. Программа Lead Finder

Молекулярный докинг – хорошо разработанный метод моделирования связывания лиганд–белок, с широкой областью применения от чисто фундаментальных исследований до внедрения практически важных биологически активных соединений. В то время как значительные достижения в развитии подходов к докингу имели место на протяжении двух последних десятилетий, существующее программное обеспечение для молекулярного докинга по-прежнему не может удовлетворить потребности исследователей, особенно тех, кто связан с промышленностью. По всей вероятности, это является одной из главных причин более редкого применения компьютерного скрининга лигандов в процессе разработки новых лекарств по сравнению с экспериментальным скринингом.

Эта преимущественная синтетическая наработка веществ для получения одного, активного, и есть не устраивающий нас принцип работы в медицинской химии и противоречит идее, согласно которой первой стадией в планировании любого синтеза должно быть математическое моделирование и модельная структура для такого синтеза.

В настоящее время ключевыми направлениями в развитии молекулярного докинга являются:

- 1) Поиск оптимального представления энергии внутримолекулярных взаимодействий (так называемой скоринг–функции);
- 2) Развитие алгоритмов докинга, выполняющего глобальную оптимизацию на молекулярной электростатической поверхности (см. описание QSAR выше).

Рассмотрим первое направление. Хотя основные силы, которые вносят вклад в связывание белок–лиганд (а именно, Ван-дер-Ваальсовы, электростатические, гидрофобные, водородные связи и др.) хорошо известны, число реализаций

скоринговых функций примерно равно количеству существующих программ для молекулярного докинга, что означает необходимость поиска лучших вариантов таких реализаций [164].

Идеальная скоринговая функция должна возвращать свободную энергию связывания лиганд–белок (мишень в общем случае) и, следовательно, должна объяснять многопараметрические изменения, ассоциирующиеся с белком, лигандом и окружающей средой растворителя при таком связывании.

Эта задача не так очевидна, как одноточечные энергетические расчеты, которые составляют основу молекулярной механики (а также квантовой химии); следовательно, повышение точности скоринговой функции является серьезной научной проблемой.

Другой аспект молекулярного докинга, а именно, алгоритм, сводится к математической задаче глобальной оптимизации. Целый ряд современных математических подходов был использован в программах молекулярного докинга, таких как алгоритмы Монте-Карло (ICM [165], GlamDock [166]), генетические алгоритмы (AutoDock [167,168], Gold [169] MolDock [170]), инкрементное построение оптимального лиганда (DOCK [171], FlexX [172,173], Surfex [174]), систематический анализ возможных минимумов с использованием графов (eHits [175,176]), и алгоритмов с использованием иерархической функции скоринга для более точной оптимизации положения лиганда (QXP [177], LigandFit [178], Glide [179]).

С практической точки зрения, качество методов молекулярного докинга определяется ровно теми же параметрами, что и качество классического QSAR-прогнозирования, а именно, *процентом успешного моделирования позиции лиганда в известных (контрольная выборка) комплексах лиганд–белок.*

Существует методологическая задача превращения исследований методами докинга в более стандартные и осмысленные. К стандартам, которые могут быть унифицированы, относятся: правильное оценивание эффективности конкретного метода докинга [180], построение тестовых наборов комплексов белок–лиганд [181,182] общий доступ к данным [183], методы подготовки исходных структур для моделирования [184], стандарты сравнения результатов виртуального скрининга [185,186], построение наборов активных и неактивных лигандов для виртуальных

скрининговых экспериментов [181,187,188], стандартизация выдачи результатов исследований [181].

Однако разработчики до сих пор не начали в достаточной мере применять унифицированный набор стандартов в своих исследованиях. В частности, совершенно недостаточно внимание, уделяемое точности привязки энергетических оценок, обеспечиваемых программным обеспечением докинга. Хотя некоторые программы, такие как AutoDock и Glide XP, оценивают свободную энергию связывания белок–лиганд, обычно полагают, что прогнозирование энергии связывания – это постдокинговая процедура, и для этой цели должны использоваться более ресурсоемкие методы, такие как Free Energy Perturbation (FEP) / возмущение свободной энергии (ВСЭ); Linear Interaction Energy Approximations (линейная энергия взаимодействия приближений) и другие.

Lead Finder [189] – новое программное обеспечение молекулярного докинга, которое обеспечило ученых быстрым и точным докингом лиганда и расчетом энергии связывания. Система сочетает в себе классический генетический алгоритм с многоуровневой локальной оптимизацией. Энергетические расчеты выполняются с использованием полуэмпирической функции молекулярной механики и реализуются в нескольких дискретных формах с изменяющимися соотношениями скорости/точности. Реализуются три специализированные скоринговые функции, предназначенные для:

- распределения по ранжиру предсказанных положений лиганда,
- оценки энергии связывания в рассчитанной докированной позиции лиганда,
- распределение по степени активности индивидуальных соединений в виртуальных скрининговых экспериментах.

До настоящей работы, эффективность Lead Finder была проверена на следующих объектах:

- а) прогнозирование энергии связывания более 330 комплексов лиганд–белок и подтверждение расчетов экспериментальными измерениями;
- б) успешный докинг 407 комплексов лиганд–белок известных тестовых наборов таких программ, как FlexX, Gold, Glide, LigandFit, MolDock, и Surflex; and
- в) виртуальное определение активности 34 белковых мишеней.

В Экспериментальной части подробно описаны алгоритмы системы Lead Finder (и, соответственно, использованные нами в данной работе, см. Обсуждение результатов): а) алгоритмы докинга; б) скоринговые функции.

Определение ионизационных групп.

На первом этапе алгоритма определяются фрагменты, изменяющие или способные изменять свое состояние протонирования в процессе координации, а также фрагменты, влияющие на такое изменение. Для этого используется библиотека ионизационных свойств фрагментов *prot_ion.lib*. Во главе списка находятся фрагменты, способные проявить несколько состояний ионизации, далее – имеющие лабильные протоны (например, Ser и Thr) или так называемые флипы (меняющие торсионный угол фрагмента на 180°) (Asn, Gln). Далее – молекулы воды, образующие более одной связи с белком («консервативная» вода), при этом другие молекулы воды не рассматриваются. Все, что не вошло в рассматриваемые группы, принимается к расчетам одним блоком, с одним конформационным состоянием.

Перебор состояний и расчет парных энергий взаимодействия.

После этапа построения списка ионизационных групп следует этап сэмплирования (перебора дискретных состояний системы). Для каждой из групп формируется структурный набор возможных конформаций, в которых переменными величинами являются как лёгкие атомы водорода, так и боковые фрагменты структуры.

Для каждой из конформаций рассчитываются энергии взаимодействия с конформациями соседних групп и фрагментами, не вошедшими в группы. Для каждого ионизационного состояния отбирается 5 конформаций, наиболее энергетически выгодных в таких парных взаимодействиях. Результатом построения схемы парных энергий является граф с вершинами – ионизируемыми группами, а ребра – таблицы значений парных энергий.

Упрощение системы.

Графы минимизируются по принципу удаления связей с малыми (меньше 0,1 ккал/моль) и вершин с самыми большими значениями парных энергий взаимодействий. Если этого недостаточно для упрощения графа, из него дополнительно удаляются связи с наименьшими парными энергиями, пока мультиплетность вершин

графа не превысит 10^6 конформаций. Последовательно проводится перебор состояний вариативности вершин и ребер, т.е. нахождение оптимальной энергии системы.

Расчет полной термодинамики системы.

Отделяем от полученного графа вершины, имеющие несколько состояний протонирования. Фиксируем одно произвольно выбранное состояние протонирования и перебираем состояния остальных вершин графа, суммарно это и есть полная энергия системы. По этим значениям вычисляется заселенность состояний системы в зависимости от рН. рН, при котором заселенности протонированного и депротонированного состояний равны, считается рКа данной группы.

Энергетическая (скоринговая) функция для поиска оптимальной конфигурации боковых радикалов аминокислот.

Функция для расчета энергий состояния молекулярной системы (скоринговая функция) включает в себя электростатические взаимодействия, водородные связи, взаимодействие с металлами, штрафы на перекрытие атомов и штрафы на отсутствие водородных связей (там, где они могли быть).

Для расчета электростатических взаимодействий используется экранированный кулоновский потенциал. Взаимодействия рассчитываются по формуле

$$E_{elec,n}(r_{ij}, q_i, q_j) = \frac{q_i q_j}{D_n(h_i, b_i, r_{ij}) r_{ij}}$$

где q – точечные заряды на атомах i и j , r_{ij} – межатомное расстояние, h и b – параметры, описывающие гидрофильность и погруженность атомов, а D – функция, описывающая зависимость диэлектрической проницаемости от расстояния, точный вид которой приведен в работе [189].

При расчетах электростатического взаимодействия для стандартных атомов белка используются точечные заряды из силового поля CHARMM. Заряды на атомах лиганда определяются по алгоритму Гастайгера.

Для расчета водородных связей используется потенциал с дополнительной зависимостью энергии от взаимной ориентации донора, водорода и акцептора:

$$E_{HB} = \sum_{\substack{i \in \text{ligand} \\ j \in \text{protein}}} k(D, A) E_{HB,ij}$$

$$\text{где } E_{HB,ij} = c_{AHD,ij} \cdot c_{LP,ij} \cdot LJ_{ij}$$

c_{LP} – ориентационный коэффициент, определяющийся направлением электронных пар акцептора по отношению к атому водорода, c_{AHD} – ориентационный коэффициент, определяющийся величиной угла донор–водород–акцептор, $k(D,A)$ – коэффициент, зависящий от типа донора и акцептора (этот коэффициент можно задать в параметрическом файле), LJ – стандартный потенциал Ван-дер-Ваальсова взаимодействия. Более подробное описание расчета энергии водородных связей приведено в работе [163].

Аналогичным образом рассчитывается энергия взаимодействия акцептора с металлом:

$$\Delta G_{Me} = k_{Me} \sum_{\substack{i=\text{ligand} \\ j=\text{metal}}} \alpha_{i,j} LJ_{ij}(r_{ij})$$

Для расчета энергии перекрывания атомов используется модифицированный потенциал Ван-дер-Ваальса, в котором удалены области, соответствующие отрицательным значениям энергии. В расчетах используются параметры Ван-дер-Ваальсова взаимодействия из силового поля CHARMM.

Избыточная энергия при отсутствии водородных связей доноров и акцепторов рассчитывается по следующей формуле:

$$\Delta E_{HB,lig-penalty} = \sum_{\substack{i \in \text{ligand} \\ i \in D,A}} (hb_p - f \cdot hb_w)$$

Где hb_p – количество водородных связей, образованных с атомами белка, hb_w – количество водородных связей, образуемых с водой. При расчете потерянных водородных связей для каждой потенциальной связи определяется наиболее выгодная позиция молекулы воды, которая могла бы эту связь образовать. Если молекула воды не может располагаться в такой позиции из-за перекрывания с атомами белка, то hb_p увеличивается на величину, специфичную для каждой группы. Коэффициент масштабирования избыточной энергии можно задать для каждой группы в параметрическом файле.

В чем же коренное отличие метода алгоритмизации, примененного в системе Lead Finder, от других методов прогнозирования биологической активности?

В общем плане:

- в моделировании структуры белка,
- в алгоритмизации,

В рамках общей терминологии:

- в определении дескрипторных особенностей структуры лиганда и мишени.

Сравнение алгоритма моделирования, примененного в системе Lead Finder, с известными алгоритмами.

В Таблице 4 приведены используемые в настоящее время программы молекулярного моделирования (как коммерческие, так и открытые).

Таблица 4. Широкоизвестные алгоритмы и компьютерные программы для моделирования структуры белка

Компьютерная программа/алгоритм	Принципы алгоритма моделирования структуры белка (боковых радикалов и петель)	Общий вид скоринговой функции для энергетических оценок	Примерное время расчетов на персональном компьютере
Prime [190,191] (Schrodinger Inc.)	Итеративный перебор множества допустимых состояний системы (конформаций боковых функциональных групп и петель) с пошаговым уточнением допустимых границ при помощи оригинального протокола локальной оптимизации с фиксированными внутренними и декартовыми координатами. Боковые функциональные группы оптимизируются последовательно и итеративно. Полная	Белок задается полной атомной моделью в явном виде; учитываются соседние фрагменты. Для расчета энергии межатомных взаимодействий используется поле OPLS AA. Растворитель учитывается неявно по обобщенной Борновской модели (для полярной составляющей сольватации); с пересчитываемыми пара-	Несколько минут на один полный цикл оптимизации – от 10 до нескольких сотен часов на моделирование петли из 4–12 остатков.

	градиентная оптимизация определенного числа минимальных по энергии состояний системы.	метрами и оригинальной модели для неполярной составляющей.	
	Точность моделирования: достаточно корректно (с точностью до 40° в конформации основной цепи) моделируется от 90% четырехзвенных петель из опубликованного тестового набора, до 60% двенадцатизвенных петель. Порядка 90% смоделированных боковых радикалов имеют достаточно корректный (с точностью до 40°) первый двугранный угол, и порядка 75% – первые два угла.		
ChiRotor [192] (Accelrys Inc.)	Конструируются стабильные конформации каждого аминокислотного остатка, осуществляется полная градиентная минимизация энергетики конформаций с учетом только атомов основной цепи белка, и отбирается лучшая конформация. На втором этапе минимизации в расчет включаются все боковые функциональные группы, и проводится несколько последовательных итераций оптимизаций каждого из остатков.	Белок задается моделью в явном виде, неполярные атомы водорода объединены в одну группу. Межатомные взаимодействия рассчитываются в поле CHARMM19. Растворитель учитывается неявно по обобщенной Борновской модели с фиксированными параметрами.	До нескольких минут для белка среднего размера.
	Точность моделирования: RMSD 0,77Å для погруженных остатков, RMSD 1,73Å для всех остатков		
Andante [193] (Tropos)	Конформации каждого из остатков генерируются в соответствии с вероятностным	В расчетах используется эмпирическая ско-	Нет данных

Inc.)	распределением, полученным из статистического анализа конформаций базы данных по белкам. Конечное состояние системы оптимизируется по методу Монте Карло.	ринговая функция, содержащая отталкивательные слагаемые (Ван-дер-Ваальсово перекрытие с основной цепью и боковым функционалом), а также вклад от статистического вклада конформаций боковых групп.	
Точность моделирования: RMSD 1,66 Å для всех остатков			
ICM [194,195, 196] (MolSoft LLC)	Глобальная оптимизация осуществляется с помощью метода Монте Карло во внутренних координатах со смещенными (по оригинальной методике) вероятностями сэмпирования определенных конформаций.	Белок задается полноатомной моделью в явном виде. Межатомные взаимодействия рассчитываются с использованием силового поля MMFF94. Растворитель учитывается неявно по оригинальной модели MIMEL (модифицированный электростатический потенциал отображения), аппроксимирующей решения Пуассона–Больцмана.	Нет данных
Точность моделирования: RMSD ~1,5 Å для петель до 5 звеньев, ~2Å для петель до 6 звеньев, ~4 Å для петель до 8 звеньев, ~6,5 Å для петель до 12 звеньев.			

SCWRL [197,198] (Institute for Cancer Research, USA)	Стадии алгоритма: 1) минимизация числа состояний функциональных групп с помощью теорем тупиковых ветвей эволюции и вероятности нахождения данной конформации в базе конформеров. 2) поиск оптимального состояния системы разбиением системы на двухсвязные группы и явного поиска минимума как каждой из групп, так и системы в целом.	Белок задается координатами тяжелых атомов, (атомы водорода и растворитель не учитываются). Используется эмпирическое силовое поле: статистический вес данного ротамера (в библиотеке ротамеров) суммируется со стерическими перекрытиями с атомами основной цепи и других боковых групп.	Несколько минут на белок среднего размера.
	Точность моделирования: для 85% смоделированных боковых групп корректен первый двугранный угол, и для 73% – первые два угла.		

2.5. Объекты применения классического молекулярного докинга

Объекты применения молекулярного докинга, как наиболее точного метода прогнозирования биологической активности – это вся медицинская химия. Поэтому в данном обзоре ограничимся тематикой собственных исследований и, соответственно, прогнозированием биологической активности исследованных объектов – ингибиторов PARP1 и Суk-киназ, а также механизмами и прогнозированием продуктов каталитических реакций.

2.5.1. Ингибиторы поли(ADP-рибозо)полимеразы (PARP1)

Ингибиторы поли(ADP-рибозо)полимеразы (PARP1), инициирующей восстановительные процессы раковых клеток – противоопухолевые препараты, от которых отказались многие исследователи в 2011 году. Ингибирование PARP1 вызывает саморазрушение раковых клеток, тогда как активные PARP1 и PARP2 облегчают восстановление их ДНК. PARP1 распознают и связывают одонитевые

разрывы в ДНК и катализируют образование разветвленных цепей поли(ADP-рибозы), которые в свою очередь активируют ферменты репарации ДНК раковых клеток в местах повреждения.

В последние 2–3 года ингибиторы ПАРП1 возвращены в орбиту разработки лекарств против рака яичников и молочной железы, и полученные вещества проходят доклинические испытания.

Фирма Санофи, как родоначальник исследований ингибиторов ПАРП1, признала неэффективность своего препарата Инипариб. AstraZeneca, выпускающая препарат Олапариб, возрождает программу по данному виду рака. Еще три компании – в процессе медицинских испытаний собственных препаратов (Таблица 5). Однако, небольшие дефекты в результатах всех имеющихся клинических испытаний могут перечеркнуть медицинское применение препаратов.

Таблица 5. Ингибиторы ПАРП1, находящиеся на стадиях внедрения

Разработчик	Препарат	Вид активности и стадия испытаний (клинических)
AstraZeneca KuDOS Pharmaceuticals acquisition)	Олапариб (AZD-2281)	Рак яичников с мутациями BRCA Поддерживающая терапия
Tesaro (по лицензии Merck)	Нирапариб (MK4827)	Ген рака груди, серозный рак яичников. Ожидает решения
BioMarin Pharmaceuticals	BMN 673	Метастатический рак молочной железы
Clovis Oncology (licensed from Pfizer)	Рукапариб (CO288)	Ген рака груди, серозный рак яичников
AbbVie	Велипариб (ABT-888)	Состояние внедрения неизвестно

Начиная с 1990-х годов, компании начали разрабатывать ингибиторы ПАРП1 в качестве хемотабilizаторов, поскольку многие химиотерапевтические препараты вызывают повреждение ДНК, которое восстанавливается активным ПАРП1, что приводит к лекарственной сопротивляемости. Но, хотя ингибиторы ПАРП1 проявляли противоопухолевую активность, они оказались высокотоксичными в комбинации с химиотерапией.

Много позже, в 2005 году, две прорывных работы показали, что ингибиторы ПАРП1 самостоятельно могут ликвидировать раковые клетки с уже восстановленными ДНК, в частности, могут быть эффективными против мутаций рака молочной железы 1 (BRCA1) и BRCA2 [199].

Убийство клеток, которое происходит после объединения двух условий (известно как синтетическая/синергетическая летальность), само по себе недостаточное условие влияния на жизнеспособность раковых клеток. Однако, эксперименты показали, что ингибирование ПАРП1 и мутантные BRCA были синтетически летальными в доклинических моделях, что вылилось в целенаправленный и минимально токсичный способ лечения данных типов рака.

Что мы имеем на данный момент?

Несколько разновидностей ингибиторов ПАРП1 (см. Таблицу 5). Инипариб, не универсален и токсичен [200,201]. Остальные пока под вопросом. Перспективен, по всем параметрам, Велипариб. VMN 673 является более мощным средством, чем остальные, но это не обязательно трансформируется в клиническое преимущество. Как и другие ингибиторы ПАРП1, его токсичность, ограничивающая эффективную лекарственную дозу, является существенным препятствием к внедрению в медицинскую практику.

Синергетические смеси имеют существенный разброс в активности. Хотя все они нацелены на ПАРП1 и ПАРП1, они имеют различную активность по отношению к 15 другим членам группы ПАРП1, а некоторые из которых проявляют новые функции. Побочная токсичность (вне действия мишени) также варьируется в широких пределах. Поэтому три основных компании – AstraZeneca, Tesaro и Clovis – тестируют свои лекарственные кандидаты в режиме одиночных ингибиторов для поддерживающей терапии с целью отсрочки или предотвращения рецидива болезни после химиотерапии. Хотя 50–60% пациентов с серозным раком яичников излечимы

первичной химиотерапией, однако большинство в результате попадают на рецидивы; а рецидивы рака яичников неизлечимы.

AbbVie, которая еще не раскрыла свои планы фазы III медицинских испытаний для своего кандидата Veliparib, позиционирует агент для комбинированной терапии, возможно, из-за его более низкой эффективности. Основная проблема – в получении разрешения американского федерального бюро по медицинским препаратам (FDA).

Такое подробное объяснение последовательности внедрения ингибиторов ПАРП1 приведено для оценки последующих достижений в этой области в результате оптимального планирования синтетического эксперимента, т.е. предварительного моделирования направленного синтеза с применением системы Lead Finder, а также усовершенствованной версии с применением FEP (BCЭ).

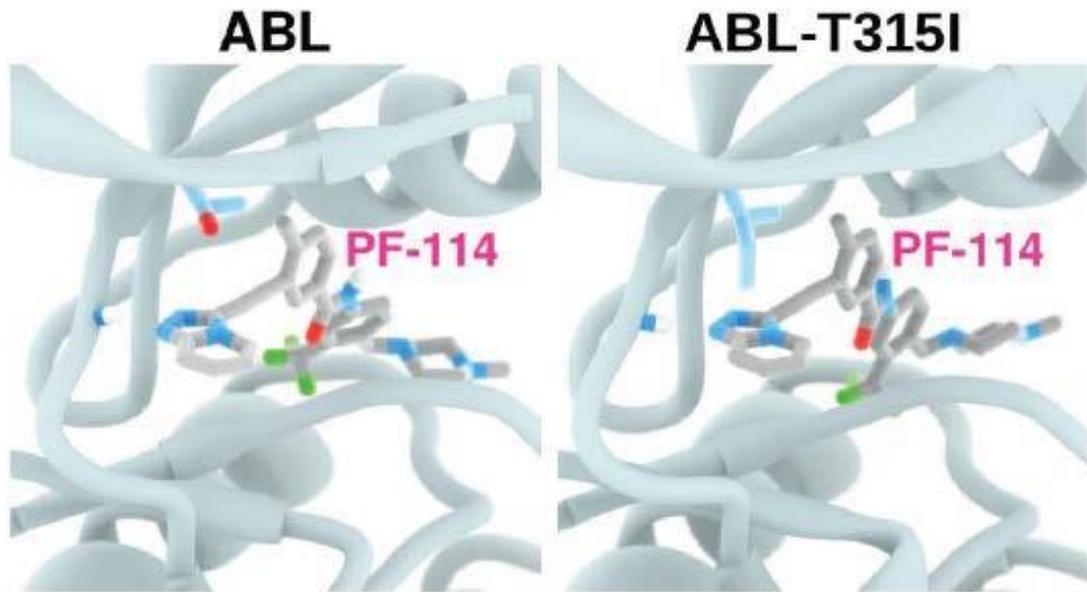
По результатам докинга (Lead Finder) и последующего направленного синтеза удалось добиться одобрения FDA в процессе медицинских испытаний ингибитора киназы ABL, способствующей увеличению резистивности к химиотерапии миелолейкоза.



Селективный ингибитор тирозинкиназ PF-114 (Рисунок 7), предназначенный для лечения резистентных форм хронического миелолейкоза [202]. Новый лекарственный кандидат ингибирует T315I и другие мутантные формы BCR/ABL, вызывающие резистентность по отношению к существующим препаратам. PF-114

высокоселективен к данным киназам, поэтому практически не дает побочных эффектов.

Рисунок 7. Визуализация комплекса ингибитора PF-114 и киназ ABL



2.5.2. Ингибиторы Syk-киназы

Syk-киназа играет одну из ключевых ролей в иммунном ответе и, в частности, в аллергических реакциях. Она является одним из ключевых медиаторов активации иммунной системы в ответ на введение иммуногена в организм. Показано, что ингибиторы Syk-киназы являются эффективными в различных аутоиммунных заболеваниях, таких как ревматоидный артрит. Кроме того, поскольку Syk-киназа является критичной для созревания В-лимфоцитов, применение её ингибиторов оправдано в лечении заболеваний кроветворной системы, связанных с неконтролируемым ростом данного вида клеток. Показана эффективность ингибиторов Syk-киназы для лечения неходжкинской лимфомы. Таким образом, выбранная в качестве объекта моделирования Syk-киназа является мишенью противоревматоидной и противораковой терапии.

В настоящее время поиском новых ингибиторов Syk-киназы занимается более тридцати коммерческих компаний, наиболее активные из них – Rigel, Astra Zeneca и Portola. Более ста патентов подано только на один класс ингибиторов Syk-киназы – 2,4-диаминопиримидины.

Согласно результатам опубликованных исследований, ингибиторы Syk-киназы 2,4-бис(3-гидроксифениламино)-5-фторпиримидин и [6-({5-фтор-2-[(3,4,5-триметоксифенил)амино]пиримидин-4-ил}амино)-2,2-диметил-3-оксо-2,3-дигидро-4Н-пирид[3,2-*b*][1,4]оксазин-4-ил]метилфосфорная кислота (фостаматиниб), для которых проводились клинические испытания, имеют достаточно неплохие характеристики биологической активности. Тем не менее, существует достаточно большое количество соединений, потенциально обладающих лучшими терапевтическими показателями (эффективность, селективность).

Анализ литературы по предложенным модификациям фостаматиниба (ингибиторы на основе 2,4-диаминопиридина, 2,4-диаминотриазина, 2-аминохиназолина, 4,6-диаминопиримидина, 6-амино-1*H*-пиразоло[4,3-*c*]пиридина) показал, что во многих случаях близкие по структуре соединения обладают ингибиторной активностью в отношении киназ, однако ни одно из производных указанных соединений не защищено патентами.

По результатам проведённых исследований представляется целесообразным проведение НИР по разработке новых ингибиторов Syk-киназы. Разрабатываемые ингибиторы содержат циклы 2,4-диаминопиридина, 2,4-диаминотриазина, 2-аминохиназолина, 4,6-диаминопиримидина и 6-амино-1*H*-пиразоло[4,3-*c*]пиридина, которые позволяют добиться улучшения тех или иных характеристик веществ (эффективность ингибирования, селективность) [203].

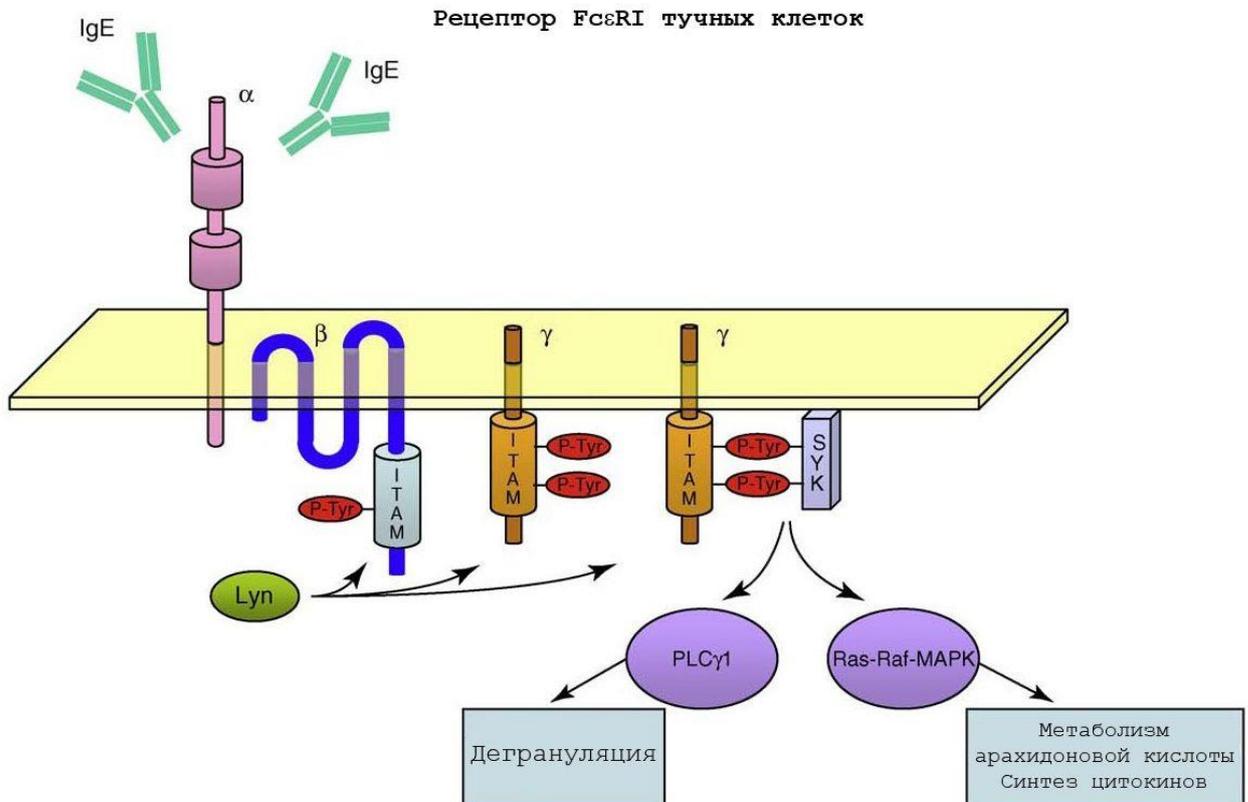
Биологические функции Syk-киназы

Syk-киназа является нерецепторной тирозинкиназой и участвует в передаче сигнала антигенными и Fc рецепторами. Syk содержит тандемные SH2 (Src Homology 2) домены, которые специфически связываются с фосфорилированным тирозиновым активирующим агентом (ITAM, Immunoreceptor Tyrosine-Based Activation Motif), присутствующим в цитоплазматическом регионе иммунорецепторов (например, Fc рецепторов, Т- и В-клеточных рецепторов, и естественных киллеров (NK, Natural Killers).

Syk киназа экспрессируется в гематопоэтических клетках (таких как макрофаги, тучные клетки, лейкоциты, тромбоциты и эритроциты), в меньшей степени – в эпителиальных клетках, фибробластах, нейрональных клетках, гепатоцитах и т.д.

Syk-киназа играет одну из ключевых ролей в иммунном ответе и, в частности, в аллергических реакциях. При аллергических патологиях происходит продуцирование большого количества иммуноглобулина Е (IgE) в ответ на воздействие антигена. На поверхности тучных клеток и базофилов находится $F_{\epsilon}R1$ рецепторы, способные связываться с IgE. Связывание IgE с внешнеклеточным доменом рецептора приводит к агрегации внутриклеточного домена, последующей активации цитоплазматической Src-киназы, Lyn, и фосфорилированию ITAM домена рецептора этой киназой. Затем связывание SH2-доменов Syk-киназы с фосфорилированным ITAM-доменом активирует киназную активность Syk, которая, в свою очередь, запускает клеточный каскад, приводящий к синтезу и высвобождению большого количества модуляторов воспаления, приводящих к острым аллергическим реакциям (Рисунок 8). Аналогичным образом Syk участвует в активации В и Т клеток, индуцированной связыванием BCR (B-Cell Receptor) и TCR (T-Cell Receptor), соответственно, с антигенами [204].

Рисунок 8. Каскад Syk-киназы



Ингибирование Syk-киназы в различных патологиях

В настоящее время достоверно установлена эффективность ингибирования Syk-киназы в двух заболеваниях – ревматоидном артрите и неходжкинской лимфоме. Ревматоидный артрит – это хроническое воспалительное заболевание периферических суставов, характеризующееся экспрессией аутоантител, синовиальным воспалением и эрозией хрящей и костей [205,206]. В настоящее время считается, что ревматоидный артрит запускается активацией врожденного иммунитета, инфекцией или факторами среды и затем поддерживается постоянной выработкой аутоантител. Связывание аутоантител с рецепторами на поверхности иммунокомпетентных клеток приводит к развитию воспалительного ответа, опосредованного Syk-киназой [207].

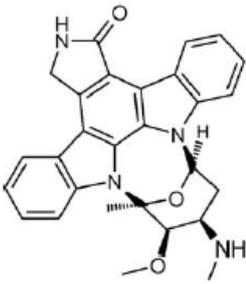
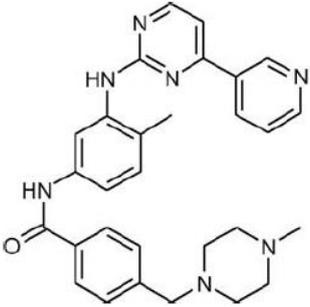
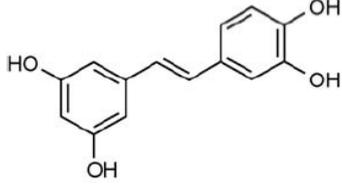
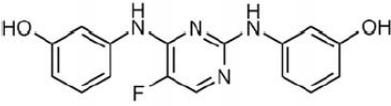
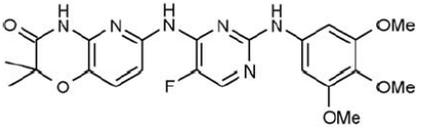
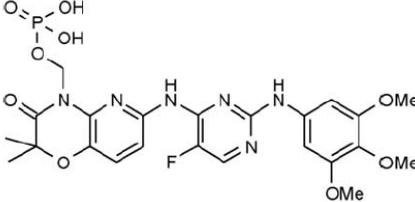
Неходжкинская лимфома – это гетерогенная группа злокачественных новообразований, состоящих главным образом из зрелых В-клеток, экспрессирующих BCR. BCR состоит из константной внешнеклеточной антиген–связывающей субъединицы (IgM), связанной с двумя инвариантными цитоплазматическими сигнальными субъединицами, Ig α и Ig β . Эти субъединицы подвергаются фосфорилированию киназами семейства Src по ITAM-домену после агрегации BCR, вызванной связыванием антигена. Затем Syk-киназа связывается с ITAM-доменом, активируется и фосфорилирует различные мишени, включая BLNK и PLC γ ₂, что приводит к дальнейшему распространению BCR-каскада. Эксперименты показывают необходимость Syk-киназы для поддержания зрелых В-клеток. У мышей, не имеющих гена Syk-киназы, происходит практически полная блокировка пре-про-В-клеточного этапа развития В-клеток. С помощью технологии shRNA было показано, что нарушение экспрессии Syk-киназы в мышинной модели неходжкинской лимфомы (НХЛ) приводит к блокированию развития опухолевых В-клеток и нейтрализует их канцерогенность при имплантации здоровым животным [208].

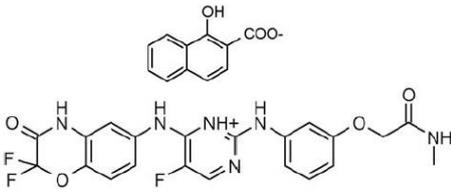
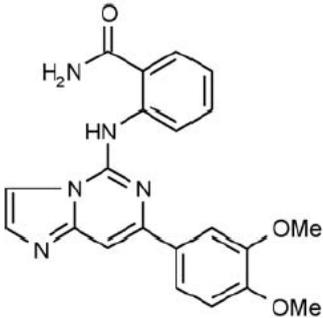
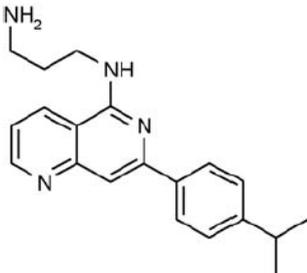
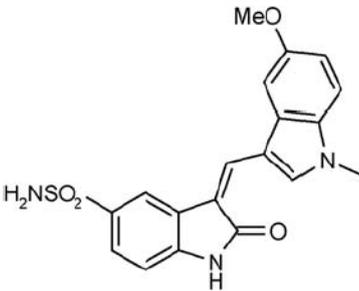
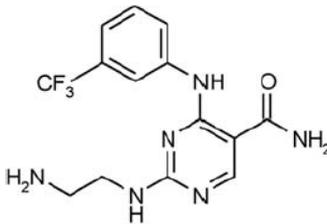
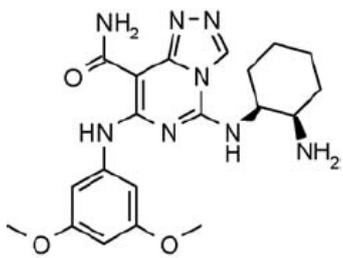
Существующие ингибиторы Syk-киназы

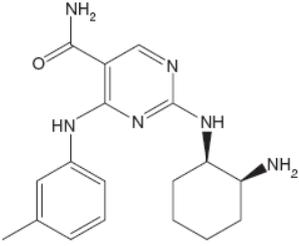
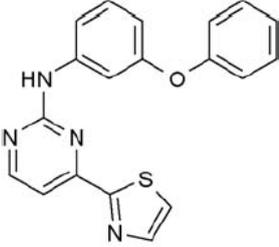
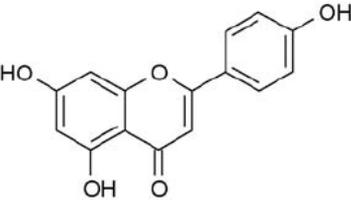
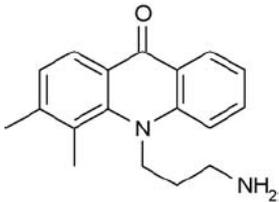
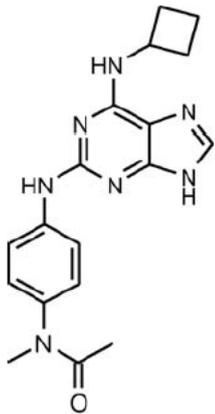
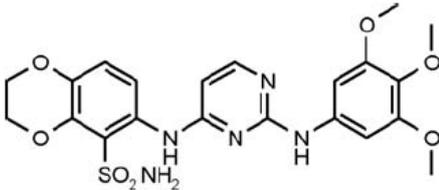
Установление функции Syk-киназы в развитии аутоиммунных заболеваний и лимфом привело к активному поиску её селективных ингибиторов. На сегодняшний день известно более ста низкомолекулярных веществ, способных ингибировать Syk [209]. Действие практически всех этих веществ основано на их высоком сродстве к

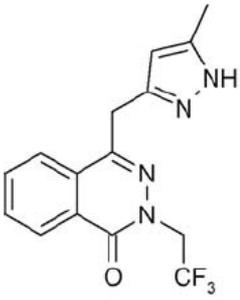
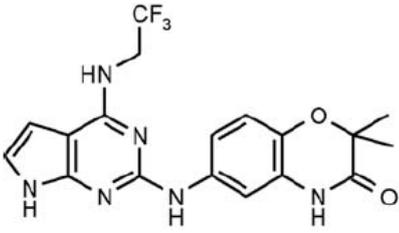
АТФ-связывающему домену киназы. Структуры представителей основных классов ингибиторов СуК-киназы приведены в Таблице 6.

Таблица 6. Структуры известных ингибиторов СуК-киназы

	Название	Компания разработчик	Структура	IC ₅₀ , нМ	Источник
1	Стауро-спорин	Природное соединение	 99	13	[210]
2	Иматиниб	Novartis	 100	5000	[211]
3	Пицеатаннол	Природное соединение	 101	10000	[212,213]
4	R112	Rigel	 102	226	[214,215, 216,217]
5	R406	Rigel	 103	41	[218]
6	R788	Rigel	 104	30	[219,220, 221,222]

7	R343	Pfizer	 <p style="text-align: right;">105</p>	?	[223]
8	BAY61-3606	Bayer	 <p style="text-align: right;">106</p>	7.5	[224,225]
9		Boehringer Ingelheim	 <p style="text-align: right;">107</p>	90	[226,227]
10	–	Aventis	 <p style="text-align: right;">108</p>	5	[228]
11	–	Astellas	 <p style="text-align: right;">109</p>	41	[229]
12	–	Kissei	 <p style="text-align: right;">110</p>	9	[230,231,232]

13	YM193306	Astellas	 <p style="text-align: right;">111</p>	5–20	[229]
14	–	Vertex	 <p style="text-align: right;">112</p>	630	[233,234]
15	Апигенин	Bayer Ya- kuhin	 <p style="text-align: right;">113</p>	4200	[235]
16	ER-27319	Eisai	 <p style="text-align: right;">114</p>	10	[236]
17	NVP- QAB205- AA	Novartis	 <p style="text-align: right;">115</p>	10	[237]
18	–	Novartis	 <p style="text-align: right;">116</p>	–	[238]

19	–	Hoffmann-La Roche	 <p style="text-align: right;">117</p>	–	[239]
20	–	Glaxo Group Limited	 <p style="text-align: right;">118</p>	–	[240]

Большинство найденных ингибиторов Syk-киназы относятся к одному из следующих классов:

- 2-аминопиридины и 2-аминопиримидины
- 5-пиридинкарбоксамиды
- Флавоноиды
- Оксиндолы

Первые три структуры, представленные в Таблице 6 (стауроспорин, иматиниб и пицеатаннол), имеют ряд недостатков и явились стартовыми для дальнейшей разработки более селективных и эффективных ингибиторов. Так, высокая прочность связывания стауроспорина нивелируется его крайне низкой селективностью (он связывается практически со всеми киназами), в то время как прочность связывания иматиниба и пицеатаннола недостаточна для их терапевтического применения.

Наибольший интерес из существующих ингибиторов Syk-киназы представляют вещества R112, R343, R406 и R788, находящиеся на различных этапах клинических исследований. Эти структурно схожие производные 2,4-диаминопиримидина были разработаны компанией Rigel Pharmaceuticals, Inc. Первый представитель этого семейства, R112, связывается с Syk с $K_i=96$ нМ [213]. Улучшение оральной биодоступности было достигнуто в соединении R406 и его пролекарственной форме, R788 [219]. Стоит отметить, что R788 в организме практически мгновенно распадается, и действующим веществом, по сути, является R406.

R343 был разработан на основе R112 в качестве удобной для ингаляции формы [231]. В настоящее время R788 проходит клинические испытания в ревматоидном артрите (II фаза) [43] и неходжкинской лимфоме (II фаза) [241].

Другой обширный класс ингибиторов Syk-киназы составляют 5-пиримидин-карбоксамиды. Вещество **109** в Таблице 8, принадлежащее этому классу, было найдено с помощью высокопроизводительного скрининга. Оно достаточно хорошо ингибирует ферментативную активность Syk с IC_{50} 41 нМ, и обладает высокой селективностью. Для исследования антиаллергической активности этого соединения была изучена его способность ингибировать пассивную кожную анафилактическую реакцию у мышей. Ингибирование являлось дозозависимым с IC_{50} 13,2 мг/кг [229].

Ингибиторы Syk-киназы, являющиеся производными 5-карбоксамидов, исследовались и другими учёными. Так, введение *cis*-циклогексилдиаминогруппы привело к созданию соединения **111**, ингибирующего Syk с IC_{50} 5–20 нМ [29]. На основании той же стартовой структуры были разработаны производные 1,2,4-триазоло[4,3-с]пиримидина и 1,2,4-триазоло[1,5-с]пиримидина, обладающие высокой эффективностью *in vitro*. Например, соединение **110** имеет IC_{50} 9 нМ [229].

Высокопроизводительный скрининг позволил идентифицировать новый структурный класс ингибиторов – оксииндольный (структура 10, IC_{50} 5 нМ). Тем не менее, EC_{50} на клетках оказалось достаточно высоким (1400 нМ), что связано с очень большой полярной площадью поверхности соединения. Для увеличения клеточной активности сульфамидная группа была заменена на значительно менее полярную метиламидную [227].

Флавоноиды, а особенно флавоны, являются хорошо известными ингибиторами киназ. В исследовании, опубликованном в 2003 году Bayer Yakuhin совместно с Gifu Pharmaceuticals University [234], была отмечена их активность и в отношении Syk-киназы и дегрануляции тучных клеток. Эффективное ингибирование было обнаружено для 7-гидроксибензопирона с карбонильной группой в 4 положении и 3- или 5-гидроксильной группой в 2-фенильном кольце. Более того, ингибиторная активность исследованных флавоноидов относительно других типов киназ (а именно PKC и PI3K) была существенно ниже, что говорит о некоторой селективности по отношению к Syk.

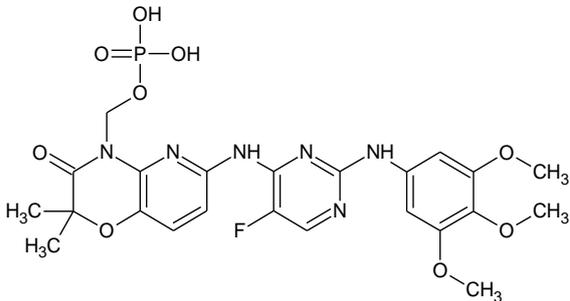
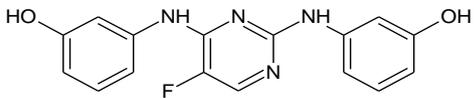
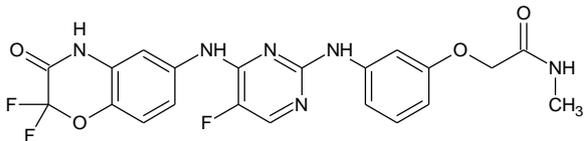
Проведённый анализ выявил 4 основных класса ингибиторов Syk-киназы,

представители которых обладают активностью *in vitro* (на ферменте и на клетках): 2-аминопиридины и 2-аминопиримидины, 5-пиридинкарбоксамиды, флавоноиды и оксиндолы. Флавоноиды являются наименее привлекательным классом ингибиторов в силу невысокой прочности связывания с мишенью, а также низкой оральной биодоступности [242]. Для анализа терапевтической эффективности ингибиторов *Syk*-киназы была дополнительно собрана информация по клиническим испытаниям этих соединений.

Клинические испытания известных ингибиторов *Syk*-киназы

Клинические испытания проводились для 3 ингибиторов *Syk*-киназы: R788 (фостаматиниб), R112 и R343 (Таблица 7). Эти соединения были разработаны компанией Rigel Pharmaceuticals, Inc. Клинические испытания R112 были прекращены, так как вторая фаза не показала эффективности препарата по сравнению с плацебо.

Таблица 7. Ингибиторы *Syk*-киназы, проходящие клинические испытания.

Название	Структура	K_i , нМ	Болезнь
R788 (фоста- матиниб)		30 [218]	Ревматоидный артрит (II фаза) Неходжкинская лимфома (II фаза) [243]
R112		96 [206]	Аллергические риниты (испытания остановлены после II фазы)
R343		–	Астма (I фаза) [243]

Экспериментально определенная клиническая эффективность существующих ингибиторов Syk-киназ.

Изучение клинической эффективности фостаматиниба при ревматоидном артрите (РА) проводилось в течение 6 месяцев на выборке из 457 пациентов, предварительно проходивших продолжительное лечение метотрексатом.

Проведённые исследования показали, что ингибиторы Syk-киназы могут применяться для лечения ряда заболеваний, включающих различные аутоиммунные заболевания (ревматоидный артрит, системная красная волчанка), а также неходжкинскую лимфому. В настоящее время существует порядка ста ингибиторов Syk-киназы, большинство из которых относится к одному из 4 классов химических соединений. Тем не менее, клинические испытания в настоящее время проводятся только для одной группы соединений – производных 2,4-диаминопиримидина R343 и R788. Соединение R112, относящееся к той же группе, оказалось неэффективным при аллергическом рините. Таким образом, наиболее перспективными ингибиторами Syk-киназы являются R343 и R788, и наиболее вероятно их терапевтическое применение в ревматоидном артрите и неходжкинской лимфоме.

2.6. Преодоление принципиальных ограничений классического докинга

Очевидные ограничения докинга в классическом варианте – отсутствие учета влияния удаленных элементов структуры; высокой лабильности молекул в активном центре, учета молекул, не относящихся к координирующимся субстратам (например, растворителя).

Для решения данной проблемы необходим достаточно точный метод, который, вместе с тем, обладает приемлемой вычислительной производительностью. Применение методов QSAR или молекулярного докинга не дает достаточно точных оценок относительного сродства ингибитора. С другой стороны, использование методов квантовой химии для обсчета серий соединений на массивах сотен белков было бы невозможно в силу вычислительных ограничений.

В отличие от методов экспериментально-эмпирического скрининга или статистических методов QSAR, математика точного определения энергии взаимодействия прямо зависит от точности описания комплекса лиганд–белок. Как

уже сказано выше, такие методы – Free Energy Perturbation (FEP), (BCЭ – возмущение свободной энергии) [244,245], Linear-response approximation (аппроксимация линейного отклика) [246] и совместные ММ–расчеты с учетом сольватации (Molecular Mechanics/Poisson–Boltzmann surface area) [247].

2.6.1. FEP (Free Energy Perturbation) / BCЭ (Возмущение свободной энергии)

Необходимость учета посторонних влияний приводит к необходимости расчета интегральных энергетических параметров – флуктуаций общей энергии системы в зависимости от лабильности и положения координирующихся субстратов, наличия и вида растворителя, наличия и участия в процессе других молекул либо удаленных фрагментов активных субстратов [248].

Наиболее точным методом расчёта биокаталитической системы, а именно, энергии связывания лигандов с терапевтическими мишенями, является метод возмущения свободной энергии (FEP/BCЭ) [249], основанный на молекулярно-механическом описании системы лиганд–белок с явным учётом молекул растворителя. Этот метод позволяет учесть ряд факторов, таких как лабильность мишени, конформационное равновесие лиганда, сольватацию комплекса, т.е. то, что не входит в сферу применения полуэмпирических и неэмпирических методов, что позволяет считать BCЭ наиболее эффективным методом оптимизации структуры (моделирования) лекарственных кандидатов [250].

Метод BCЭ строго математически обоснован, и при проведении моделирования активных структур дает точное значение динамического изменения свободной энергии в процессе связывания лиганда с белком. Однако, исходные параметры для определения BCЭ формируются с помощью молекулярной динамики, т.е. точность моделирования определяется точностью исходных молекулярно-динамических расчетов, которые зависят от корректности силового поля ММ и частоты дискретизации (качества сэмпирования) структурного образования. А эти параметры, основанные на фиксированных базовых значениях, далеко не всегда точны сами по себе.

В статистическом методе, QSAR, мы наблюдали аналогичный эффект – качество сэмпирования в QSAR определяет точность сравнения поверхностей

молекулярного электростатического поля молекул обучающей и контрольной выборок.

Таким образом, и в классическом QSAR, и в классическом докинге параметр «качество» четко зависит от преобразования непрерывной функции в дискретную, т.е. от точности описания поверхности второго рода 1D (точками), 2D (отрезками и кривыми, описываемыми одной функцией), 3D (трехмерный граф). Последний вариант – по определению, самый точный, и, в случае переменной плотности описывающих точек (увеличение плотности описательных поверхностных точек при увеличении значения производной, т.е. отклонения сложной поверхности от плоскости), может, как и в случае QSAR, давать максимальное качество сэмпирования.

Следовательно, задача увеличения точности представления комплекса лиганд–мишень прямым образом зависит от точности представления геометрии комплекса (т.е. перехода от точечного, дискретного описания к поверхностным интегралам 2-го рода), непрерывности описания геометрии и свойств полей этого комплекса, в первую очередь электростатического.

Существует существенный дополнительный фактор, влияющий на динамику и энергию связывания системы лиганд–мишень – это молекулы воды, непосредственно контактирующие с комплексом, образующие с ним водородные связи и изменяющие его геометрию. Соответственно, неточный расчет мостиковых молекул воды критичен, и может приводить к погрешностям порядка нескольких ккал/моль, что неприемлемо при дизайне лекарств.

Для решения проблемы влияния молекул растворителя расчетчики пошли традиционным дискретным способом – каждую молекулу воды считать отдельно. Систематический же интегральный анализ координации воды в комплексе так и не был проведен. А такой анализ позволил бы уточнить расчеты свободной энергии и определить конформационную вариативность мостиковых молекул, кроме того, обозначить активные графы связности (сайты сольватации) и (основная задача) определить, как надстроить лиганд для увеличения энергии связывания с мишенью. Таким образом, динамический график сольватации комплекса белок–лиганд от времени является актуальной проблемой.

Замедленная во времени конформационная лабильность свободного лиганда

в окружении молекул воды также может приводить к недостаточной точности сэмплирования. Если не учесть конформационную лабильность лиганда, можно получить ошибку в определении энергии связывания в несколько ккал/моль.

К этому добавляется проблема заторможенных конформаций белка-мишени, когда появляется еще одна переменная в уравнении энергии связывания. Решению обеих проблем посвящены методы двухтопологического (одновременно рассчитываемого, зависимого) ВСЭ, потенциала средней силы и методов общего уточнения сэмплирования системы.

Решение проблемы лишних переменных – в увеличении вычислительной мощности компьютера, а это, как уже было сказано, путь тупиковый. Все методы, пытавшиеся снизить ошибку определения свободной энергии связывания (т.е. учесть и конформационную составляющую, и участие растворителя в образовании комплекса лиганд–мишень), отличаются низким быстродействием (в сочетании с имеющимися техническими возможностями). Этот факт обуславливает необходимость конструирования эффективного (во всех смыслах) метода молекулярного моделирования, учитывающего конформационное равновесие при образовании комплекса, воздействия растворителя, а также исходную лабильность – как лиганда, так и мишени.

Обычно ВСЭ рассчитывает относительное сродство лигандов L и L' – моделирование перехода несвязанного L в связанный L' относительно воды ($\Delta G_{aq}(L \rightarrow L')$) и белка ($\Delta G_{protein}(L \rightarrow L')$). Полученные значения вписываются в термодинамический цикл:

$$\Delta \Delta G_{связ}(L \rightarrow L') = \Delta G_{связ}(L') - \Delta G_{связ}(L) = \Delta G_{protein}(L \rightarrow L') - \Delta G_{aq}(L \rightarrow L') [250].$$

Этот подход был успешно применен при моделировании ингибиторов ВИЧ-1 транскриптазы [251,252], фосфолипазы A2 [253] и p38 α MAP киназы [254], теоретических исследованиях ингибиторов катехол-O-метилтрансферазы (catechol-O-methyl transferase (COMT)) [255], ингибиторов нейраминидазы гриппа N1 [256], ингибиторов глюкозидазы [257] и т. д., и в настоящее время признается в качестве стандартного алгоритма подобных расчетов.

При расчетах возникает еще одна проблема: *конформация связанного в комплекс лиганда обычно не является самой низкой энергетической конформацией несвязанного лиганда.* Процесс расчета свободной энергии, которая требуется, чтобы

преобразовать существующие конформеры в растворе в связанные, называется «конформационная фокусировка» (КФ) [258]. Энергия КФ может быть достаточной, чтобы внести системные ошибки в энергетические оценки, игнорирующие этот параметр.

Например, при изучении 150 комплексов белок–лиганд связанный конформер показал энергию на 4–5 ккал/моль больше, чем у несвязанного наиболее энергетически выгодного конформера в растворе [259].

Конформационное пространство лиганда для оценки энергии КФ было рассчитано при помощи квантово-механических и молекулярно-механических (QM–MM) расчетов в неявном присутствии растворителя. Авторы [258] установили, что отсутствие явного учета для энергии конформационных превращений даст погрешность в 0–10 ккал/моль итогового расчета прогнозируемых значений энергии связывания.

Расчет энергии конформационной фокусировки $\Delta G_{\text{КФ}}$ требует знания доли каждого конформационных состояний как для несвязанного, так и для связанного лиганда.

Один из эффективных способов учета конформационных переходов всех видов – включить энергию диэдральных переходов (которые регулируют нужные конформационные переходы) в общий расчет параметров возмущения ВСЭ [260,261]. Для этого введен параметр возмущения свободной энергии, называемый λ сэмплирование.

В 2010 году был предложен [262] протокол молекулярной динамики с учетом кинетического контроля связывания конформаций лиганда на белковой мишени при расчетах абсолютной свободной энергии связывания лиганда. В этом алгоритме λ и температурные зависимости считаются вместе и при этом независимо, но он требует существенных вычислительных мощностей. Оптимизирована динамическая составляющая в алгоритме FEP/REST [263], где оба параметра оптимизируются одновременно: λ плавно растет в процессе моделирования, а температура поднимается при $0 < \lambda < 0,5$ и падает при $0,5 < \lambda < 1$, тем самым давая эффективные результаты при меньшей требуемой вычислительной мощности. Оба метода не требуют предварительной информации степенях свободы структуры; однако во втором методе принято, что расчет ведется по степеням свободы рядом с сайтом связывания

лиганда, а не по всем подряд. Относительное сродство разных конформаций одной и той же молекулы лиганда можно оценить с помощью двойной топологии ВСЭ [264]

2.6.2. Механистическая (динамическая) составляющая молекулярного моделирования

Абстрагируемся от конечной цели – молекулярной модели для синтеза соединения, обладающего заданными свойствами, и постулируем очевидную лемму:

– в основе молекулярного моделирования находится гипотетический (либо экспериментально подтвержденный) механизм воздействия активного вещества на мишень, и это полностью определяет фундаментальные процессы предполагаемого вида и степени активности.

В этом механизм-ориентированном подходе прогнозируемая активность валидируется экспериментально определенными значениями. Соответствие между рассчитанными и измеренными значениями активности свидетельствует о корректности предложенного механизма.

Процесс моделирования в варианте прогнозирования с использованием докинга заключается в определении энергии связывания молекулы и мишени, выбора наиболее комплементарных объектов, и снова – прогнозирования активности как конечной цели.

В обоих случаях мы имеем статическую картину, которая отражает:

- в случае QSAR повторяемость определяющих свойств (дескрипторов) статических молекул;
- в случае докинга – энергетику конечного продукта – комплекса координирующейся молекулы и мишени.

В обоих случаях никак не учитывается динамическая составляющая процесса координации. В случае QSAR нам удалось в какой-то степени этот недостаток компенсировать, учитывая в качестве дескриптора электростатическую поверхность молекулы, и, тем самым, внося влияние динамической составляющей координации и учитывая ее на дистанции дальнего взаимодействия.

В случае докинга расчет также идет по конечным структурам: степень комплементарности определяется по энергии связывания. Конечно, FEP частично

решает эту задачу, поскольку учитывает свободную энергию системы в любой момент превращения – но мы использовали этот метод в ограниченном объеме, опять же для расчета статических систем.

И в обоих случаях *статические методы весьма ограниченно пригодны для моделирования направленности химических процессов*. А ведь такая задача – неотъемлемая часть молекулярного моделирования, поскольку знание направленности химической реакции позволяет определить состав ее продуктов, т.е. провести синтез соединений с заданными свойствами.

Для решения такой задачи необходимо определить энергетические параметры активных реагирующих частиц, таких как комплекс катализатора с промежуточными или конечными продуктами реакций, энергетику переходного состояния, в том числе с участием катализатора.

Существуют тысячи работ, в которых предрасчитаны энергетические параметры химической реакции, поэтому в настоящем обзоре мы ограничимся только целями и объектами, исследованными нами, а именно, компьютерным моделированием асимметрического каталитического синтеза.

2.6.2.1. Подходы к динамическому моделированию направленного синтеза.

QSAR. Использование зависимостей QSPR (Quantitative Structure–Property Relationships, количественная корреляция структура–свойство) для прогнозирования влияния структур реагентов и растворителей на константу скорости и направление реакции с момента появления QSAR/QSPR. Однако, как правило, в QSPR моделях отсутствует линейная зависимость активности химических соединений от их структуры. Первые попытки моделирования в рамках QSPR с целью прогнозирования константы скорости реакции описаны в работах [265,266]. Изучалась либо зависимость скорости конкурирующих реакций (и, соответственно, состава продуктов) от изменения структуры реагентов и/или от природы растворителя. Касаемо исследования влияния структуры реагентов, Чаудри и Попелиер [265] разработали систему прогнозирования константы скорости гидролиза сложных эфиров с использованием квантово-химических дескрипторов. Эстрада и Матамала [267] использовали топологические индексы для прогнозирования направления газофазных реакций алканов

и циклоалканов с OH, Cl и NO₃-радикалами.

CAMD (Computer-Aided Molecular Design) позволяет оценить и влияние природы растворителя на направление химической реакции. Struebing и др. [268]. разработали методологию дизайна растворителя с использованием упрощенных моделей в квантово-химических расчетах. В работе [269] исследовали влияние растворителей на скорость реакции Кольбе–Шмитта.

Таким образом, модель QSPR, которая бы отражала совместное воздействие структур реагентов и растворителей на скорость и продукты реакции, вполне актуальна. Такая модель служит решению двух задач: первая должна была бы быстро предсказать константу скорости без осуществления эксперимента, а второй будет моделирование комплекса, включающего реагенты, продукты реакции и растворители.

Методология расчетов с использованием энергетических профилей была представлена в работе [270]. Т.е. существует методология, которая моделирует наиболее устойчивый комплекс из реагентов, продуктов и растворителя, как результат оптимизации пути реакции. В рамках QSPR с использованием квантово-химических дескрипторов рассчитаны [271] количественные соотношения структура – активационный барьер для реакции Дильса–Альдера. Взаимосвязь между активационной энергией и структурами используемых реагентов и растворителей – это ключевая проблема предсказания структуры продуктов реакции. Однако в наборе данных [196] отсутствовало структурное разнообразие растворителей. В другой работе рассчитывались [272] различные растворители для реакции Дильса–Альдера, но при этом все растворители применялись к одной и той же реакции. Т.е., объединив результаты [196] и [272], получим и набор реагентов, и набор растворителей для моделирования реакции Дильса–Альдера.

2.7. Моделирование стереохимического синтеза (переходных состояний, конечных продуктов)

Реакция Дильса–Альдера



– благодатный пример для изучения механистических (динамических) зависимостей структура–свойство/активность/состав продуктов реакции. Эванс и Джонсон [273] представили обзор этой реакции с этих точек зрения. Влияние растворителя на скорость реакции изучено в работе [274]. Многие из исследователей замечают, что в реакции Дильса–Альдера влияние растворителя на скорость и состав продуктов минимально. Это качество реакции делает ее еще более привлекательной для моделирования.

2.7.1. Стереохимия реакции Дильса–Альдера. Термодинамический/кинетический контроль

Химические реакции часто дают совершенно разные распределения продуктов в зависимости от условий, при которых они выполняются. В частности, высокие температуры и длительное время реакции благоприятствуют наиболее стабильным («термодинамическим») продуктам, в то время как низкие температуры и короткое время реакции благоприятствуют наиболее легко формируемым («кинетическим») продуктам. В частности, в типовой реакции Дильса–Альдера циклопентадиена и малеинового ангидрида образуются *эндо* или *экзо*-аддукты (Схема 3).

Схема 3

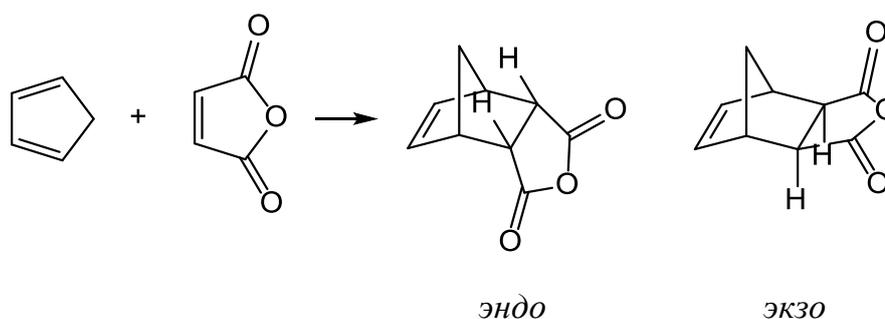


Рисунок 9. Координата типовой реакции Дильса–Альдера

Кинетический продукт имеет меньшую энергию активации, термодинамический продукт – меньшую суммарную теплоту образования (Рисунок 9).

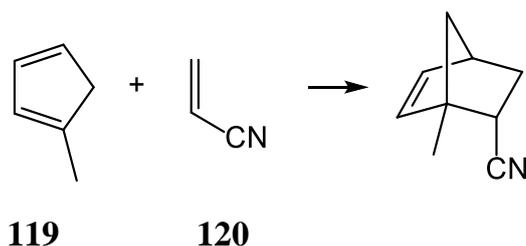
Задача – сравнить энергетические профили, чтоб понять, какой из аддуктов образуется в данной реакции, выявить специфические взаимодействия, которые влияют на переходное состояние (ПС) реакции. Вариант 1 – сравнить общий энергетический эффект реакции, т.е. принять термодинамический контроль, и сравнить энергии двух аддуктов. Вариант 2 – сравнить энергии переходного состояния *эндо* и *экзо* переходного состояния. Была ли реакция под термодинамическим контролем, какой из аддуктов будет основным продуктом и каким будет соотношение основных и второстепенных продуктов?

Еще более интересен состав продуктов в случае участия катализатора в реакции Дильса–Альдера. С точки зрения вычислительного процесса, это добавление еще одного многопараметрического объекта в граф, состоящий из координирующихся в реакции исходных веществ (и растворителя, который тоже хорошо бы учесть).

По данным теории граничных орбиталей (Frontier Molecular Orbital (FMO)), реакция Дильса–Альдера между электрононасыщенным диеном и электронодефицитным диенофилом предполагает взаимодействие между ВЗМО диена и НСМО диенофила. Чем лучше перекрытие ВЗМО/НСМО и чем меньше их энергетическая разница, тем благоприятнее взаимодействие и тем быстрее реакция.

Кислоты Льюиса снижают энергию НСМО орбиталей диенофила и тем самым катализируют реакцию, например, приведенную на Схеме 4.

Схема 4. Реакция 1-метилциклопентадиена (диен) (**119**) и акрилонитрила (диенофил) (**120**)



Энергия НСМО “свободного” и комплексованного акрилонитрила: 65 и 56 ккал/моль, соответственно. Трифторид бора, снижая энергию комплексованного диенофила, катализирует реакцию, что показано квантово-химическими методами и подтверждено экспериментально [275].

На таком простом примере видно, что выяснение соотношений между кинетическим и термодинамическим контролем (и, соответственно, прогнозирование состава продуктов реакции) дает максимальный эффект, если действовать квантово-химическими методами. То, что в терминах QSAR-QSPR называется «квантовохимическими дескрипторами», является нормальным *переходом от статистической модели QSAR к пространственной механистической модели квантовохимических расчетов.*

2.8. Квантово-химические методы молекулярного моделирования

Существует множество доступных универсальных вычислительных инструментов, каждый из которых имеет свои сильные и слабые стороны [276]. Однако, часто конкретный инструмент разрабатывается для конкретного его приложения.

При изучении реакционной способности малых органических и металлоорганических молекул (~60 атомов, не считая водорода), квантово-механические подходы показывают неплохую точность и требуют все меньше машинного времени, т.е. в скрининге могут составить конкуренцию и быть дополнением к эксперименту. Проиллюстрируем прогресс в возможностях: для квантовохимических расчетов образования и разрыва связей в реакции Дильса–Альдера обычно нужно оптимизировать переходное состояние (ПС) в статическом режиме. Что требовало около 6 месяцев в 1980-е годы, с использованием метода Хартри–Фока и суперкомпьютера того времени. Сейчас то же самое можно сделать за пару минут на персональном компьютере.

Перечень оболочек для квантово-химических вычислений [50]:

Расчеты ab initio:

Name	URL
ABINIT	http://www.abinit.org
ACES	http://www.qtp.ufl.edu/aces
BigDFT	http://bigdft.org
CP2K	http://www.cp2k.org
DACAPO	https://wiki.fysik.dtu.dk/dacapo
ErgoSCF	http://www.ergoscf.org
ERKALE	https://github.com/susilehtola/erkale
GPAW	https://wiki.fysik.dtu.dk/gpaw
HORTON	http://theochem.github.io/horton
JANPA	http://janpa.sourceforge.net
MPQC	http://www.mpqc.org
NWChem	http://www.nwchem-sw.org
Octopus	http://www.tddft.org/programs/octopus
OpenMX	http://www.openmx-square.org
Psi4	http://www.psicode.org
pyquante	http://sourceforge.net/projects/pyquante
PySCF	https://github.com/sunqm/pyscf
QMCPACK	http://qmcpack.org
Quantum espresso	http://www.quantum-espresso.org
RMG	http://rmgdft.sourceforge.net
Siam Quantum	https://sites.google.com/site/siamquantum

ABINIT может рассчитать полную энергию комплекса, распределение заряда и электронной структуры молекул и твердых тел с помощью теории функционала плотности (Density Functional Theory (DFT)) и многочастичной теории возмущений (Many-Body Perturbation Theory (MBPT)). ABINIT также может оптимизировать геометрию, выполнить молекулярно-динамические симуляции, сформировать

динамические матрицы связности, диэлектрические тензоры и др. свойства.

ACES выполняет расчеты точечных энергетических параметров, аналитических градиентов, распараллелен, в расчетах включает поддержку графических процессоров. Умеет использовать MBPT для расчета связанных кластеров для уточнения электронной корреляции.

BigDFT выполняет расчеты неэмпирическим методом, используя линейное масштабирование. Моделирует (с соответствующими граничными условиями) периодические структуры, поверхности и изолированные системы. Входит в состав ABINIT.

CP2K выполняет моделирование твердотельных, жидких, молекулярных и биологических систем. Его особое внимание массово-параллельной и линейное масштабирование методов электронной структуры и *ab initio* и молекулярной динамики (AIMD) моделирования. Он оптимизирован для метода смешанных Гауссовых и плоских волн с использованием псевдопотенциалов и может работать параллельно и на графических процессорах.

Dacapo – это программа расчета полной энергии системы, которая использует теорию функционала плотности. Он может моделировать молекулярную динамику/структурные переходы при решении уравнения Шредингера. Поддерживает параллелизацию и используется в среде Atomic Simulation Environment (ASE).

ErgoSCM – это квантово-химический инструмент для масштабируемых самосогласованных расчетов поля. Он выполняет электронные расчеты структуры с использованием методов Хартри–Фока и Кона–Шэма теории функционала плотности. Обеспечивает линейное (а не квадратичное, как во многих других системах) масштабирование загрузки процессора и использования памяти.

ERKALE предназначен для вычисления рентгеновских свойств, таких как плотности электронного импульса и профилей Комптона, рентгеновского поглощение и рентгеновского комбинационного рассеяния, а также электронных спектров возбуждения атомов и молекул.

GPAW – модуль DFT, использующий метод Projector-Augmented Wave (PAW) и интегрирующийся в Atomic Simulation Environment (ASE, оболочка атомной симуляции).

HORTON (Helpful Open-source Research TOol for N-fermion systems): основная

цель – расширяемость для исследования новых методов в теории электронной структуры *ab initio*.

JANPA вычисляет атомные орбитали из приведенной матрицы плотности одной частицы.

MPQC (massively parallel quantum chemistry program) предлагает множество функций, включая интерфейс для расчета энергии и градиентов методом Хартри–Фока, визуализацию расчетов теории функционала плотности энергии и градиентов второго порядка, интерфейс теории возмущений и теории энергетических возмущений.

NWChem предоставляет полный набор методов для моделирования классических и квантово-механических (QM) систем. Его возможности включают моделирование молекулярной электронной структуры, QM/MM, псевдопотенциала электронной структуры и молекулярной динамики. Система поддерживает многопроцессорные вычисления.

Octopus оболочка базовых расчетов с использованием зависящих от времени DFT (TDDFT) и псевдопотенциалы. Включен в проект *libxc*, который является самодостаточной библиотекой обменно–корреляционных функционалов для DFT (выпускается под лицензией LGPL).

OpenMX (Open source package for material eXplorer) разработан для моделирования наноматериалов на основе DFT, псевдо-атомной локализации базисных функций. OpenMX способен осуществить расчет таких физических свойств, как магнитные, диэлектрические, и оптимизирован для широкомасштабных параллельных вычислений.

Psi4 представляет собой набор программ для неэмпирических расчетов квантовой химии, поддерживающих широкий спектр вычислений (напр., Хартри–Фока, MP2, связанных кластеров) и общих процедур, таких как оптимизация геометрии и частотный анализ с использованием более чем 2500 базисных функций.

PyQuante и **PySCF** – это набор модулей, в основном написанных на Python, для выполнения расчетов методами Хартри–Фока и DFT. Хорошо спроектированный набор инструментов. Очередная версия находится в стадии разработки (<https://github.com/rpmuller/pyquante2>).

QMCPACK является реализацией неэмпирических расчетов методом Монте-Карло для вычисления свойств электронной структуры молекулярных, квази-2D и твердотельных систем. Стандартные форматы файлов, используемые для ввода и вывода – XML и HDF5.

QUANTUM ESPRESSO предназначен для моделирования на наноразмерных частиц с использованием DFT, включая расчеты основного состояния, структурную оптимизацию, переходных состояний и минимизацию энергии пути реакции, неэмпирические расчеты в молекулярной динамике, теорию возмещения в ДПФ, спектроскопические свойства и квантовый транспорт.

RMG – DFT оболочка, использующая мультипроцессорные системы и графические ускорители для оптимизации структур и молекулярной динамики.

Siam Quantum (SQ) оптимизирован для параллельных вычислений и ее возможности включают в себя энергетические расчеты Хартри–Фока и MP2, локализацию энергетических минимумов, оптимизацию геометрии, анализ заселенности уровней и квантовой молекулярной динамики. DFT доступен с различными функционалами.

2.8.1. Некаталитические трансформации

Несколько десятилетий исследования механизмов органических реакций в области физической органической химии обеспечили химикам интуитивное понимание потенциала и механизмов многих органических реакций [277]. Однако, для объяснения селективности пути реакции или для различения электронной вариативности (например, полярных или радикальных механизмов) экспериментальные результаты пригодны в весьма ограниченном объеме. Причина этому – необходимость глубокого понимания природы переходного состояния, которое является ключевым контролирующим фактором механизма реакции.

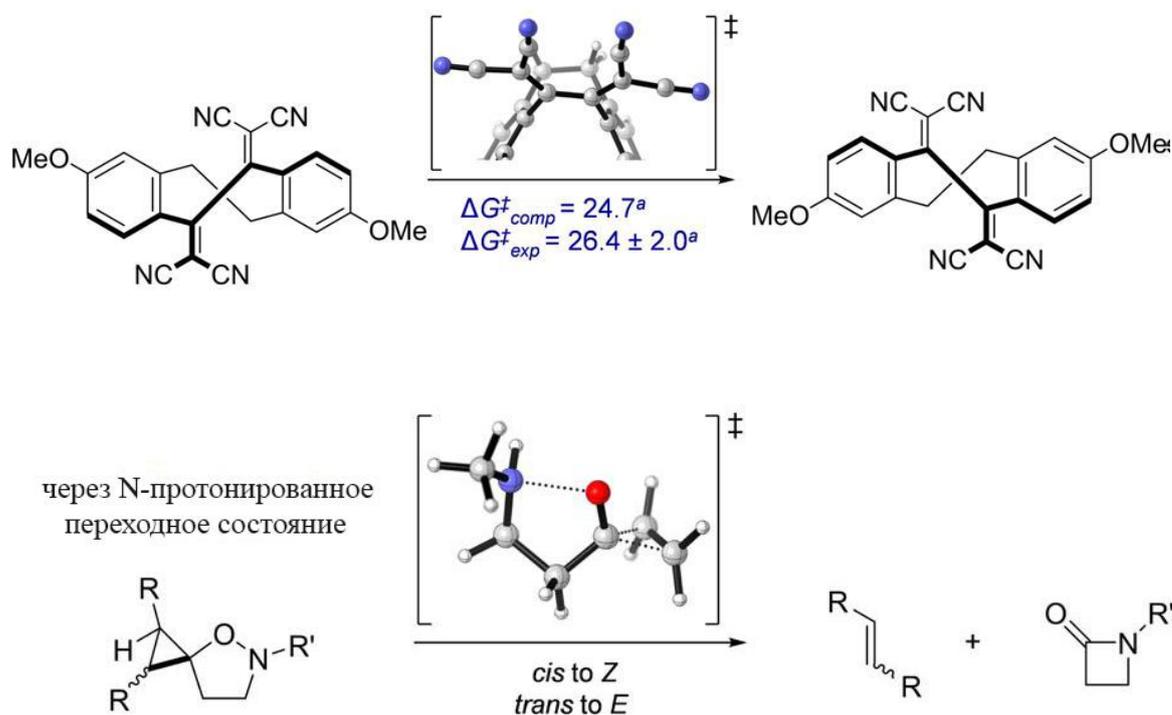
Именно при изучении переходного состояния вычислительная химия даёт ценную информацию о механизме. Примеры: детали механизма пуш-пульной перегруппировки хромофоров [278], в том числе фуллеренов [279]), процессы энантиомеризации хиральных органических циклоаддуктов [280], механистические аспекты и селективность при образовании амидной связи [281,282], основы селективности при синтезе природных веществ [283], расчеты в области медицинской химии –

направленном синтезе биологически активных соединений [284].

Некоторые расчеты легли в основу сотрудничества академической науки и индустрии. Например, изучение механизма энантиомеризации напряженных циклобут-1,3-диенов [280].

Это проект направленного синтеза оптически активных соединений, которые могут использоваться в качестве блоков или датчиков хиральной памяти. При этом молекула остается хиральной. Таким образом, понимание (а также количественная оценка) механизма энантиомеризации легло в основу предсказания продуктов реакции с заданными свойствами. Вопреки химической интуиции, расчеты показали, что энергетически наиболее осуществимый путь реакции включает в себя переход кресло–ванна, с последующей изомеризацией. Все впоследствии было подтверждено экспериментальными данными (Схема 5) [285].

Схема 5



2.8.2. Каталитические трансформации

Механистическое изучение каталитических реакций требует детализированного временного профиля свободной энергии, включая энергии всех переходных

состояний и интермедиатов, с достаточным уровнем точности. Компьютерные методы, использованные в этих исследованиях, с 1990 года практически полностью базируются на DFT-методологии. Наиболее популярен гибридный функционал B3LYP [286], который предлагает разумный баланс между скоростью и точностью расчетов и широко используется в этих расчетах. В некоторых случаях используется функционал M04 [287].

Геометрия, как правило, оптимизируется со средним базисным набором, а более точно энергии могут быть определены с помощью Single Point расчётов с расширенным базисным набором. Влияние растворителя обычно модулируется с использованием неявных методов расчета сольватации, как, например, CPCM [288,289] и SMD [290].

Коррекция свободной энергии часто включается в общий расчет при помощи жесткой роторной модели гармонического осциллятора (другими словами, эквивалента шаростержневых моделей Драйдинга). Что, понятно, не слишком повышает точность расчетов.

Селективность реакции определяется идентификацией наиболее выгодного в профиле свободной энергии переходного состояния. Из разницы энергетике переходных состояний следует селективность образования продуктов реакции. Таким образом, определение селективности обычно сводится к изучению разницы в малых энергиях переходных состояний (в диапазоне 1–2 ккал/моль). Это важно для расчетов гетерогенного катализа, обсуждаемого во второй части данной работы.

Необходимо отметить что, хотя эта общая последовательность компьютерной методологии принята на вооружение многими исследователями, каждый элемент схемы постоянно уточняется. Но всегда остается неприятная проблема: по мере увеличения количества атомов используемой модели становится всё более сложным расчет информационных зависимостей и минимизации конформационных энергий профиля реакции, существенным образом влияющих на состав конечных продуктов как каталитического, так и любого другого синтеза.

2.8.3. Моделирование переходного состояния внутримолекулярной реакции Дильса–Альдера

2.8.3.1. Реакция внутримолекулярной циклизации Дильса–Альдера и ее катализ ферментом SpnF (*Spinosa*)

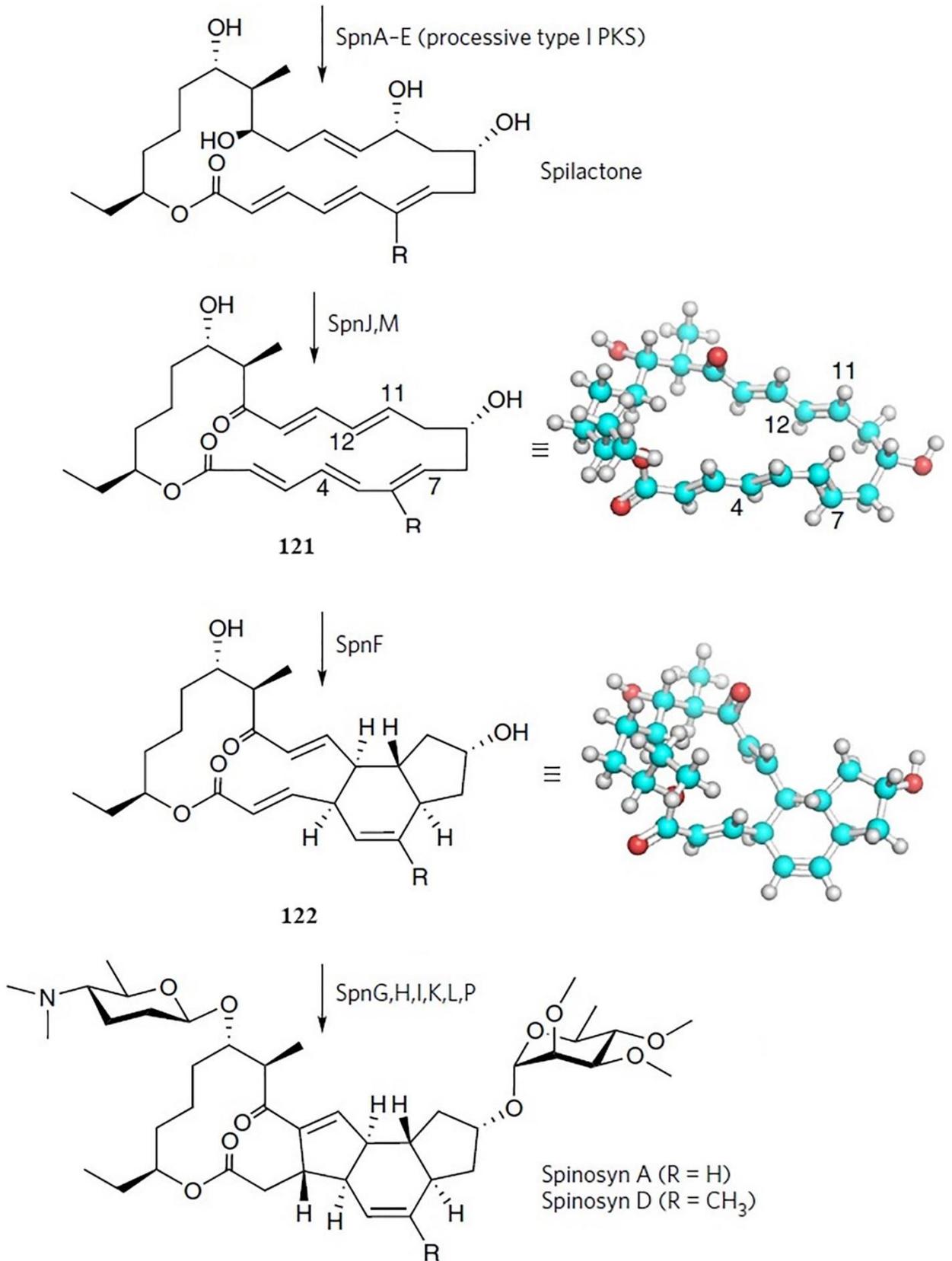
В биосинтезе инсектицида spinosin, фермент SpnF катализирует [4 + 2] циклоприсоединение 22-членного макролактона с образованием циклогексенового кольца (Рисунок 10) [291].

SpnA–E генерирует спилактон, который после отщепления воды подвергается [4 + 2] циклизации в реакции Дильса–Альдера между C4–C7 диена (C5–C6 в номенклатуре *S-цис* геометрии) и двойной связью C11–C12 диенофила той же молекулы.

В дальнейших преобразованиях (инициируемых другими белками) продукт реакции образует пару продуктов – Спинозины A и D.

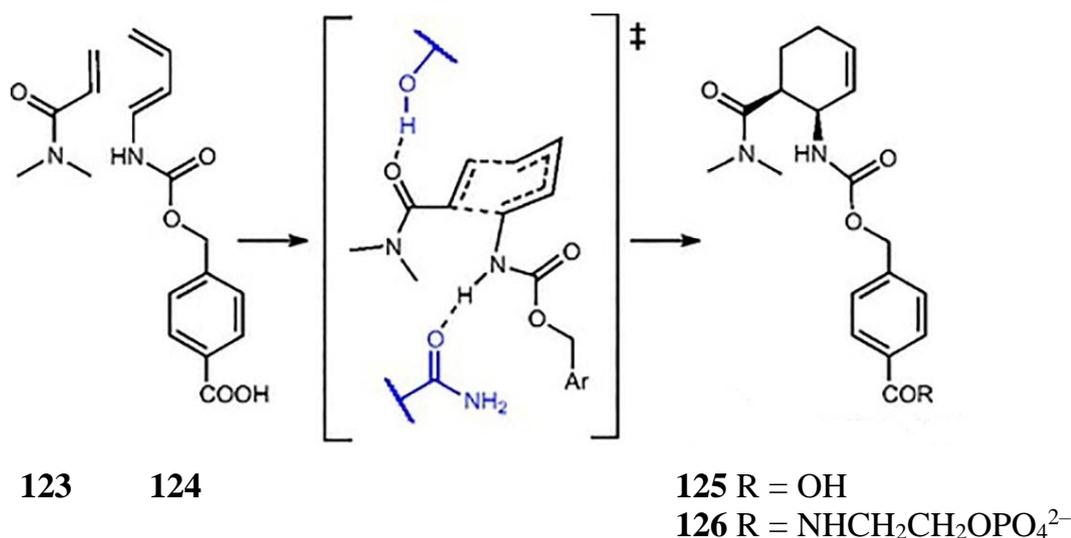
И SpnF, и SpnL похожи на *S*-аденозилметионин (SAM)-зависимые метилтрансферазы (MT), но участие SAM в этих реакциях циклизации не изучено. Несмотря на то, что другие варианты биосинтеза (например, биосинтез ловастатина, эквисетина, соланапиранов, спиротетранатов, рибофлавина и тиоциллина L) также могут включать реакцию [4 + 2] циклоприсоединения [292], SpnF является первым ферментом, который может быть смоделирован, поскольку его единственная функция заключается в катализе такого [4+2]-циклоприсоединения [293]. Если циклизация, катализируемая SpnF, действительно протекает через механизм Diels–Alder (Рисунок 10), то особенности, которые здесь проявляются, представляют фундаментальный интерес для медицинского синтеза и энзимологии [294].

Рисунок 10. Реакция Дильса–Альдера, катализируемая SpnF



На Схеме 6 представлен компьютерно смоделированный пример катализа природными ферментами реакции, по-видимому, проходящей по механизму Дильса–Альдера.

Схема 6. Реакция 4-карбоксибензил-*транс*-1,3-бутадиен-1-карбамата и *N,N*-диметилакриламида [295]



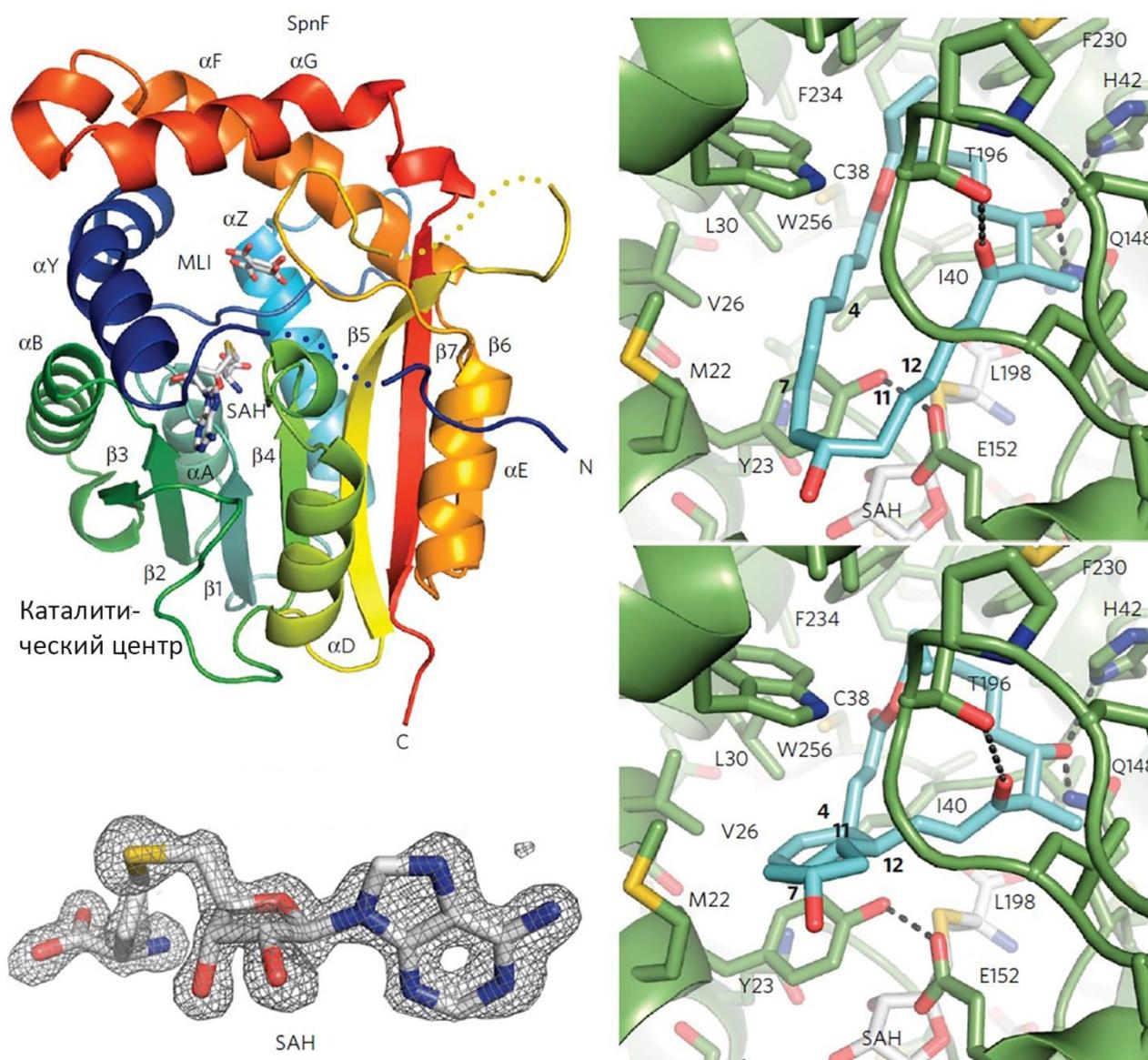
Имеются также другие примеры компьютерного моделирования ферментативного катализа в живой природе [296,297].

Ни у кого из исследователей не возникло ни малейшего сомнения в том, что реакция протекает по самому простому пути – с образованием (4+2)–переходного состояния. В последнем случае, с локализованными связями диена и диенофила – все, очевидно, так и есть.

На Рисунке 10 стадия циклизации также выглядит как типичная реакция Дильса–Альдера; и естественно, что у исследователей не возникло никаких сомнений в ее механизме. И, несмотря на то что конформации соединений были смоделированы молекулярной механикой (т.е. далеко не самым точным методом конформационной фильтрации), каталитический центр белка был определен в соответствии с этим гипотетическим механизмом циклизации, и осуществлено полное компьютерное моделирование и визуализация каталитического процесса (Рисунок 11).

Spinosin – эффективный природный инсектицид [298], биосинтез которого включает в себя реакцию Дильса–Альдера как внутримолекулярной циклизации, катализируемой ферментом SpnF [299]. Это был первый пример того, как природный фермент катализирует реакцию циклоприсоединения.

Рисунок 11. Компьютерное моделирование каталитического центра и механизма реакции циклизации, катализируемой SpnF [291]

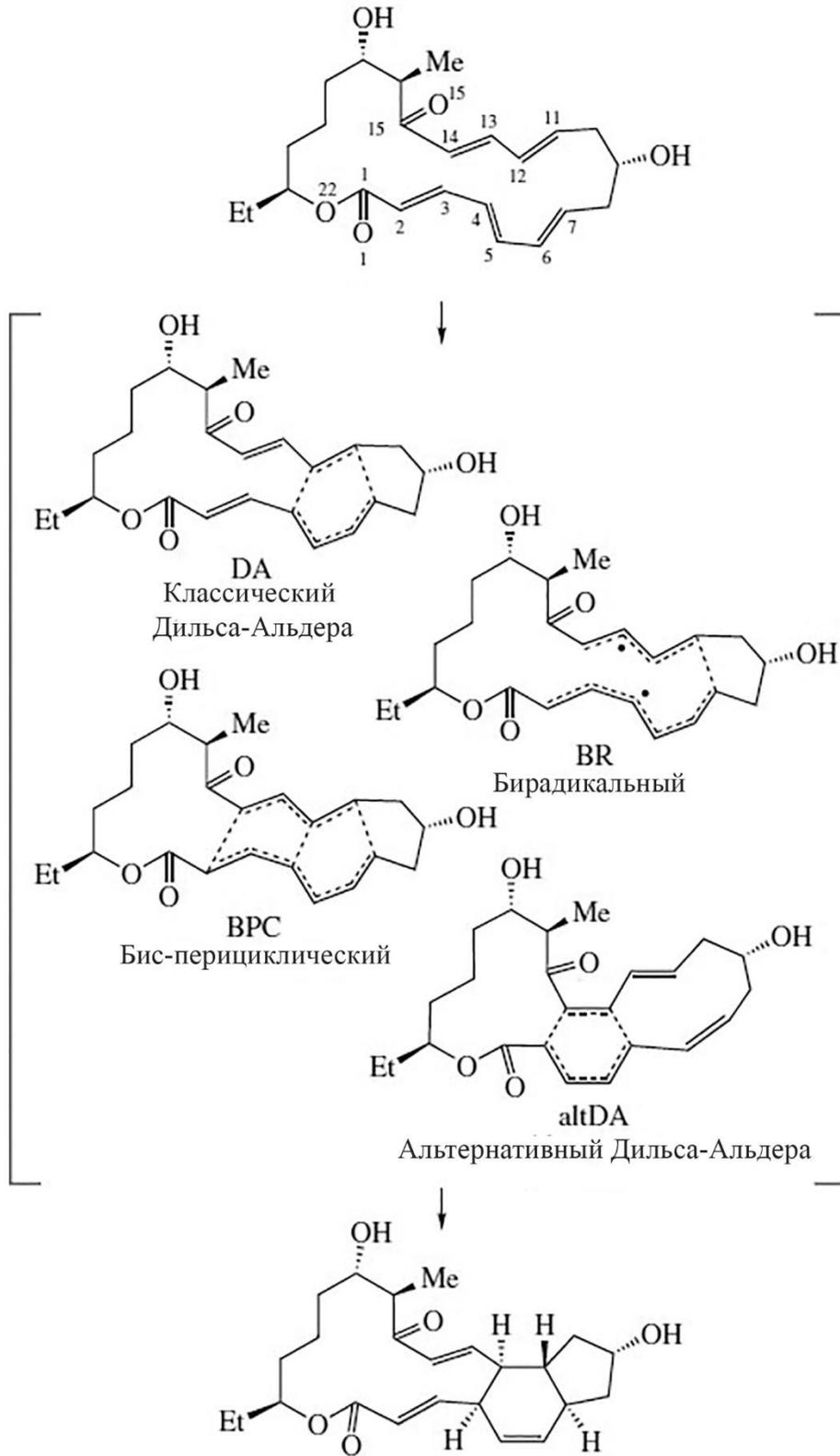


В качестве модели для моделирования каталитического центра SpnF был принят каталитический центр RebM (rebeccamycin MT, Protein Data Bank (PDB), код 3BUS).

Однако, при взгляде на структуру циклизующегося макроцикла, можно заметить, что в системе есть сопряженные дополнительные двойные связи, и логично было бы предположить, что они могут участвовать в циклическом переходном состоянии, меняя его размер цикла. В таком случае и механизм реакции циклоприсоединения, и структура каталитического сайта связывания могут быть совсем не такие, как предполагается в работе [291].

Предложены и изучены несколько механизмов вышеуказанной циклизации – во-первых, квантово-химическими способами [300 – 302], и с помощью изучения кинетического изотопного эффекта [299]. При помощи молекулярной механики и DFT изучена [303] перегруппировка электронной плотности в циклизации **121** без учета пяти атомов алифатического кольца, и сделан вывод – что SpnF катализирует согласованную асинхронную реакцию Дильса–Альдера в активной конформации субстрата и снижения энергии активации реакции стабилизацией высокополяризованной структуры переходного состояния (ПС) [301]. Исследовано также влияние заместителей [300], делокализации зарядов [303] водородных связей и межмолекулярных взаимодействий на энергию активации основной реакции. Сделан вывод, что делокализация заряда незначительна, а основная роль SpnF заключается в энергетике комплексообразования, т.е. в термодинамическом контроле реакции. Наконец, Патель и др. [302]. показали с помощью DFT и молекулярной динамики, что реакция может протекать через бис-перициклическое переходное состояние, которое через [6+4]–циклизацию после перегруппировки Коупа приводит к **122**. Этот путь не может быть корректно описан в теории переходного состояния, которая постулирует соотношение "один к одному" между ПС и продуктами. Таким образом, было предложено [302], что механизм заключается в контроле реакции через "нестатистические динамические эффекты". В Обсуждении результатов описан наш подход к изучению этой реакции, где с помощью исчерпывающего поиска всех возможных конформаций переходного состояния из четырех предложенных ПС: реакции Дильса–Альдера [292] (**DA**), бис-перициклического (**BP**), бирадикального (**BR**), плюс альтернативные реакции Дильса–Альдера [304] (**altDA**) (Схема 7).

Схема 7. Возможные механизмы реакции Дильса–Альдера, катализируемой SpnF



3. Обсуждение результатов

3.1. Этапы оптимизации моделирования направленного синтеза

Расположим, как это уже сделано в Обзоре литературы, систематическую (она же и хронологическая) последовательность наших действий в рамках алгоритма моделирования активных структур – в соответствии с увеличением точности результата математического и физического представления молекулярных ансамблей.

Глобальное разделение методов поиска потенциально активных структур для последующего направленного синтеза этих соединений состоит в различных подходах к моделированию их активности.

Вариант 1. В распоряжении исследователя нет никаких экспериментальных данных о механизме (или хотя бы о структуре обоих составляющих – активного вещества и мишени) проявления искомой активности. В этом случае исследователь не может сформировать физическую картину связывания активной структуры и мишени, взаимодействие с которой индуцирует эту активность. Остается формализовать структурную информацию о соединениях, достоверно проявляющих искомую активность, и найти общие для всей выборки черты, отвечающие за это свойство.

Методы, основанные на формализации структурной информации молекул (без учета структуры мишени), в литературе принято называть QSAR/QSPR (Quantitative Structure–Activity/Structure–Property Relationship).

Классический 3D-QSAR оперирует жесткими каркасами структур, описываемых либо одномерными матрицами связности, либо двумерными отрезками/треугольниками, прилегающими к 3D-молекулярной поверхности. Во всех случаях описание представляет собой дискретный граф, каждый элемент которого обладает координатой и каким-либо свойством (в нашем случае – зарядом). Точки могут быть привязаны к атомам, а могут располагаться, например, на молекулярной электростатической поверхности, облегающей молекулу. Расчет модели заключается в нахождении общих характеристик обучающей выборки структур, описанных одинаковым образом.

Такое сравнение пространственных дискретных графов является задачей топологии, математической статистики и, как прикладная задача, – QSAR в химии. Мы в своей работе последовательно перешли от сравнения точечного и/или и двумерного (дискретного в любом случае) описания замкнутой (3D) поверхности к работе в нечеткой логической системе сравнения молекул, с элементами машинного обучения, т.е. от дискретных описаний поверхности молекул к функционально-непрерывным (*этап прогресса моделирования 1*).

Вариант 2. На следующем этапе исследований для отдельных видов активности появилась технологическая возможность прогнозирования не на основе обучающей выборки соединений, обладающих искомым свойством, а на основе моделирования физического контакта молекулы активного вещества (называемой в общем случае лигандом) и молекулой с экспериментально определенной структурой, в общем случае называемой мишенью (*этап 2, докинг*). Это может быть белок, ответственный за какое-то конкретное заболевание, либо другая мишень фиксированной известной структуры, чувствительная к молекуле лиганда, либо катализатор, селективно координирующийся с лигандом. Зная механизм такой статической координации, мы можем понять направление течения исследуемой реакции.

Но *этап 2* также обладает характерным недостатком – расчет ведется по взаимодействию двух (или нескольких) объектов, и к проблеме *этапа 1* (обязательная жесткость структур) добавляются следующие:

а) неизвестно, в какой из конформаций взаимодействуют конформационно-лабильный лиганд (а в биологически активных структурах других практически не существует);

б) как мишень, так и система лиганд–мишень находится не в вакууме, а в растворе, и растворитель существенным (часто решающим) образом влияет на прочность связывания в этой системе. Из чего следует, что появляются:

а) необходимость конформационной фильтрации (дополнительно к *этапу 2*). Назовем – *этап 3*;

б) учет влияния молекул растворителя на прочность связывания лиганд–мишень, т.е. существенное усложнение системы уже не двух объектов, а столько, сколько позволят используемые вычислительные мощности; *этап 4*.

И последнее – как упомянуто выше, докинг как система статична. Расчет ведется по фиксированным во времени комплексам лиганд–мишень, арифметическим суммированием энергий взаимодействия связей, диполей, энергии несвязывающих взаимодействий и т.д. Априори при этом, как и в классическом QSAR, принимается, что скорость-определяющая стадия реакции – это и есть она координация. Конформация лиганда, которую учитываем в расчетах, выбрана по минимуму относительной энергии, что, в силу множественности конформаций и легкости переходов между ними, далеко не всегда соответствует истине. Поэтому на *этапах 1–3* исследователи обычно занимаются перебором всех возможных пространственных конформаций. В сочетании с необходимостью увеличения числа объектов на *этапе 4* это заводит как *статический (без учета динамики процесса координации) QSAR*, так и *статический/дискретный (определение энергии связывания арифметическим сложением энергий связывания функциональных структурных фрагментов, без учета динамики процесса координации, участия растворителя) докинг* в тупик увеличения требуемых вычислительных мощностей.

Однако, на *этапе 3* уже можно применить известный метод, называемый FEP/BCЭ (free energy perturbation, возмущение свободной энергии), в котором происходит переход от арифметики докинга к некоторому обобществлению расчетов по системе в целом. Т.е. мы можем, как это происходит в широко распространенных программах юстирования структуры в силовых полях энергетических расчетов, «покрутить» систему в координатах «свободная энергия – молекулярно-механические или полуэмпирические параметры – время», и вычленивать локальные энергетические минимумы в такой кривой потенциальной энергии. Это будет означать, что мы провели конформационную фильтрацию, т.е. поняли, в какой конформации лиганд взаимодействует с мишенью. При этом еще и мишень в какой-то степени может быть лабильной, т.е. появляются дополнительные конфигурации комплекса. Такое вычленение стабильных комплексов все равно останется набором статических данных, но, поскольку мы его последовательно укладываем на квази-непрерывный график от времени, назовем метод *квази-динамическим*. Это еще не полный учет переходного состояния реакции, но уже динамическая юстировка конформационных вариаций молекул под (лабильную) конфигурацию мишени, потому и приставка «квази».

Примененный на *этапе 3* метод отличается тем, что количество учитываемых

при расчетах элементов системы (элементов молекулярного графа), не обязательно состоящих из одной (двух, трех) молекул или молекулярных комплексов, ограничено только вычислительной мощностью, и уже можно перейти к введению в систему молекул растворителя, т.е. при помощи ВСЭ решить проблему *этапа 4*.

В моделировании прочности связывания лиганда и мишени как в биокаталитическом процессе, так и в гомогенном и гетерогенном каталитическом синтезе основным методом является построение временной зависимости энергетики процесса. Из такой зависимости однозначно следует состав продуктов реакции, т.е. рассчитанные структуры являются целью направленного синтеза.

При моделировании как биологической, так и любой другой активности необходимо понять (и оцифровать) природу и энергетику скорость-определяющей стадии процесса, которой, в большинстве каталитических реакций, является зависимость от времени энергии образования и распада каталитического комплекса. Естественно, с рассмотрением всех конформационных переходов и возможных вариантов распределения зарядовой плотности образующихся переходных состояний. Здесь необходим переход от стандартно используемой с этой целью молекулярной динамики к методам квантовой химии (*этап 5а*).

И, что существенно в методологическом плане, как ВСЭ, так и квантовохимические методы оперируют энергетическими параметрами, т.е. для них не имеет значения, какие структуры мы вводим в расчет комплекса. Можно оперировать любыми энергетическими взаимодействиями как дальнего порядка, так и ближнего (например, взаимодействиями катализатора и реагирующих веществ в каталитическом синтезе, при этом катализатор, если удастся его упростить до «single-atom», может быть и гетерогенным).

Очевидно, что не следует ограничиваться квантовохимическими методами при моделировании природы каталитического комплекса. Совершенствование спектральных и других прямопоказывающих методов (*этап 5б*), а, тем более, применение спектральных и расчетных методов в комплексе (*суммарный этап 5*), вполне может привести к более точным результатам, поскольку оперирует и математическими методами, и реальными физическими объектами.

Перейдем к объектам исследований. Здесь также происходит, в соответствии с планом, разделение на

а) Направленный синтез биологически активных соединений, т.е. либо статистическое моделирование 3D-QSAR, либо расчет координации лигандов на биологической мишени, построение модельной структуры, по результатам расчетов обладающей заданными свойствами, синтез и проверка заданной активности полученных веществ.

б) Расчет продуктов гомогенного каталитического синтеза, т.е. расчет координации лигандов на гомогенном катализаторе в растворе либо расчет каталитического переходного состояния; и прогнозирование продуктов реакции с требуемыми свойствами.

в) Подходы к моделированию гетерогенного катализа, т.е. начальный этап аналогичных п. б) расчетов на гетерогенной системе.

Все три объекта исследований моделируются для последующего синтеза с использованием общих математических и компьютерных методов. Направленный синтез соединений, смоделированных этими методами, призван валидировать эффективность прогнозирования, а объекты исследований для направленного синтеза выбраны по признаку практической важности. Такая методологическая связность, а также задача – сделать целевым любой синтез в органической химии – объясняет соединение разнородных объектов в одной работе.

3.2. Классический QSAR, расчет дискретных взаимодействий атомов и функциональных групп

Направленный синтез соединений, обладающих заданными свойствами, имеет большое значение. Для проведения направленного синтеза желательно установить реальные связи между структурами химических соединений и их свойствами.

В описанной в литературном обзоре последовательности и проводились наши исследования. Приведенная выше логика совершенствования алгоритмов расчета модельной структуры и хронология выполнения данной работы совпадают.

Моделирование соответствия структура–свойство начиналось в конце 80-х, со структур душистых веществ, рассчитанных в рамках молекулярной механики. Степень детализации и точности расчетов ограничивалась только вычислительными возможностями.

Расчеты точечных взаимодействий в классическом QSAR связаны с системами матричных представлений [305,306] и, как правило, рассчитываемые параметры привязаны к атомам активного вещества.

Идея замены скалярных и матричных описаний молекул и комплексов пространственными параметрами была *впервые изложена Н.С.Зефировым* на семинаре по QSAR в ИОХ РАН в сентябре 1989 г., а суть ее заключалась в одной фразе: «Пора уходить от привязки к атомам к молекуле в целом, т.е. от атомарных и молекулярных матриц связности к реальным физическим объектам с реальными физическими параметрами – геометрией, молекулярной поверхностью, суммарным электростатическим потенциалом». Этот семинар можно считать днем рождения 3D-QSAR (как минимум, в России).

Для выяснения степени комплементарности “ключ–замок” молекулы и рецептора используются топологические параметры [307], либо физические объекты, геометрия и прочие физические факторы, которые рассчитывают при помощи квантовой химии [308]. Этими физическими объектами могут быть произвольно выбранные элементы молекулярной структуры.

Очевидно, что второй метод даст наиболее полную картину пространственной комплементарности взаимодействующих объектов.

При расчете такого взаимодействия необходимо учитывать пространственную электростатическую комплементарность двух или нескольких молекул, либо молекул и рецепторов (дальнодействие). Для этой цели необходимо рассчитать напряженность электростатического поля, создаваемого молекулой; а затем дополнить модель результатами определения геометрической комплементарности (ближнего действия).

Нами в 1992 г. впервые была предложена такая методика, основанная на предварительном расчете объемного электростатического поля, создаваемого молекулой, в сочетании с геометрическими ограничениями системы молекула–рецептор (или молекула–молекула).

Важно в этой связи отметить, что подобные электрические поля могут быть созданы молекулами, различающихся по структуре и по типу атомов, которые они содержат.

Подход базируется на следующих основных ограничениях, присущих любым QSAR расчетам:

1. Никакая химическая реакция не происходит при координации молекулы и рецептора, т. е. старые связи не рвутся и новые связи не образуются – как в самой молекуле, так и между молекулой и рецептором. Это означает, что минимальное расстояние, разделяющее молекулы лиганда и рецептора, равно сумме Ван-дер-Ваальсовых (ВДВ) радиусов контактирующих атомов.

2. На этих расстояниях энергия взаимодействия между молекулой и рецептором регулируется электростатической составляющей. Таким образом, предполагается, что интегралы перекрытия орбиталей молекулы и рецептора равны 0; дисперсионные взаимодействия не учитываются.

3.2.1. 3D-QSAR, описание поверхности взаимодействия

Анализируя приведенный набор описаний координации вещество–связующая подложка (рецептор, молекула), видим, что они, за исключением информационной составляющей («топологические дескрипторы»), являются производными электростатического (диполь–дипольного) взаимодействия. В соответствии с классическими представлениями [309], водородную связь считаем частным случаем таких взаимодействий.

Основу предлагаемого подхода составляет постулат, что биологическая активность химических веществ определяется эффективностью комплексообразования с рецептором-мишенью [310,311]. Также постулируется, что для эффективного взаимодействия биологически активной молекулы (лиганда) и рецептора необходима комплементарность геометрии как обеих составных частей образующегося комплекса [312 – 315], так и их электростатических полей [316 – 318].

3.2.2. Молекулярная электростатическая поверхность (МЭП)

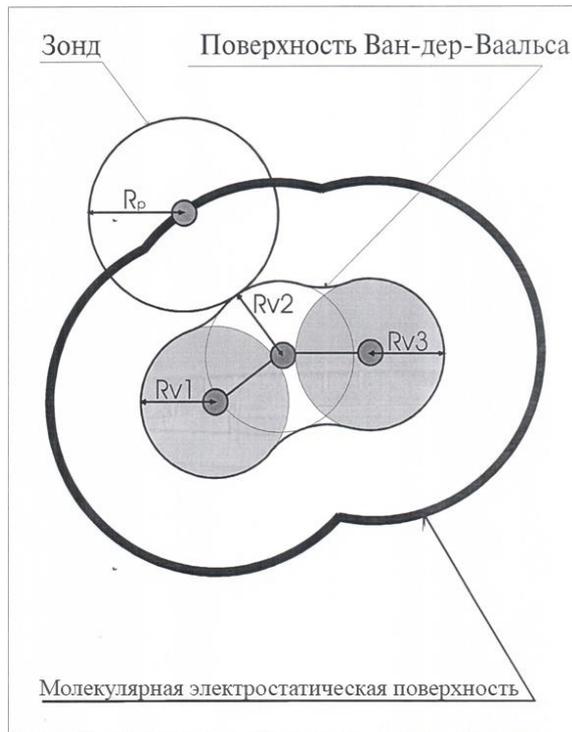
Совместимость геометрии и электростатического поля молекул означает, что комплекс лиганд–мишень соответствует следующему критерию: молекула лиганда отстоит от мишени на точном расстоянии суммы Ван-дер-Ваальсовых радиусов атомов контактирующих субстратов, или больше – но не меньше, т.е. не происходит химическая реакция между координирующимися молекулами [319].

Очевидно, что, если не происходит химической реакции, внутри облегающей («покрывающей») поверхности молекулы не может быть атомов другого вещества.

В то время как жесткие пространственные ограничения действуют на малых расстояниях порядка суммы молекулярных радиусов любого вида (Ван-дер-Ваальсовых, ковалентных), электростатические поля, генерируемые атомами молекулы, взаимодействуют между собой на больших расстояниях, и энергетическая составляющая этого взаимодействия увеличивается по мере сближения взаимодействующих субстратов. Соответственно, минимальное расстояние (т.е. сумма радиусов покрывающих поверхностей) означает максимальную энергию взаимодействия. Расчет характеристик электростатического поля покрывающей поверхности (у каждой точки которой два параметра – координата, обеспечиваемая относительным расположением атомов молекулы, и заряд) позволит учесть и геометрию, и электростатику как отдельных молекул, так и (в случае комплементарности электростатических полей!) комплекс молекула–рецептор.

Описание «покрывающей поверхности» может быть дискретным (точечным) и интегральным (функция). В Обзоре литературы мы описывали усложнение расчетных методов от описания поверхности как точечного графа к интегральному. Начнем с точек. Присвоим каждой точке (координата у нее уже есть) такой поверхности величину электростатического потенциала и назовем полученную двухпараметрическую (параметр 1 – координата, параметр 2 – заряд) поверхность “молекулярная электростатическая поверхность” (МЭП) [153]. Анализ обоих параметров поверхности МЭП (координаты и распределение электростатического потенциала на ней) – это база для нахождения закономерностей, общих для МЭП обучающей выборки молекул, обладающих одним и тем же видом активности. Построить МЭП легко, пространственно обойдя сферой радиусом атома водорода молекулу, как показано на Рисунке 12.

Рисунок 12. Построение МЭП модельной молекулы



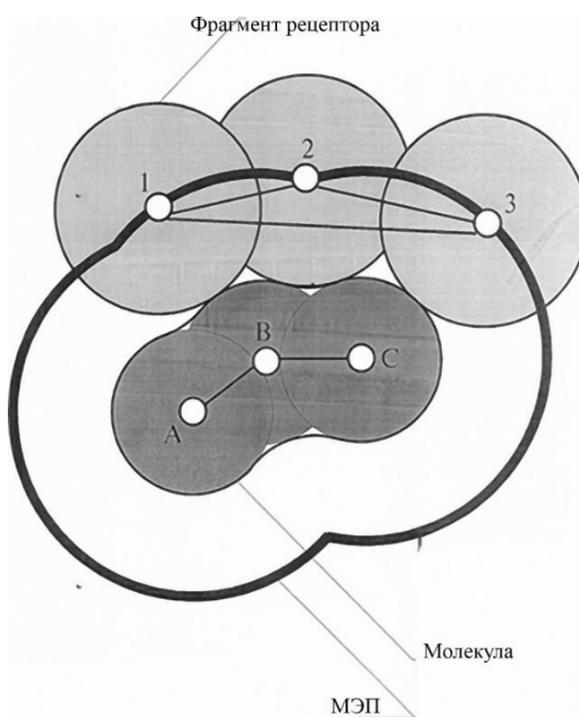
Выявленные статистическим методами закономерности и будут представлять математическую (*не* физическую) модель вида активности.

Как упоминалось в предыдущем разделе, прямое сопоставление величин потенциала на поверхности молекул не является целесообразным, поскольку сходные по типу проявляемой активности молекулы могут иметь большие отличия в форме своей поверхности, что в значительной степени усложняет задачу нахождения общих признаков у двух или нескольких молекул. Кроме того, при координации молекул на рецепторе, центры атомов и рецептора могут находиться лишь в определенных точках на МЭП, что повышает важность именно этих точек при анализе. На Рисунке 13 для демонстрации этого положения приведен условный пример.

Этот факт естественным образом определяет способ представления МЭП в виде набора точек, равномерно распределенных по МЭП. Поскольку расстояния между центрами атомов рецептора, расположенными на МЭП при координации молекулы на рецепторе, определяются геометрией рецептора и не могут быть произвольными, то становится очевидной необходимостью анализировать расстояния между точками на МЭП, что, в свою очередь, определяет тип молекулярного дескриптора, используемого для установления зависимости

между структурой и биологической активностью – простейшие геометрические элементы (отрезки, треугольники, тетраэдры), характеризуемые величинами потенциала на своих вершинах и длинами ребер между ними.

Рисунок 13. Взаимная координация молекулы и рецептора



При использовании таких структурных дескрипторов МЭП может быть представлена в виде списка геометрических элементов, вершины которых помечены величинами молекулярного электростатического потенциала. Иными словами, каждую молекулу мы представляем целым “спектром” структурных дескрипторов. Назовем полученное описание *структурным спектром молекулы*. Далее, сравнивая между собой списки дескрипторов для всей обучающей выборки, мы можем выделить общие геометрические элементы, присутствующие во всех активных молекулах и отсутствующие в неактивных молекулах. Тогда совокупность этих элементов и будет определять те закономерности в структуре молекул, которые отвечают за проявление рассматриваемого вида биологической активности.

Таким образом, мы можем изучать комплементарность молекулы и рецептора путем анализа молекулярных поверхностей электростатического потенциала, которые являются функциями геометрии молекул и их электростатических полей.

Поскольку локальная картина создаваемого различными молекулами электростатического поля может быть подобна, то использование точек на МЭП позволяет не связывать биологическую активность с конкретным типом атомов, входящих в состав молекул. Отличающиеся структуры могут образовывать сходные по обоим параметрам МЭП, и тем самым иметь сходные параметры активности. *Эта особенность легла в основу наших исследований в области классического QSAR.*

Разрабатываемый в настоящей работе подход является обобщением метода “пустых цепочек” системы БИБИГОН [320,321], описанного в обзоре литературы, при этом топологические индексы заменяются реальным 3D-точечным графом с расстояниями-ребрами графа. Наш подход основывается на представлении формы (в широком плане, не только геометрии) поверхности молекул в виде структурного спектра (упорядоченного набора однотипных структурных дескрипторов), что является более общим методом и позволяет применять простые математические методы [320,321] обработки таких спектров к решению задачи выявления общих закономерностей в геометрической форме и пространственном распределении скалярного электростатического потенциала вокруг молекул.

Таким образом, *начало всей работе в 1989–1996 гг. по прогнозированию биологической активности при помощи молекулярного моделирования 3D-QSAR было положено на основе дискретного подхода (точка–точка) к описанию молекулярной электростатической поверхности и сравнению таковых поверхностей в молекулах обучающей выборки. В рамках подхода речь не может идти о взаимодействующих поверхностях, поскольку о второй поверхности – МЭП мишени – ничего не известно.*

Общая схема подхода:

- Молекула рассматривается как 3D-объект, описываемый тремя параметрами: атомной структурой, таблицей связности атомов и МЭП.
- На МЭП обозначаются "ключевые точки", характеризующие геометрическую и электростатическую структуру МЭП.
- У каждой ключевой точки (как и всех остальных точек МЭП) имеются пространственные координаты и, согласно экспертной оценке, степень участия в активном сайте молекулы.

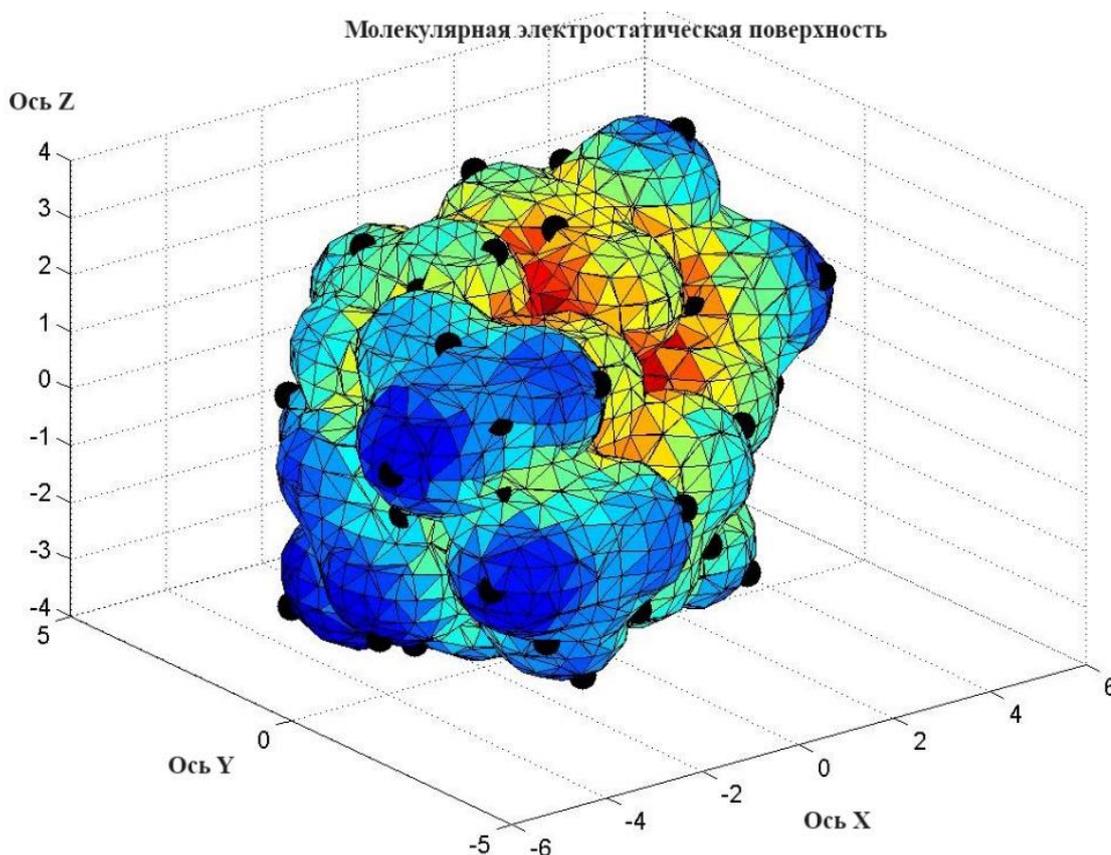
- Для таких характеристичных точек молекулы строится матрица расстояний $D=\{d_{ij}\}$, где d_{ij} – расстояние между i -ой и j -ой точками. Расстояния разбиваются на интервалы, величина которых зависит от степени дискретизации описания поверхности, и строится матрица интервалов $P=\{p_n\}$, где p_n – номер интервала, в который попало расстояние d_{ij} .
- Интервальная матрица служит базой описания объекта путем перечисления всех пар (или троек, или четверок) ключевых точек. Формируется таблица повторов всех связанных групп точек в виде $(T_1, T_2, T_k, P), N$, где: T_1, T_2, T_k – имена ключевых точек, входящих в ансамбль точек, P – номер интервала расстояния, N – число повторений фрагмента (T_1, T_2, T_k, P) в объекте. Символьная запись (T_1, T_2, T_k, P) принимается как имя признака. Совпадение имеет признаков означает их тождественность. Ключевые точки и номера интервалов расстояний перечисляются в имени признака в возрастающем лексикографическом порядке.
- Ансамбли ключевых точек составляют граф описания молекулы обучающей выборки. Формируется полный список признаков, характерных для данной выборки, и на его основе строится матрица описания “структура–признак” $X=\{x_{ij}\}$, где x_{ij} – число повторений j -ого признака в i -ой молекуле.
- Далее проводится отбор существенных признаков и строится ансамбль моделей, достоверно описывающих искомое свойство в терминах типов ключевых точек на МЭП. QSAR модель строится на основе существенного набора признаков, входящих во все или большинство молекул обучающей выборки, определяются особенности формы модельной МЭП и модельной геометрической конфигурации молекул, обладающих заданным свойством.

Первым (как хронологически, так и логически) в нашей работе был расчет душистых веществ. Начало 90-х – т.е. вынужденное применение имеющихся в наличии вычислительных методов (процессоры, полуэмпирика для вычисления геометрии молекул и расчетов МЭП. Только дискретное описание МЭП двумерными графическими примитивами), но результат получен вполне достоверный.

3.2.3. Формализация проблемы QSAR

Для каждой молекулы в наборе изучения строим ее триангулированную молекулярную поверхность без учета растворителя [322]. Выбираем критические (ключевые) точки на молекулярных поверхностях, используя методологию Коннолли [323] для описания локальных выпуклостей и дырок. В конечном счете, вычисляем простой электростатический (кулоновский) потенциал каждой ключевой точки путем суммирования эффектов электростатических полей, создаваемых отдельными атомами (Рисунок 14).

Рисунок 14. Молекулярная электростатическая поверхность (МЭП) и построенные компьютером ключевые точки



На данный момент мы можем переформулировать проблему QSAR следующим образом: каждый объект (молекула) в наборе изучения представлен множеством N критических точек $(x_i, y_i, z_i, S_i, Q_i), i = 1..N$, где (x_i, y_i, z_i) являются 3D-координатами критической точки, S_i описывает форму (0 для дырки и 1 для выпуклости), и Q_i является электростатическим потенциалом. Нашей предполагаемой целью было построение трехмерных структурных дескрипторов, представляющие

критические точки, их пары и триплеты, и проверка их значения для моделирования QSAR.

3.2.4. Математическое решение проблемы QSAR

Сначала проводится кластерный анализ электростатических потенциалов всех критических точек во всех молекулах для разделения на три кластера – А, В и С. Затем помечается каждая ключевая точка в соответствии с ее геометрическими и электростатическими свойствами. Метка критической точки состоит из двух символов, где первый описывает электростатические свойства, а второй – геометрические (Q_i). Таким образом, есть метки А0, А1, В0, В1, С0 и С1.

Далее вычисляется d_{\max} – максимальное эвклидово расстояние между всеми критическими точками всех молекул в нашем образце. Мы разделяем сегмент $[0, d_{\max}]$ на три равные части $D = [0, \frac{d_{\max}}{3}]$, $E = [\frac{d_{\max}}{3}, 2\frac{d_{\max}}{3}]$, $F = [2\frac{d_{\max}}{3}, d_{\max}]$.

Создается трехуровневый «алфавит дескриптора». Первый уровень алфавита дескриптора состоит из символьных меток всех критических точек. Второй уровень состоит из символьных дескрипторов в форме $F^2 = (P_1, P_2, d) = P_1 + P_2 + d$, где P_1, P_2 являются 3-символьными метками критических точек, а d обозначает сегмент расстояния (D, E или F) между P_1 и, «+» – операции конкатенации. Значение дескриптора (P_1, P_2, d) представляет собой число вхождений строки F^2 в молекуле. Аналогично, мы создаем дескрипторы третьего уровня $F^3 = ((P_1, P_2, d_1), P_3, d_2) = F^2 + P_3 + d_2$, где d_2 показывает расстоянием между серединой точки $[P_1, P_2]$ и P_3 .

Нумеруются все пары и триплеты для вычисления дескрипторов F^2 и F^3 и их значений для каждой молекулы. Всего имеется 703 дескриптора с ненулевыми значениями. В результате формируется матрица «молекула–дескриптор» с 50 строками и 703 столбцами. Затем применяется линейная версия группового метода обработки данных (GDMH) [324] для создания набора линейных моделей. С помощью GDMH формируются дескрипторы, установленные для определения расстояния на молекулах, сгруппированные в два кластера с 28 и 10 точками (молекулами). На первом кластере получена модель линейки с 6 дескрипторами и перекрестной проверкой 0,79, на втором кластере – с двумя дескрипторами и перекрестной проверкой 0,8.

Итак, мы можем попытаться предсказать запах амбры для тех новых молекул, которые расположены в окрестности первого или второго кластера. Если новая молекула находится далеко от центров кластера, следует отказ от прогноза.

Преимущество метода заключается в том, что описание трехмерных молекул инвариантно к пространственным вращениям молекулы, что устраняет необходимость согласования молекул в трехмерном пространстве (как в методе CoMFA).

3.3. 3D-QSAR исследование амбрового запаха в ряду норлабдановых соединений

Задача оценки свойств душистых веществ решается формированием обучающей выборки соединений – химической структурной базы данных (СБД). Используемая методика и разработанная в начале 90-х годов в ИОХ РАН программная система описаны в [105]. В основе метода лежит построение и анализ структурных символьных спектров пространственных форм веществ, имеющих амбровый запах.

Нами исследовались различные варианты дискретизации расстояний между точками представления особенностей молекулярных форм, а также более консервативный метод – варианты дискретизации зарядов на атомах. Применялся оригинальный 3D-QSAR подход (с использованием пакета BIBIGON) для построения прогностических корреляционных уравнений.

Преимущества изучения именно амбрового запаха очевидны: богатый фактический материал в классе обладающих этим запахом лабдановых дитерпеноидов. Изучение этих молекул удобно ввиду того, что соединения этого класса имеют жесткий бициклический декалиновый скелет. Это позволяет избежать трудностей, связанных с неоднозначностью трехмерного представления молекул, и необходимостью конформационной фильтрации. И, наконец, амбровый и мускусный запахи – наиболее интенсивные и однозначно характеризуемые в общем спектре одорантов.

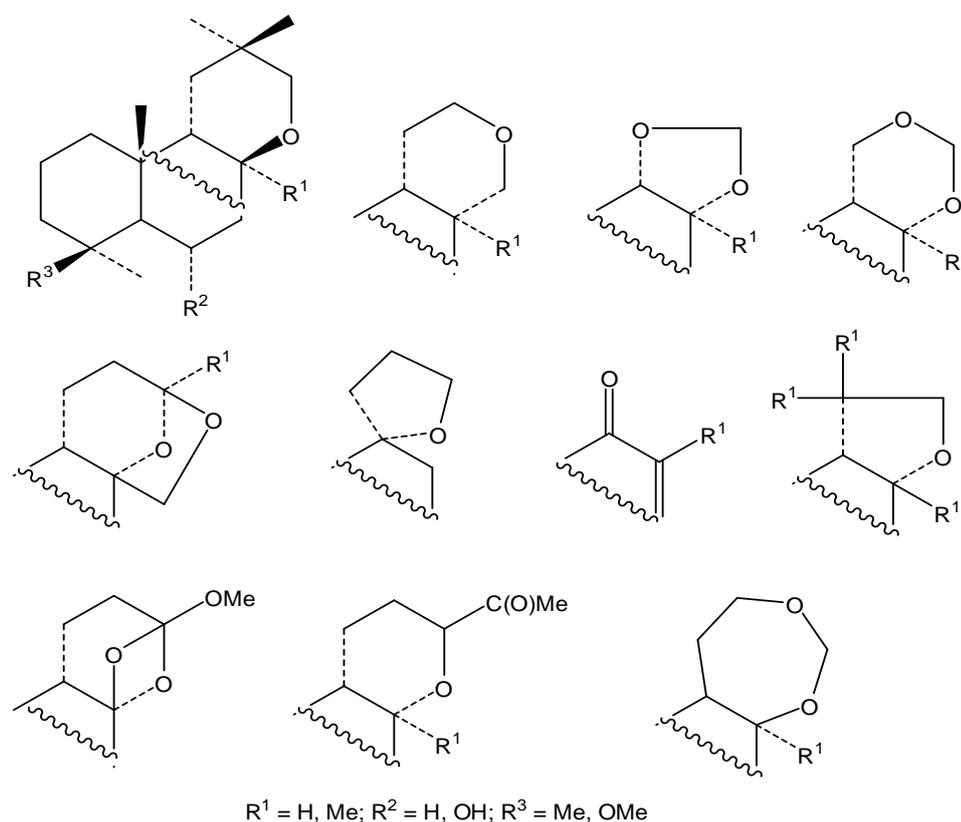
Первоначальная оптимизация геометрии после построения трехмерных структур проводилась при помощи силового поля MMX+ программой PC MODEL v. 4.0 (единственный набор программного обеспечения (ПО), бывший в распоряжении расчетчиков на тот момент). При последующей оптимизации использовался полуэмпирический гамильтониан AM1 программы MOPAC v.7.0. Выбор данного

гамильтониана основан на сравнении точности результатов оптимизации геометрии в ряду MNDO, AM1, PM3 с рентгеноструктурными данными 4 родственных соединений, 2 из которых относятся к классу лабданов. Реализация QSAR-подхода описана в более ранних работах и заключается в генерации дескрипторов вида $[N1_I1-N2_I2]L$ (здесь $N1,2$ – имена атомов, $I1,2$ – заряды остовов, L – расстояние между двумя атомами) и последующем построении с помощью программного комплекса БИБИГОН [105] корреляционных уравнений.

Для описания свойства использовалась обучающая выборка (Рисунок 15), содержащая 76 молекул, 34 из которых обладают запахом (активность = 1), а оставшиеся 42 – нет (активность=0). При последующих экспериментах выборка подвергалась модификации. Полученные модели, построенные на 10–11 дескрипторах, правильно предсказывают активность 74 соединений обучающей выборки. Кроме того, результирующая модель была использована для анализа 22 активных соединений, в обучающую выборку не входящих, и правильно описала 18 из них.

Для каждой молекулы мы считаем, что известны 3D-координаты узлов (атомов) и их количественные характеристики (частичные атомные заряды).

Рисунок 15. Примеры обучающей выборки соединений



Геометрическая оптимизация и расчеты зарядов выполнялись с помощью Gaussian03. Молекула представляли в виде молекулярной поверхности (расстояния от атомов равны 1 ВдВ радиусу), которая «окрашивается» локальным физическим параметром (ЛФП), например, зарядом, липофильностью, способностью отдавать или принимать электрон (донорно–акцепторные свойства).

В Таблице 3 и на Рисунке 15 приведены структурные формулы соединений, составляющих тестовую и контрольную выборки. Свойство (запах) определяется дискретной величиной: активность = 1 (звездочка в Таблице 3) означает, что данное соединение обладает запахом; = 0 (отсутствие звездочки) – не обладает или имеет слабый запах либо слабый смешанный запах (например, древесно-амбровый). Эти значения активностей и используются при построении QSAR-уравнений.

Тестовая выборка строилась на основе ранее используемой другими исследователями [114] таким образом, чтобы в нее входили соединения, не содержащие подвижных заместителей (при наличии основного жесткого бициклического декалинового скелета), как то: винильных, оксиметильных и т.п. фрагментов. Так же соблюдалось небольшое преобладание в обучающей структурной базе данных (СБД) неактивных структур, позволяющее получать модели, более удобные для физической интерпретации (конечные регрессионные уравнения содержат, в основном, положительные члены).

Учитывая то, что модели, построенные на подобных дескрипторах, имеют склонность ошибаться в предсказании активных соединений (активность ближе к 0), в качестве контрольной выборки использована совокупность молекул, обладающих запахом, но не включенных в обучающую выборку.

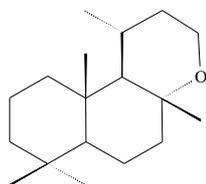
После прохождения процедур, описанных в экспериментальной части и методологии, были получены гистограммы распределения зарядов на атомах и расстояний для совокупности молекул обучающей выборки. Первоначальная задача сводилась к получению такой классификации, QSAR-модели которой отличались бы высокими значениями параметров R^2 (квадрат коэффициента корреляции) и cross validation R^2 (R^2 скользящего контроля). Подобная классификация и была получена. Ее численные значения представлены ниже:

D00: (0,000_1,997)	C00: (–9,0000_–0,3125)
D01: (1,997_2,284)	C01: (–0,3125_–0,2578)

D02: (2,284_3,002)	C02: (-0,2578_-0,2266)
D03: (3,002_3,104)	C03: (-0,2266_-0,1680)
D04: (3,104_3,207)	C04: (-0,1680_-0,1328)
D05: (3,207_3,309)	C05: (-0,1328_-0,1055)
D06: (3,309_3,535)	C06: (-0,1055_-0,0664)
D07: (3,535_3,638)	C07: (-0,0664_-0,0312)
D08: (3,638_3,740)	C08: (-0,0312_+0,0078)
D09: (3,740_3,843)	C09: (+0,0078_+0,0664)
D10: (3,843_3,945)	C10: (+0,0664_+0,0781)
D11: (3,945_4,048)	C11: (+0,0781_+0,0898)
D12: (4,048_4,191)	C12: (+0,0898_+0,1016)
D13: (4,191_4,807)	C13: (+0,1016_+0,1133)
D14: (4,807_5,422)	C14: (+0,1133_+0,1641)
D15: (5,422_7,473)	C15: (+0,1641_+9,0000)

Численные значения расстояний между парами атомов (D) приведены в ангстремах, зарядов атомных остовов (C) в а.е.

QSAR-модели, построенные на всех 76 моделях обучающей выборки, обнаруживали лишь одну грубую ошибку предсказания, когда соединение **127**, обладающее запахом, систематически относилось к классу неактивных.



127

Применение полученных моделей к контрольной выборке показало, что они не предсказывают активность у 8 соединений, обладающих запахом. Последующее исключение этого соединения из обучающей выборки и построение новых моделей позволило избежать грубых ошибок предсказания для обновленной обучающей выборки и понизить число ошибок при анализе контрольной выборки до 6. Аналогичным образом было получено третье семейство моделей, также безошибочно предсказывающее активности соединений обучающей выборки и позволяющее снизить число ошибок предсказания для контрольной выборки до 4.

В Таблице 8 схематически представлены описанные результаты.

Таблица 8. Результаты применения трех различных моделей в прогнозировании амбрового запаха

N	R2	Cr. R2	Err1	Err2
76	0.78	0.71	1	8
75	0.80	0.75	0	6
74	0.83	0.81	0	4

N – число соединений в обучающей выборке.

Err1 – количество ошибок предсказания для обучающей выборки.

Err2 – количество ошибок предсказания для контрольной выборки.

В результате анализа моделей всех трех экспериментов было обнаружено, что совокупности эффективных дескрипторов (т.е. входящих в модели с коэффициентами $>+0,2$) пересекаются на 80%. Ниже представлено множество эффективных дескрипторов, играющих в своих моделях главную роль и встречающихся как минимум в моделях двух экспериментов.

[C_C06–C_C03]D03	[H_C15–H_C10]D11
[H_C10–C_C04]D08	[H_C11–C_C09]D16
[H_C10–C_C09]D03	[C_C09–C_C05]D02
[H_C10–C_C15]D06	[H_C09–C_C08]D02
[H_C11–C_C12]D11	[H_C12–H_C12]D08
[H_C13–C_C08]D05	[C_C12–C_C04]D08
[O_C01–H_C12]D03	[H_C12–C_C04]D08
[O_C01–H_C13]D08	[C_C09–C_C04]D07

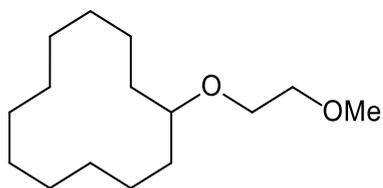
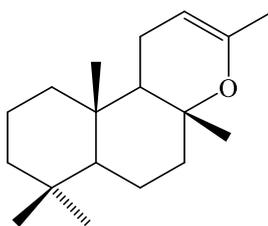
Различные сочетания этих дескрипторов друг с другом (от 2 и более) в молекуле обеспечивают наличие исследуемого свойства.

Таким образом, набор дескрипторов, определенный в вышеописанной дискретизации, позволяет описывать все соединения, входящие в тестовую выборку, а также 18 из 22 активных соединений, составляющих контрольную выборку. Полученные результаты позволяют идентифицировать молекулярные фрагменты, ответственных за появление запаха, что и является предметом исследований направленного синтеза соединений с заданными свойствами.

В процессе прогнозирования амбрового запаха мы получили модель, включающую известные два вещества с амбровым запахом, но никак не сочетавшихся с

моделями, построенными П.Ф. Владом и сотрудниками [116,330].

Это были амбролигнан **128**, имеющий древесно–амбровый запах [325], и оксид склареола **129**.

**128****129**

Наша модель вполне описывает оба эти вещества (второе – в наиболее энергетически выгодной конформации), на 85 и 79% соответствия.

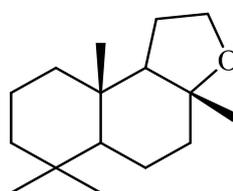
Результат косвенно подтверждает предложенный «электростатический» подход к выбору дескрипторов – ведь структура амбролигнана существенно отличается от структур сесквитерпенов ладанового ряда, и система ключ–замок в варианте теории Берсукера [116] здесь не работает.

Молекулярные дескрипторы для амбрового запаха были построены в виде строк символов из анализа особых точек на молекулярной поверхности без учета растворителя, каждая точка которой имеет определенное значение электростатического потенциала, и расстояниях между ними. Групповой метод расчета обработки данных (Group Method of Data Handling (GMDH)) применялся для отбора репрезентативных наборов дескрипторов и для построения семейства линейных предсказательных функций для запаха, который имеет коэффициент перекрестной проверки для амбрового запаха 0,8 (GMDH).

Предложенный метод можно использовать для учета пространственного электростатического комплементарного параметра для двух или нескольких молекул, или молекулы и рецептора (дальнее взаимодействие). Для этого необходимо вычислить электростатическое поле, созданное молекулой, а затем дополнить полученную

таким образом модель зеркальной (заряд, геометрия) комплементарностью (ближнее взаимодействие).

В тот момент, когда мы конструировали модель амбрового запаха, текущая задача состояла не только в нахождении новых, патентно чистых веществ, обладающих этим свойством, но и в максимально простом синтезе уже известных соединений, имеющих весьма важное промышленное значение – к тому моменту был запрещен отлов кашалотов, и прекратились поставки амброна (природного животного продукта), а основным применяемым веществом с амбровым запахом стал амброксид, самый распространенный и наиболее активный источник амбрового запаха:

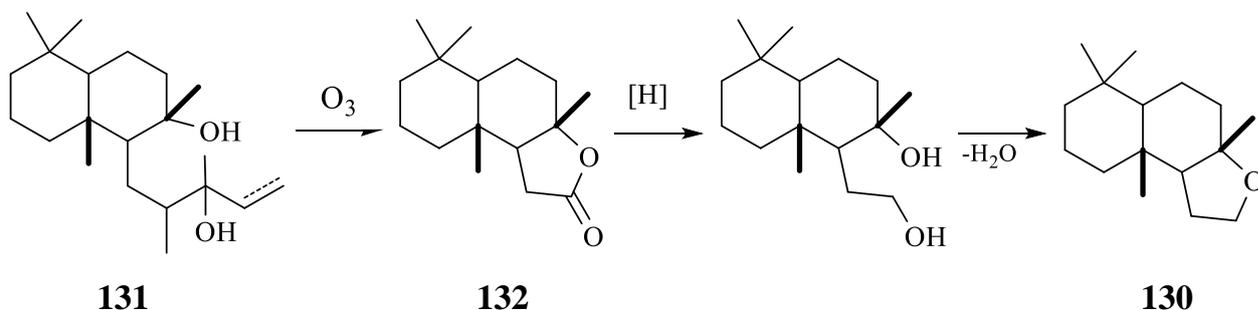


130

Амброксид (3а,6,6,9а-тетрамилдодекагидронафто[2,1-*b*]фуран), $C_{16}H_{28}O$, мол. масса 236,40, бесцветное кристаллическое вещество с сильным и очень стойким амбровым запахом; т. пл. 75–76°C.

Получают его окислительным расщеплением склареола **131**, восстановлением образующегося норамбреинолида **132** комплексными гидридами металлов и дегидратацией полученного при восстановлении диола (Схема 8).

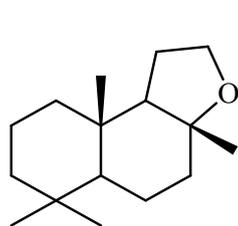
Схема 8



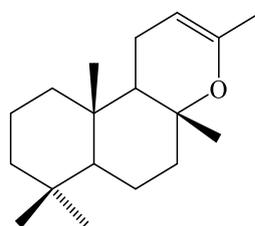
Нам удалось сократить этот процесс до одной стадии и получить этот продукт в смеси с оксидом склареола прямым электролитическим окислением склареола [326].

Реакцию проводят в бездиафрагменном электролизере в среде метанола или этанола в присутствии электролита – гидроксида щелочного металла при температуре 60–80°C и пропускании 25–30 F электричества на 1 моль склареола (плотность тока 0,1 А/см²). В качестве анода для электролиза склареола используют платину, графит или другие высокостойкие анодные материалы, в качестве катода – металлы с низким перенапряжением водорода или графита. При токе 1А и площади поверхности электродов 10 см² время реакции составляет 2,5–15 ч на 1/60 моль склареола.

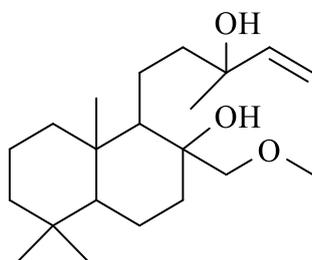
Готовый продукт представляет собой смесь веществ с амбровым запахом (**129**, **130**, **135**) и лишенных запаха (**133**, **134**, исходный склареол **131**), не мешающих парфюмерному использованию смеси и не влияющих на парфюмерную оценку.



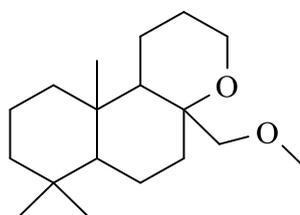
130 Амброксид (10–15%)



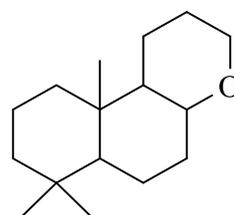
129 Оксид склареола (20–30%)



133



134



135

Запах полученной смеси был охарактеризован парфюмерным советом Института душистых веществ Союзпарфюмерпрома как амбровый, чистый, пригодный к парфюмерному применению.

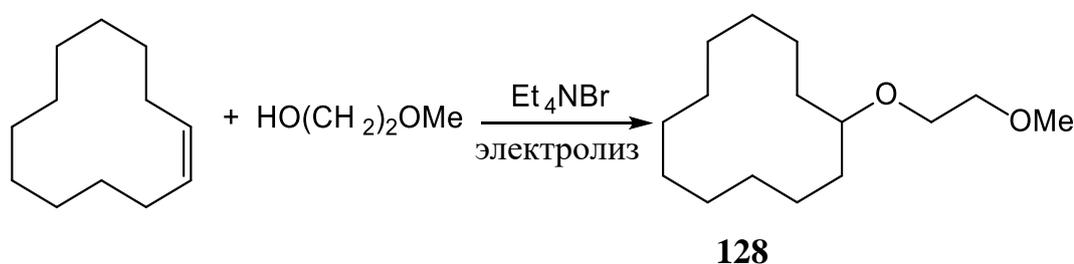
Потенциальный экономический эффект при цене 1 г амброксида \$450 был бы вполне ощутимым, но, к сожалению, опытное производство в Молдавии (сырьевой источник склареола), где было в 1990 г. организовано электрохимическое окисление, вскоре закрылось по причине разрушения экономических связей.

Важен также упомянутый выше еще один заменитель натуральной амбры; в 90-х годах – мало используемый, а сейчас – повсеместно в большинстве амбровых

композиций – амбролигнан (**128**). Он существенно дешевле амброксида, который остался на данный момент, по собственным наблюдениям, только в духах Шанель №5 и №19.

Амбролигнан (**128**) мы также синтезировали в одну стадию, электрокаталитическим способом в присутствии каталитических количеств Et_4NBr , используя открытую в группе М.Н.Элинсона (ИОХ РАН) реакцию электрокаталитического присоединения спиртов к циклоолефинам [327] в условиях, аналогичных получению амброксида (**129**) (Схема 9).

Схема 9. Электрокаталитическое присоединение метилцеллозольва к циклододецену



В последующих работах, используя вышеописанный алгоритм, мы построили модели

- 1) мускусного запаха [328];
- 2) новых структур бицикломочевин с нейротропной активностью [329] (успешно синтезированы).

3.4. Моделирование мускусного запаха

Сформулируем еще раз что нужно сделать для моделирования и прогнозирования запаха:

- Найти геометрические взаимосвязи (система ключ–замок) между элементами структур, обеспечивающих “фиксирующую” комплементарность лиганд–рецептор [330,331] при этом вещества желательно представить максимально точно, т.е. рассчитать геометрию неэмпирическими методами, а комплекс желательно смоделировать молекулярной динамикой, а если это невозможно, то по крайней мере молекулярной механикой.
- Математические методы должны использовать пространственные факторы,

т.е. элементы общей топологии [332], которые вполне адекватны в QSAR расчетах отдельных структур, но межмолекулярные взаимодействия описывают слишком сложным функционалом. Очевидно, что «фиксирующая» комплементарность молекулы и рецептора, которая соотносится с биологической активностью [333,334,335], не определяется статистической обработкой данных обучающей выборки. Более того, невозможно подтвердить комплементарность молекулы и рецептора исключительно на основе молекулярной топологии; дополнительно нужно учитывать:

- 3D-электростатическую комплементарность двух или нескольких молекул, или молекулы и рецептора (дальнее взаимодействие). По нашему мнению, наличие этих параметров и определяет математическую модель этого свойства.

Как уже было сказано, одинаковые по параметрам координата–заряд электрические поля могут быть созданы молекулами, существенно отличающимися по структуре и типу составляющих атомов.

Так что, используя тот же подход, что и в случае амбрового запаха, а именно предположение, что интеграл перекрывания орбиталей молекулы рецептора равны 0 (химической реакции не происходит); дисперсионные взаимодействия также не учитываются; а сравнивать необходимо электростатические поля молекул обучающей выборки, мы провели моделирование мускусного запаха.

Построим поверхность с оптимальным размером, где можно расположить активные центры рецептора. Для этой цели мы применим следующие определения:

(а) модель представляет собой сферу с радиусом 1,4–1,5Å, что соответствует средним Ван-дер-Ваальсовым радиусам атомов, наиболее часто встречающихся в органических соединениях;

(b) Ван-дер-Ваальсова поверхность является поверхностью, полученной покрытием молекулы жесткими взаимопроникающими сферами, построенными вокруг атомных центров и имеющими радиусы, приблизительно равные Ван-дер-Ваальсовым радиусам для соответствующих атомов [336];

(с) поверхность покрытия является кривой точек, создаваемых центром сферы образца, когда он окружен Ван-дер-Ваальсовой поверхностью [336].

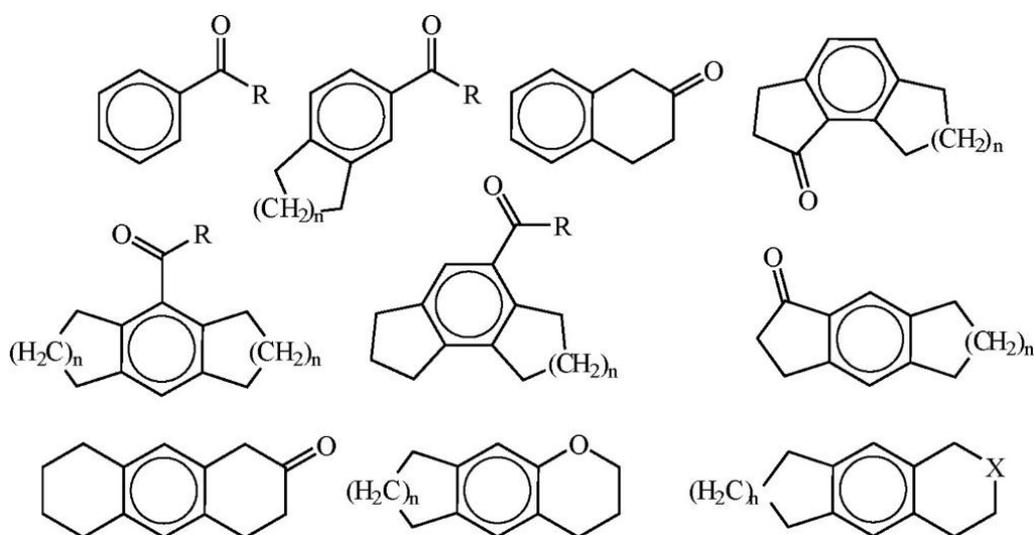
Полученная покрывающая поверхность определяет геометрические ограничения, налагаемые на взаимодействие молекулы и рецептора, поскольку центры атомов рецептора не могут быть расположены внутри поверхности.

После этого для электростатического поля, создаваемого молекулой, устанавливается молекулярный электростатический потенциал всех описывающих точек этой поверхности. Полученная двухпараметрическая поверхность и есть поверхность МЭП. На этой поверхности расположены точки, соответствующие локальным экстремумам МЭП. Комбинация ключевых точек такой поверхности, или «точек представления» молекулы, описывает ее исчерпывающим образом.

Предположим, что появление комбинаций точек с определенными координатами и потенциалами на поверхности потенциальной энергии определяет желаемую электростатическую комплементарность, т. е. выявляет существенное свойство; другими словами, фактор, определяющий заданное свойство молекулы. Это и будет моделью свойства.

Был выбран набор из 110 соединений, 35 из которых обладают выраженным мускусным запахом, тогда как другие – не обладают. Вещества, входящие в набор, являются структурами с различными заместителями, показанными на Рисунке 16. В качестве заместителей использовали следующие группы: H, Alk, NO₂, OAlk, Hal, CHO, C(O)Alk, CN, C(O)Hal.

Рисунок 16. Примеры обладающих и не обладающих мускусным запахом соединений, использованных в обучающей выборке



$n=5,6$; $R = H, Alk, NO_2, OAlk, Hal, CHO, C(O)Alk, CN, C(O)Hal$; $X = O, NH$

Для каждой молекулы, включенной в набор, выполняли оптимизацию геометрии. Оптимизация состояла из двух этапов: начальная оптимизация методами молекулярной механики с использованием силового поля ММХ; конечной оптимизации полуэмпирическими методами (AM1).

Каждая молекула была окружена покрывающей поверхностью, используя сферу с радиусом в $1,5\text{Å}$. Поверхность определялась точками, расположенными на таком расстоянии, что расстояние между ближайшими соседями составляло $0,2\text{Å}$ (эмпирический выбор). Значение МЭП рассчитывалось для каждой точки поверхности. Были найдены точки, где МЭП имеет минимальные и максимальные значения, как в самой точке, так и в ближайших соседях. Набор таких ключевых точек для каждой молекулы использовался для представления формы МЭП.

Диапазон значений МЭП в точках представления был разделен на интервалы, зависящие от частоты появления различных значений МЭП, и был выбран оптимальный шаг дискретизации диапазона $0,2\text{Å}$.

Для каждой молекулы был построен набор дескрипторов типа (E1, E2, R), N, где E1 и E2 – номер интервалов диапазона МЭП для первой и второй точек в паре; R – количество интервалов в диапазоне между точками представления; и N – частотность данного дескриптора для молекулы. Дескрипторы были построены для каждой пары точек представления для каждой молекулы.

Комбинации дескрипторов определяли для всех молекул, принадлежащих к испытываемому набору. Каждому дескриптору присваивалось уникальное имя. Таким образом, вектор (X_1, \dots, X_N) , где X_i – это частотность i -го дескриптора для молекулы.

Процедуру пошаговой регрессии выполняли, используя систему VIBIGON [337] для выбора значимых дескрипторов и построения линейных моделей, описывающих свойство молекулы:

$$P_i = a_0 + \sum (a_j X_{ij}), \quad i = 1 \dots n,$$

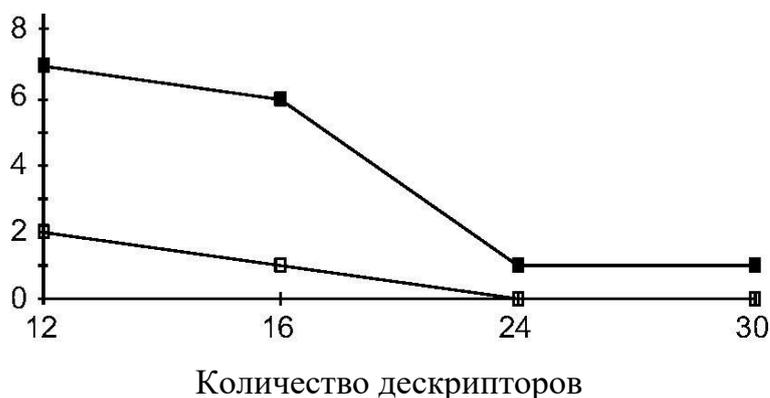
где j – номер критерия, используемого для построения модели.

Классификация предполагала, что j -е соединение обладает запахом, если $P \gg 0,5$, в противном случае соединение считалось без запаха. Ошибка типа 1 – неверное присвоение душистого вещества классу соединений без запаха, в то время

как ошибка типа 2 является неверным присвоением вещества без запаха в класс душистых.

Рисунок 17. Диаграмма распределения погрешности,

■ количество погрешностей типа 1; □, количество погрешностей типа 2



Процедура моделирования была испытана на запахе мускуса (рассматриваем запах как пример биологической активности).

В результате модели, приведенные на Рисунке 17, были получены с различным количеством (оптимизация точности прогноза) дескрипторов.

Сравнительные расчеты показали, что вычисление повторяемости только геометрических параметров (учет пространственных координат и зарядов атомов) без учета электростатических параметров приводит к гораздо меньшей вероятности предсказания желаемого свойства.

Таким образом, полученные результаты предполагают, что появление аналогичных точек (имеющих идентичное взаимное расположение и потенциалы) на поверхности потенциальной энергии задает заданный параметр набора веществ на достаточном уровне вероятности, а метод, предложенный для формирования дескрипторов, адекватно описывает тип изучаемой биологической активности.

3.5. Определение связи структура бицикломочевин (гликольурилов) – нейротропная активность

Тема относится к Этапу 1 – прогнозирование биологической активности конформационно жестких структур, когда структура мишени неизвестна.

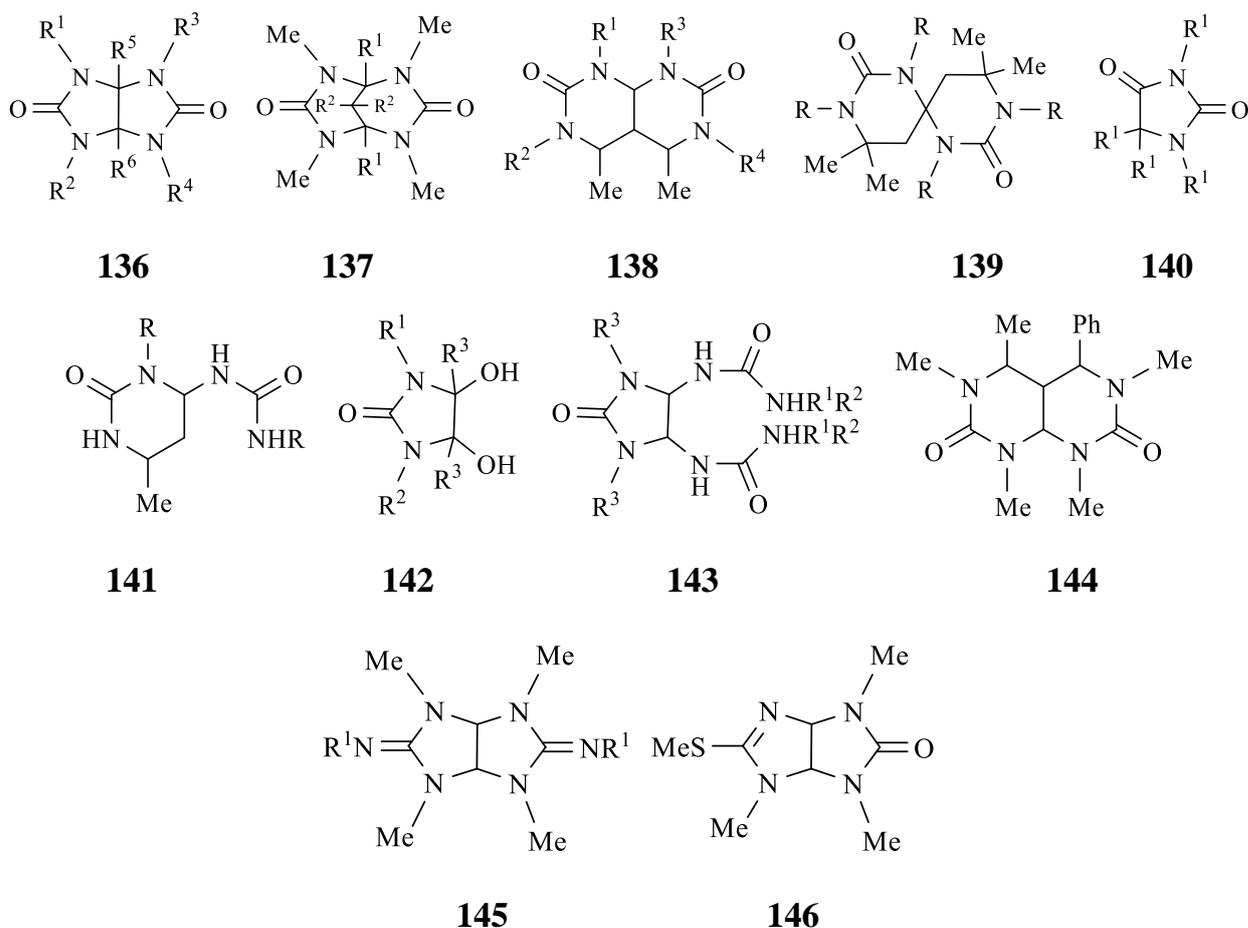
Построена математическая модель психотропной активности на основе расчетов электростатического потенциала молекул соединений, обладающих активностью такого типа; был сделан вывод о том, что этот метод имеет общую применимость для решения проблемы «структура–свойство».

В конце 70-х годов исследователи из Института органической химии им. Н. Д. Зелинского РАН и Казанского медицинского института обнаружили новый класс соединений, обладающих психотропной активностью, а именно конденсированные насыщенные бициклические бис-мочевины (ББМ). С тех пор синтезированы многие ББМ и их гетероаналоги. Изучена их психотропная активность, но зависимость активности этих соединений от структуры до сих пор полностью эмпирическая [338].

Использовали ранее описанную методик [339], чтобы попытаться выявить взаимосвязь «структура–свойство» касательно психотропной активности соединений, содержащих мочевину, гуанидин или сульфониламидные фрагменты. Использовался этот более сложный метод, поскольку попытки выявить связь между структурой и психотропной активностью, предпринятую авторами с использованием простой и эффективной системы со встроенными дескрипторами VIBIGON [340], не позволили получить ожидаемые результаты.

Исследование проводилось в соответствии с общим методом [339] и его конкретной реализацией [105]. Ниже представлена последовательность логических шагов, практическая реализация которой мало отличалась от использованной при анализе мускусного запаха.

- Использовали обучающую выборку из 47 соединений, девять из которых обладают выраженной активностью, 26 соединений неактивны, а остальные 12 соединений имеют промежуточную активность. Используемые соединения разделяли на три класса: соединения класса «1» с ярко выраженной активностью, класс «0» с низкой активностью или отсутствие активности и класс «Z», включающий соединения с промежуточной активностью. Все соединения, входящие в набор, относятся к типам, показанным на Рисунке 18. Кроме того, была сформирована контрольная выборка, которая содержала 21 соединение с неизвестной активностью.

Рисунок 18. Основные структуры обучающей выборки

R^1 – R^4 : H, Me, Et, Pr^n , Pr^i , Bu^n , Ph, CH_2OMe , CH_2OH , CH_2CH_2OH

- Для каждой молекулы исходного набора выполняли оптимизацию геометрии, во-первых, ММХ [341], во-вторых, квантовохимическими методами, используя гамильтониан РМ3 [342].
- Каждая молекула охвачена поверхностью, образованной центром сферы образца (радиус 1.6 Å) [105], как она обтекает поверхность Ван-дер-Ваальса. Поверхность представлена точками с средним расстоянием 0,2Å между ближайшими соседями. Потенциалы на поверхности МЭП рассчитаны в каждой точке поверхности. Затем найдены точки, в которых МЭП имел максимальные или минимальные значения для самой точки и ее ближайшего окружения. Такой набор точек для каждой молекулы использовался для последующего анализа. Мы будем называть эти точки, ключевыми точками (КТ), согласно работе [105].

- Диапазон потенциалов МЭП в КТ был разделен на интервалы в зависимости от того, как часто выполняются конкретные значения МЭП. Для этой цели построена гистограмма распределения МЭП. Затем в точках минимумов и максимумов гистограммы были выбраны необходимые границы интервалов. Кроме того, высокие и широкие пики на гистограмме были дополнительно равномерно разделены.
- Диапазон расстояний между КТ (0–18 Å) разделяли на интервалы (средний шаг 0,2 Å).
- Для каждой молекулы все возможные пары КТ были сведены в таблицу, и создали набор дескрипторов типа (E1, E2, R), N, где E1 и E2 – номера интервалов диапазона МЭП для первой и второй точки в паре; R – количество интервалов в интервале между точками представления; и N – частотность данного дескриптора для молекулы.
- Определяли набор дескрипторов для всех молекул. Каждому дескриптору, принадлежащему набору, присваивался уникальный номер.
- Отбирали значимые качественные признаки, и построили линейные модели, используя систему VIBIGON [340]. В результате мы получили ряд моделей, различающихся по сложности, которые описывают зависимость свойства–структуры с различной точностью (см. Таблицу 9).

В Таблице 9 N_d – количество дескрипторов в модели; 0^Z – количество неправильных признаков соединений типа «0» для типа «Z»; ...; 1^Z – количество неправильных признаков соединений типа «1» с типом «Z»; Ошибки – общее количество ошибок. В каждой ячейке Таблицы 9 находятся два числа. Они показывают количество ошибок типа, указанного в заголовке столбца, при таком количестве дескрипторов в модели, как указано в первом столбце таблицы. Номера, не указанные в круглых скобках, относятся к тестированию модели на обучающем наборе. Числа в круглых скобках взяты из тестирования на испытуемом наборе с использованием метода перекрестной проверки. Ниже приведен метод: во-первых, одно соединение исключается из испытуемого набора и другие вещества используются для построения модели; затем активность исключенного соединения прогнозируется на основе данной модели и сравнивается с фактической активностью этого вещества. После

этого, другое соединение исключается из набора, и повторяется та же процедура. Эта методика повторяется для каждого соединения испытуемого набора. Звездочки в определенных ячейках таблицы указывают на «неисчезающие» ошибки, которые не исчезают по мере увеличения сложности модели.

Таблица 9. Зависимость количества ошибочных признаков от числа дескрипторов, включенных в модель

N_d	0^Z	0^1	Z^0	Z^1	1^0	1^Z	Ошибка
22	0(0)	0(0)	3(3)*	0(0)	0(2)	0(0)	3(5)
20	0(0)	0(0)	3(3)*	0(2)	0(2)	0(0)	3(7)
16	0(0)	0(0)	3(3)*	0(1)	0(2)	0(0)	3(7)
12	1(3)	0(0)	3(4)	0(0)	0(2)	0(1)	4(10)
10	2(3)	0(0)	4(4)	0(0)	0(2)	0(1)	6(10)
8	3(3)	0(0)	5(5)	0(1)	0(2)	1(2)	9(13)

Детальный анализ моделей показал, что три из пяти ошибочно назначенных структур (две ошибки Z^0 и одна ошибка 1^0) имеют высокую конформационную подвижность. Поэтому разумно предположить, что конформация с оптимальной энергией, используемой в расчетах, неактивна, а соединение проявляет активность в другой, менее благоприятной конформации. Это позволяет нам объяснить, почему активность этих соединений недооценивается. Аналогичное объяснение применимо также к оставшейся неправильной структуре (тип 1^0) общей формулы 1, содержащей четыре группы Et в атоме азота. Вышеприведенное предположение подтверждается тем фактом, что количество ошибочных классификаций резко сократилось, когда эти структуры были исключены из испытуемого набора.

Очевидно, что количество соединений с завышенной активностью очень невелико (см. Столбец «Ошибки» в Таблице 10), и они полностью отсутствуют, когда используется достаточное количество дескрипторов. Однако могут существовать структуры, активность которых будет завышена, независимо от того, насколько сложна модель. Это должно указывать на то, что молекула, активная часть которой

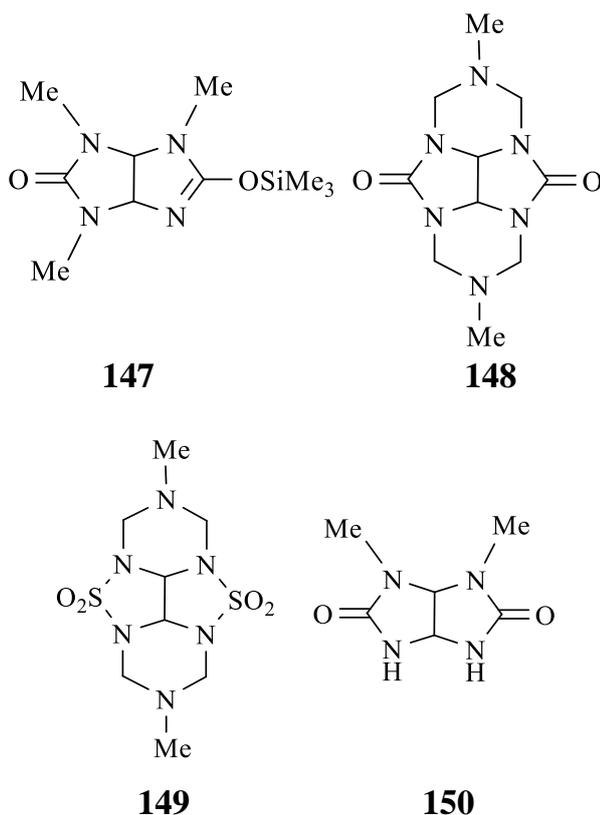
находится на поверхности потенциальной энергии, не может координироваться на рецепторе из-за стерического затруднения. Тот факт, что это явление не наблюдается в изученном множестве, указывает на то, что приведенный выше метод построения дескрипторов достаточно хорошо учитывает не только энергетические, но и геометрические характеристики молекул.

Таблица 10. Количество ошибочных качественных признаков на количество дескрипторов, включенных в модель

N_d	$0 \wedge Z$	$0 \wedge 1$	$Z \wedge 0$	$Z \wedge 1$	$1 \wedge 0$	$1 \wedge Z$	Ошибка
14	0(1)	0(0)	0(1)	0(0)	0(0)	0(1)	0(3)
8	0(0)	0(0)	1(2)	0(0)	0(0)	0(2)	1(4)

Важно отметить, что построение классификации осложнялось отсутствием подробной количественной информации об активности веществ: была доступна только приблизительная классификация с довольно размытыми границами между классами. Например, коэффициент активности был меньше 2,5 для класса «0», более 3,5 для класса «1» и в пределах 2,5–3,5 для класса «Z».

Полученные модели затем использовали для прогнозирования активности других веществ, не включенных в обучающую выборку. Анализ 21 соединения привел к четырем структурам, для которых можно было прогнозировать активность (Рисунок 19). Показано, что структуры **147** и **149** не должны проявлять нейротропную активность, а структуры **148** и **150** – должны.

Рисунок 19. Модельные структуры с предсказанной активностью

Последующий синтез всех четырех структур показал, что отнесение сделано корректно, и структуры **148** и **150** показали активность около 70 и 80% по сравнению с мебикаром, соответственно.

Успешное использование вышеуказанного метода для выявления связи между структурой ББМ и их психотропной активностью позволяет предположить, что этот метод, который по существу является методом определения дополнительных поверхностных потенциалов молекул, также может быть полезен для изучения других видов биологической активности и физико-химических свойств.

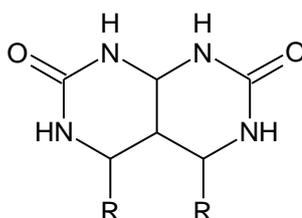
Промежуточный вывод:

3D-QSAR показывает вполне реальные результаты прогнозирования запаха и моделирования тех биологически активных соединений, для которых неизвестны терапевтические мишени.

Логично проверить такие соединения на наборе терапевтических мишеней, показавших участие в координации с лигандами близкой биологической активностью, например, координацию бицикломочевин (гликольурилов, общая формула

представлена на Рисунке 20) с мишенями, связанными с противораковой активностью, из базы данных, содержащей более 1000 терапевтических мишеней, при помощи классического докинга.

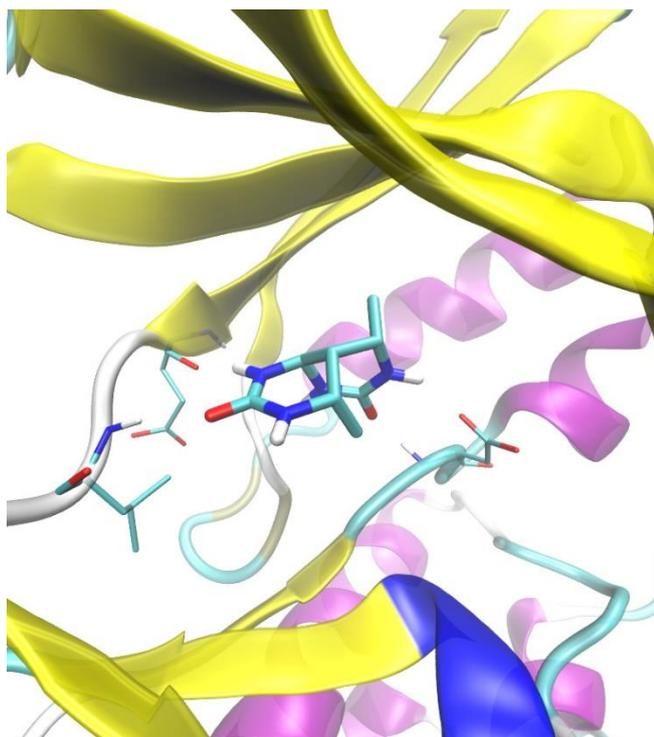
Рисунок 20. Общая формула производных бицикломочевин с потенциальной противораковой активностью



151

Компьютерное моделирование спектра биологической активности виртуальной библиотеки синтетически доступных бицикломочевин проводилось с использованием докингового программного пакета LeadFinder и базы данных, содержащей более 1000 полноатомных трехмерных моделей белков–терапевтических мишеней. Дополнительно использовалась методология структурной фильтрации полученных результатов, позволяющая выявить паттерны связывания, характерные для каждого конкретного белка.

Рисунок 21. Связывание бицикломочевины в активном центре киназы CDK2



При экспертной оценке полученных результатов прогнозировалось, что некоторые производные бицикломочевины (Рисунок 20) способны прочно связываться с циклин-зависимыми киназами CDK2 и CDK5. Известно, что эти киназы напрямую участвуют в цикле клеточного деления и являются терапевтическими мишенями многих онкологических заболеваний. Обнаружено, что при связывании бицикломочевин в активном центре CDK2 и CDK5 происходит образование двух водородных связей, характерных для всех уже известных ингибиторов этих киназ, а также двух дополнительных водородных связей (Рисунок 21).

Результаты компьютерного моделирования (докинг) указывали на то, что производные бицикломочевины могут проявлять ингибирующую активность по отношению к циклин-зависимым киназам, т.е. являются потенциальными противораковыми препаратами.

Однако, *никакой комплементарности в экспериментальной проверке* к мишеням нейро-группы данные соединения не проявили. Возможно, мишени для данной активности пока не определены. Что говорит о том, что *в отдельных случаях не докинг, а именно 3D-QSAR является единственным методом моделирования.*

3.5.1. 3D-QSAR моделирование в направленном синтезе биологически активных замещенных тиогликольурилов с учетом конформационной лабильности лиганда

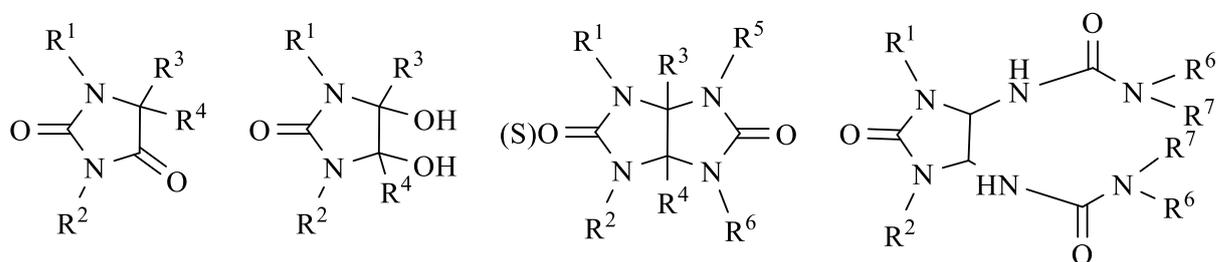
Выше нами показано, что возможность предварительных расчетов потенциальных биологически активных структур может быть обеспечена оптимизацией свободной энергии их связывания с биологическими мишенями [343 – 346]. Объектами исследования в данной части работы выбраны соединения ряда бициклических бисмочевин октанового ряда – гликольурилов, биологическая активность которых (см. выше) никак не коррелируется с имеющимися терапевтическими мишенями, и докинг в данных системах не дал никаких результатов. Единственный способ моделирования структур при отсутствии достоверной конфигурации мишени – классический 3D-QSAR. В данной работе использовался вариант 3D-QSAR, состоящий в квантово-химическом расчете молекулярной электростатической поверхности (МЭП) молекул обучающей выборки, определении экстремумов поверхностей, сравнении точек экстремумов («ключевых точек») [105,347].

Учет конформационных вариаций лиганда при контакте с мишенью, когда неизвестна конфигурация мишени означает, что при расчете необходим перебор всех стабильных конформаций как независимых соединений.

Модель, состоящая из общих точек МЭП структур соединений обучающей выборки, как показано ранее, позволяет оценить перспективность тех или иных структур для направленного синтеза[329,348].

Обучающая выборка состояла из 33 активных и неактивных моно- и бициклических мочевинов, которые относятся к структурам четырех типов (Рисунок 22).

Рисунок 22. Типы структур обучающей выборки гликольурилов



152

153

154

155

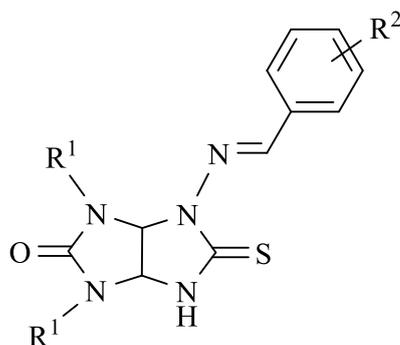
$R^1, R^2, R^5, R^6 = H, Me, Et, Pr, Bu, Ph, CH_2OH, Ac; R^3, R^4 = H, Ph; R^7 = Me, CH_2CH_2OH$

Для них были выполнены расчеты геометрического строения с оптимизацией МЭП. На поверхности МЭП были выделены ключевые точки, соответствующие локальными “вершинам” и “впадинам”. Эти точки были разбиты на несколько классов по величине электростатического потенциала и их положению (на вершинах или впадинах МЭП). Для каждой структуры были определены все возможные пары ключевых точек, которые затем были сгруппированы по множествам, включающим подобные классы точек и расстояний между ними. Количество групп при классификации подбиралось по распределению значений так, чтобы можно было определить границы плотных групп значений (границы проводились по местам, где плотность значений меньше, чем в других местах). Таким образом, для каждой структуры был получен набор отвечающих ей пар.

Были зафиксированы 10 пар точек, описывающих седативную активность активных соединений обучающей выборки и не встречающихся в структурах

неактивных. Это и есть математическая модель свойства, т.е. набор структурных дескрипторов данного свойства.

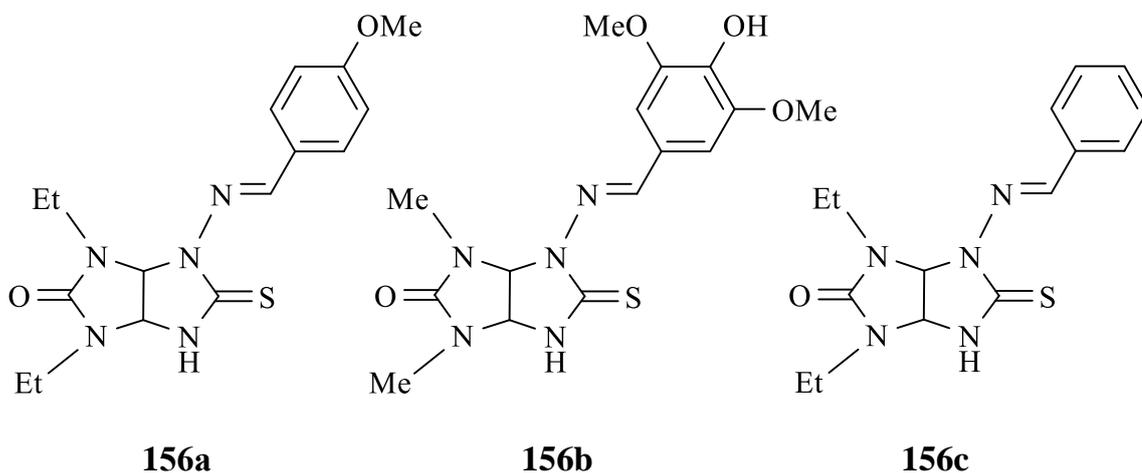
С целью проверки эффективности модели были проверены на совместимость с моделью структуры двенадцати не описанных в литературе производных тиогликольурилов общей формулы **156**.



$R^1 = \text{Me, Et}; R^2 = \text{H, 4-Br, 3(4)-OMe, 4-NO}_2, 2\text{-OH, 4-OH-3,5-(OMe)}_2$

156

В результате для синтетического эксперимента были отобраны три структуры: **156a** – наиболее соответствующая модели; **156b** – с несколько меньшей степенью соответствия; и контрольная **156c**, дававшая неоднозначное совпадение с моделью.



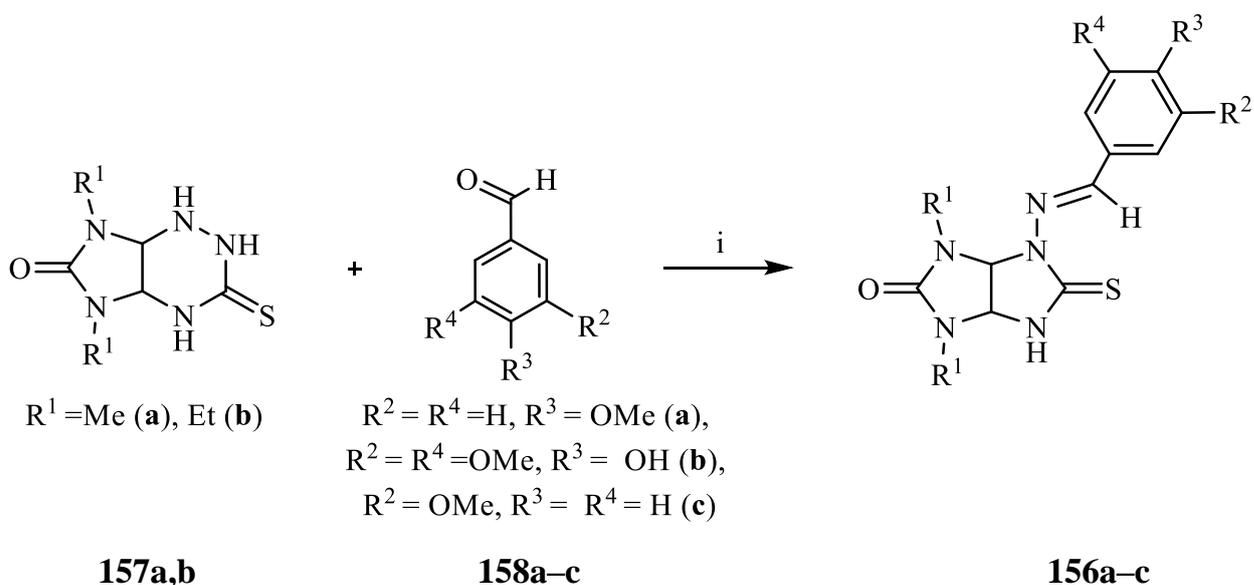
156a

156b

156c

Соединения **156a–c** были синтезированы реакцией 3-тиоксопергидроимидазо[4,5-*e*]-1,2,4-триазин-6-онов **157a,b** с ароматическими альдегидами **158a–c** в метаноле в условиях катализа HCl, аналогично разработанной ранее методике (Схема 10) [349].

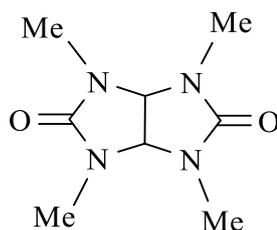
Схема 10. Синтез потенциально активных гликольурилов



Реагенты и условия: i, MeOH, HCl_{конц}, кипячение, 1 ч

Строение синтезированных соединений подтверждено данными элементного анализа, ЯМР спектроскопии и масс-спектрометрии.

Исследованы психофармакологические свойства синтезированных соединений **156a–c**. В качестве препарата сравнения выступал мебикар (1,3,4,6-тетраметилтетрагидроимидазо[4,5-*d*]имидазол-2,5(1*H*,3*H*)-дион) **154a** – дневной транквилизатор, оказывающий, наряду с седативным, также и анксиолитическое действие [350].



154a ($R^3, R^4 = \text{H}$; $R^1, R^2, R^5, R^6 = \text{Me}$)

Влияние соединений на двигательную и исследовательскую активность оценивали в тесте «открытое поле», анксиолитическую активность – в тесте «приподнятый крестообразный лабиринт» на беспородных белых мышах-самцах массой 22–24 г. Все данные обрабатывались статистически по критерию Стьюдента и сравнивались с данными контрольной группы. Разница в оценках считалась достоверной при $p < 0,05$.

Полученные результаты показали полное совпадение предсказанной и измеренной экспериментально седативной активности синтезированных соединений. **156a** проявило выраженный седативный эффект, снижая двигательную и исследовательскую активность мышей как в тесте «открытое поле», так и в «приподнятом крестообразном лабиринте». Соединение **156b** обладает средним влиянием на центральную нервную систему, снижая двигательную и повышая исследовательскую активность мышей, но только в тесте «открытом поле». Соединение **156c** не влияет на поведенческие характеристики животных.

Препарат сравнения мебикар снижает количество пересеченных квадратов в открытом поле в 1,5 раза и общее количество заходов в рукава крестообразного лабиринта в 2 раза, то есть оказывает угнетающее действие на двигательную активность животных. Вместе с тем, он не влияет на исследовательское поведение животных и проявляет анксиолитический эффект, увеличивая время нахождения животных на открытых рукавах лабиринта. По сравнению с мебикаром все исследованные соединения не обладают анксиолитической активностью в тесте «приподнятый крестообразный лабиринт».

Таким образом, продемонстрирована эффективность методов QSAR для предварительного моделирования структуры потенциально эффективных соединений с целью последующего направленного синтеза биологически активных бициклических бисмочевин ряда тиогликольбурилов.

3.6. Докинг, дискретные взаимодействия

3.6.1. Методология и последовательность действий

Подготовка и валидация полноатомных моделей структур терапевтических мишеней в части данной работы, посвященной взаимодействиям точка–точка (классический докинг) производится с помощью программного пакета Lead Finder, разработанным в ИОХ РАН. В пакет входит утилита `build_model`, которая рассчитывает ионизационное состояние аминокислот белка при заданном значении pH, оптимизирует положения подвижных атомов водорода, достраивает неразрешенные (в рентгеновской структуре) боковые радикалы аминокислот. Обычно подготовке

моделей белка уделяется второстепенное внимание, и достройка недостающих атомов водорода осуществляется с помощью программ, которые не проводят детальных расчетов ионизационных свойств белка (например, такие программы, как SwissPDB Viewer, Gromacs и т.д.). Методология подготовки полноатомных моделей программы build_model была ранее валидирована расчетом ионизационных свойств свыше ста белков, для которых эти свойства были определены экспериментально, и было получено хорошее согласование расчетных и экспериментальных данных.

Для валидации подготовленных моделей к каждой из терапевтических мишеней докируется набор соответствующих нативных лигандов (соединений, обладающих сродством к данному белку, согласно имеющимся в литературе экспериментальным данным) и рассчитывается энергия образования комплекса белок–лиганд.

Набор нативных лигандов (не менее 10 для каждой мишени) берется из открытых научных источников: баз PDB, AffinityDB, BindingDB, PDBbind. В случае правильного докирования нативных лигандов делается вывод о применимости модели лиганд–терапевтическая мишень для поиска соответствующих лекарственных прототипов. В случае необходимости (конформационной подвижности белка) для некоторых мишеней используется несколько структурных моделей. Описанный метод валидации моделей является малораспространенным в научной практике в силу его трудоемкости, однако он необходим для получения надежных результатов.

Методы виртуального скрининга библиотек химических соединений реализованы также при помощи программного пакета Lead Finder, зарекомендовавшего себя как одной из самых точных на сегодняшний день программ для молекулярного докинга [351].

Библиотеки химических соединений, синтетически доступных в выбранном классе соединений, конструируются с помощью специально разработанных программ на языке Perl. Процесс фильтрации данных, предоставляемых библиотеками химических структур:

- классический QSAR, регрессионный анализ
- байесова статистика
- группировка

- нейронные сети
- дерево решений.

На основании виртуального скрининга сконструированных библиотек относительно выбранных мишеней отбираются соединения, обладающие наибольшей потенциальной активностью для каждой мишени. Ранжирование соединений по потенциальной активности (прочности комплекса с белком–мишенью) определяется на основании специально разработанной экспертной системы. Для каждой (найденной методом молекулярного докинга) структуры комплекса белок–лиганд экспертная система определяет вероятность образования такого комплекса, более точно – оценку вероятности того, что рассчитанная методом докинга свободная энергия образования комплекса не является результатом ложноположительного предсказания.

Подобная экспертная оценка необходима для того, чтобы уменьшить (идеально – устранить) ложноположительные предсказания, когда комплекс, энергия которого пренебрежимо мала (в условиях эксперимента связывание лиганда с белком невозможно детектировать) получает выгодную энергетическую оценку от докинг-программы (что является общим принципиальным ограничением методов докинга).

Экспертная система включает две численные оценки:

- а) значение дополнительной скоринговой функции, приписывающую действительное число – аналог энергии связывания – каждому комплексу белок–лиганд
- и б) значение дискретной функции, определяющей является ли данное соединение истинно активным, или ложно активным.

Новыми подходами, предлагаемыми в данной последовательности действий, по сравнению с общепринятыми методиками, являются

- а) тренировка дополнительной скоринговой функции на рядах известных активных соединений для каждой мишени. Тренировка дополнительной скоринговой функции для отсева ложноположительных предсказаний также была апробирована на примере одного фермента.

- б) созданием т.н. структурных фильтров – структурных критериев образования комплексов белок–лиганд для каждой выбранной мишени. Метод применения

структурных фильтров для многократного повышения точности виртуального скрининга был апробирован на примере нескольких терапевтических мишеней.

Совмещение обоих подходов в сочетании с наиболее точным из доступных на сегодняшний день методов докинга продемонстрировало ожидаемо высокие результаты.

Практическая проверка достоверности экспертных оценок.

Все соединения, заложенные в начальные группы для экспертных оценок, синтезированы и практически проверены на заданную активность. Тем самым схема докинга, как и схема QSAR, не заканчивается на построении модели, как большинство теоретических изысканий в этой области, а имеет 100% практического приложения в виде органического синтеза соединений с заранее заданными свойствами.

Биохимические методы

Испытания синтезированных соединений проводятся в моделях *in vitro* с использованием очищенных белковых препаратов. Данный этап необходим для однозначного подтверждения активности синтезированных соединений, а, значит, и качества разработанной экспертной системы. Также эксперименты на белках *in vitro* необходимы для дальнейшего планирования более сложных экспериментов, включающих испытания на клеточных моделях и животных, интерпретация результатов которых может быть более сложной.

Экспериментальная проверка синтезированных соединений на выбранных мишенях выполнена в организациях, специализирующихся на подобных работах (ИБХ РАН, РОНЦ РАМН, и др.), где подбираются необходимые мишени и экспериментальные методики.

В качестве объектов исследования методом возмущения свободной энергии были выбраны терапевтические мишени: Полиаденинрибозополимераза (ПАРП1) и Сук-киназы. Сук-киназы – это мишень ревматоидного артрита и неходжкинской лимфомы. ПАРП1 – это мишень сердечной недостаточности и различных онкологических заболеваний:

В настоящее время ингибиторы этих мишеней, найденные вслепую зарубежными компаниями, проходят медицинские испытания. И поэтому создание

отечественных аналогов, полученных с использованием молекулярного моделирования, имеет большое значение.

3.6.2. Ингибиторы ПАРП1, определенные на базе классического (дискретного) докинга

Поли(АДФ-рибозо)-полимераза-1 (ПАРП1, PARP1) – фермент, который, локализуясь в ядре клетки, катализирует поли-АДФ-рибозилирование белков [352]. NAD⁺ выступает в качестве доноров АДФ-рибозы. В каталитическом комплексе ковалентная модификация ПАРП1 регулирует активность белков, участвующих в регенерации повреждений ДНК и поддержании стабильности генома, транскрипции генов, пролиферации и дифференцировки клеток, и другие процессы [353,354] таким образом, что подавление действия ПАРП1, т.е. подавление репарации и механизмов выживания раковых клеток в ходе химио- и радиотерапии, является перспективной стратегией для лечения рака.

Профилактика некроза, вызванного сокращением ресурсов NAD⁺, которая происходит в инфаркте миокарда и других патофизиологических состояниях, является еще одним терапевтическим применением ингибиторов ПАРП1 [355]. Наконец, ингибирование ПАРП1 стимулирует транскрипцию провоспалительных генов, и может быть полезным в терапии сердечно-сосудистых заболеваний [352]. В литературе известны восемь ингибиторов ПАРП1, прошедшие клинические испытания – как монотерапевтические препараты, так и комбинации с другими противоопухолевыми препаратами [356]. Однако безопасность длительного использования этих соединений не подтверждена, поэтому возникает необходимость в создании широкого спектра новых ингибиторов ПАРП1.

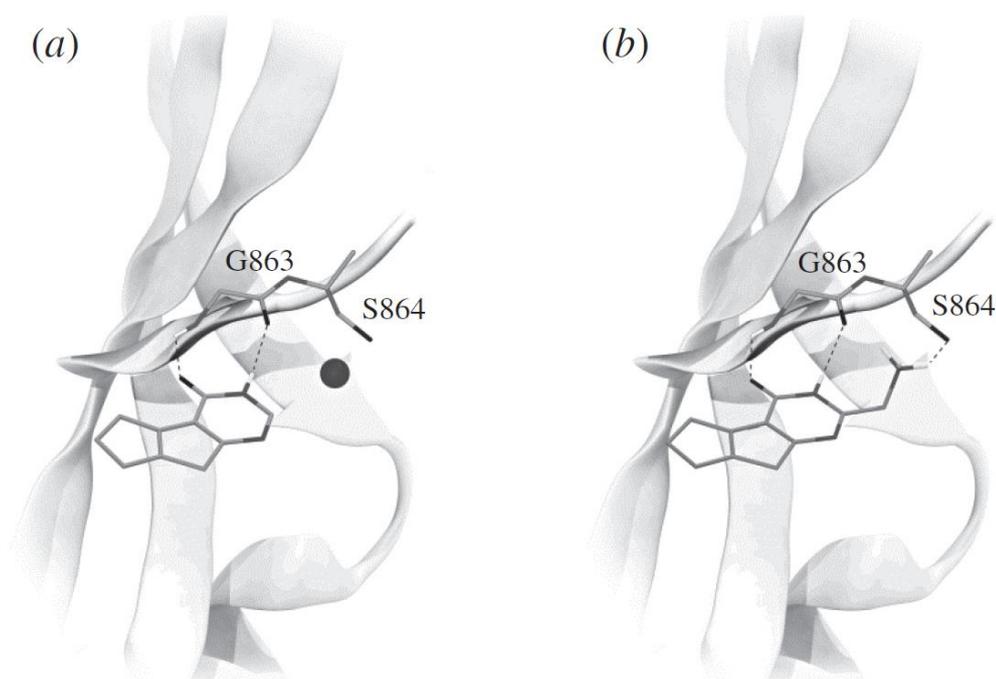
Мы использовали высокопроизводительный виртуальный скрининг библиотеки фрагментов для поиска новых ингибиторов ПАРП1. Скрининг фрагментов – это парадигма рационального дизайна лекарственных средств в *этапе 3* (дискретный докинг), когда комплементарные связывающие фрагменты сначала отбираются из библиотеки низкомолекулярных соединений с мольной массой 150–300, а затем энергия связывания усиливается путем модификации или добавления дополнительных фрагментов [357,358]

Фрагмент, который наиболее сильно привязан к ПАРП1, был выбран во время

виртуального скрининга. Рассчитанная структура комплекса **159** с киназой (Рисунок 23(a)) характеризуется корреляционными водородными связями между карбонильной и NH группами пиримидинового кольца и остатком G863. Остаток G863 не изменяется в различных изоформах ПАРП1; он входит в набор структурных фильтров, описывающих ферментный каталитически активный участок [359]. В дополнение к корреляциям через водородные связи, обнаружено комплексообразование циклических систем через π -взаимодействия с Y896 и Y907 группами.

Рисунок 23. (a) Картина связывания соединения **159** в активном сайте ПАРП1 по результатам моделирования. Молекула воды показана как черная сфера.

(b) связывание соединения **164** с ПАРП1

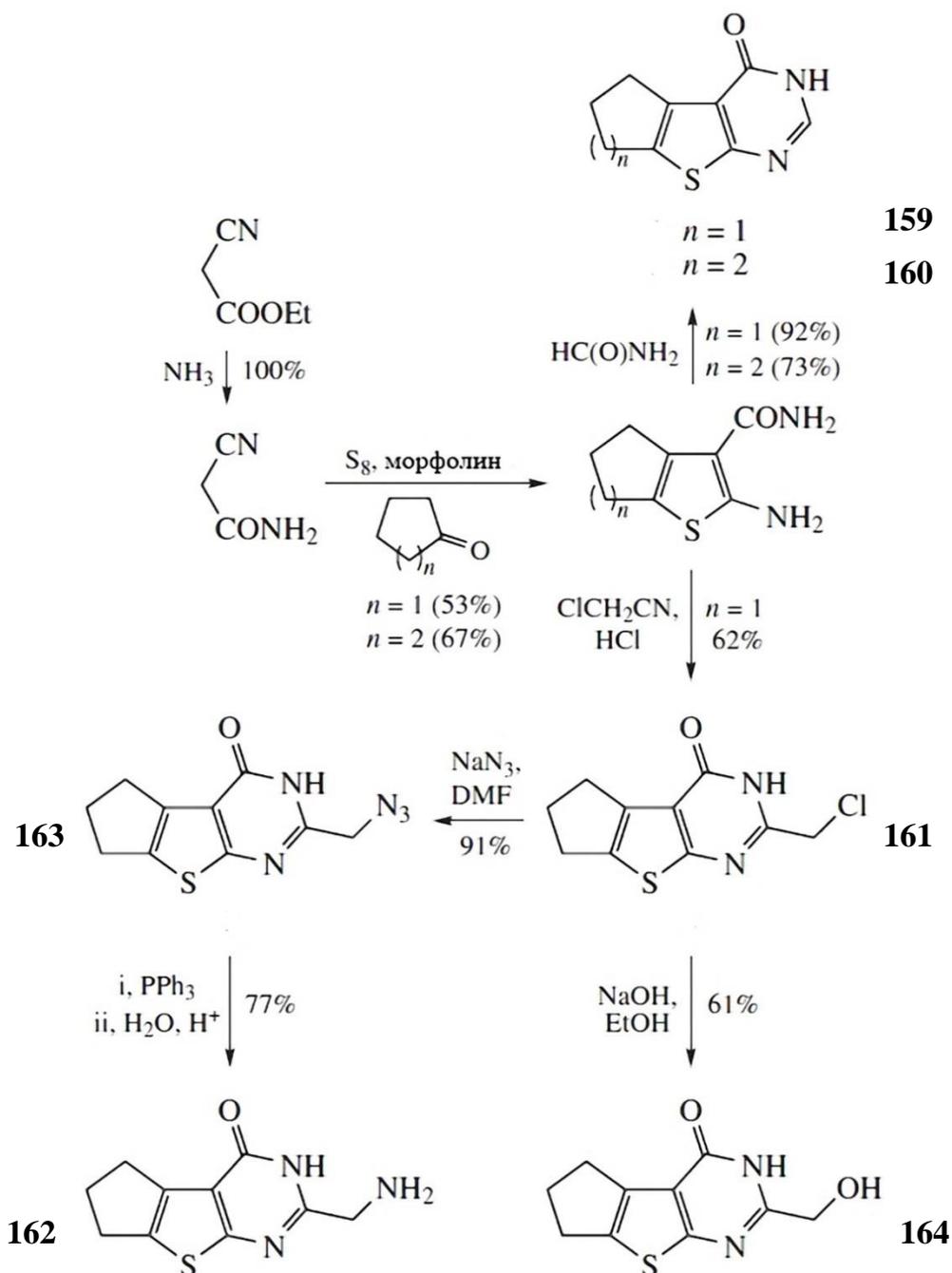


Анализ известных структур ПАРП1 с различными ингибиторами показал, что связывание молекулы воды происходит вблизи основного фармакофорного сайта связывания группы. Основная роль в связывании воды принадлежит водородным связям с остатками G863 и S864. Пространственное отклонение атомов кислорода воды при этом составляет всего $0,5\text{\AA}$, что свидетельствует о сравнительно жесткой структурной локализации молекулы. Известно, что подобные консервативные привязки позиций молекул воды играют очень важную роль в комплексообразовании

белок–лиганд [360]; следовательно, они должны быть учтены при рациональном дизайне ингибиторов. В зависимости от белкового окружения и природы вытесняющей воду группы замещение фиксированных на белке молекул воды может привести как к улучшению сродства лиганда, так и к увеличению энтропии и потере энергии, т.е. уменьшению энергии связывания. По результатам молекулярного докинга мы выяснили, что гидрофобные алкильные заместители в соединениях **161–164** вытесняют молекулы воды из фиксированной позиции (Рисунок 23(b)); и, согласно результатам расчетов, имеет место второй вариант – их связывание с ПАРП1 происходит более эффективно, чем у базового фрагмента **159**.

В результате моделирования на основе активного фрагмента 1 были предложены к синтезу структуры **160–164**, которые, по расчетам, должны показать эффективное ингибирование активности ПАРП1. В частности, молекулярный докинг показал, что циклопентановый фрагмент базового соединения **159** связан в полости, которая образована двумя гидрофобными остатками Y896 и Y907, с одной стороны, и заряженными остатками E928 и K843, с другой. По результатам моделирования соединение **160**, в котором пятичленное кольцо заменяется шестичленным, имеет большую площадь гидрофобных контактов с ферментом; следовательно, оно более прочно связано в активном центре мишени.

На основе молекулярного моделирования структур были синтезированы соединения **159–164** (Схема 11) и определены дозозависимые концентрации полумаксимального ингибирования для ПАРП1 (IC_{50}). Синтезы соединений **159** и **160** проводились с использованием опубликованных процедур [361,362] Методики, используемые для синтеза соединений **161–164**, описаны в [363 – 366].

Схема 11. Синтез соединений **159–164**

Рассчитанные константы ингибирования ПАРП1 для соединений **159–164**: 80, 27, 11, 20, 23 и 20 мкМ соответственно.

По данным измерений IC_{50} , эффективность соединения **160** (IC_{50} 160 мкМ, 95% доверительный интервал 141–181 мкМ) значительно ниже, чем у эталонного соединения **159** (IC_{50} 17 мкМ, 95% доверительный интервал 6–39 мкМ). Измерения IC_{50} (Таблица 11) показали, что смещение консервативно связанной воды, по-видимому, не приводит к существенным изменениям в энергии связывания. Сравнение с

константами ингибирования известных ингибиторов ПАРП1, с известными данными, обобщает ингибирующие свойства **159–164** (Таблица 11).

Таблица 11. Новые ингибиторы ПАРП1

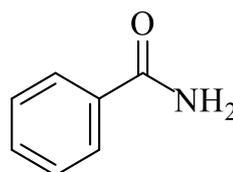
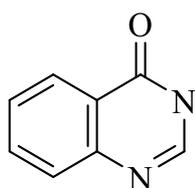
Вещество	IC ₅₀ ^a /мкМ	K _i ^b /мкМ
159	17 (6; 39)	1,21
160	160 (141; 181)	12,3
161	47 (нет данных)	3,6
162	27 (18; 42)	2,1
163	26 (8; 92)	2,0
164	12 (6; 24)	0,9

^a Экспериментально определенные значения IC₅₀ (доверительный интервал – 95%).

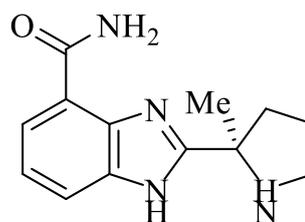
^b Экспериментальные значения K_i, рассчитанные из значений IC₅₀.

Данные, приведенные на Рисунке 24, показывают, что новые соединения по эффективности связывания сопоставимы с бензамидом и производными хиназолина, что подтверждает эффективность предварительного моделирования в органическом синтезе биологически активных соединений.

Рисунок 24



K_i = 1,8 мкМ [367] K_i = 1,8 мкМ [367]



K_i = 0,095 мкМ [368]

K_i = 0,004 мкМ [369]
Велипариб (стандарт сравнения)

3.6.3. Ингибиторы Syk-киназы, классический (дискретный) докинг

Протеинкиназы представляют собой один из наиболее важных классов терапевтических мишеней, особенно в области онкологии. Ингибиторы протеинкиназ получили наибольшее распространение среди новых лекарственных препаратов в последние 5–7 лет. При этом роль подавляющего большинства протеинкиназ как в норме, так и в патофизиологических процессах, в том числе в онкогенезе, далека от исчерпывающего изучения. Поэтому исследование новых ингибиторов протеинкиназ как лекарственных средств будет одним из основных направлений фармакологии в ближайшие годы. Основной проблемой при разработке ингибиторов протеинкиназ является недостаточная селективность их действия. Геном человека насчитывает свыше 500 различных протеинкиназ, поэтому ингибирование (большого количества) побочных киназ для заданного ингибитора определенной – целевой – киназы является распространенным явлением. Вместе с тем, недостаточная селективность приводит к побочным эффектам, зачастую неприемлемых даже для терапии жизнеугрожающих – онкологических – заболеваний.

В дополнение, ингибирование побочных киназ может приводить к противоположному фармакологическому эффекту, поскольку осуществляемое киназами фосфорилирование (белковых) субстратов может носить как активирующее (пролиферацию опухолевых клеток), так и ингибирующее действие.

Благодаря недостаточной селективности применение ряда ингибиторов киназ было в последнее время ограничено или полностью отменено. В связи с этим большое значение имеет возможность моделирования профиля селективности киназного ингибитора в ходе оптимизации его структуры. Для решения данного вопроса необходим достаточно точный метод, который, вместе с тем, обладает приемлемой вычислительной производительностью.

Начало такой работе было положено применением классических методов докинга при моделировании ингибиторов Syk-киназ.

Одним из ключевых медиаторов активации иммунной системы в ответ на введение в организм иммуногена является Syk-киназа [370].

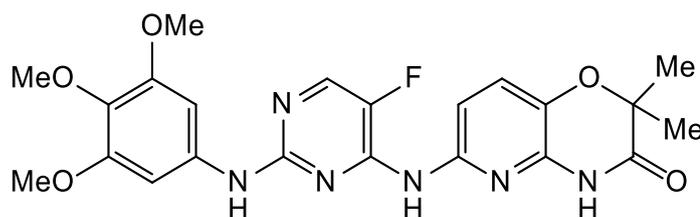
Syk-киназа участвует в различных каскадах сигнальной трансдукции, которые встречаются во множестве типов клеток [371]. Он играет важную роль в иммунном

ответе, являясь перспективной терапевтической мишенью для аутоиммунных заболеваний, таких как ревматоидный артрит и идиопатическая тромбоцитопеническая пурпура [372,373]. Кроме того, активность Syk-киназы необходима для созревания злокачественных В-клеток, и показано, что ингибитор Syk дозозависимо ингибирует пролиферацию нескольких линий клеток неходжкинской лимфомы (НХЛ) [208].

Наиболее перспективная небольшая молекула, нацеленная на Syk-киназу, фостаматиниб (пероральная депо-форма метиленфосфата ингибитора R406 Syk-киназы) (**165**), в настоящий момент проходит клинические испытания III фазы против ревматоидного артрита [373] и клинические испытания II фазы для НХЛ [241]. Несмотря на свою впечатляющую эффективность, которая доказывает терапевтический потенциал ингибирования Syk-киназы, использование токсичных предшественников в виде хлорметилдиалкилфосфата в химическом синтезе молекулы депо-формы и высвобождение формальдегида при превращении в активный метаболит *in vivo*, способствует поиску новых ингибиторов Syk.

В ряде работ [374,375] показано, что ингибиторы Syk-киназы являются эффективными средствами для лечения различных аутоиммунных заболеваний, таких как ревматоидный артрит. Кроме того, поскольку активность Syk-киназы является критичной для созревания В-лимфоцитов [376] оправдано применение её ингибиторов в лечении заболеваний кроветворной системы, связанных с неконтролируемым ростом клеток данного типа.

Клинические испытания ингибитора Syk-киназы [6-(5-фторо-2-[(3,4,5-триметоксифенил)амино]пиримидин-4-ил}амино)-2,2-диметил-3-оксо-2,3-дигидро-4H-пирид[3,2-b,1,4]оксазин-4-ил]метилфосфорной кислоты (фостаматиниб) подтвердили его эффективность для лечения ревматоидного артрита, однако применение препарата, по-видимому, приводит к ряду побочных эффектов, предположительно связанных с недостаточной селективностью таматиниба (R-406, **165**) [374].



165

Таким образом, возникает необходимость создания широкого спектра новых патентно чистых ингибиторов Syk-киназы с улучшенными профилями селективности.

Молекулярное моделирование. В данной части работы для направленного поиска низкомолекулярных соединений, способных к координации в активном центре Syk-киназы, был применен метод молекулярного докинга, использующий пространственные полноатомные модели биологической мишени для геометрической и энергетической оптимизации комплекса лиганд–мишень на базе молекулярно-механических потенциалов [377,378]. Для улучшения результативности моделирования использован метод структурной фильтрации [359].

Идеей метода является дополнительная фильтрация результатов молекулярного докинга наложением комплекса специфических взаимодействий – водородных связей (специфический критерий наличия водородной связи – расстояние менее 3,5 Å между донором и акцептором водорода [309]), транс-влияния, координированных комплексов металлов и др. Лиганды, образующие такие специфические взаимодействия, вероятно, являются вероятностно более активными.

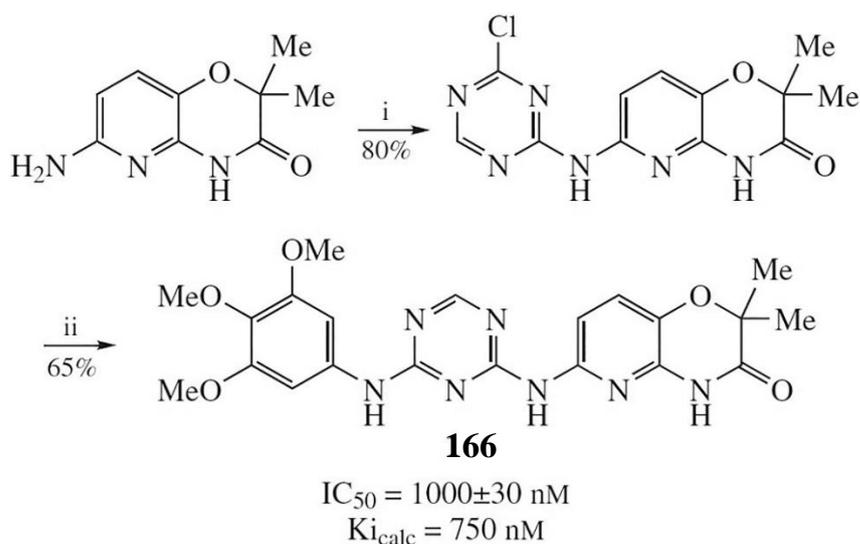
По итогам молекулярного докинга и структурной фильтрации был сформирован набор из 18 возможных ингибиторов Syk-киназы.

Два соединения из этого набора: 2,2-диметил-6-({4-[(3,4,5-триметоксифенил)амино]пиримидин-4-ил}амино)-2*H*-пирид[3,2-*b*,1,4]оксазин-3-(4*H*)-он **166** и метил-2-метил-2-[(2-нитропиридин-3-ил)окси]пропаноат **167** показали наибольшую эффективность ингибирования (Таблица 12, синтез – Схемы 12 и 13). Дозозависимые кривые ингибирования Syk-киназы и соответствующие значения IC₅₀ были получены для обоих соединений.

Таблица 12. Новые ингибиторы Syk-киназы

Соединение	IC ₅₀ , нМ	K _i расчит., нМ
166	1000±30	750
167	230±10	470

Схема 12. Схема синтеза соединения **166**



Реагенты и условия: i, 2,4-дихлор-1,3,5-триазин. ДМФА, 60°C, 8 ч; ii, 3,4,5-(MeO)₃C₆H₂NH₂, ДМФА, 110°C, 6 ч.

Анализ известных рентгеноструктурных данных о комплексах фермента Су-киназы с различными ингибиторами показал, что во всех случаях низкомолекулярные химические соединения формируют водородные связи с –C=O и –NH₂ группами остатка Ala 451 (см. Рисунок 25). В соответствии с идеей структурной фильтрации, эти связи признаны обязательными для образования комплекса лиганд–мишень, и лиганды, не образующие коррелированные водородные связи с остатком Ala 451, считались неактивными.

Рисунок 25. Связывание **166** в активном центре Су-киназы

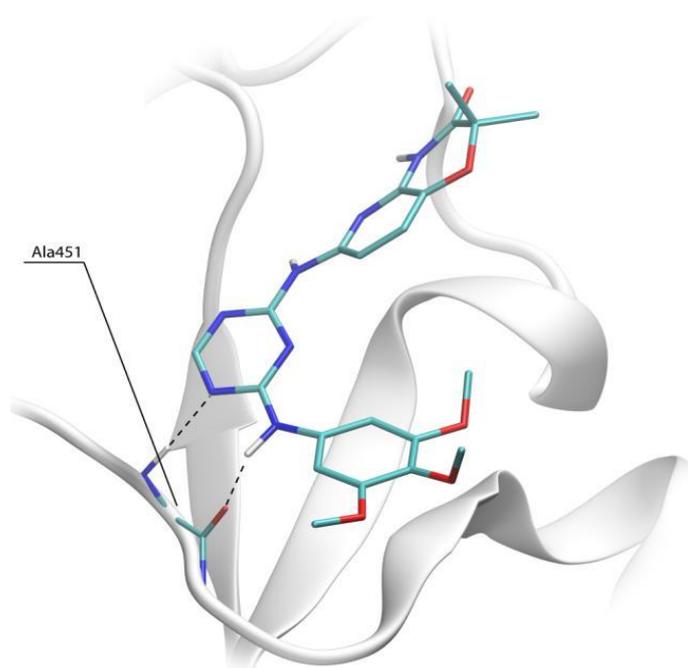
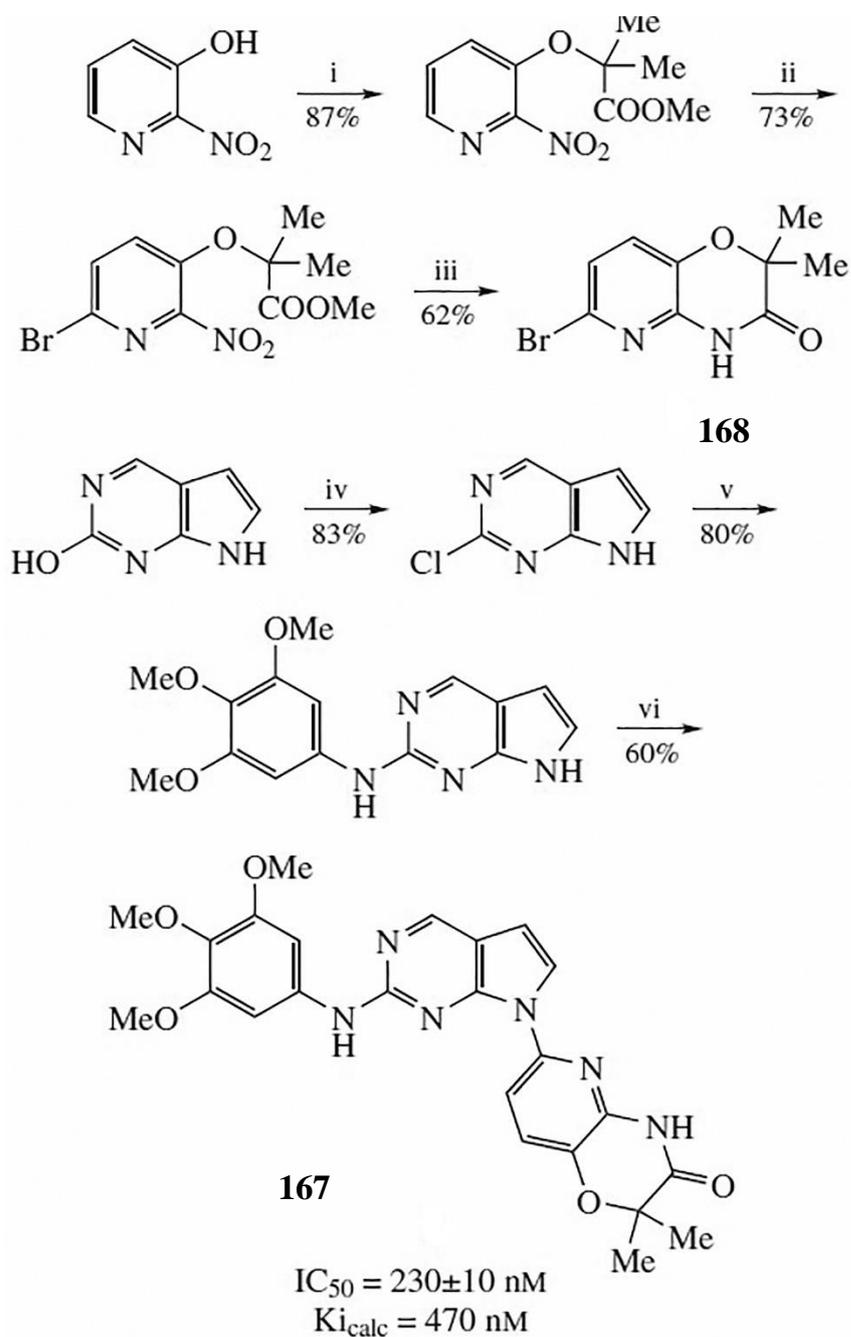
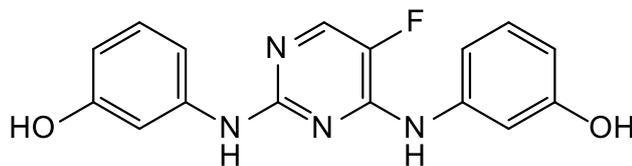


Схема 13. Схема синтеза соединения **167**

Реагенты и условия: i, NaH– ДМФА, далее $\text{Me}_2\text{C}(\text{Br})\text{CO}_2\text{Me}$; ii, Br_2 , AcOH ; iii, SnCl_2 , HCl , $\text{H}_2\text{O}/\text{MeOH}$, далее NaHCO_3 ; iv, POCl_3 , PhNMe_2 60°C ; v, 3,4,5- $(\text{MeO})_3\text{C}_6\text{H}_2\text{NH}_2$, ДМФА, 110°C , 6 ч; vi, ДМФА, **168**, 60°C 8 ч.

Сравнение эффективности ингибирования для известных ингибиторов тама-
 тиниба **165** (IC_{50} , 41 нм) [379] и 5-фтор-2,4-бис(3-гидроксифениламино)пиридина
169 (IC_{50} , 226 нм) [214] с результатами, полученными в настоящей работе (Таблица
 13), показывает, что соединение **167** обладает исключительной эффективностью.

Последнее соединение было выбрано для будущей оптимизации связывания Syk-киназы.



169

Таким образом, предварительное молекулярное моделирование при направленном синтезе биологически активных соединений показало свою эффективность на очередном примере – ингибиторах Syk-киназы.

Таблица 13. Сравнение эффективности ингибирования Syk для синтезированных (**166, 167**) и известных (**165, 169**) соединений.

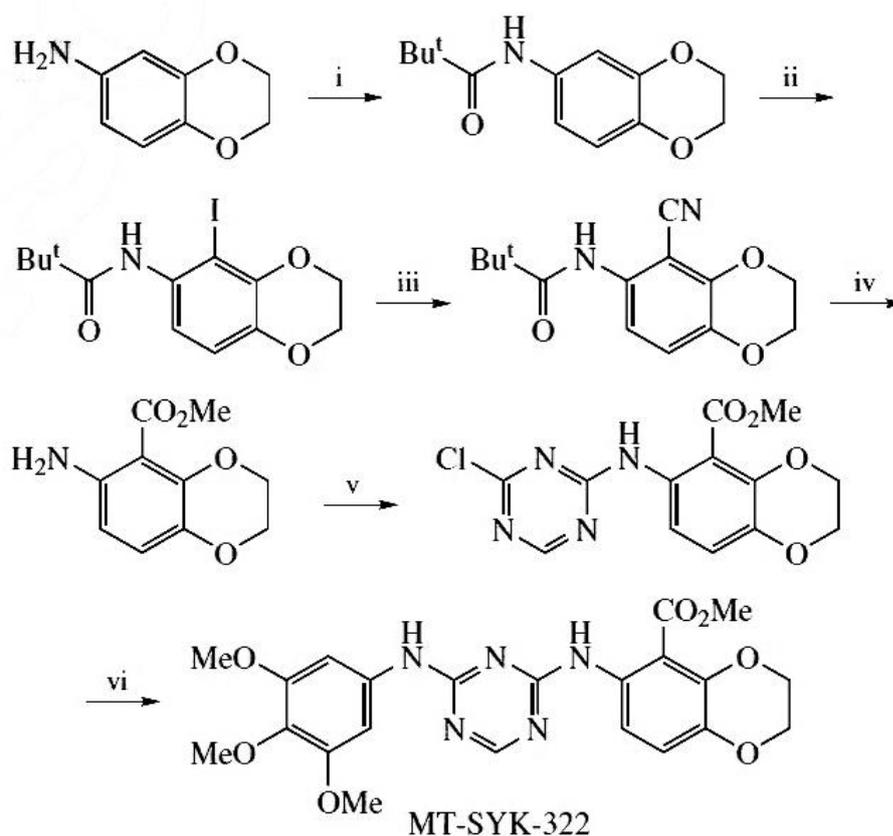
Соединение	IC ₅₀ , нМ
166	1000±30
167	230±10
165	41 [379]
169	226 [374]

Таким образом, предварительным моделированием с последующим направленным синтезом мы получили ингибитор Syk-киназы MT-Syk-03 (**166**), продемонстрировавший равную эффективность с R406 (активный метаболит – фостаматиниб) в клеточных моделях аутоиммунитета и рака со значениями EC₅₀ в субмикромольном диапазоне. Синтезированное позднее модельное соединение MT-Syk-322 было менее активно, чем **165** и **166**.

Была изучена *in vitro* эффективность нового ингибитора Syk-киназы MT-Syk-03 и недавно синтезированного соединения MT-Syk-322 (Схема 14) с помощью клеточных анализов и сравнили его с результатами R406. MT-Syk-03 и MT-Syk-322 были смоделированы с использованием молекулярного докинга с последующей структурной фильтрацией [189,380] и, как ожидается, будут обладать подходящей

фармакокинетикой без пролекарственной формы выпуска из-за повышенной гидрофильности (сLogP для MT-Syk-03 составляет 1,88, тогда как сLogP для R406 – 4,32). MT-Syk-03 продемонстрировал значительную активность в ферментативном анализе *in vitro*, сравнимую с анализом R406 [344].

Схема 14. Синтез MT-Syk-322



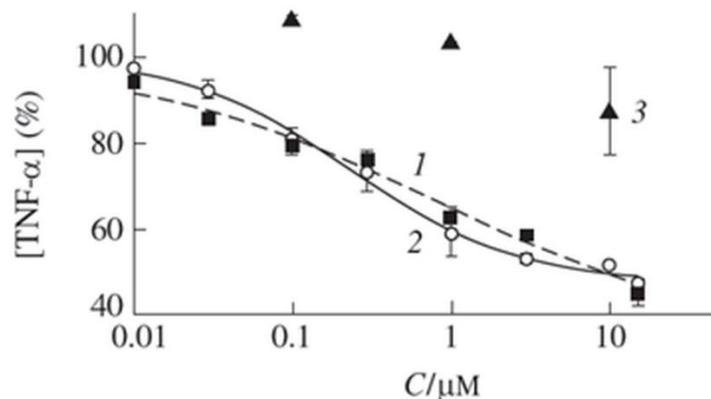
Реагенты и условия: i, $\text{Bu}^t\text{C}(\text{O})\text{Cl}$, Pr^i_2NEt , MeCN ; ii, Bu^tLi , ТГФ, -100°C , далее I_2 ; iii, $\text{Zn}(\text{CN})_2$, PdCl_2 , 60°C , 85 ч; iv, HCl (H_2O), Δ , далее MeOH/H^+ ; v, 2,4-дихлор-[1,3,5]триазин, ДМФА, 60°C , 12 ч; vi, 3,4,5-(MeO) $_3\text{C}_6\text{H}_2\text{NH}_2$, ДМФА, 110°C , 6 ч.

Синтезированные MT-Syk-03 и MT-Syk-322 исследованы во многих дополнительных клеточных анализах, которые имеют значение для ревматоидного артрита и неходжкинских лимфом.

Первый клеточный анализ использовали для оценки эффективности веществ в блокировании аутоиммунного ответа, опосредованного Syk. Syk-киназа, как известно, частично опосредует аутоиммунную реакцию через сигнальный каскад $\text{FC}\gamma\text{R}$, который инициируется IgG-связыванием с поверхностью клетки и, наконец,

вызывает дегрануляцию и выделение цитокинов в различных иммунных клетках [381]. Полученный клеточный ответ в присутствии ингибитора Syk влияет на оценку эффективности ингибитора [382]. В исследованиях *in vitro* (Институт биохимии им. А.Н.Баха РАН, Институт биоорганической химии им. М.М.Шемякина и Ю.А.Овчинникова РАН) были использованы IFN γ -примированные моноциты ТНР-1, стимулированные IgG, для инициации передачи сигналов FC γ R и измерено высвобождение фактора некроза опухоли альфа (TNF- α). Как R406, так и MT-Syk-03 продемонстрировали дозозависимое ингибирование высвобождения TNF- α с EC $_{50}$ 0,24 и 0,85 мкМ, соответственно (Рисунок 26), в то время как MT-Syk-322 ингибировал незначительно. Данные для R406 находятся в хорошем соответствии с ранее известными в литературе, EC $_{50}$ = 0,171 мкМ [382].

Рисунок 26. Ингибирование анти-IgG стимулированного TNF- α , вырабатываемые дифференцированными ТНР-1 моноцитами
(1) R-406, (2) MT-Syk-03 и (3) MT-Syk-322



Во втором исследовании были оценены противораковые свойства MT-Syk-03, MT-Syk-322 и R406. Ингибиторы Syk-киназы могут нарушить путь В-клеточного рецептора (BCR), который необходим для созревания как нормальных, так и злокачественных В-клеток [381]. Подтипы НХЛ имеют морфологию В-клеток и, следовательно, чувствительны к ингибиторам Syk-киназы. Выбраны следующие линии клеток НХЛ для настоящего исследования: NAMALWA, РЗНЗ, РЗНР-1, исходя из доступности в банке клеток Института цитологии РАН (RCCC, <http://www.rccc.cytspb.rssi.ru/>). Влияние веществ на жизнеспособность клеток приведено в Таблице 14 и на Рисунке 27.

Рисунок 27. Влияние ингибиторов (1) R-406, (2) MT-Syk-03 и (3) MT-Syk-322 на жизнеспособность клеток NAMALWA, P3H3, P3HR-1 (a, b и c, соответственно)

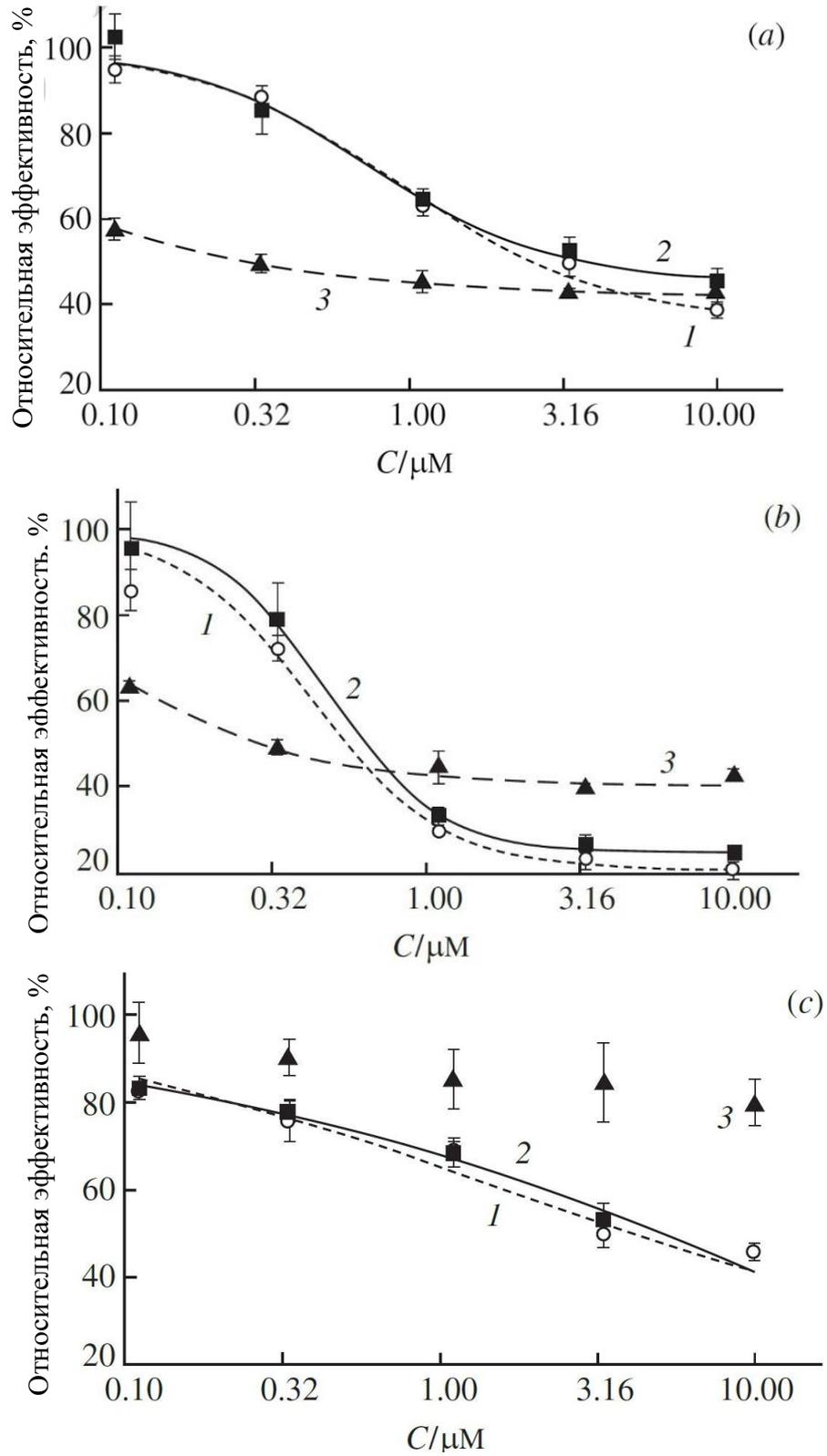


Таблица 14. Влияние ингибиторов (EC_{50}) на жизнеспособность клеток линии NHL.

Линия	MT-SYK-03	MT-SYK-322	R406
NAMALWA	0.74±0.12	0.04±0.007	0.95±0.15
P3H3	0.48±0.08	0.06±0.02	0.43±0.05
P3HR-1	4.48±2.09	> 10	1.61±1.21

Как R406, так и MT-Syk-03 продемонстрировали дозозависимую цитотоксичность с EC_{50} в субмикромольном диапазоне во всех исследованных клеточных линиях. MT-Syk-322 дозозависимо ингибирует пролиферацию NAMALWA и P3H3 (сублинии P3HR-1), но не клеточную линию P3HR-1. Такое различие потенциально может быть использовано для выявления дополнительных клеточных целей MT-Syk-322, если появится возможность подробного сравнения клеточных линий P3H3 и P3HR-1.

В итоге всех исследований новых ингибиторов Syk-киназы, новое соединение MT-Syk-03 (часто применяется название, данное этому ингибитору в испытаниях биологической активности его структурных производных – MT-030) продемонстрировало сходную эффективность как R406 в клеточных моделях аутоиммунитета и рака и, следовательно, было намечено для дальнейшей доклинической проверки.

3.7. Докинг, интегральные взаимодействия. Метод возмущения свободной энергии (ВСЭ)

3.7.1. Моделирование протеинкиназ методом ВСЭ

Описанные выше методы докинга [344,359,380,383 – 385] в направленном синтезе биологически активных соединений, в том числе ингибиторов протеинкиназ [385,386], показали, что для успешного прогнозирования биологической активности необходим более точный подход к расчету значений энергии связывания белок–лиганд.

Ранее были описаны достижения в области принципиально новых подходов к моделированию статистическо-термодинамических свойств молекулярных ансамблей (заселенность определенных конформаций белка и лиганда, заселенность определенных состояний ионизации белка, вклад различных конформеров лиганда в свободную энергию связывания с белком) [387,388]. В частности, был предложен принципиально новый граф-теоретический алгоритм для расчета заселенностей определенных (конформационных, ионизационных и т.д.) состояний молекулярной системы (белок, комплекс белок–лиганд и т.д.), насчитывающей астрономическое число микросостояний. Разработанный подход – алгоритм TSAR (Thermodynamic Sampling of Amino acid Residues).

Алгоритм TSAR был успешно применен для моделирования ионизационных свойств белков [387]. Этот уникальный алгоритм был усовершенствован для расчета свободной энергии комплексов лиганд–мишень для наших объектов – когда лиганд имеет слаборабильные химические связи, вращение заместителей вокруг которых происходит в микросекундном и более длительном интервалах времени, что находится за пределами практических возможностей моделирования (наносекундный интервал).

В этом усовершенствованном подходе впервые была реализована идея, в которой возмущаемым параметром выступало равновесное значение торсионного угла (прилегающего к вращаемой связи) [388]. Это позволило осуществить гладкую трансформацию всех конформеров молекулы друг в друга и определить суммарное значение свободной энергии связывания как вклад от индивидуальных конформеров. Данный подход – теоретически точный для учета вклада как конформеров, так

и других медленных степеней свобод – имеет огромное практическое значение [248].

Применение методов QSAR или молекулярного докинга не дает достаточно точных оценок относительного сродства ингибитора протеинкиназ. С другой стороны, использование методов квантовой химии для обсчета серий соединений на массиве сотен белков было бы невозможно в силу вычислительных ограничений. Наиболее оправданным подходом с точки зрения точности и вычислительных затрат было бы применение методов теории возмущения свободной энергии в молекулярно-механическом приближении для нахождения разности свободной энергии связывания родственных соединений, исследуемых в ходе оптимизации структуры лекарственного кандидата. Несмотря на применение данных теоретических подходов к различным задачам вычислительной химии, до сих пор не было попыток моделировать селективность ингибиторов киназ с его помощью.

Мы имеем задел в разработке новых методик и применении методов возмущения свободной энергии для решения разных задач в моделировании органических и биоорганических задач. Подобное исследование применительно к моделированию профиля селективности киназных ингибиторов может иметь важное фундаментальное значение для расширения инструментария методов возмущения свободной энергии, и впоследствии – прикладное значение для дизайна и направленного синтеза новых лекарственных препаратов на основе ингибиторов протеинкиназ.

С помощью усовершенствованного метода моделирования прочности связывания органических соединений с биомолекулами – метода возмущения свободной энергии – были определены прочности связывания известных ингибиторов ПАРП1 и Суk-киназы, а также смоделированы структуры новых биологически активных соединений. Дополнительное повышение точности при моделировании достигается путём явного учёта растворителя в системе белок–лиганд и детального анализа конформационного равновесия ингибитора в растворе.

3.7.1.1. Моделирование ПАРП1 с использованием метода возмущения свободной энергии.

Как описано в соответствующем разделе обзоре литературы, метод точечного классического докинга был усовершенствован, и разработан новый способ подготовки полноатомной структуры мишени для молекулярного докинга и последующих

расчётов методом возмущения свободной энергии (ВСЭ), или FEP, Free Energy Perturbation. Данный подход учитывает ряд факторов, критично влияющих на точность последующих вычислений: состояние протонирования ионизируемых остатков белка, геометрия боковых радикалов белка, а также сольватация комплекса белок–лиганд.

Для первичного приближения ΔG связывания при координации электростатически комплементарных объектов достаточно минимизировать полную энергию комплекса в равновесном состоянии, т.е. определить зависимость полной энергии координирующихся частиц от их расстояния и положения в пространстве методом ВСЭ/FEP.

В рамках данного подхода сначала определяется ионизационное состояние ионизируемых остатков белка и проводится реконструкция плохо разрешённых боковых радикалов. Для решения данных задач применяется алгоритм TSAR [248], разработанный в нашей лаборатории ИОХ РАН.

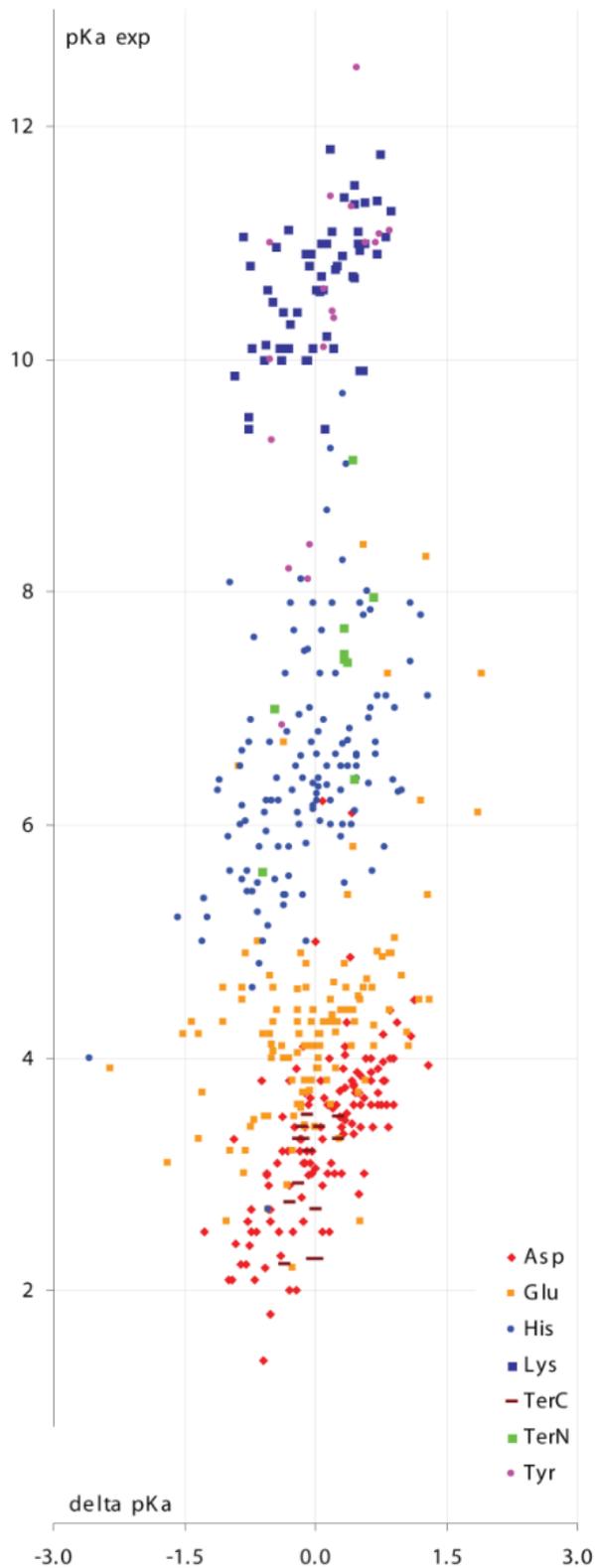
Детализация базовых принципов и последовательности выполнения алгоритма TSAR описана в Экспериментальной части (раздел 5).

В рамках алгоритма белок рассматривается как независимая система лабильных функциональных групп на жесткой полипептидной основе.

Белку сопоставляется граф, вершинами которого являются структурно независимые группы белка (функциональные группы мишени, лиганды или вода). Каждой вершине соответствует сумма параметров (пространственных или электростатических), описывающих конфигурационное пространство белка. Ребрам графа соответствует таблица парных энергий взаимодействия параметров вершин, объединенных данным ребром. Полученный граф представляет собой Байесову сеть (Belief Network), в которой конкретные параметры узла зависят от набора параметров соседних узлов.

Суммарная ошибка расчета констант ионизации определённых значений pK_a по набору 484 аминокислотных остатков из 96 различных белков составила менее 1 ед. pK_a (Рисунок 28).

Рисунок 28. Соответствие расчетных и полученных экспериментально значений рКа для набора 484 аминокислотных остатков



Точность реконструкции боковых функциональных групп с помощью алгоритма TSAR составила 88,3% по углу χ_1 и 80,4% по углам χ_1 и χ_2 (Таблица 15).

Таблица 15. Точность моделирования геометрий боковых функциональных групп на структурах из тестового набора

Тип остатка	Кол-во остатков	Все		Поверхностные		Погруженные	
		χ_1	χ_1 и χ_2	χ_1	χ_1 и χ_2	χ_1	χ_1 и χ_2
Аргинин	616	81.5%	68.2%	74.6%	56.9%	86.2%	75.7%
Аспарагин	646	89.2%	80.2%	81.4%	71.0%	97.5%	88.0%
Аспарагиновая кислота	804	88.9%	77.9%	85.6%	73.4%	94.3%	82.9%
Цистеин	248	95.6%		91.7%		95.6%	
Глутамин	522	82.8%	68.8%	72.8%	49.4%	94.4%	87.2%
Глутаминовая кислота	651	73.3%	60.1%	65.2%	45.1%	84.8%	80.0%
Гистидин	293	94.9%	86.0%	88.4%	84.1%	98.6%	88.4%
Изолейцин	725	97.0%	85.2%	97.3%	77.0%	97.3%	88.5%
Лейцин	1011	93.5%	86.7%	89.4%	83.2%	96.0%	89.7%
Лизин	727	81.8%	65.7%	73.6%	53.8%	91.4%	77.4%
Метионин	218	84.9%	72.0%	71.1%	50.0%	90.9%	80.3%
Фенилаланин	520	96.7%	93.8%	86.8%	84.2%	97.6%	95.2%
Серин	874	70.3%		62.5%		78.9%	
Треонин	884	91.5%		89.4%		93.8%	
Триптофан	232	94.8%	94.8%	88.9%	88.9%	96.2%	96.2%
Тирозин	478	97.1%	94.1%	87.3%	82.5%	99.0%	96.9%
Валин	929	91.3%		81.5%		93.5%	

Для определения положения консервативных молекул воды предложен способ явного учёта сольватации (анализ сольватационных карт растворителя). Данный способ включает проведение молекулярно-динамических расчётов системы белок–

лиганд, кластеризацию полученного набора положений молекул растворителя в белке и выделении наиболее стабильных сайтов сольватации. Пример корреляции карт растворителя приведен на Рисунке 29, пример сайтов сольватации показан на Рисунках 30 и 31.

Рисунок 29. Пример кластеризованной матрицы корреляции карт плотности молекул воды в ходе молекулярной динамики комплекса ПАРП1 с ингибитором

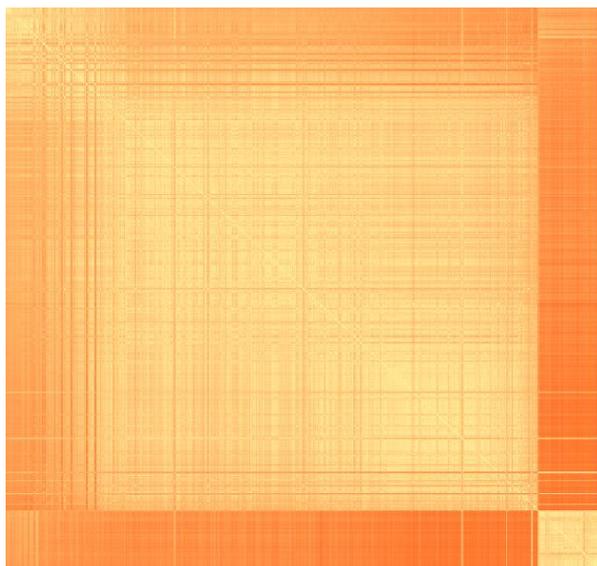


Рисунок 30. Альтернативные сайты сольватации комплекса ПАРП1 с ингибитором, определённые с помощью предложенного метода кластеризации

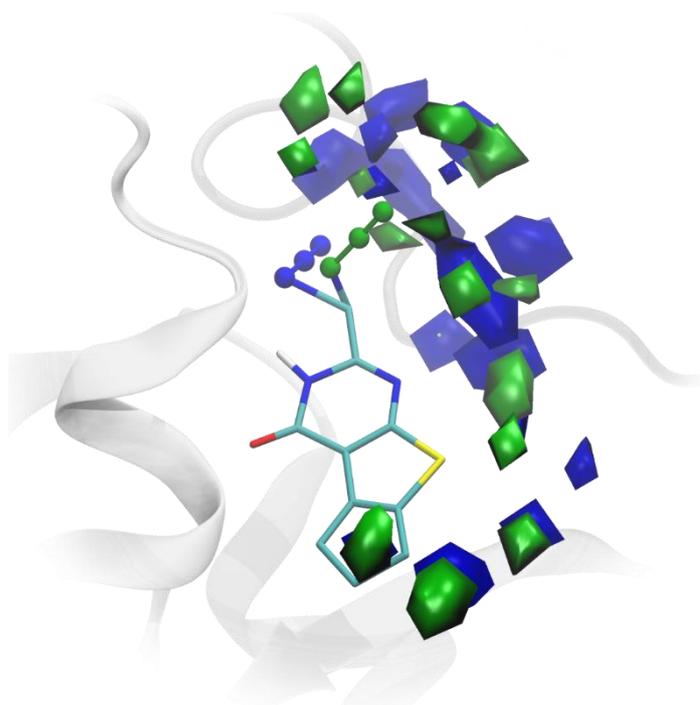
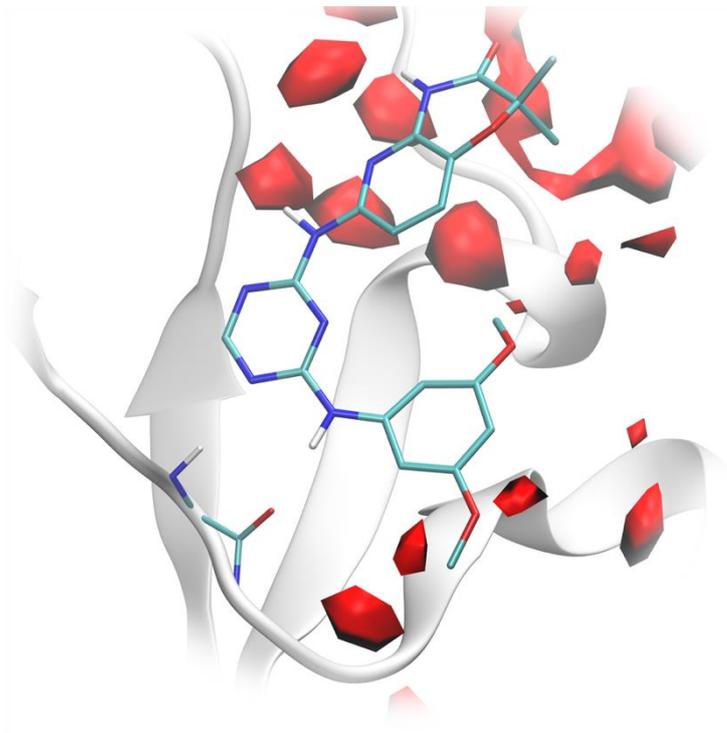


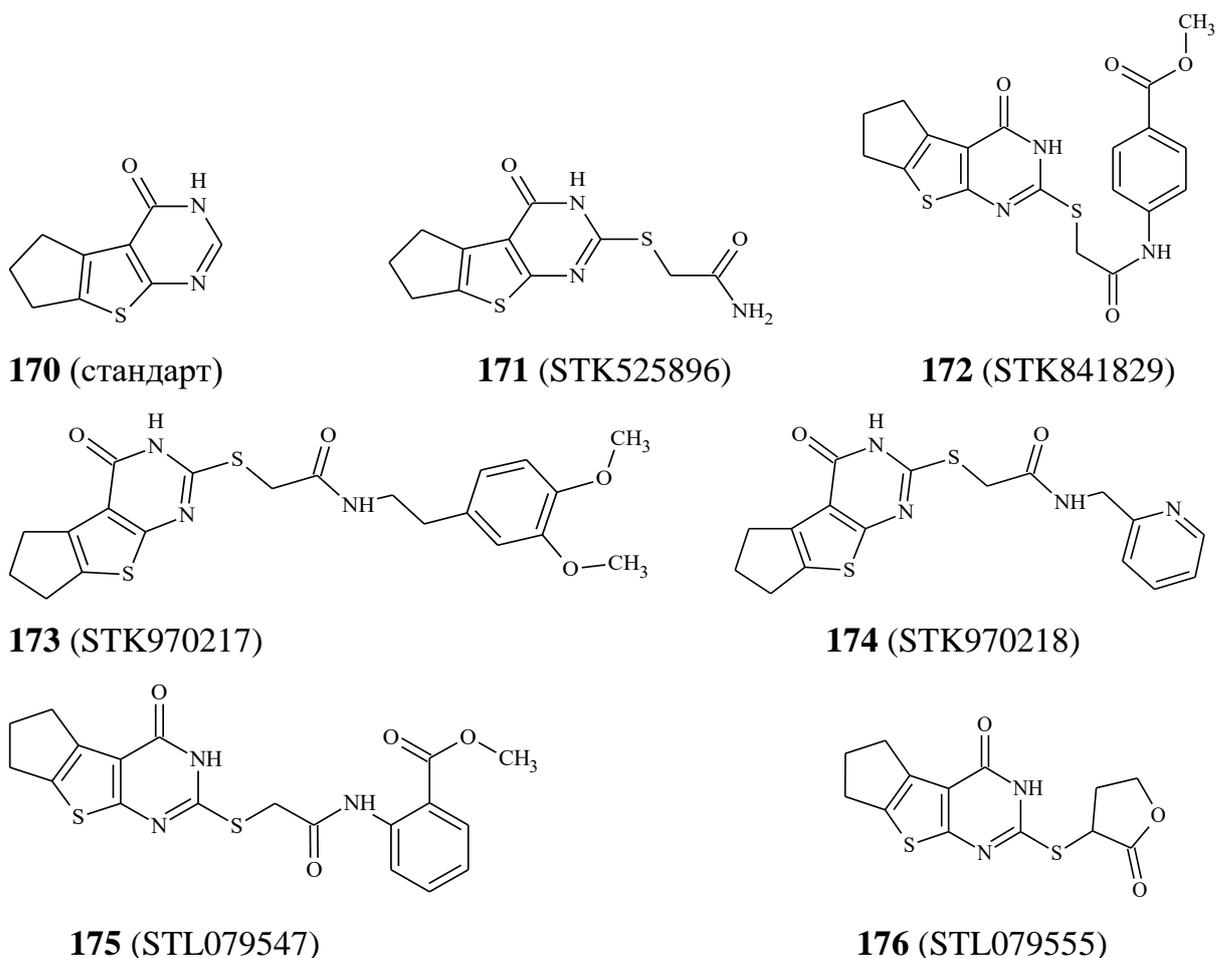
Рисунок 31. Основные сайты сольватации комплекса Syk-киназы с низкомолекулярным ингибитором



Таким образом, алгоритм поиска новых биологически активных веществ на основании известной (определенной PCA) структуры мишени включает в себя следующие этапы:

1. Определение состояния протонирования ионизируемых остатков белка с помощью алгоритма TSAR.
2. Реконструкция отсутствующих в рентгеновской структуре белка боковых радикалов с помощью алгоритма TSAR.
3. Молекулярно-динамический расчёт холо- и апо-форм белка с явным учётом растворителя.
4. Анализ сольватационных карт с помощью предложенного метода кластеризации карт растворителя и определение наиболее вероятных сайтов гидратации.
5. Молекулярный докинг в альтернативные структуры мишени без консервативных молекул воды и с молекулами воды, характерными для апо- и холо-фермента.
6. Структурная фильтрация результатов докинга.
7. Молекулярно-механический расчёт комплексов наиболее перспективных ингибиторов с мишенью.

Рисунок 32. Потенциальные ингибиторы ПАРП1



С помощью структурного поиска с последующим молекулярным докингом в библиотеке VitasM было выявлено 6 соединений (**171–176**), связывающихся аналогично **170** и потенциально ингибирующих ПАРП1 (Рисунок 32). Для данных соединений была изучена индивидуальная цитотоксичность (Рисунок 33), а также цитотоксичность в комбинации с доксорубицином (антрациклиновый антибиотик, применяемый в химиотерапии злокачественных опухолей по свободнорадикальному механизму с блокированием синтеза нуклеиновых кислот раковыми клетками, т.е. ингибитор общего действия).

Доксорубин ингибировал рост клеток НерG2 с IC_{50} 1,5–2 мкг/мл. Изомерные соединения **172**, **175** и **176** также проявили цитотоксичность в исследованном диапазоне концентраций 0–20 мкМ в отсутствие доксорубина. В комбинации с доксорубицином соединения **172**, **173**, **175** и **176** увеличивали его цитотоксичность, в то время как соединения **171** и **174** не влияли на неё (Рисунок 34).

Рисунок 33. Индивидуальная цитотоксичность соединений 171–176 на клетках HepG2

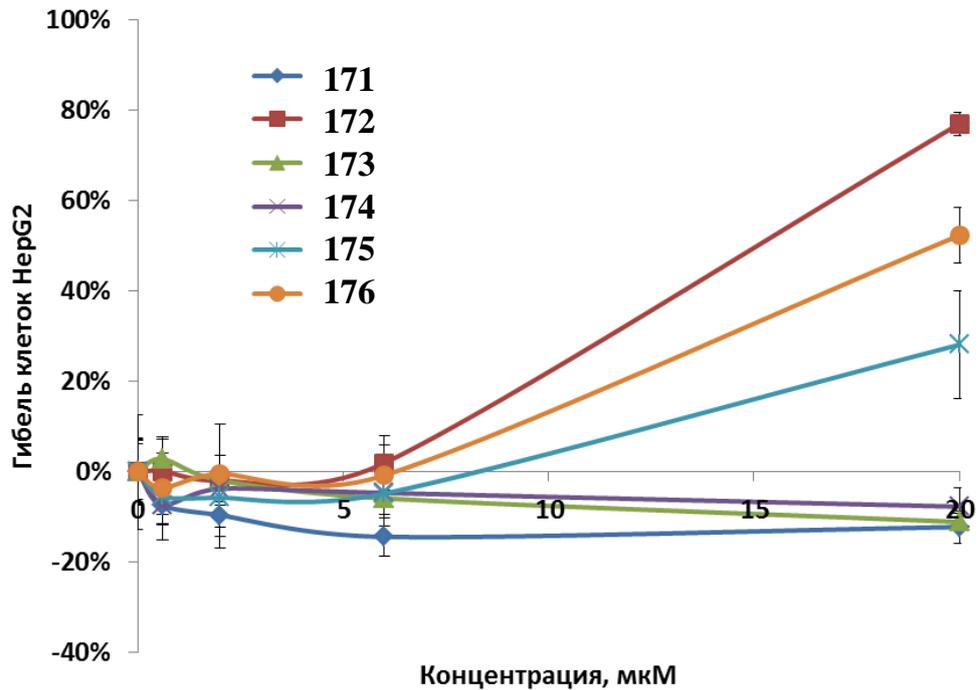
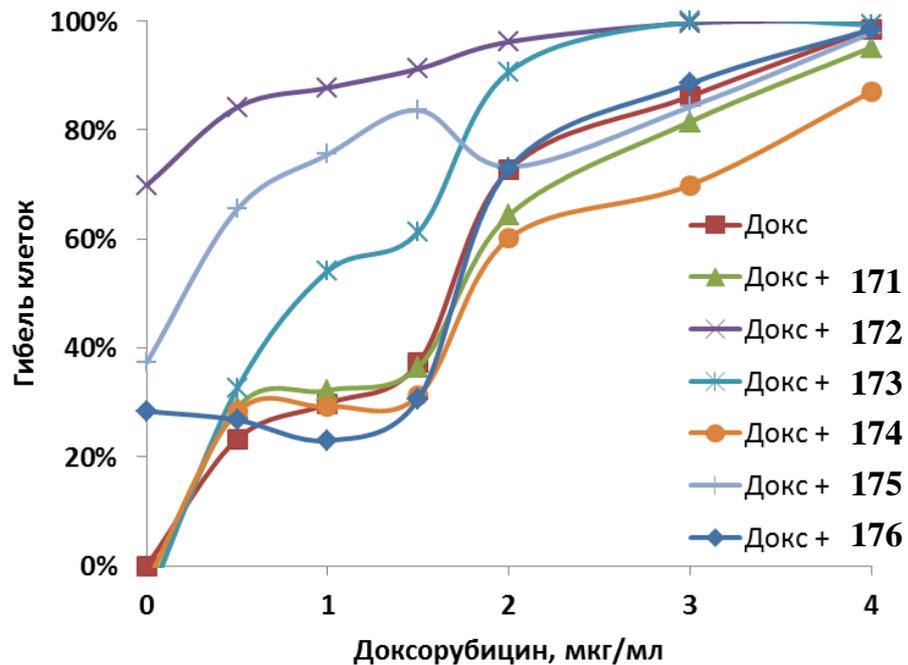


Рисунок 34. Комбинированная цитотоксичность доксорубицина и соединений 171–176 (20 мкМ) в клетках HepG2

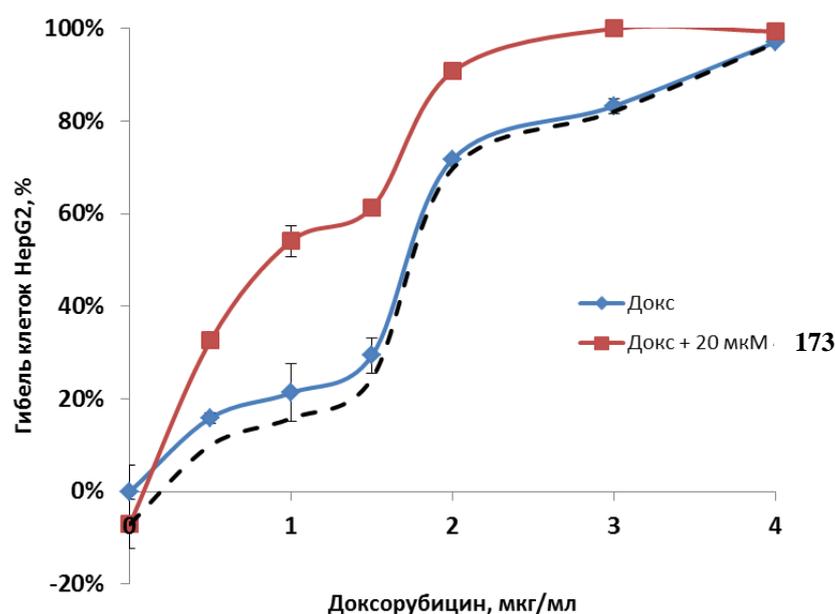


Согласно полученным данным, соединение 173 повышает цитотоксичность доксорубицина, не проявляя активности в индивидуальном состоянии. Соединения

172, **175** и **176** цитотоксичны как в индивидуальном виде, так и в комбинации с доксорубицином. Для анализа аддитивности данной комбинации была использована модель Блисса [389]. Согласно данной модели, аддитивный эффект, соответствующий независимому действию препаратов, определяется формулой $GI_{Add} = GI_{Dox} + (100\% - GI_{Dox}) * GI_{cpd}$, где GI_{Dox} – % гибели клеток при выбранной концентрации доксорубицина, GI_{cpd} – % гибели клеток при выбранной концентрации ингибитора ПАРП1. В случае синергизма наблюдаемый эффект больше предсказанного аддитивного, в случае антагонизма – меньше.

Экспериментальная цитотоксичность комбинации **173** (20 мкМ) с доксорубицином во всём диапазоне изученных концентраций выше, чем цитотоксичность аддитивного эффекта, что однозначно указывает на наличие синергического действия (Рисунок 35). Для соединений **172**, **175** и **176** практически не наблюдалось отклонения от аддитивности.

Рисунок 35. Цитотоксичность доксорубицина в комбинации с веществом **173** на клетках НерG2. Пунктиром показана линия аддитивного эффекта от доксорубицина и вещества **173**



Таким образом, с помощью поиска подструктуры и молекулярного докинга мы идентифицировали 6 новых соединений, которые потенциально ингибируют ПАРП1, 4 из которых увеличивают цитотоксичность доксорубицина в клетках

HerG2. Одно из изученных соединений (**173**) не обладало цитотоксичностью в качестве единственного агента и поэтому перспективно для дальнейшего развития безопасного ингибитора ПАРП1 для комбинации с ДНК-разрушающими агентами.

Алгоритмы отдельных этапов подготовки структуры белка соответствуют современным представлениям мирового уровня [390,391]. В то же время, подобный комплексный подход, сочетающий учёт конформационно-зависимых свойств белка с сольватацией, не применялся на практике к дополнительному анализу и оптимизации рентгеновских структур до настоящего момента.

Найденные потенциальные ингибиторы ПАРП1 по эффективности не уступают мировым аналогам [392].

Из вышесказанного можно сделать вывод, что применение в молекулярном моделировании методов докинга, в котором учитывается структура мишени, имеет значительные преимущества перед 3D-QSAR, в котором структура мишени неизвестна. В дополнение к этому, известная волюнтаристичность в выборе дескрипторов QSAR приводит к неточностям в прогнозировании активности.

Однако, возможности вычислительной математики в учете максимального количества вариантов, улучшение технических возможностей для расчетов 3D-QSAR вполне могут привести к увеличению точности прогноза.

Мы попытались доказать, используя как нечеткую логику в расчетах, так и сразу несколько вариантов математических приближений при определении активных молекул методами QSAR, что *молекулярный докинг и 3D-QSAR представляют собой два взаимодополняющих подхода к предварительному молекулярному моделированию в направленном синтезе.*

3.7.1.2. Совместное применение методов QSAR и молекулярного докинга в моделировании ингибиторов ПАРП1

Цель этой части работы заключалась в сравнении результатов моделирования методами QSAR и стандартными и усовершенствованными методами молекулярного докинга; а также в изучении того, производят ли они сводные прогнозы. Для этого мы провели поиск потенциальных ингибиторов ПАРП1 в тестовом наборе 196 соединений с последующей экспериментальной проверкой наилучших прогнозируемых лигандов.

Молекулярный докинг проводился с помощью Lead Finder 1.1.15 [189] с параметрами по умолчанию.

Для повышения точности рационального поиска ингибиторов на основе докинга была использована структурная фильтрация [380]. Анализ имеющихся рентгеноструктурных данных комплекса лиганд–мишень показали, что активные лиганды образуют пары коррелированных водородных связей с С=О и NH₂ группами фрагментов G863 (Рисунок 23). Предположительно, эти водородные связи характерны для режима связывания активных лигандов. Поэтому из результатов докинга были выбраны только лиганды, сформировавшие эти взаимодействия с белком.

В качестве альтернативы при расчетах тех же объектов методом *3D-QSAR* был проанализирован тот же тестовый набор из 196 соединений [393] с использованием методов *3D-QSAR*. Тестовый набор составил базу обучающей выборки соединений.

Для решения задачи сравнения структур тестового набора с точки зрения биологической активности используется математическая теория распознавания образов. Отдельные методы и классификаторы в рамках этой теории имеют ограниченное применение, поэтому мы использовали суммарный классификатор. Базовые алгоритмы описаны в алгебраическом подходе [394]. Развитие метода суммарных классификаторов описано в работах [395 – 397]. Наша проблема – сравнение физических объектов – решается в рамках подхода, называемого «смеси алгоритмов», основанного на понятии «область компетенции» [396], и включающего три базовых компонента [397]:

- набор классификаторов (базовых алгоритмов),
- набор шлюзов, делящих 3D-пространство объектов на нечеткие области, для которых учитывается вклад каждого классификатора,
- суммарную вероятностную модель, объединяющую шлюзы и классификаторы; модель представляет собой взвешенную сумму классификаторов (каждый классификатор имеет свой коэффициент вклада с весами, соответствующими значениям шлюзов).

Концепция смеси алгоритмов обладает следующими важными свойствами:

- позволяет выбрать классификаторы для узкой части общей проблемы;

- оперирует нечеткой логикой, используя нечеткое разбиение исходных данных;
- допускает оптимизацию дискретного описания пространства сравниваемых объектов.

Для проверки концепции был сформулирован итеративный алгоритм мульти-плетно кластеризующий обучающую выборку, впервые описанный в [398,399], а локальный классификатор сформулирован на основе метода ближайших соседей (k-NN) [400].

Применен двухэтапный классификатор на основе метода опорных векторов (также двухфазная схема решения) [401], в котором ограничения допустимости (шлюзы) формулируются как фильтр второго уровня, задача которого состоит в уточняющей обратной связи – разделении верно и неверно отнесенных структур моделью первого уровня.

Обучающую выборку соединений составляли 86 активных и 34 неактивных ингибиторов ПАРП1 – те же соединения, которые входили в расчет методами молекулярного докинга в описании выше. Остальные (из 196) вещества составили контрольную выборку.

Для математического описания химических структур мы использовали выделение линейных фрагментов с маркировкой вершин молекулярного графа [321]. Фрагментные дескрипторы кодируют дискретную информацию о молекулярных структурах. Их важным преимуществом является простота расчета, сортировки промежуточных результатов и интерпретации конечных данных.

В соответствии с вышеизложенными принципами описания молекулы с точки зрения «структурного спектра», или числового графа со структурными параметрами, отбор повторяющихся ключевых параметров графа проводили по вершинам – пары, тройки и четверки вершин группировались по параметрам описания: длине линейных фрагментов ($k = 2,3,4$) и меткам, включающим информацию о химической связи и ее месте в кольцевой структуре). Подбор этих параметров при помощи машинной логики привели к построению 24 моделей описания; еще 72 модели были построены с использованием различных алгоритмов распознавания.

Алгоритм селекции дескрипторов на основе k-NN

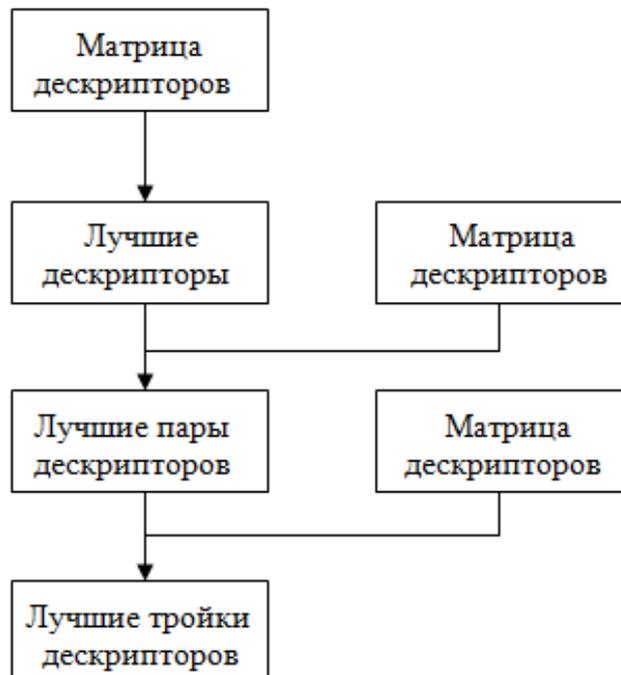
На входе алгоритма формируется матрица A дескрипторов и параметр n . На выходе формируется набор дескрипторов, для которого функциональный параметр качества $\varphi(F)$ максимален.

С помощью метода k-NN по каждому отдельно взятому дескриптору из матрицы A производится классификация и по значению параметра функции качества $\varphi(F)$ отбирается n лучших представителей. Этот отбор назовем селекцией.

К каждому набору дескрипторов, полученному на предыдущем шаге, по очереди добавлен еще один дескриптор матрицы A , проводится классификация, и по значению функционального параметра $\varphi(F)$ отбирается n лучших, получаем n^2 мультипараметрических дескрипторов.

На Рисунке 36 приведена схема алгоритма. Он позволяет обрабатывать матрицы с большим количеством дескрипторов M , на которых невозможно использование полного перебора из-за чрезмерных требований к вычислительным возможностям оборудования

Рисунок 36. Схема селективного отбора дескрипторов



При использовании описанного выше алгоритма главный член в оценке сложности nmN^2M , где N – количество молекул. Для построения моделей использовался данный селективный алгоритм отбора дескрипторов на k -NN. Как результат нескольких селекций, отбирали 3–4 лучшие модели (Рисунок 36). Ниже приведены таблицы с оценкой качества прогноза по проценту успешных предсказаний на построенных моделях.

Из Таблицы 16 видно, что наилучшее качество прогноза составляет 98% и достигается почти на всех матрицах, при построении которых использовались маркеры b и r .

Таблица 16. Качество прогноза моделей, построенных с помощью селективного алгоритма отбора дескрипторов на k -NN по дескрипторам первого уровня

Маркер	***	**r	*b*	d**	*br	d*r	db*	dbr
Длина фрагмента	k=2							
Качество прогноза	0.8916	0.9500	0.9333	0.9333	0.9833	0.9333	0.9666	0.9833
Длина фрагмента	k=3							
Качество прогноза	0.9333	0.9833	0.9583	0.9666	0.9833	0.9666	0.9750	0.9833
Длина фрагмента	k=4							
Качество прогноза	0.9250	0.9666	0.9333	0.9500	0.9750	0.9666	0.9533	0.9833

Таблица 17 показывает, как улучшилось качество прогноза по сравнению с топологическими дескрипторами.

Таблица 17. Качество прогноза моделей, построенных с помощью селективного алгоритма отбора дескрипторов на k-NN по дескрипторам второго уровня

Маркер	***	**r	*b*	d**	*br	d*r	db*	dbr
Длина фрагмента	k=2							
Качество прогноза	0.9747	0.9916	0.9666	0.9833	0.9833	1.0000	0.9833	0.9833
Длина фрагмента	k=3							
Качество прогноза	0.9916	0.9916	0.9750	0.9833	0.9833	0.9833	0.9916	1.0000
Длина фрагмента	k=4							
Качество прогноза	0.9833	1.0000	1.000	0.9750	1.0000	0.9916	0.9916	0.9916

Прогнозирование активности новых соединений

Выборка соединений с неизвестными значениями активности была проверена на комплементарность к моделям, построенным по описанным выше алгоритмам. Обозначенные, соответственно, как активные и неактивные (+1 и -1), соединения оценивались по простой сумме параметров для суммы моделей.

В результате такой операции, проведенной для выборки 196 соединений с неизвестной активностью, было выделено несколько предположительно активных соединений.

Сравнение результатов, полученных как молекулярным докингом, так и 3D-QSAR, дало в целом хорошо согласующуюся картину. Большая часть (165 из 196) тестовых соединений была оценена одинаково обоими методами; т.е. в данном конкретном случае оба метода могут быть использованы для виртуального скрининга.

Прогнозирование активности:

- методом классического докинга показало 12 активных соединений;
- по данным 3D-QSAR – 26.

Для экспериментальной характеристики были дополнительно отобраны три подмножества:

- соединений, прогнозируемых двумя методами (7 структур),
- молекулярным докингом, но не 3D-QSAR (5 структур)
- и наоборот, 3D-QSAR, но не докингом (19 структур).

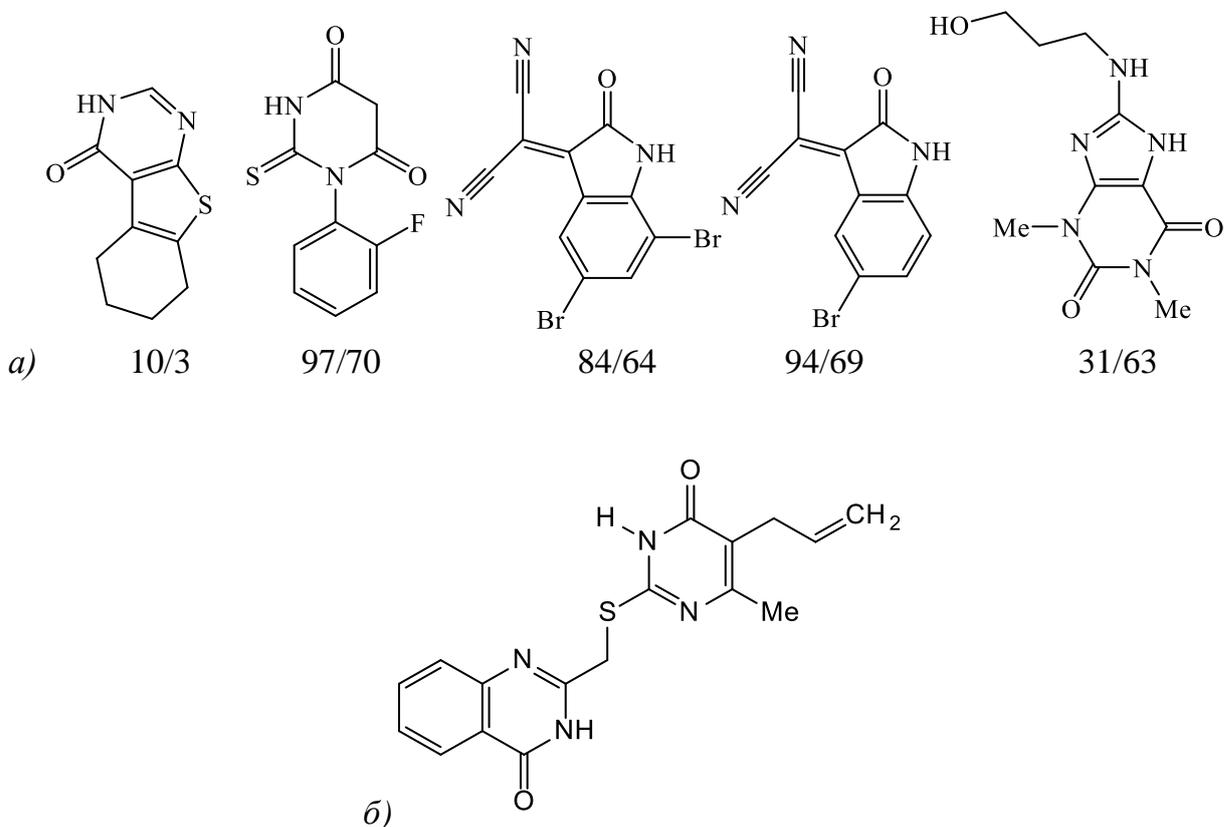
В Новосибирском институте химической биологии и фундаментальной медицины была проведена оценка ингибирующей активности по отношению к ПАРП1.

Совместное применение метода 3D-QSAR и молекулярного докинга существенно повышала вероятность нахождения биологически активных соединений. Среди веществ, для которых связывание с ПАРП1 было предсказано обоими методами, доля экспериментально активных ингибиторов *in vitro* (10 мкМ) составляла 57%, в то время как среди соединений, активных по молекулярному докингу, но не активных по 3D-QSAR, их доля составила всего лишь 40%. Полученные данные подчёркивают важность и невзаимозаменяемость обоих методов моделирования.

Из 7 веществ, связывающихся с ПАРП1 по результатам обоих методов, 4 продемонстрировали значительную ингибирующую активность в концентрации 10 мкМ. Только 2 из 5 соединений, активных по молекулярному докингу, но не по QSAR, подтвердили расчетные данные экспериментально (Рисунок 37а).

Экспериментальная проверка показала, что из 19 QSAR-активных, докинг-неактивных веществ одно имело активность, и дала ингибирующую концентрацию на данном соединении в 200 мкМ (0,2 мМ) (Рисунок 37б). Это говорит о том, что данное соединение является ингибитором ПАРП1.

Рисунок 37. а) Примеры структур с идентичной активностью по расчетам молекулярным докингом и по расчетам 3D-QSAR (IC_{50} , расчет/эксперимент);
 б) структура соединения, неактивного по расчетам с применением докинга, при этом активного по QSAR и подтвердившего активность в эксперименте



Приведенные данные свидетельствуют о важности и несвязанности обоих применяемых методов моделирования. Принципиальная разница между 3D-QSAR и молекулярным докингом состоит в том, что первый опирается только на структуру лиганда, а второй непосредственно оценивает полноатомное взаимодействие белок–лиганд.

Ранее было показано, что молекулярное моделирование может быть эффективно использовано при синтезе новых ингибиторов поли(АДП-рибозо)полимеразы-1 (ПАРП1) [346], которая локализована в клеточном ядре и катализирует поли-АДП-рибозилирование различных белковых и ДНК субстратов. Молекулярный докинг (особенно в сочетании со структурной фильтрацией) позволяет уверенно отвергать лиганды, которые не образуют критических контактов в месте связывания белка и, таким образом, имеют тенденцию быть идентичными, что приводит к низкой общей скорости ложных негативных прогнозов. Тем не менее,

число ложноположительных прогнозов докинга (50% в данном исследовании) по-прежнему велико, и это, безусловно, может быть улучшено, в том числе с применением 3D-QSAR.

Таким образом, показано, что совмещение методов статистической обработки и формирования дескрипторов для обучающей выборки соединений с известной активностью (QSAR) и метода прямого моделирования взаимодействия лиганд–мишень (докинг) является актуальным направлением, а 3D-QSAR – эффективным уточняющим дополнением методов прямого моделирования; выводы о взаимодополняемости данных методов подтверждаются исследованиями других систем [402].

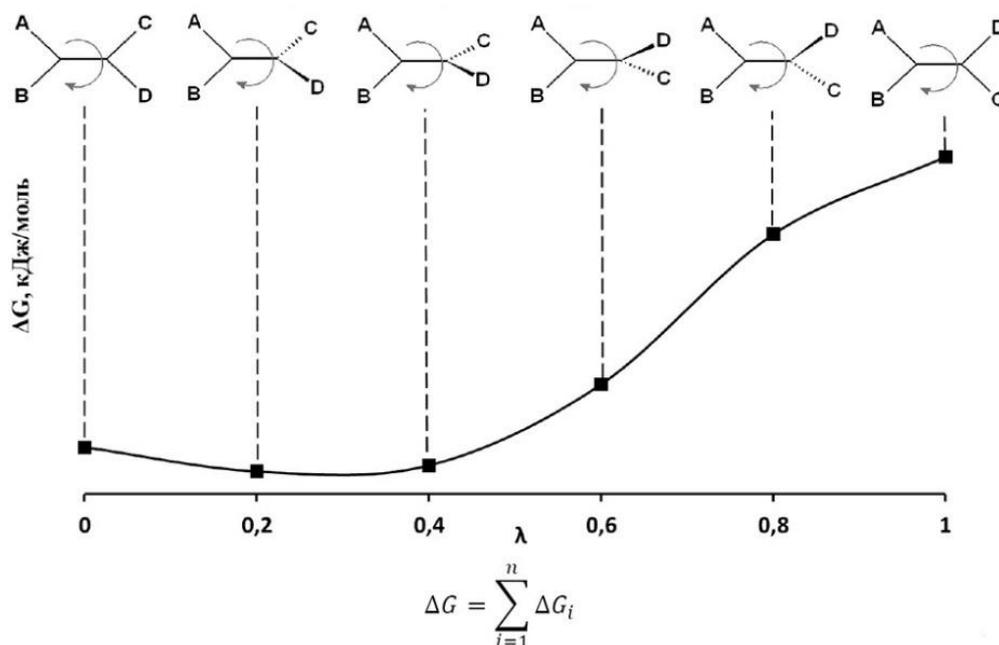
3.7.2. Моделирование с учетом конформаций. Конформационная фокусировка

Для того, чтобы рассчитать конформационные переходы какой-либо молекулы, мы разработали свой алгоритм, который состоит в постепенном изменении угла вращения мобильных конформационных связей в молекуле. При этом молекула переходит из одного конформационного состояния в другое, и на протяжении всего этого процесса мы определяем энергетику перехода, затем суммируем свободную энергию по всем этапам перехода и определяем энергетическую выгодность одного конформера по отношению к другому, тем самым получая относительные концентрации конформеров (Рисунок 38).

Нами впервые проведены расчеты энергии конформационных переходов методом ВСЭ. Для этого потребовался модельный эксперимент – построение графика зависимости изменения полной свободной энергии конформационных переходов *N*-метилформамида от шага перехода λ по кривой свободной энергии возмущения.

Изменения свободной энергии на каждом перемещении по λ было не более 1–2 кДж/моль, что указывает на перекрывание фазовых пространств (непрерывность перехода) между соседними, $\lambda_k \dots \lambda_n$ состояниями. Отсутствие максимума при $\lambda=0,5$ (заместители перпендикулярны) на графике переходов объяснялось постоянным смещением минимума двугранного потенциала от -180° к 180° , который при $\lambda=0,5$, соответственно, составляет 90° . Для корректности представления $\Delta G \lambda=0,5$ включен в график Рисунок 38.

Рисунок 38. Расчет заселенности конформеров с явным учетом растворителя

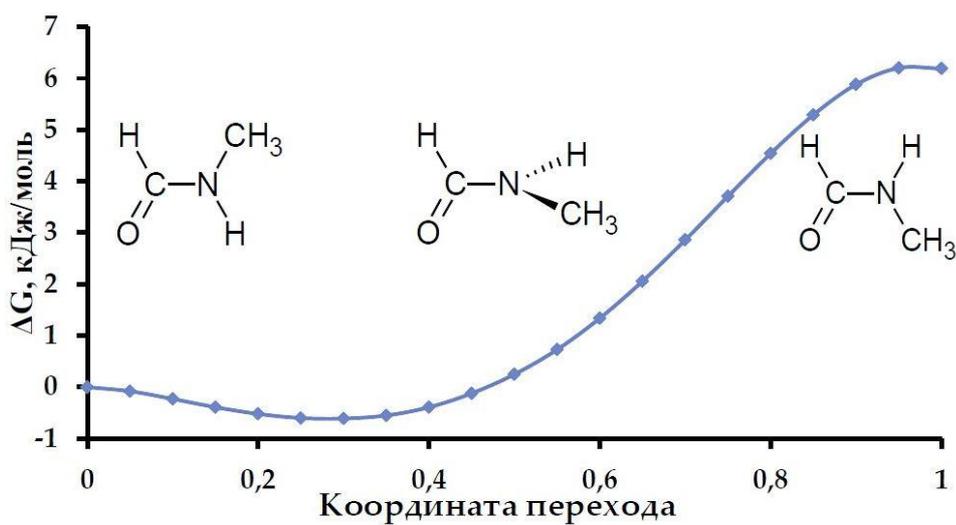


Экспериментальные данные для модельной системы *N*-метилформамид–вода: 8% *цис*- и 92% *транс*-конформации. Расчетные – в *цис*-конформации находится 7,7%, в *транс* – 92,3%, то есть наблюдается хорошее согласование расчетов с экспериментом (Рисунок 39).

Рисунок 39. Подтверждение модельных расчетов простых молекул

Эксперимент: 8% *цис*, 92% *транс*

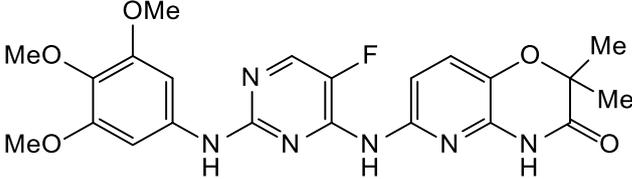
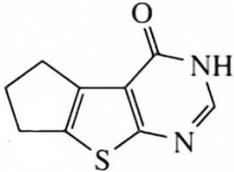
Расчет: 7,7% *цис*, 92,3% *транс*

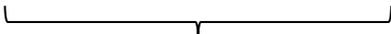


3.7.3. Конформационное равновесие лигандов Syc-киназы

Полученные другими исследователями ингибиторы терапевтических мишеней Syc-киназы и ПАРП1 весьма лабильны (Рисунок 40), однако, координируется с мишенью только один из конформеров. Ошибки в определении активного конформера – одна из базовых проблем классического докинга. В Таблице 18 приведены проблемы, с которыми мы столкнулись при расчетах ингибиторов ПАРП и Syc-киназы. Как результат, пришли к выводу о необходимости конформационной фильтрации при любых расчетах комплекса лиганд–мишень.

Таблица 18. Проблемы в расчетах ингибиторов ПАРП1 и Syc-киназы

Syc-киназа	ПАРП1
Ингибитор (R406) проходит клинические испытания от ревматоидного артрита	Мишень ревматоидного артрита и неходжкинской лимфомы
Мишень сердечной недостаточности и различных видов рака	Ингибиторы во II фазе испытаний от рака молочной железы и от меланомы
	
Имеют 4 заторможенных связи и 16 конформеров в растворе	Комплексы с белком сильно гидратированы и содержат мостиковые молекулы воды

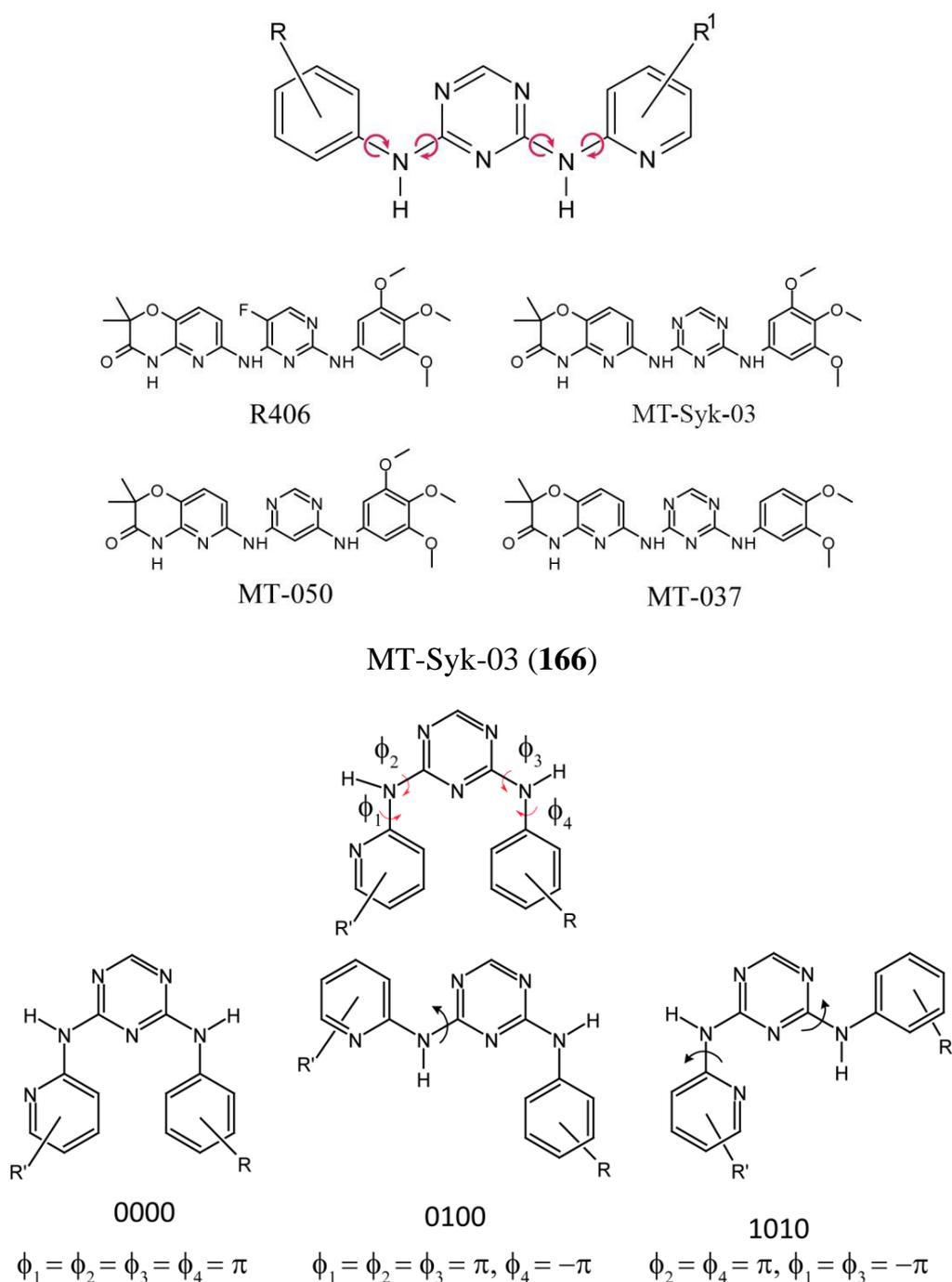

 Ошибка в расчетах классическим докингом >10 кДж/моль
 (отличие IC₅₀ в 10–100 раз)

Смоделируем конформационное равновесие для группы ингибиторов Syc-киназы с известной активностью, затем применим закономерности модельной системы к набору ингибиторов Syc-киназы с известной активностью для разработки и анализа алгоритма моделирования конформационного равновесия методом ВСЭ.

Модельные структуры были получены молекулярным докингом; некоторые проявили активность *in vitro* в наномолярной концентрации. Структуры (Рисунок 40) содержат 2,4-диариламиноазинный скелет, характеризующийся фиксированными конформациями, образующимися вращением вокруг связей C(Ar)–N–C(Ar), и

находятся в растворе в виде 16 лабильных конформеров. На 2 нс траекториях потенциальной энергии этих соединений в растворе не происходит конформационных переходов, поэтому вводится понятие конформационной фокусировки.

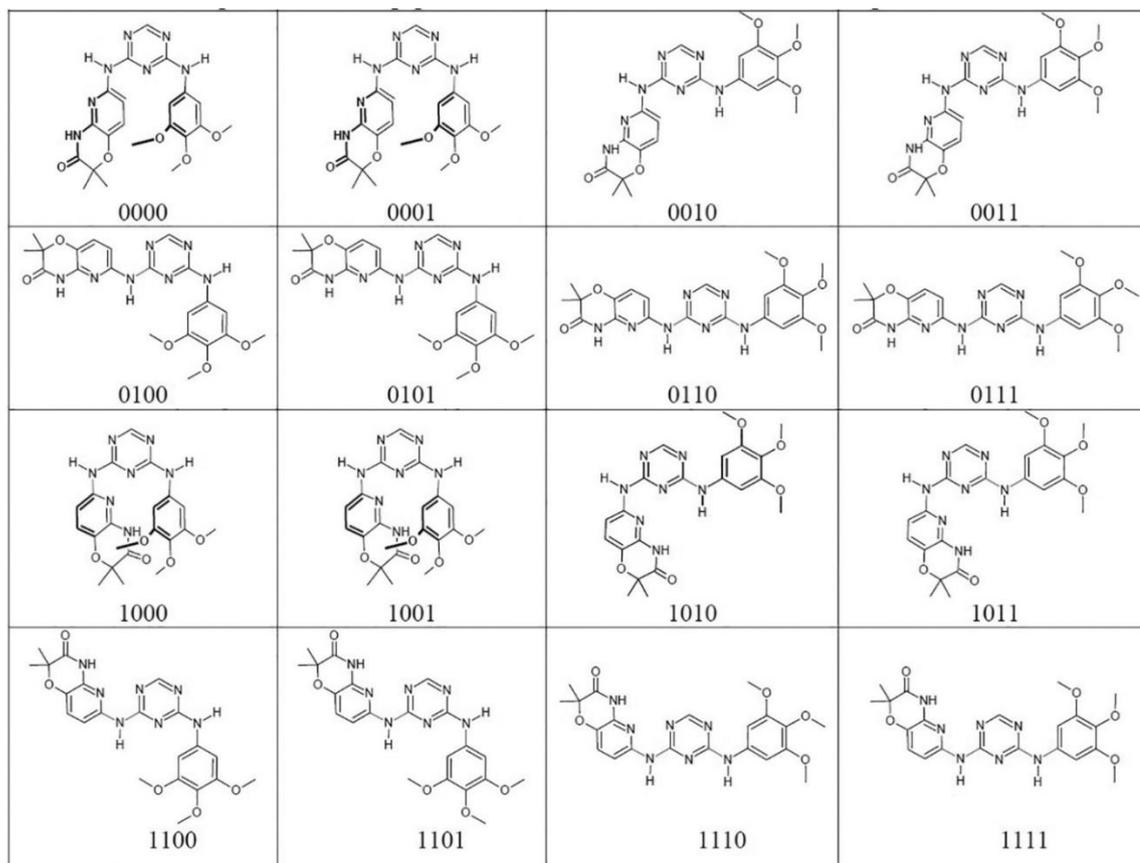
Рисунок 40. Структуры изученных ингибиторов Syk-киназы. Вращение вокруг связей Ar–NH–Ar приводит к образованию 16 конформеров



Вращение вокруг межкольцевых связей несколько затруднено. Обозначим конформеры 4-значным бинарным маркером – уникальным идентификатором конкретного

двугранного угла φ_1 – φ_4 конкретной конформации. При этом «0» соответствует первой конформации лиганда в комплексе с белком, «1» – второй. Конформер, связанный с белком, обозначается «0000». Кластеризация конформеров MT-Syk-03 приведена на Рисунке 40, структуры конформеров – в Таблице 19.

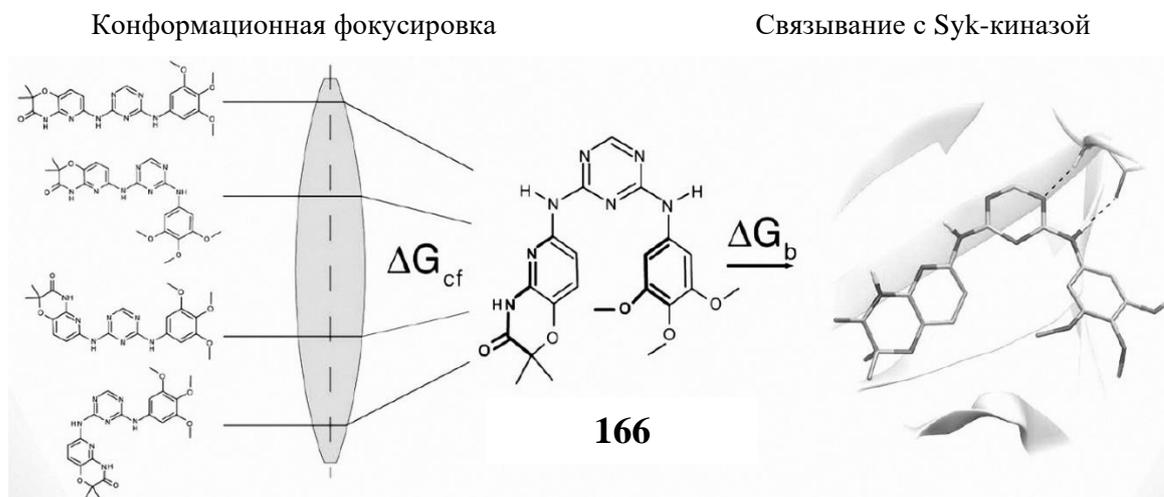
Таблица 19. Возможные заторможенные конформации (0000 – 1111) соединения MT-Syk-03 (166)



Однако, как показали данные рентгеноструктурного анализа, с белком они связываются только в одной конформационной форме. Таким образом, чтобы предсказать прочность связывания данного соединения с другой молекулой, нужно четко знать какая доля биологически активной конформации присутствует в сайте связывания молекулы и рецептора в растворе (Рисунок 41).

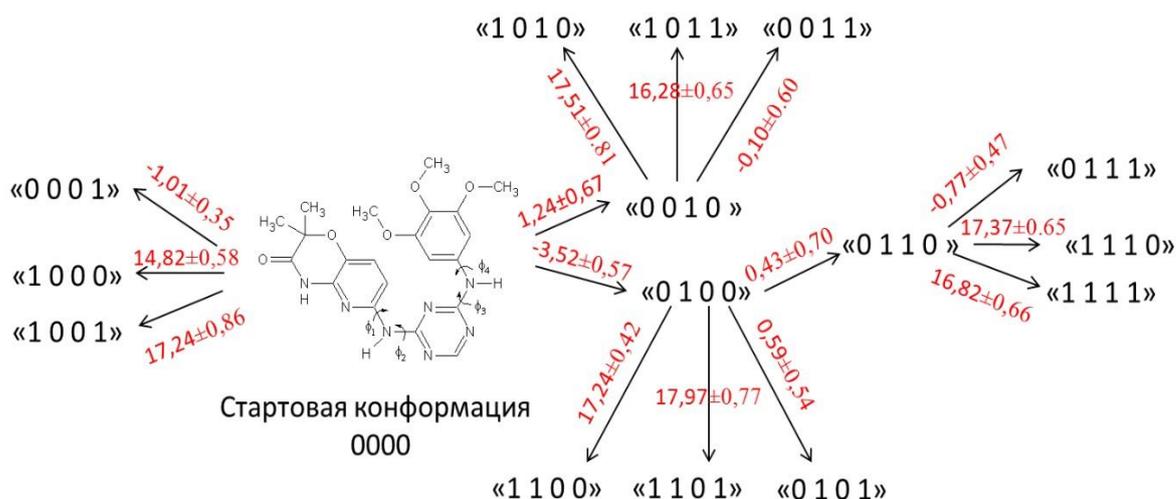
Конформационная фокусировка (КФ) – переход конформаций лиганда, представленных в растворе, в связывающуюся с белком конформацию. КФ необходимо учитывать отдельно при наличии у лиганда заторможенных степеней свободы.

Рисунок 41. Конформационная фокусировка MT-Syk-03 при связывании с Syk-киназой



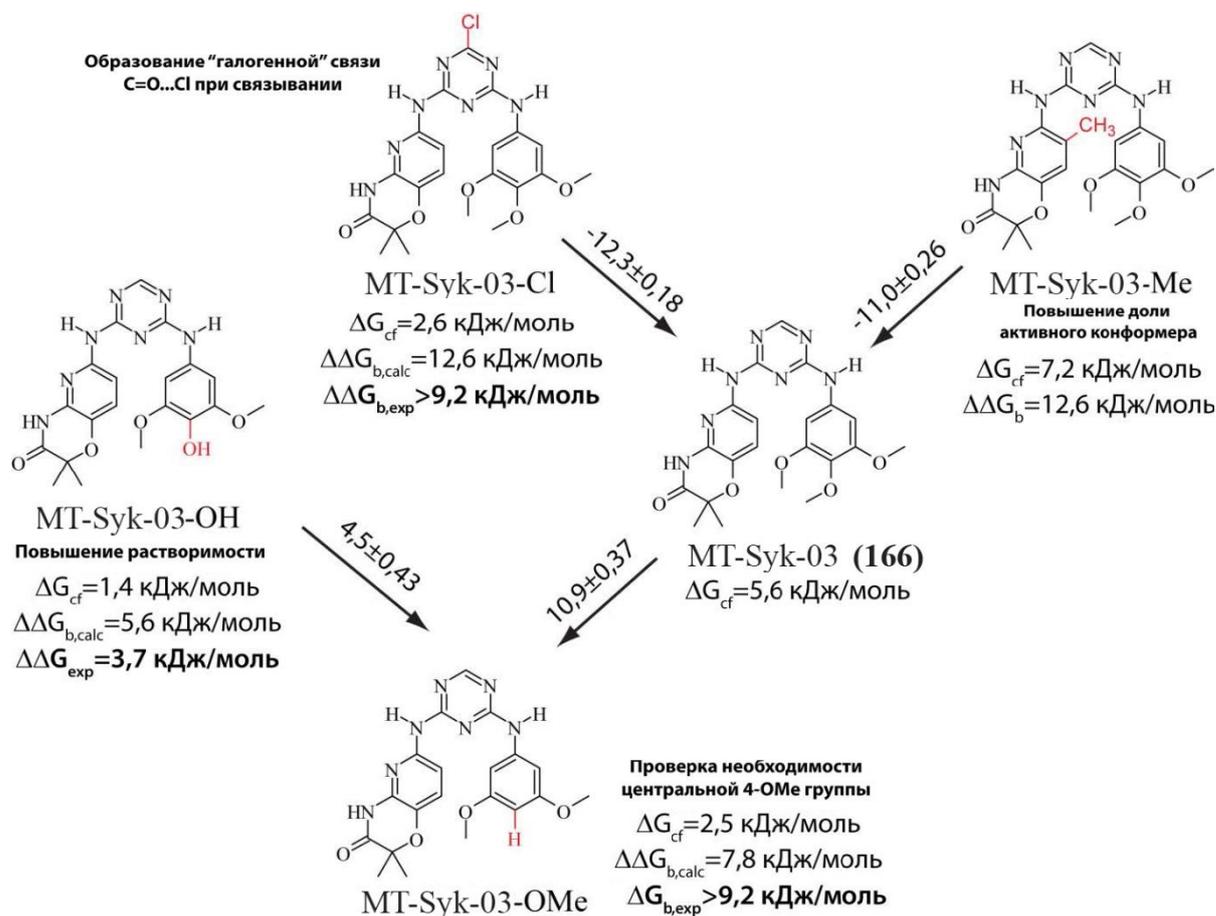
Для того, чтобы выяснить долю биологически активной конформации, необходимо посчитать энергию всех конформационных переходов этих соединений. Это можно сделать методами квантовой химии. Однако методы квантовой химии не учитывают явным образом молекулы растворителя, а также не учитывают отдельные подвижки других частей молекулы, которые могут влиять на конечный результат. Полученные значения представлены на Рисунке 42.

Рисунок 42. Энергетика конформационных переходов



В подтверждение рассчитанных конформационных переходов произведен расчет различных замещенных ингибиторов с каркасной структурой MT-Syk-03 и их синтез для подтверждения метода (Рисунок 43).

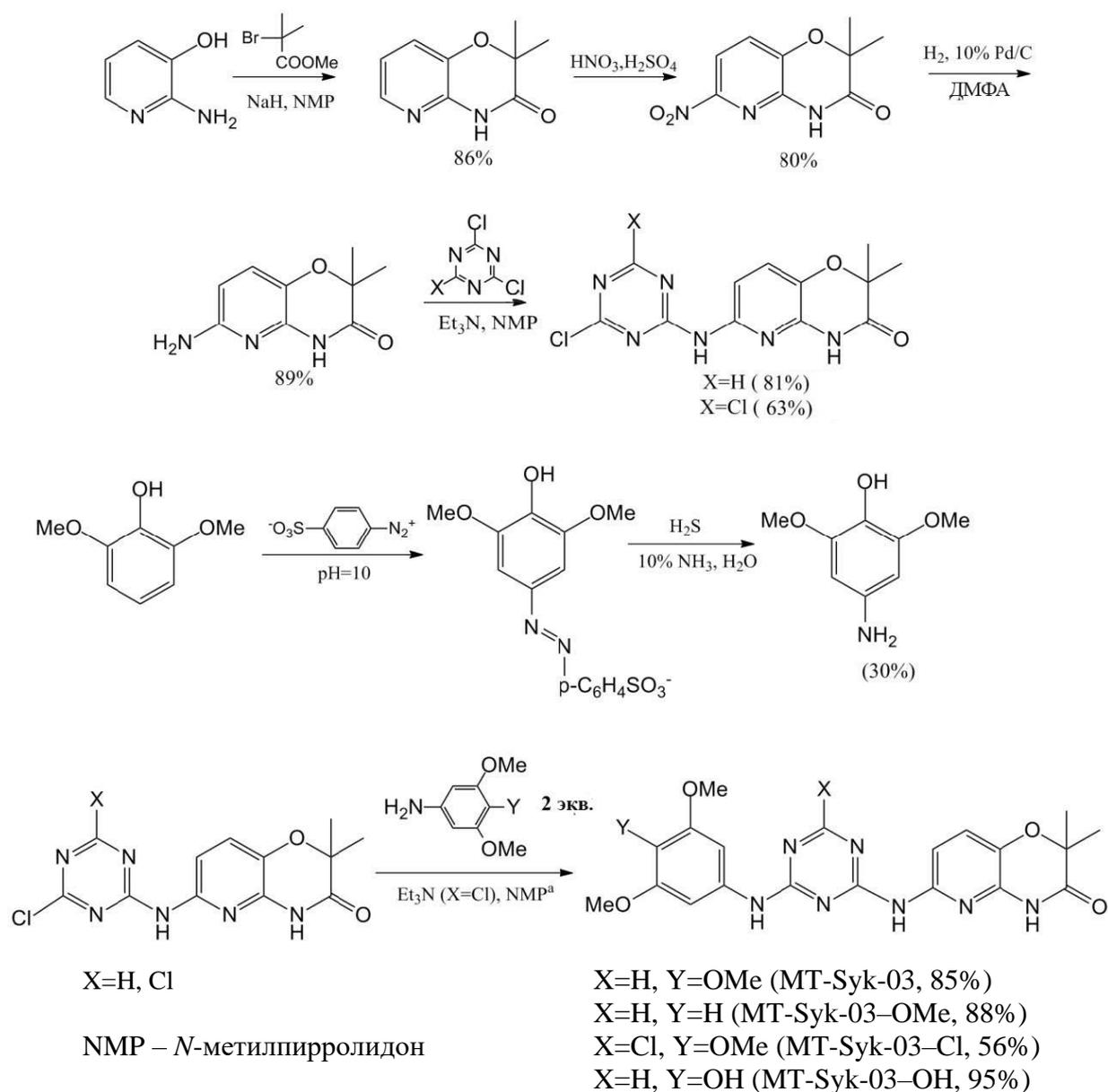
Рисунок 43. Рассчитанные и экспериментально определённые* величины относительной свободной энергии связывания (в сравнении с **166**) ингибиторов Syk-киназы – структурных модификаций базового ингибитора MT-Syk-03 (**166**)



* cf – конформационная фокусировка, calc – расчетные значения, exp – экспериментально определенные.

Как видно из Рисунка 43, для структурных модификаций MT-Syk-03 (**166**), рассчитанных на предмет увеличения прочности связывания с белком (MT-Syk-03-OMe, MT-Syk-03-Cl) либо увеличения гидрофильности (MT-Syk-03-OH), и синтезированных для экспериментальной проверки (Схема 15), было предсказано снижение энергии конформационной фокусировки, изначально неочевидное.

Схема 15. Синтез структурных модификаций MT-Syk-03



Экспериментально определенная активность синтезированных соединений соответствовала прогнозу (разница не более 2 кДж/моль): MT-Syk-03-OH проявило несколько меньшую активность, чем MT-Syk-03, в то время как MT-Syk-03-OMe и MT-Syk-03-Cl оказались неактивны в изученных концентрациях, как и предсказывали расчёты.

Таким образом, на основании предсказанных значений активности выбран и синтезирован набор новых ингибиторов Syk-киназы, эффективно блокирующих данный фермент в экспериментах *in vitro* на ферментном препарате и на клеточных моделях неходжкинской лимфомы и ревматоидного артрита в субмикромольном

диапазоне концентраций. Полученные экспериментальные данные по эффективности ингибиторов согласуются с расчётными данными (погрешность <2 кДж/моль). Синтезированные соединения являются крайне перспективными для дальнейшего предклинического и клинического изучения.

3.7.4. Применение квантовохимических подходов в моделировании ингибиторов киназ

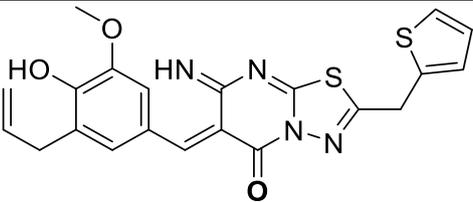
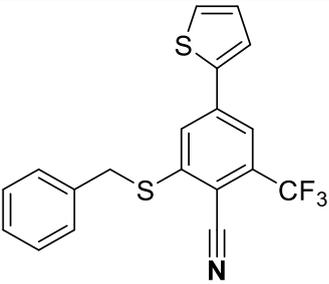
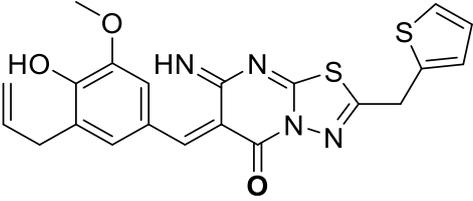
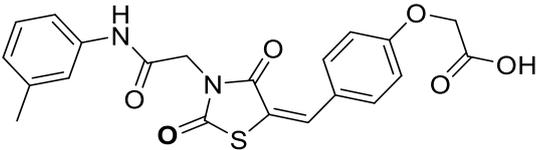
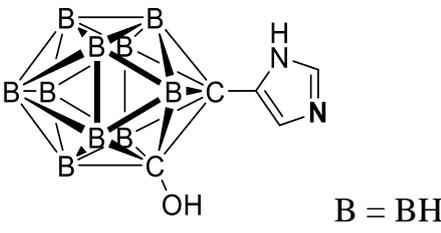
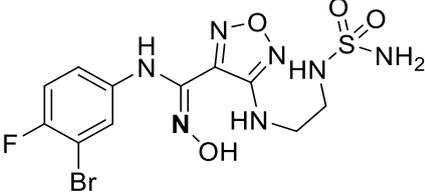
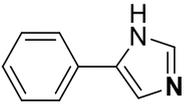
Разработка алгоритмов моделирования ингибиторов индол-2,3-диоксигеназы (ИДО). Изучение связывания комбинаторно сшитых фрагментных ингибиторов ИДО методами докинга и возмущения свободной энергии (ВСЭ).

Выше описаны разработанные нами подходы к моделированию биологически активных соединений методом ВСЭ, а именно, (1) протокол фрагментного скрининга в статическом варианте докинга, (2) метод учета конформационной фокусировки и (3) метод явного учета растворителя в ВСЭ (подходы к динамике процессов координации).

Имеющиеся и усовершенствованные подходы использованы для разработки новых лекарственных кандидатов, таких как ингибиторы индол-2,3-диоксигеназы (ИДО), которая экспрессируется злокачественными опухолями для защиты от иммунного ответа. На первой стадии моделирования необходимо понять механизм действия уже известных ингибиторов ИДО – для прогнозирования новых.

Поиск фрагментных ингибиторов ИДО был проведён с использованием предложенного метода «докинга по поверхности». Фрагментные ингибиторы были «сшиты» с получением фрагментно-сшитых ингибиторов, наиболее активные из которых представлены в Таблице 20. Для них, а также для предложенных ранее ингибиторов ИДО (эпакадостат и 4-фенилимидазол), был проведён докинг и рассчитано сродство к ИДО с использованием программного комплекса Lead Finder. Результаты докинга показали, что ни один из предложенных фрагментно-сшитых ингибиторов не связывается с ИДО сильнее, чем предложенный ранее компанией Incyte ингибитор эпакадостат. Данный вывод был также подтверждён результатами квантовохимических расчётов, которые наиболее адекватно описывают различия в связывании различных лигандов с гемом.

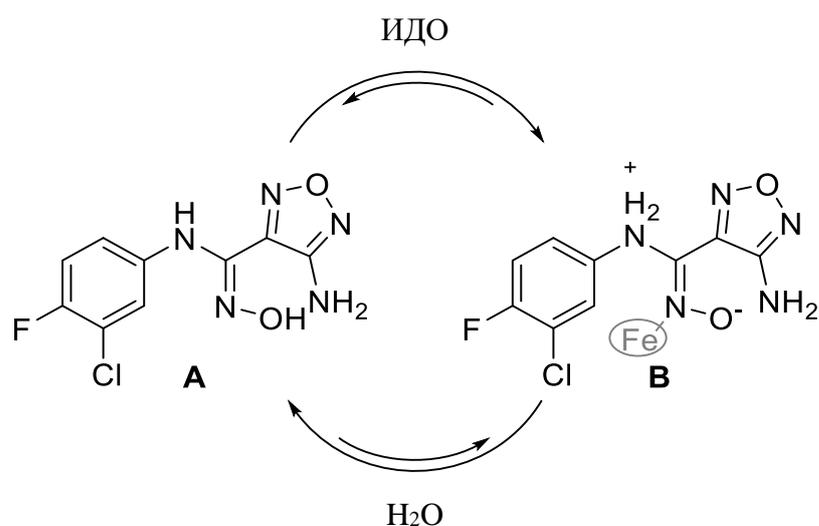
Таблица 20. Наиболее активные фрагментно-сшитые ингибиторы и два предложенных ранее ингибитора ИДО: эпикадостат и 4-фенилимидазол. Полу жирным начертанием выделены атомы, связывающиеся с атомом железа гема

Название лиганда	Структура	Свободная энергия связывания с ИДО (ккал/моль)
Ligand-109		9,5
Ligand-2771		10,0
Ligand-6216		9,8
Ligand-11445		9,0
Ligand-16125		10,5
эпикадостат		11,4
4-фенилимидазол		8,1

Для дальнейшего поиска новых, более эффективных ингибиторов ИДО и родственных ферментов, было принято решение детально исследовать связывание эпакадостата с ИДО методом квантовохимического моделирования. Моделирование проводилось на уровне теории PBE1PBE-D3/SDD PCM(EtOH), в качестве модели был взят сам лиганд эпакадостат, гем вместе со связанным с ним фрагментом гистидина, а также часть цепи белка, формирующая его активный центр.

Квантово-химические расчёты показали, что наблюдающаяся в рентгенодифракционном анализе комплекса ИДО с эпакадостатом координация эпакадостата с атомом железа гема с помощью атома азота $C=N-O^-$ группы не является оптимальной, и комплекс с эпакадостатом, в котором он координируется с атомом железа гема с помощью атома кислорода $C=N-O^-$ группы, должен быть стабильнее. Однако, полученная с помощью рентгенодифракционного анализа структура не оставляет сомнений, что координация происходит именно через атом азота. Для объяснения этого несоответствия, были проведены расчёты рKa свободной и связанной с гемом молекул эпакадостата, которые показали, что перед связыванием с ИДО она существует в форме А (Рисунок 44), тогда как после связывания наиболее устойчивой становится форма В (на 16 ккал/моль).

Рисунок 44. Две таутомерные формы эпакадостата, одна из которых (А) реализуется в растворе, и вторая (В) – в активном центре ИДО



При этом, до образования связи с гемом, форма В очень неустойчива – соответственно, в активный центр ИДО попадает форма А, которая связывается с гемом

атомом азота – и затем уже переходит в форму В. Это объясняет наблюдаемую координацию эпакадостата. Однако, рядом с эпакадостатом в активном центре отсутствует основная группа, которая могла бы переносить протон от гидроксильной группы. Единственным возможным переносчиком протона является амидная группа основной цепи ИДО, принадлежащая аланину-264. Проведённое квантово-химическое моделирование данного процесса показало, что такой перенос протона действительно, может происходить; он представляет собой двухстадийный процесс, в котором на первой стадии амидная группа присоединяет протон с гидроксиламинового фрагмента эпакадостата, а на второй стадии один из протонов метильной группы переносится на аминогруппу эпакадостата. Энергия активации двух стадий переноса протона составляет, соответственно, 15,2 и 8,5 ккал/моль.

Таким образом, нами был установлен молекулярный механизм связывания эпакадостата с ИДО, который включает в себя перенос протона через амидную группу основной цепи фермента – уникальный процесс, который, насколько нам известно, ранее не был описан. Установленный механизм связывания будет использован нами в дальнейшем для разработки новых, более эффективных ингибиторов ИДО и родственных ему ферментов.

3.8. Моделирование катализа

Как ясно из литературы, никакое моделирование катализа, а, следовательно, направления каталитического синтеза, т.е. прогнозирования активных структур для направленного каталитического синтеза невозможно без представления динамики этого химического процесса. В этом – основное отличие алгоритмов и методов, применённых в первой части работы (прогнозирование биологической активности) и во второй – прогнозировании каталитической активности.

Вторая часть работы поставила задачу учета как

- дополнительных статических факторов (участия растворителя, конформационной фильтрации – т.е. того, что учтено при интегральном подсчете колебаний энергии взаимодействия, FEP/BCЭ; назовем это квази-динамическим приближением), так и
- динамических факторов – механизма образования и распада начальных комплексов реагирующих соединений с мишенью (катализатором в частном случае) (предмет исследований QSAR и докинга), т.е. энергетических переходов промежуточных переходных состояний (что непременно требует применения квантовохимических методов) и комплексов мишени (катализатора) с продуктами реакции (что также возможно квази-динамическими методами).

В рамках такой обобщенной модели катализатором, в соответствии с определением из учебника, может являться любое вещество, гомогенное или гетерогенное по отношению к реакционной системе, влияющее на ход реакции и остающееся без изменений в этой реакции.

Например, далее рассмотрено моделирование асимметрического синтеза с участием асимметрического индуктора.

3.8.1. Моделирование асимметрического синтеза с участием индуктора

В настоящее время различные методы асимметричного синтеза широко используются в реакциях тонкого органического синтеза и в промышленности, включая хиральные катализаторы [403,404], хиральные растворители [405,406], в

частности, хиральные ионные жидкости [407], временные хиральные вспомогательные реагенты [408] и даже спонтанный асимметричный синтез [409]. Вычислительная химия потенциально позволяет достичь желаемой селективности без тестирования всех возможных условий реакции. В данной работе мы использовали молекулярную механику в методе возмущения свободной энергии (ВСЭ) и квантово-механические расчеты DFT для оптимизации условий реакции Дильса–Альдера и получения наилучшей стереоселективности.

Мы исследовали циклоприсоединение (*E*)-3-(4-нитрофенил)-1-(пиридин-3-ил)проп-2-ен-1-она к цикlopентадиену-1,3 по Дильсу–Альдеру (Схема 16), который приводит к двум энантиомерным продуктам (*R,R,R,S* и *S,S,S,R*). Для оценки потенциальной хиральной индукции хиральным растворителем мы рассчитали свободную энергию сольватации для каждого продукта в смеси этанола с хиральной ионной жидкостью Vmim-CSA или Promim-TSA (Рисунок 45) с помощью возмущения свободной энергии в рамках молекулярной динамики. Результирующие свободные энергии сольватации были по существу одинаковыми для обоих энантиомеров в обеих системах, что указывает на то, что в этом случае не ожидается никакой хиральной индукции (Таблица 21), если предполагать, что реакция протекает под термодинамическим контролем.

Схема 16. Циклоприсоединение по Дильсу–Альдеру (*E*)-3-(4-нитрофенил)-1-(пиридин-3-ил)проп-2-ен-1-она к цикlopентадиену-1,3

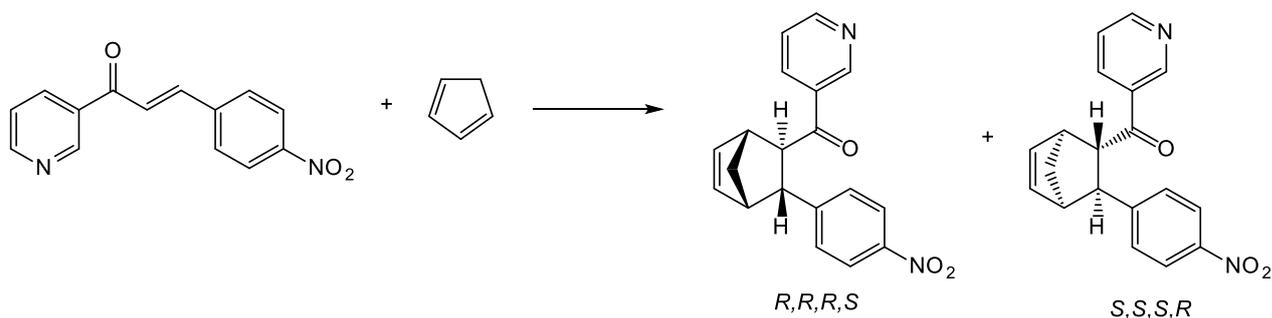


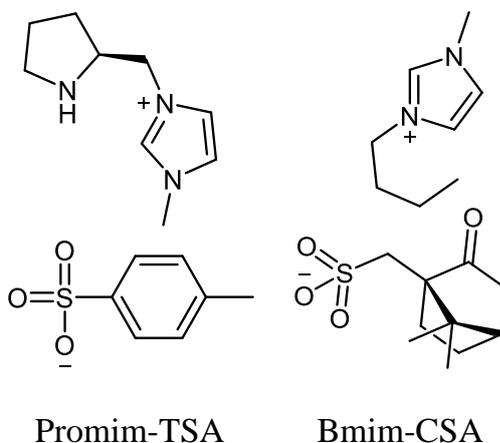
Рисунок 45. Хиральные ионные жидкости

Таблица 21. Свободные энергии сольватации энантиомерных аддуктов (*E*)-3-(4-нитрофенил)-1-(пиридин-3-ил)проп-2-ен-1-она и цикlopентадиена-1,3 в смеси этанола, ионной жидкости (0,5 М) и CuCl₂ (0,02 М), рассчитанные с использованием подхода ММ FEP/BCЭ

Изомер	Ионная жидкость	Кулоновская ΔG, кДж/моль	ВдВ ΔG, кДж/моль	Общий сольволиз ΔG, кДж/моль
<i>RRRS</i>	Vmim-CSA	-24,34±0,99	-58,66±0,19	-83,00±1,00
<i>SSSR</i>		-25,55±1,71	-58,28±0,22	-83,83±1,71
<i>RRRS</i>	Promim-TSA	-25,31±0,62	-57,39±0,40	-82,70±0,74
<i>SSSR</i>		-24,97±0,16	-57,07±0,46	-82,04±0,49

Отсутствие результата в расчете хиральной индукции хиральными жидкостями привело к мысли оценить влияние хирального оксазаборолидинового катализатора (Рисунок 46) на состав продуктов реакции. Геометрическая оптимизация методами квантовой химии (DFT) структуры исходного комплекса продукт-катализатор привела к диссоциации каталитического комплекса (расстояние между бором и кислородом изменилось с 1,55Å в исходном предположении до 3,61Å в конечной структуре). Однако, когда атом азота оксазаборолидина, с учетом протонирования при геометрической оптимизации, равномерно приблизился к тетраэдрическому бору с длиной связи В-О 1,42Å (Рисунок 47), свободная энергия этого комплекса бора для продукта *R,R,R,S* оказалась на 5,75 кДж/моль ниже, чем

для продукта S,S,S,R , что указывает на то, что катализируемая реакция должна происходить энантиоселективно.

Рисунок 46. Индуктор оксазаборолидин для реакции Дильса–Альдера

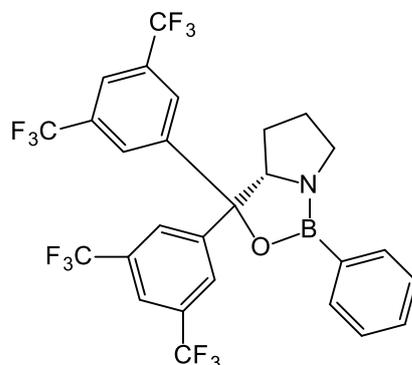
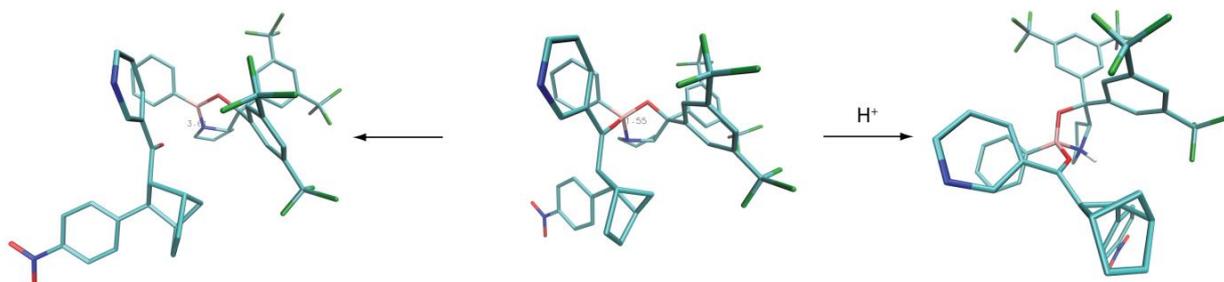


Рисунок 47. Эффект протонирования в катализе оксазаборолидином реакции Дильса–Альдера



Дополнительно энантиоселективность смоделированной реакции экспериментально изучалась в хиральной ионной жидкости или с хиральным катализатором. Как и ожидалось, добавление 0,5 М хиральной ионной жидкости к этанолу привело к образованию рацемического продукта ($ee = 0,04\%$ для Promim–TSA и $3,26\%$ для BMIM–CSA). Добавление 1 мол.% хирального катализатора приводило к умеренной, но различной энантиоселективности ($ee = 11\%$ в Vmim–CSA и $ee = 16\%$ в Promim–TSA), что соответствовало моделируемому предпочтению по отношению к R,R,R,S -энантиомеру.

Стоит отметить, что протонирование азота в оксазаборолидине было абсолютно необходимо для каталитической активности, как это было недавно описано в [410] и независимо показано в данной работе.

Таким образом, мы успешно применили современные методы вычислительной химии для изучения факторов, влияющих на энантиоселективность реакции

Дильса–Альдера. Наша работа обосновывает широкое применение методов молекулярного моделирования для оптимизации условий органических реакций.

С помощью молекулярной механики и квантовой химии мы смоделировали и экспериментально изучили энантиоселективность реакции Дильса–Альдера между (*E*)-3-(4-нитрофенил)-1-(пиридин-3-ил)проп-2-ен-1-оном и циклопентадиеном-1,3 в хиральных ионных жидкостях: BMIM-CSA и Promim-TSA; или в хиральном катализе с хиральным оксазаборолидином.

Расчеты находились в хорошем соответствии с экспериментом, оказалось, что не было стереоселективности с хиральной ионной жидкостью, используемой в качестве соразтворителя, и наблюдали значимую стереоселективность при хиральном катализе.

3.8.2. Дальнейшее продвижение методов динамического моделирования в катализе. Применение ВСЭ и методов квантовой механики в моделировании гомогенного катализа

Гомогенный металлокомплексный катализ с точки зрения моделирования его механизма – благодатный объект исследований. А реакция Сузуки в этом плане изучена весьма обстоятельно. Однако, даже в таких изученных механизмах появляются моменты, принимаемые априори – например, что промежуточный переходный комплекс на последней стадии реакции Сузуки распадается при атаке атомов кислорода воды на атом палладия. Расчет показывает, что это, вероятно, неверное допущение.

Мы сравнили явные и неявные подходы к сольватации при моделировании профиля свободной энергии на заключительном этапе реакции сочетания Сузуки–Мияуры (Suzuki–Miyaura). Оба подхода дали аналогичные ΔG^\ddagger во всех изученных растворителях (бензол, толуол, ДМФА, этанол и вода). Свободные энергии сольватации отдельных компонентов реакции приемлемо скоррелированы для явных и неявных моделей в апротонных растворителях (RMSE = 30–50 кДж./моль⁻¹, R² > 0,71). Однако для этанола и воды корреляция была низкой. Мы связали эту разницу с образованием водородной связи Pd···H–O с Pd(PPh₃)₂, что было неожиданно обнаружено при явном моделировании. Дальнейшие квантово-механические расчеты системы Pd(PPh₃)₂–H₂O подтвердили направление (Pd···H) и стабильность этой связи. Поэтому мы подчеркиваем необходимость рассмотрения явной сольватации

для моделирования Pd-катализируемых реакций в протонных растворителях.

Будучи обнаруженными почти 40 лет назад [411 – 413], реакции перекрестного сочетания, катализируемые металлом, коренным образом изменили конструкцию связи C–C, обеспечивая элегантные и простые пути синтеза соединений, труднодоступных в иных условиях. Несмотря на простоту общей трансформации (которая формально представляет собой реакцию замещения), лежащие в ее основе молекулярные механизмы могут быть чрезвычайно сложными и включать в себя большое количество промежуточных продуктов, распределенных по множеству альтернативных путей реакции [414,415]. Реакция перекрестного сочетания Сузуки–Мияуры протекает через три основных этапа:

- (1) окислительное присоединение R–X к PdL₂,
- (2) реакция трансметаллизации между RXPdL₂ и R'B(OH)₂, приводящая к RR'PdL₂ и
- (3) восстановительному элиминированию, в результате которого получается R–R' и PdL₂ [416].

Две из этих стадий – восстановительное элиминирование и окислительное присоединение – также представлены в ряде других металлорганических реакций. Экспериментальные исследования этапа восстановительного элиминирования затрудняются коротким временем жизни промежуточных продуктов, но методы вычислительной химии могут предоставить информацию об электронной структуре [417] и размере [418] реагирующих молекул, и позволяют выбрать правильный механизм реакции из нескольких альтернатив [419], объяснить влияние лигандов на каталитическую активность [420 – 422], а также влияние растворителя на результат реакции [423]. Фактически свободная энергия сольватации может быть сопоставима с свободной энергией активации [416], а селективность реакции можно контролировать с помощью подходящего растворителя [423], однако проблема выбора модели сольватации мало привлекала внимание. В ONIOM (Собственная N-слойная интегрированная молекулярная орбитальная молекулярная механика) расчетов [424,425] реакций, катализируемых металлом, обычно используется модель континуального растворителя (ONIOM–PCM) [426]. Несмотря на высокую вычислительную производительность, неявные модели сольватации могут приводить к серьезным

ошибкам, что было продемонстрировано в нескольких макромолекулярных симуляциях [427 – 429]. С другой стороны, явное квантово-механическое рассмотрение молекул растворителя имеет неприемлемо высокие вычислительные затраты, в то время как чистые расчеты с помощью ММ оправдываются только для моделирования нековалентных взаимодействий.

Таким образом, конкретный протокол, сохраняющий преимущества обоих подходов, мог бы использовать квантовую механику для расчета энергетического профиля реакции в вакууме и молекулярную механику FEP/BCЭ [430,431] для оценки эффекта сольватации. Этот протокол был первоначально предложен Йоргенсеном для изучения реакций SN2 [432,433], а затем дополнительно расширен до ферментативных реакций [434]. Однако, насколько нам известно, такой подход до сих пор еще не применялся к процессам, катализируемым металлом.

Нами был применен метод QM/ММ–FEP для изучения смоделированного этапа восстановительного элиминирования двух Ph-лигандов из Pd(PPh₃)₂Ph₂ (ключительный этап реакции перекрестного сочетания Сузуки–Мияуры) [411,416,435], и мы сравнили результаты с традиционной континуальной моделью сольватации.

1) Квантово-механические расчеты были использованы для определения параметров свободной энергии и параметров равновесия (атомные координаты, атомные заряды, силовые константы и т. д.) реагентов и продуктов в вакууме.

2) Затем для каждой молекулы свободную энергию сольватации рассчитывали с использованием метода возмущения свободной энергии путем усреднения изменения энергии в статистическом ансамбле, генерируемого в ходе моделирования с помощью молекулярной динамики.

Такой подход несколько отличается от первоначального метода Йоргенсена (который подразумевал BCЭ-переход из реагента в ПС (переходное состояние) и продукты при явном учете растворителя, а не полную десольватацию каждого компонента реакции), поскольку основной целью работы было оценить свободную энергию абсолютной сольватации.

Профиль свободной энергии реакции получен путем суммирования его профиля свободной энергии в вакууме и соответствующих надбавок сольватации. Этот

метод можно использовать для сравнения профиля свободной энергии реакции в разных растворителях и в разных средах путем непосредственной оценки взаимодействий растворенного комплекса и растворителя, что может представлять особый интерес для оптимизации условий реакции.

Методы моделирования:

Реакция была смоделирована в наборе с различными растворителями (бензол, толуол, ДМФА, этанол и вода) с использованием методов QM+MM/FEP и PCM SMD [436] (SMD является стандартным методом расчета свободной энергии сольватации в Gaussian09). С использованием двухслойного подхода ONIOM, как описано в [437] с Gaussian09 [438], рассчитывали равновесные геометрии исходного состояния, ПС (переходное состояние) и продуктов, а также соответствующие свободные энергии Гиббса в вакууме.

Комплексы были разделены на два слоя: первый слой включал палладий, атомы фосфора и связанные с палладием фенильные кольца (и воду для комплексов $\text{Pd}(\text{PPh}_3)_2\text{-H}_2\text{O}$), второй слой включал заместители в фосфиновом лиганде. Первый слой обрабатывался на высоком уровне B3LYP / BSI (B3LYP обычно используется для изучения кросс-сочетания по Сузуки [416], базовый набор BSI состоит из базового набора SDD для Pd и 6-311G(d) для других атомов). Второй слой рассчитывали на относительно низком уровне HF / Lanl2mb. Энергии сольватации SMD были рассчитаны для варианта геометрии с оптимизацией растворителя. Параметризацию по CHARMM [437,439]. MM выполняли с помощью плагина Paratools для VMD [440] для воспроизведения молекулярных энергий QM для 10 различных атомных координат, генерируемых MD.

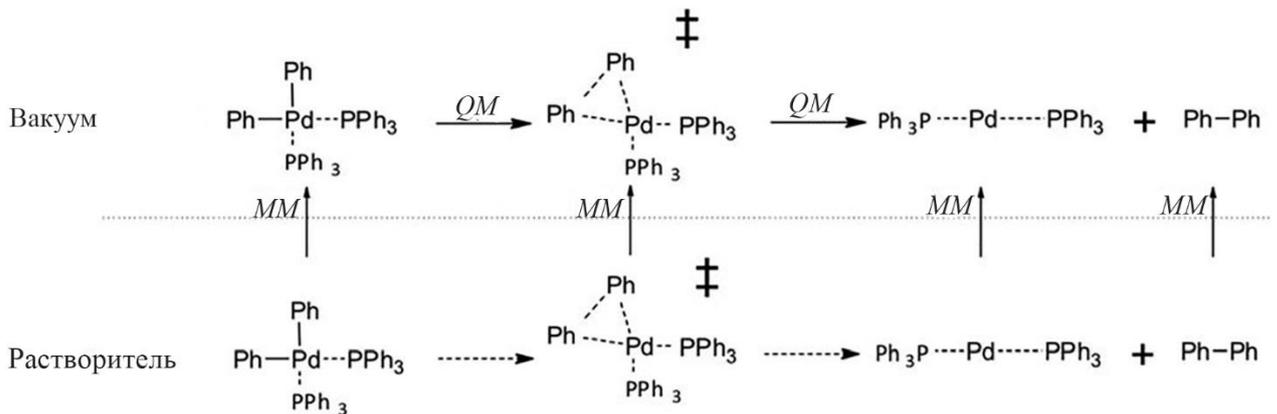
Заряды были определены программным пакетом RESP, параметры Lennard-Jones для Pd взяты из работы [441]. Использовали водную модель TIP4P с 4 сайтами [442]. В Gromacs [443] были рассчитаны свободные энергии сольватации MM-FEP/BCЭ. Вкратце, молекулу растворенного вещества помещали в кубический (с ребром 70Å) ящик растворителя, используя пакет Gromacs. Затем BCЭ использовали для отключения взаимодействия растворитель-растворитель.

Расчеты BCЭ проводились с шагом 10 Å (отдельно для Ван-дер-Ваальсовых и кулоновских взаимодействий), каждый шаг включал 100 ps NVT-уравновешиваний, 500 ps NPT-уравновешиваний и 10 ns типовой динамики NPT. Перекрытие фазового

пространства – это параметр управления, который показывает, насколько вероятны соседние λ -состояния, которые были правильно отобраны, – оценивался на основе относительных энтропий состояний [444].

Независимо друг от друга был применен классический метод усреднения по шкале для оценки статистических ошибок расчетов свободной энергии, которые показывают, достаточно ли выделенной длины траектории молекулярной динамики, чтобы получить статистически устойчивое среднее значение. Полученные ΔG реагентов, ПС и продуктов были рассчитаны как сумма энергии вакуума QM и свободной энергии сольватации MM (Схема 17).

Схема 17. Механизм последней стадии реакции Сузуки–Мияуры. Расчеты QM проводились для получения ΔG^\ddagger в вакууме, а MM FEP обеспечивала свободную энергию сольватации реагентов и продуктов



Переходы ВСЭ проходили без резких флуктуаций (см. Рисунок 48 в качестве примера) и продемонстрировали хорошую стабильность на протяжении всего периода (10 ns), что подтверждается относительно небольшой статистической ошибкой (<1 кДж $^{-1}$ для всех возмущений), оцененной по блочному усреднению. Значения Π , рассчитанные для каждого λ -окна каждого перехода, были в основном положительными (за исключением небольших отрицательных значений не хуже $-0,5$ на несколько шагов), что указывает на то, что смещение в расчетах свободной энергии, потенциально вызванное недостаточным перекрытием фазового пространства, было незначительным. Несколько отдельных переходов с значением $\Pi < 0,5$ (последние этапы расщепления $\text{Pd}(\text{PPh}_3)_4$ и ПС ВдВ) были повторно имитированы с уменьшенным размером окна ($\Delta\lambda = 0,05$). Результирующие свободные энергии были по

существованию такими же, как и для $\Delta\lambda = 0,1$, в то время как значение Π было $> 0,5$, что означало, что полное перекрытие фазового пространства было наконец достигнуто.

Рисунок 48. Кумулятивное и дифференциальное изменение свободной энергии при десорбции $\text{Pd}(\text{PPh}_3)_2\text{Ph}_2$ в воде

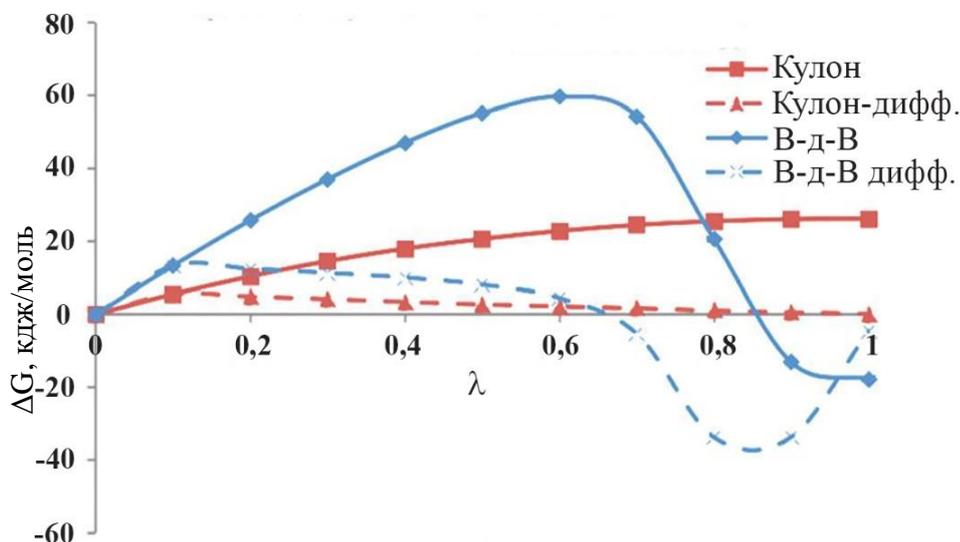
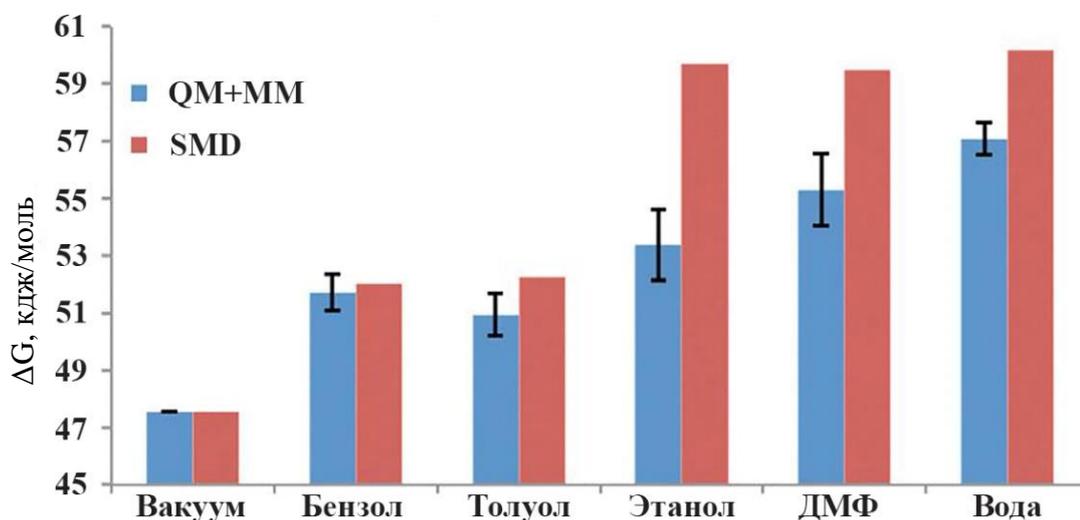


Рисунок 49. Свободная энергия активации восстановительного элиминирования PhPh из $\text{Pd}(\text{PPh}_3)_2\text{Ph}_2$ в разных растворителях. Шкалы ошибок для $\Delta G \neq \text{QM} + \text{FER}$ получены из блочного усреднения



Приемлемое соответствие между SMD и QM + FER/BCЭ для энергии активации наблюдалось в неполярном бензоле и толуоле (Рисунок 49). Для полярных растворителей разница была значительной, составляющей 3–6 кДж/моль⁻¹. Энергия активации во всех растворителях была больше, чем в вакууме, главным образом за

счет электростатической сольватации, которая значительно отличалась между исходным комплексом и менее полярным переходным состоянием ($\mu = 10,9$ против $7,2$ соответственно), в то время как неэлектростатическая составляющая практически не изменялась из-за аналогичной молекулярной площади поверхности.

Мы сравнили свободные энергии сольватации реагентов, ПС и продуктов в каждом из изученных растворителей (Рисунок 50).

Большое различие между сольватацией FEP и SMD ΔG наблюдалось для всех растворителей со средним RMSE около $30\text{--}50$ кДж/моль⁻¹ (Таблица 22). Хорошая корреляция ($R^2 = 0,71 - 0,85$) наблюдалась для апротонных растворителей (бензол, толуол и ДМФА).

Таблица 22. Корреляция между энергиями сольватации SMD и FEP для разных растворителей

	RMSE, кДж-моль ⁻¹	Ошибка, кДж/моль ⁻¹	R2:
Бензол	32,6	-25,0	0,75
Толуол	31,7	-25,3	0,81
Этанол	40,4	-29,4	0,43
ДМФА	48,0	39,6	0,71
Вода	28,7	-8,5	0,69 (a <0)

Однако хуже соответствие было для этанола ($R^2 = 0,41$) и воды (отрицательная корреляция с $R^2 = 0,69$). Анализ траекторий MD (Рисунок 50) показал, что центральный атом Pd в Pd(PPh₃)₂ образует связь Pd \cdots H-O (Рисунок 51), которая не учитывалась в расчетах SMD. Поскольку взаимодействие Pd-вода полярно, мы провели дополнительные вычисления QM, чтобы выяснить, какая именно полярность энергетически выгодна. Исходя из различных исходных геометрий и с использованием базисных наборов 6-311G, рассчитанная оптимизация системы Pd(PPh₃)₂-H₂O неизбежно приводила к образованию Pd \cdots H-O. Реальность этого взаимодействия дополнительно подтверждается исследованиями кристаллических структур комплексов металл-лиганд [445 – 447].

Рисунок 50. Свободные энергии сольватации для реагента, продуктов и переходного состояния, рассчитанные с использованием SMD или MM/FEP в наборе растворителей

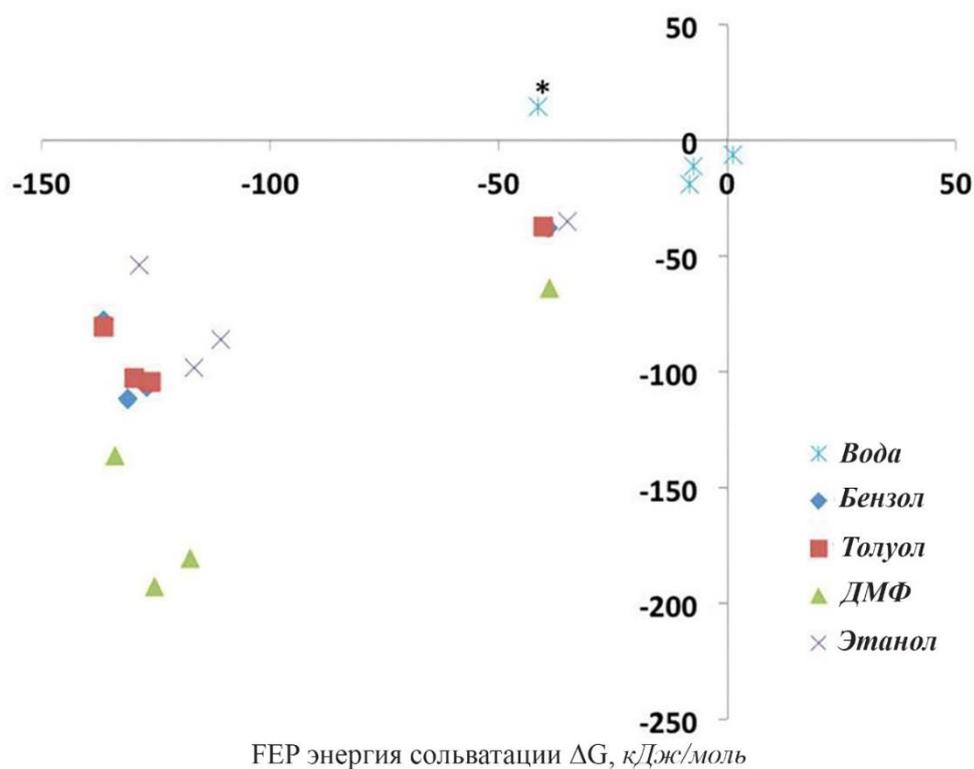
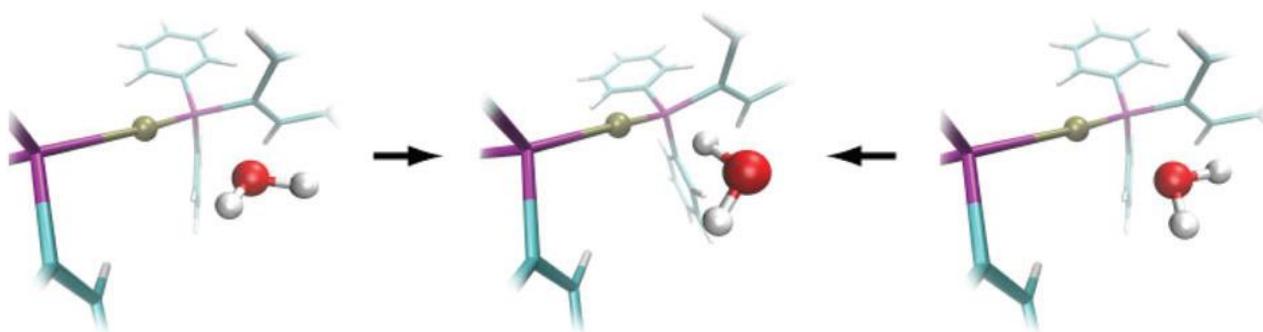


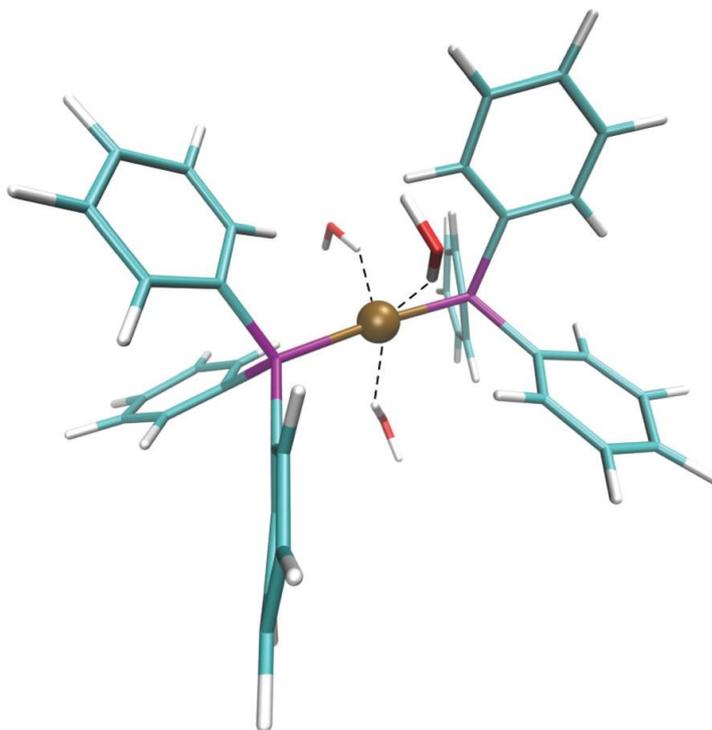
Рисунок 51. Оптимизация $\text{Pd}(\text{PPh}_3)_2 \cdot 2\text{H}_2\text{O}$ методом QM приводит к формированию связи $\text{Pd} \cdots \text{H}-\text{O}$ в моделировании реакции Сузуки–Мияуры



Чтобы продемонстрировать вклад водородной связи $\text{Pd} \cdots \text{H}$ во всю сольватирующую энергию, мы повторили SMD-расчет сольватации $\text{Pd}(\text{PPh}_3)_2$ с тремя молекулами воды первого слоя, обработанными явно (Рисунок 52). Этот подход, известный как модель континуума кластера [448], может быть использован для систем с конкретными ближними взаимодействиями растворителя [449]. Свободная энергия

сольватации оптимизированного комплекса приближалась к значению ВСЭ гораздо ближе (-66 кДж/моль $^{-1}$ относительно -41 кДж/моль $^{-1}$ по ВСЭ), чем чисто неявный подход (14 кДж/моль $^{-1}$ относительно -41 кДж/моль $^{-1}$ по ВСЭ). Поэтому это взаимодействие может способствовать наблюдаемой разности явных и неявных энергий сольватации.

Рисунок 52. В соответствии с ММ–FEP моделированием, незранированный атом палладия в $\text{Pd}(\text{PPh}_3)_2$ образует связь $\text{Pd}\cdots\text{H}-\text{O}$ с молекулами воды



Плохое количественное соответствие, наблюдаемое для ДМФА, по-видимому, было вызвано неподходящей параметризацией этого растворителя для SMD. Литературные данные по растворителю для ДМФА противоречивы и варьируются от $2,44\text{\AA}$ [450] до $2,64\text{\AA}$ [451]) и $3,48\text{\AA}$ [452], что указывает на необходимость дальнейшей проверки.

Выводы: Проведено сравнение явных и неявных протоколов для моделирования эффектов растворителя в моделированной органической реакции, катализируемой металлом. Хорошая корреляция наблюдалась для свободной энергии активации, однако возникла большая разница (в десятки кДж/моль $^{-1}$) для свободной энергии сольватации отдельных компонентов реакции, которая была частично отменена в энергии активации, но не в энергии полной реакции. Наиболее

интересным результатом, который мы получили, было отсутствие надлежащего не-явного описания специфических взаимодействий $\text{Pd}(\text{PPh}_3)_2$ с протонными растворителями, которое наблюдалось с более сложной моделью MM-FEP и подтверждено расчетами QM.

Итоговый результат: в последней стадии реакции Сузуки–Мияуры комплекс палладия образует водородную связь $\text{Pd}\cdots\text{H}$, а не $\text{Pd}\cdots\text{O}$, как принято считать в настоящий момент.

Из чего следует, что сольватация должна рассматриваться как альтернатива РСМ при моделировании Pd-катализируемых реакций в протонных растворителях.

Общий вывод этой части работы: мы показали, что, рассчитав зависимость изменения свободной энергии системы катализатор–лиганд (лиганды) от времени, можно с высокой степенью достоверности описать механизм каталитического действия.

Как в случае моделирования асимметрической индукции, так и в случае моделирования металлокомплексного катализа основным методом было построение временной зависимости свободной энергии системы на последней стадии моделируемого процесса.

Однако, хоть это и дало свои результаты, подтвержденные экспериментом, логика подсказывает, что моделировать все-таки необходимо скорость-определяющую стадию процесса, которой, в большинстве каталитических реакций, является стадия образования и распада каталитического комплекса. Естественно, с учетом участия растворителя и с рассмотрением всех конформационных переходов и возможных вариантов распределения заряда в образующихся переходных состояниях. Здесь уже необходимо более полное задействование методов квантовой химии.

Для отработки квантовохимических методов расчета переходного состояния начнем с моделирования известных реакций.

В применении к переходным состояниям химических реакций задача заключается в том, чтобы обеспечить макроскопические наблюдения на основе оптимального усреднения событий на атомарном уровне. Исторически моделирование проводилось на базе сильно усредненных физических свойств тестовых систем [453].

Однако, в последние годы все возрастающие вычислительные возможности помогли сдвинуть изучение системы с сильно усредненных расчетов к более точным и детализированным, с параллельным усложнением самих систем к более высокомолекулярным. Эту тенденцию можно легко увидеть в дизайне новых материалов [454]. Анализ больших массивов данных кристаллографических синтонов в последнее время привел к открытию первого примера супрамолекулярных стереоэлектронных эффектов, который приводит к инверсии предпочтительной конформации карбоксильной группы в случаях сильного связывания с водородом (например, в гексамере инсулина) [455].

В другом недавнем исследовании полная вычислительная оценка возможных конформаций каталитических переходных состояний позволила исследователям сконструировать новый эффективный катализатор [456].

В теоретической химии методы изучения конформационных ансамблей часто представлены молекулярной динамикой (MD) [457], которая играет решающую роль в теоретических исследованиях в биохимии, катализе и других важных областях современной химии. Ее эффективность обусловлена способностью просто описывать химические системы во всем их многообразии.

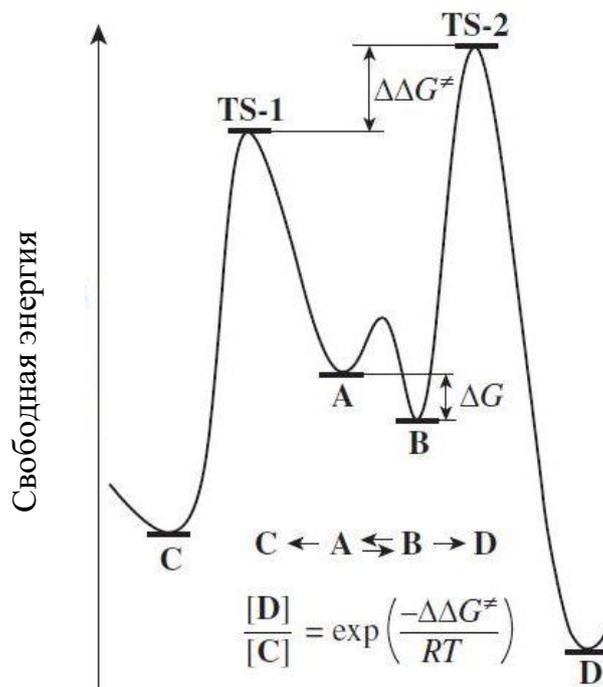
Однако, чтобы быть надежной, молекулярная динамика должна быть пролонгирована во времени, т.е. учитывать столько состояний системы, сколько требуется для сходимости макроскопических данных. Из этого факта возникает серьезный недостаток MD – ее потребность в вычислительных мощностях: даже при использовании полуэмпирических методов требуются суперкомпьютеры для изучения больших атомных массивов биомолекул.

Молекулярная динамика *ab initio* еще дороже и редко применяется к системам с большим количеством атомов. Классические молекулярные поля обычно бесполезны при изучении химических реакций, поскольку они не могут описать разрыв и образование связей (исключение составляет ReaxFF [458], которая требует очень строгой параметризации под каждую конкретную реакцию).

Альтернативный путь – учет ансамбля механизмов реакции, то есть, изучение ансамблей переходных состояний химической реакции (возможно, с поправками, вытекающими из вариационной теории переходных состояний (Variational Transition State Theory [459]) либо теории Крамера [460]), опираясь на постулат, что кинетика

реакции определяется относительной энергией переходного состояния (ПС, или transition state, TS, точка минимума градиента реакции). Хотя любая реакция может протекать через различные ПС, принадлежащие одному и тому же механизму (далее, ПС-конформеры), обычно считается, что реакция протекает через ПС, связанное с самой низкоэнергетической конформацией реагента. Однако это не совсем так, в случае высокой лабильности (скорости превращения друг в друга) конформаций реагентов [461] (эмпирически найденное соотношение – в десять раз быстрее скорости основной реакции [462,463]). В этом случае реакция регулируется принципом Кертина–Гаммета [461], согласно которому реакция протекает через энергетически выгодные переходные состояния, а это совсем не обязательно связано с низшими конформациями реагентов (Рисунок 53). В соответствии с принципом Кертина–Гаммета, соотношение продуктов $[D]/[C] = \exp(-\Delta\Delta G^\ddagger/RT)$, в частности, не зависит от ΔG .

Рисунок 53. Иллюстрация принципа Кертина–Гаммета. Если переходы А–В достаточно быстрые, реакция идет через переходное состояние TS-1



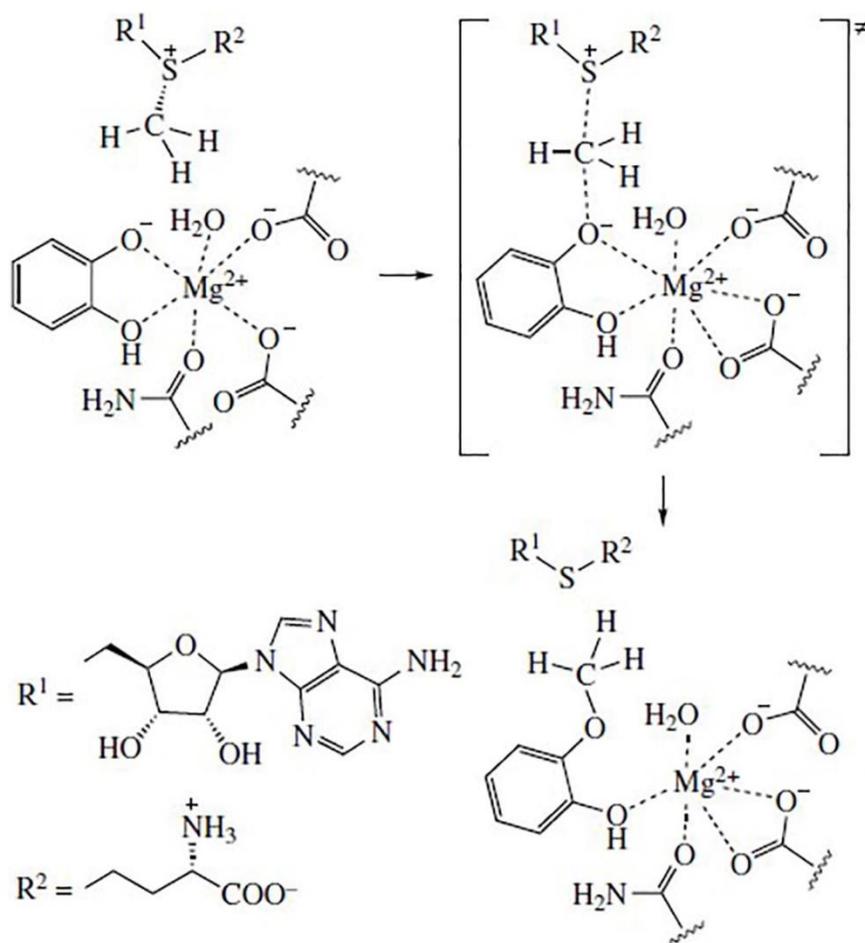
Очевидно, что для корректного моделирования реакции необходимо моделировать все конформации переходного состояния этой реакции. Однако, далеко не все программные пакеты имеют такую функциональность. Для моделирования и генерации конформационных состояний можно использовать бесплатные пакеты

OpenBabel [464] and RDKit [465] (оба требуют написания дополнительных скриптов) или коммерческий Schrödinger MacroModel [466].

Катализируемая модельная реакция показана на Схеме 18. Эта реакция интенсивно изучалась как молекулярной динамикой [467], так и методами QM/MM [468]. Мы упростили схему и смоделировали его без участия большей части фермента, заменив R^1 , R^2 и все точки связывания на Схеме 18 метильными группами. Отсутствие фермента делает комплект возможных ПС структурно разнообразным, и позволяет находить ПС, стерически невозможные в комплексе с ферментом.

Схема 18. Сайт связывания катехол-*O*-метилтрансферазы.

Обрезанные связи – точки связывания с ферментом



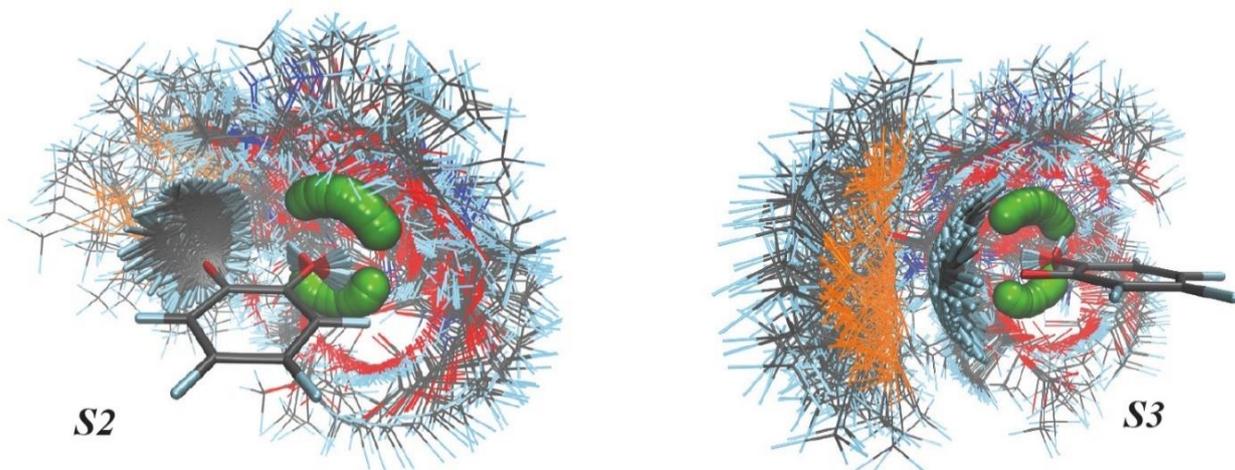
На первой стадии были сгенерированы все возможные конформации переходного состояния. Для этого была построена геометрия гипотетического ПС (Рисунок 54), а конформационный поиск был выполнен с использованием MacroModel [466] со всеми расстояниями лиганд- Mg^{2+} .

Рисунок 54. Возможные конформации ПС ингибирования катехол-*O*-метилтрансферазы



После оптимизации по энергетике ПС отфильтрованы все конформации с энергией больше 300 ккал/моль. Осталось 205 конформаций, показанных на Рисунке 55 (*S2* и *S3*). Согласно Рисунку 55, сгенерированные конформации включают как моно-, так и бидентатные связи катехола в Mg^{2+} , а метильная группа всегда расположена соответствующим образом для реакции, соответствующей механизму SN_2 .

Рисунок 55. Две проекции набора конформаций, фиксация вокруг С и О-атомов.



Катехины и метильные группы показаны как закрытые области; Ионы Mg^{2+} показаны в виде сфер, а остальные молекулы показаны как линии. Окраска: Mg – серый, S – оранжевый, O – красный, N – синий, C – серый, H – светло-голубой.

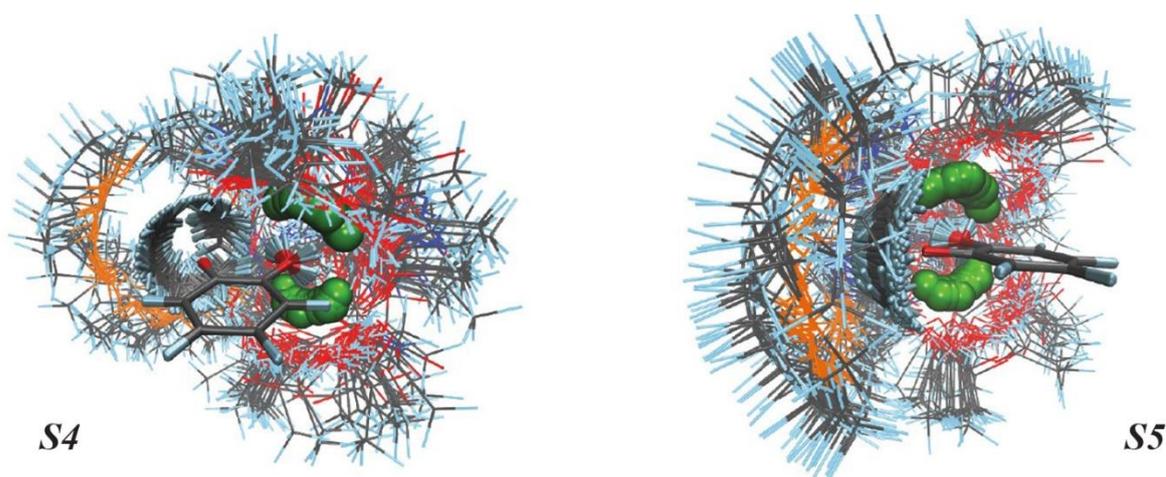
Квантово-механические расчеты осуществлены при помощи Gaussian09 [438] в полуэмпирическом варианте, описанном для СОМТ в [468]. Для уточнения результатов использовался метод DFT [469,470].

Поскольку геометрия часто отличается при расчетах методами QM и MM,

прямой поиск ПС на базе геометрии ММ может привести к неправильному ПС, что неприемлемо из-за множественности конформаций ПС. Чтобы избежать этого, оптимизация проведена последовательно методами РМ6–QM(constrained)–QM и в результате фильтрации (не сошлись 141 ПС) осталось 162 переходных состояния; их суперпозиции показаны на Рисунке 56 (*S4* и *S5*).

Рисунок 56. Оптимизированные варианты набора ПС.

Цветовая раскладка – как в Рисунке 55.



Геометрия оптимизированных ПС хорошо согласуется с базовыми теоретическими построениями [493].

Все построенные конформации имеют близкую энергию вращения, разброс не более 8 ккал/моль.

Этот подход может также использоваться вместе с методологией QM/ММ для моделирования связывания ферментов. Наш подход позволяет применить более точные методы неэмпирических расчетов больших систем, и более полного учета эффектов электростатических полей и взаимодействий, которые, как показано в предыдущих наших исследованиях, играют важную роль во взаимодействиях лиганда и мишени при образовании комплекса, в том числе и при ферментативном катализе.

3.8.3. Применение модельных принципов расчета переходных состояний к структуре переходного состояния ферментокатализируемой внутримолекулярной реакции Дильса–Альдера

Все наши последние исследования механизма реакций и состава продуктов, т.е. моделирование направленного синтеза, касаются реакции Дильса–Альдера, как характерного объекта для таких исследований. Во-первых, реакция Дильса–Альдера – краеугольный камень органического синтеза [471], и не только в лаборатории, но и в живой природе. Во-вторых, в этой реакции комплекс исходных соединений, циклическое переходное состояние и комплекс продуктов представлены одним и тем же набором атомов, и изучение касается образования и разрыва химических связей, что существенно облегчает расчет средствами молекулярной механики и квантовохимическими методами.

Это имеет особое значение для синтеза биологически активных соединений и фармацевтических препаратов в медицинской химии, там, где для выяснения структуры продуктов требуется расчет структуры и электронного строения сложных циклических промежуточных комплексов. Однако, несмотря на общую тенденцию проводить синтезы в медицинской химии «зеленым» способом, биосинтез (биокатализ) в реакции Дильса–Альдера пока практически недоступен [472]. Было показано, что среди очень немногих ферментов, катализирующих, по крайней мере, формальное циклоприсоединение (т.е. формально относящихся к реакции Дильса–Альдера), действуют как механизмы, не имеют никакого отношения к классическому циклоприсоединению [473], тогда для многих других «формально Дильса–Альдера» реакций механизмы остаются неизвестными [474 – 477].

Одним из наиболее перспективных кандидатов на роль проводника «зеленой химии» в биокатализе является SpnF, который ускоряет в 500 раз реакцию циклоприсоединения (Схема 19) в биосинтезе Спинозина А, эффективного природного, т.е. «зеленого» инсектицида [478]. SpnF примечателен тем, что он был первым обнаруженным природным ферментом, катализирующим только этап циклоприсоединения [479].

Несмотря на то, что SpnF был обнаружен семь лет назад [480], точный механизм катализируемой им реакции до сих пор не выяснен. Для SpnF были предложены несколько механизмов основной реакции циклизации (Схема 19). Они изучены квантовомеханически [479,481,482], изучен также кинетический изотопный эффект [483], но реальный механизм базовой реакции никем не был доказан.

В работе [481] изучали циклизацию усеченного субстрата (**S** на Схеме 19), убрав из рассмотрения пятичленное алифатическое кольцо и макроцикл с тремя гидроксильными группами, в расчетах DFT и MP2 заменив их метильными и этильными группами. Авторы пришли к выводу, что SpnF катализирует согласованную, асинхронную реакцию Дильса–Альдера, загоня субстрат в надлежащую конформацию и понижая его энергию активации путем стабилизации высокополяризованной структуры ПС. Гордеев и Анаников [479] изучили влияние заместителей на энергию активации базовой реакции и пришли к выводу, что главная роль SpnF заключается в стабилизации субстрата. Последние газофазные молекулярные симуляции методом DFT [482] предположили, что реакция может протекать через бис-перициклическое ПС, которое может в конечном итоге либо непосредственно привести к продукту (**P**), либо к **I**–[6+4], которое через перегруппировку Коупа переходит в **P** (Схема 19). Изучить наиболее стабильное ПС, конечно, полезно, но гибкая структура макроцикла предполагает, что одного ПС, даже очень стабильного, может быть недостаточно, чтобы получить полную картину механизма реакции.

Цель нашего исследования в этой части работы – количественно и исчерпывающе оценить возможные переходные состояния, ведущие от **S** к **P**, и определить истинную природу основной реакции в воде, как первый шаг к моделированию реакции, катализируемой SpnF. Мы стремимся отобразить весь набор возможных переходных состояний для изучаемой реакции. Как было обнаружено в [484], экстремумы свободной энергии системы для механизмов Дильса–Альдера и бис-перициклического ПС совпадают с перегибами кривой потенциальной энергии, поэтому каноническая теория ПС должна обеспечить хорошее описание основной реакции.

Соответственно, мы рассматриваем следующие механизмы основной реакции Дильса–Альдера: канонический [480] (**DA**), бис-перициклический [482] (**BPC**) и бирадикальный [483] (**BR**). Поскольку бирадикальный механизм может

осуществляться двумя разными путями (см. Схему 19), вариант **BR-1** предпочтительнее, поскольку большая делокализация электронов обеспечивает более низкую энергию ПС [485]. Мы также протестировали гипотетический механизм, который назовем «**altDA**», поскольку он начинается с альтернативного циклоприсоединения Дильса–Альдера, за которым следуют две перегруппировки Коупа (Схема 19).

Мы изучили все возможные переходные состояния всех путей реакции для катализируемого белком SpnF циклоприсоединения (реакция Дильса–Альдера) (Схема 19). Для нахождения всех конформаций реакции, были

- 1) сгенерированы все конформации продукта;
- 2) оптимизированы под ПС две формообразующие ПС связи;
- 3) оптимизированы структуры до ближайших минимумов, ограничивающие формообразующие связи;
- 4) выполнена безусловная оптимизация структур в сторону ближайших переходных состояний.

Принцип Кертина–Гаммета, примененный к найденным конформациям переходного состояния, утверждает, что эта реакция в основном протекает через бис-перициклические переходные состояния. Однако, в реакциях, где переходное состояние не зависит от субстрата или продукта, нужно явно рассчитывать конформации всех возможных переходных состояний. Использован следующий алгоритм:

1. Строится предположительное ПС и оптимизируется помощью метода квантовой механики (QM). ПС могут быть построены автоматически с использованием эвристического метода [486].

2. Фиксируются все связи и выполняется принудительный поиск всех конформеров с использованием либо молекулярной динамики (ММ– или QM), либо минимизации энергии конформеров, полученных вращением вокруг фиксированных связей с постоптимизацией. Или, наконец, комбинацию обоих методов.

3. Фильтруются структурно идентичные ПС и ПС со слишком высокой энергией.

4. С сохранением всех ограничений, каждая из оставшихся структур минимизируется методами QM, удаляются дубликаты.

5. Полученные оптимизированные структуры переводятся в ПС методами квантовой химии; удаляются дубликаты.

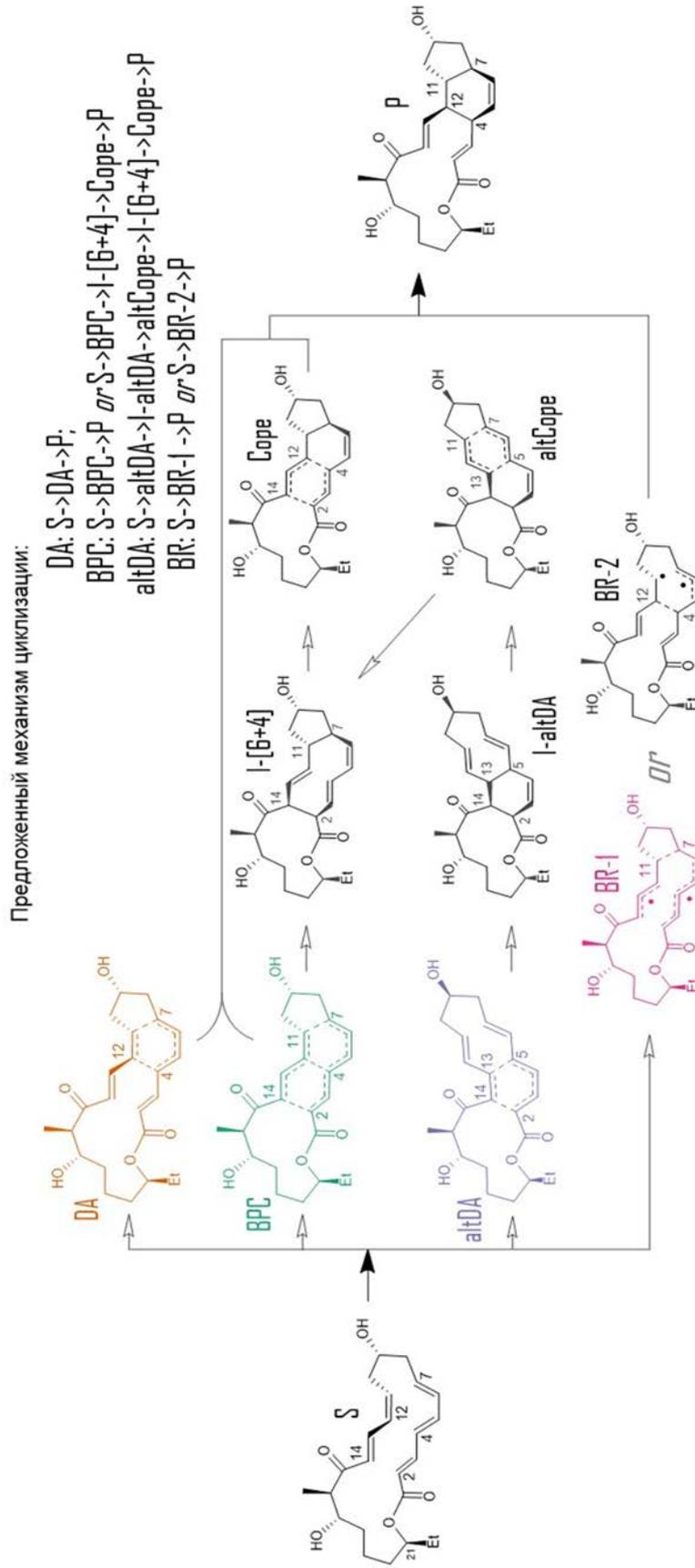
6. При необходимости корректируется поверхность потенциальной энергии с помощью методологии, описанной для реакции SN2, которая происходит в активном сайте катехол-*O*-метилтрансферазы (англ. Catechol-*O*-methyl transferase (COMT)). COMT представляет собой метилтрансферазу, которая регулирует нейротрансмиттеры посредством металлирования. Он включает *S*-аденозилметионин (SAM) и Mg²⁺ в его активном сайте.

Квантово-механические расчеты были проведены программой Gaussian09 D01 [438]. Упрощение при моделировании механизма коснулось этильной группы при C21, замененной на метильную. Для перекрестной проверки энергии найденных стационарных точек были пересчитаны с помощью метода PBE0 [487]–D3 [488], который недавно был обоснован теоретически [469]; и получили те же результаты.

На первом этапе мы создали возможные конформации **P**, используя недавно опубликованную методологию, разработанную специально для макроциклов [489]. В результате были созданы 1971 структур, которые были оптимизированы с помощью силового поля MM3. После удаления дубликатов со среднеквадратичным отклонением координат неводородных атомов ниже 0,2Å осталось 560 уникальных структур. Каждая структура была оптимизирована, и был выполнен поиск [4+2]–ПС, ведущего к текущей конформации **P** (подробнее см. Схему 19). Это привело к созданию 376 уникальных ПС с разбросом значений энергии в пределах 30 ккал/моль от минимального значения (ПС с более высокими энергиями были исключены из дальнейшего рассмотрения).

Найденные переходные состояния имеют сходную геометрию реагирующего фрагмента с углом C4–C7–C11–C12 в диапазоне от 0° до 30° для 85% из них. Они представляют собой конформации, приводящие к **P** через образование связей C4–C12 и C7–C11, поэтому мы использовали их геометрию при моделировании ПС для других механизмов. 376 ПС, сгенерированных на предыдущем этапе, относятся либо к **DA**, либо к **ВРС**–механизмам, однако некоторые пары ПС могут быть слиты вместе, если они приводят к одному и тому же продукту. Мы получили эти дополнительные конформации ПС, начиная с уже найденных ПС, сокращая связь C2–C14. Так обнаружили 8 дополнительных уникальных структур ПС. Расширенный поиск путем увеличения расстояния C2–C14 не привел к появлению новых ПС.

Схема 19. Формальноно [4+2] циклоприсоединение, катализируемое SpnF. Показаны все предложенные механизмы: Diels-Alder (DA), бис-перидицические (BPC), бирадикальные (BR) и альтернативные Diels-Alder (altDA).



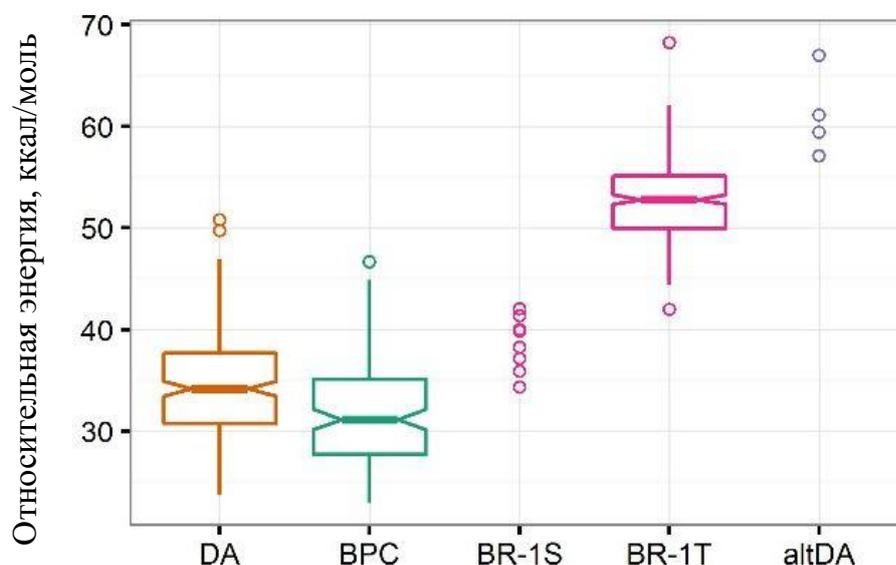
Механизм **BR-1** должен иметь меньшую энергию активации [490], чем **BR-2**, поэтому был изучен первым. На базе предварительно полученных структур 376 ПС рассчитаны 331 уникальных триплетных бирадикальных ПС (**BR-1T**), все гораздо более энергетически насыщенные, чем соответствующие синглетные ПС. Наинизшее по энергетике ПС **BR-1T** находится на 19 ккал/моль выше, чем общий низший уровень ПС. Взяв за образец аналогичную реакцию [489], и базовый набор триплетных ПС, мы провели поиск синглетных бирадикальных ПС (**BR-1S**).

Были смоделированы 9 **BR-1S** ПС, причем самый низкоэнергетический из них был на 11 ккал/моль выше самого низкого небирадикального TS. Следовательно, бирадикальные механизмы энергетически невыгодны и не способствуют реакции в воде.

Кроме того, мы провели моделирование ПС **altDA**, базирясь на 376 начальных структур ПС; и только 4 ПС оказались в рамках энергетических параметров. Самый низкоэнергетический из них – на 34 ккал/моль выше, чем энергия самого низкого по энергетике ПС (Рисунок 57).

Таким образом, единственными возможными маршрутами для основной реакции в отсутствие фермента являются **DA** и **BPC**, для которых было найдено в общей сложности 384 ПС, т.е. 384 субстрата и такое же количество продуктов Дильса–Альдера. Все субстраты соответствуют **S**, а продукты получаются двух видов: 371 соответствуют **P**, а остальные 13 – I-[6+4]. Для последнего механизма были найдены ПС Коупа (см. Схему 19), которые оказались по энергии ниже ПС предыдущего циклоприсоединения по меньшей мере на 7 ккал/моль. Следовательно, переходное состояние типа **BPC** является наиболее предпочтительным в механизме $S \rightarrow [BPC] \rightarrow I-[6+4] \rightarrow [Cope] \rightarrow P$.

Рисунок 57. Рассчитанные по методике [491] ПС механизмов **DA**, **BPC**, **BR-1S**, **BR-1T** и **altDA**) по отношению к низшей по энергии структуре **S**.



Из-за сложности механизма **BPC** найденные ПС не могут быть однозначно отнесены к механизмам **DA** или **BPC**: невозможно решить, основываясь только на составе продуктов, к какому из механизмов, **DA** или **BPC** отнести механизм данной реакции. И вручную этого сделать также нельзя, поэтому для решения этой задачи мы применили квантовую теорию атомов в молекулах (QTAIM) [492], которая идентифицирует химическое связывание как «электронный мост» («направление связывания»), связывающий атомы этой химической связи. По определению, химическая связь в терминах QTAIM требует наличия градиента электронной плотности (ЭП), связывающего максимумы электронной плотности ЭП на ядрах химической связи. Ничего нового по сравнению с классическими представлениями, но наличие строгого математического аппарата делает эту теорию полезной в наших расчетах.

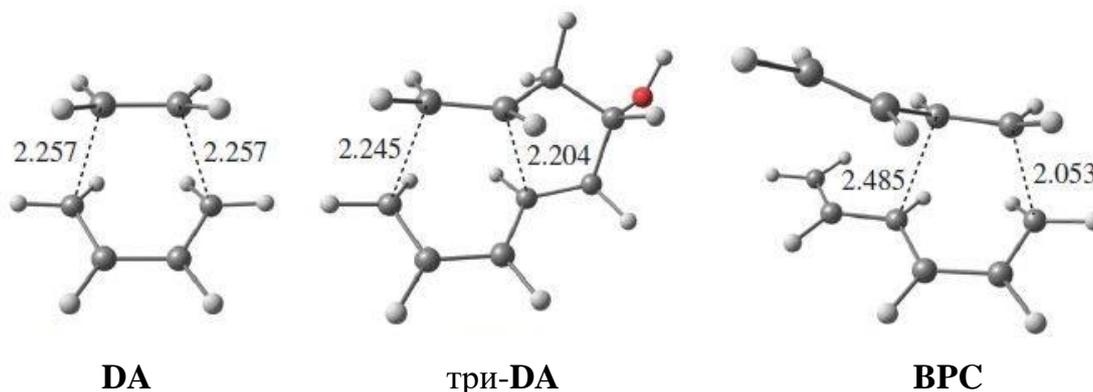
Распределение зарядов ПС играет значительную роль в ферментативном катализе [493] и ускоряет формирование связей в реакции Дильса–Альдера [494]. Мы изучили перераспределение атомных зарядов QTAIM при переходе от **S** к соответствующему ПС (**DA** либо **BPC**).

Модельное перераспределение заряда также рассчитано в трех системах отсчета:

- (1) простая реакция Дильса–Альдера между бутадиеном и этиленом (**DA**),

- (2) Трехатомные взаимодействия енового и диенового компонентов в основной реакции (три-**DA**); и
- (3) модельная бис-перициклическая реакция между гексатриеном и бутадиеном (**BPC**) (Рисунок 58).

Рисунок 58. Оптимизация геометрии (H_2O) трех модельных систем (длины связей приведены в Å).



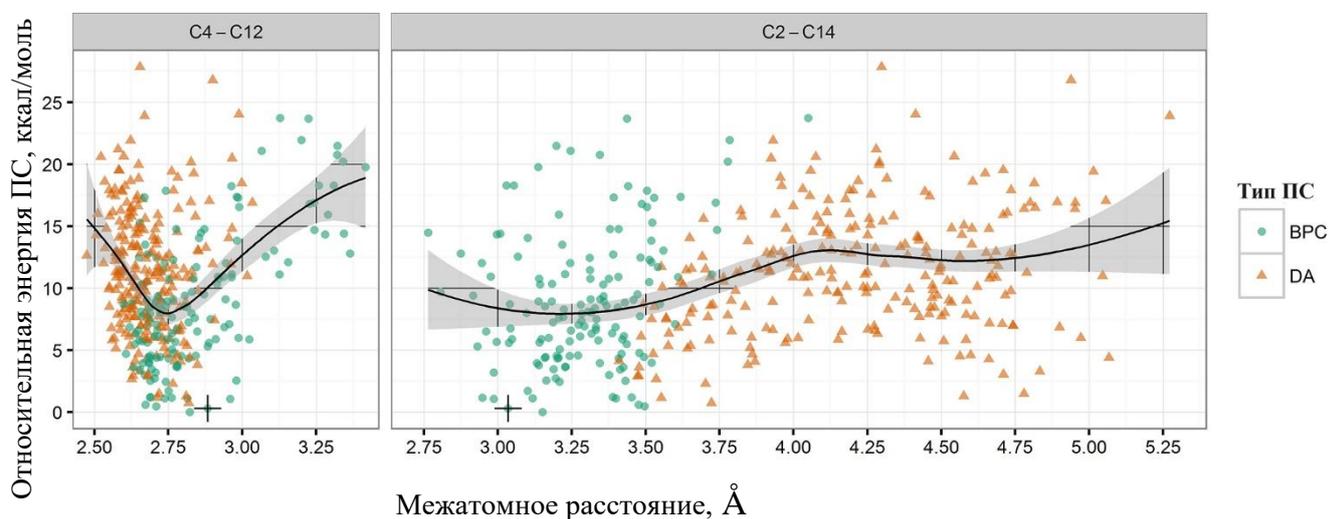
Тот же теоретический алгоритм [495] использовался для оптимизации систем отсчета, первоначальные структурные приближения для них были получены с помощью упомянутой выше методологии.

Анализ распределения зарядов **DA** и три-**DA** показал, что атомы углерода, которые образуют новые связи, действительно, приобретают относительный положительный заряд в ПС, в отличие от механизма **BPC**. В случае механизма **BPC** атомы C2, C4, C12 и C14 приобретают более отрицательный заряд, тогда как атомы C7 и C11 теряют электроны. Это коррелирует с длинами образующихся связей, показанных на Рисунке 59. Примечательно, что атом O15 всегда получает избыточный отрицательный заряд в переходных состояниях.

Модельные реакции показывают, что

- 1) Существенное перераспределение зарядов невыгодно с энергетической точки зрения, и реакция представляет собой не полярное циклоприсоединение, а катионную перегруппировку;
- 2) Бис-перициклические ПС (**BPC**) минимизируют энергию активации реакции Дильса–Альдера в тех случаях, когда они могут быть реализованы.

Рисунок 59. Зависимость энергии ПС от межатомных расстояний C4–C12 и C2–C14. ПС, найденное в работе [482], обозначено крестом. Линия показывает усредненную энергию для данной длины связи, а светло-серая область обозначает 95%-ный доверительный интервал.



Связывание атомов C7–C11 и C4–C12 были обнаружены во всех ПС, за исключением нескольких высокоэнергетических; но неожиданным было то, что наличие связывания C2–C14 ассоциировано со снижением энергии ПС, что, вероятно, объясняется благоприятным вторичным орбитальным взаимодействием [496] в этом случае. Так как образование этой связи является принадлежностью переходного состояния **ВРС**, можно заключить, что переходные состояния **ВРС** в среднем более стабильны, чем **DA**. Чтобы исключить влияние возможных ошибок в расчете, мы ассоциировали связывание C2–C14, либо любого из них с ближайшим соседом, как часть механизма **ВРС**.

Таким образом, в результате фильтрации всех найденных ПС, 144 соответствуют бис-перициклическому механизму, а оставшиеся 240 относятся к классическому механизму Дильса–Альдера. Распределение ПС по энергии относительно низшего конформера **S** и по механизмам показана на Рисунке 57 (240 **DA**, 144 **ВРС**, 9 **BR–1S**, 331 **BR–1T** и 4 **altDA**).

Однако, эта картина упрощает фактическое распределение переходных состояний, связанных с основной реакцией. Из явного рассмотрения расстояний C4–C12 и C2–C14 (Рисунок 59) мы видим, что нет резкой границы между состояниями перехода **ВРС** и **DA**.

Фактически, две практически независимые тенденции влияют на энергию переходного состояния: во-первых, укорачивание связи C4–C12 влияет неблагоприятно, ее оптимальная длина составляет $\sim 2,75 \text{ \AA}$; во-вторых, переходные состояния с частично сформированной связью C2–C14 в среднем более благоприятны. Связь C7–C11 в то же время является консервативной, с длиной $1,94 \pm 0,1 \text{ \AA}$ для 95% TS. В частности, существуют относительно благоприятные TS, соответствующие пути **DA** (с расстоянием связи C2–C14 $> 4 \text{ \AA}$), но связь C4–C12 в них еще не полностью сформирована, будучи длиннее C7–C11 по меньшей мере на $0,7 \text{ \AA}$ и превращая соответствующую реакцию Дильса–Альдера в асинхронную.

Поведение реакции в воде в целом объясняется принципом Кертвина–Гаммета [461], который гласит, что в случае быстрого взаимопревращения потенциально реакционноспособных субстратов путь реакции определяется относительной энергией переходных состояний, а не энергией активации. Применив соответствующие уравнения к нашим данным, мы обнаружили, что $\sim 95\%$ реакции должно протекать через 14 переходных состояний, суммированных в Таблице 23. Таким образом, хотя ПС **DA** более многочисленны, **ВРС** имеют больший вклад из-за их более низких относительных энергий. При 298К $\sim 83\%$ реакции протекают по механизму через ПС **ВРС** и $\sim 17\%$ – через ПС **DA**.

Таблица 23. Идентификаторы, типы, скорректированные к температуре относительные энергии ($E_{\text{отн}}$), длины образующих связей и вклады переходных состояний.

№ ПС	Тип ПС	$E_{\text{отн}}(\text{ПС}),$ ккал/моль	$d_{\text{C4-C12}}, \text{ \AA}$	$d_{\text{C2-C14}}, \text{ \AA}$	Вклад, ^a %
120	ВРС	0.00	2.822	3.151	24.0
7 ^b	ВРС	0.29	2.883	3.034	14.6
69	ВРС	0.46	2.709	3.496	11.0
231	ВРС	0.58	2.672	3.476	9.0
197	DA	0.73	2.819	3.722	7.0
85	ВРС	1.07	2.960	2.986	3.9
3	ВРС	1.11	2.815	3.117	3.7
309	ВРС	1.11	2.739	3.399	3.7

43	DA	1.15	2.713	3.553	3.4
232	BPC	1.17	2.761	2.948	3.3
0	BPC	1.17	2.688	3.452	3.3
35	BPC	1.20	2.691	3.388	3.2
36	DA	1.29	2.811	4.576	2.7
113	DA	1.48	2.719	4.779	2.0

^a рассчитанные в соответствии с принципом Кертина–Гаммета

^b ПС изучены в работе [482].

В результате проведенных исследований мы можем заключить, что только два механизма, а именно, классический Дильса–Альдера и бис-перициклический, могут реализовываться в водном окружении, в то время как энергетические барьеры для альтернативных путей Дильса–Альдера и бирадикалов существенно больше, и мы их исключаем из рассмотрения. Бис-перициклические переходные состояния менее многочисленны, но имеют более низкие энергии и составляют ~83% неферментативной реакции при 298К в воде. Учитывая, что **BPC** являются более компактными и стабильными (ниже по энергии) в среднем, ферментативная реакция вряд ли протекает через классический механизм Дильса–Альдера или синглетные бирадикальные механизмы.

Дальнейшее теоретическое моделирование этой реакции в активном сайте белка потребует учета всех полученных в данном исследовании геометрий субстратов и переходных состояний и позволит рассчитать полный механизм этого природного каталитического процесса с участием SpnF, дильс-альдеказы.

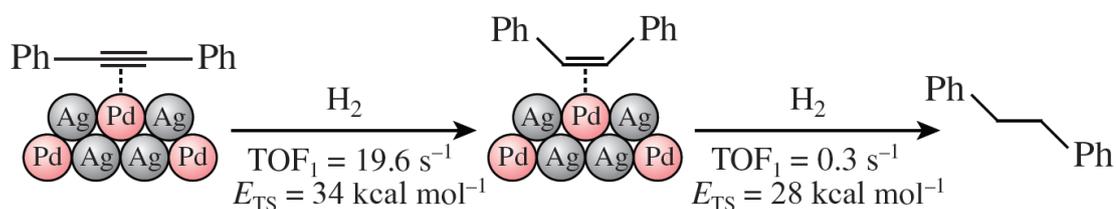
3.8.4. Подходы к моделированию гетерогенного катализа

Логичное продолжение исследований механизмов каталитических взаимодействий должно вылиться в исследованиях координации молекул, вступающих в реакцию, на пригодном для расчетов гетерогенном катализаторе – например, одноатомном (“single-atom”).

Исследовано гидрирование дифенилацетилена (ДФА) на палладиево-серебряных катализаторах с одноатомной структурой. Экспериментально показано, что скорость гидрирования алкена в алкан существенно ниже скорости полугидрирования дифенилацетилена.

Для теоретического исследования механизма протекания реакции гидрирования алкина на поверхности Pd–Ag single-atom катализатора (Схема 20) был использован метод DFT. Структура поверхности катализатора была смоделирована как кубическая упаковка чередующихся атомов Pd и Ag, геометрия которых была оптимизирована в рамках GAMESS и не менялась при дальнейших расчетах.

Схема 20. Гидрирование дифенилацетилена (ДФА) на палладиево–серебряных катализаторах с одноатомной структурой.



Протекание реакции смоделировано как ступенчатое последовательное присоединение четырех атомов (2+2) водорода к молекуле субстрата. Основной задачей теоретического исследования было определение структур переходных состояний (ПС) на всех четырех стадиях. Произведен расчет четырех переходных состояний, из которых был рассчитан энергетический профиль реакции (Рисунок 60).

Рисунок 60. Энергетический профиль ступенчатого последовательного присоединения водорода (2+2) на первой стадии гидрирования молекулы дифенилацетилена на поверхности Pd–Ag катализатора

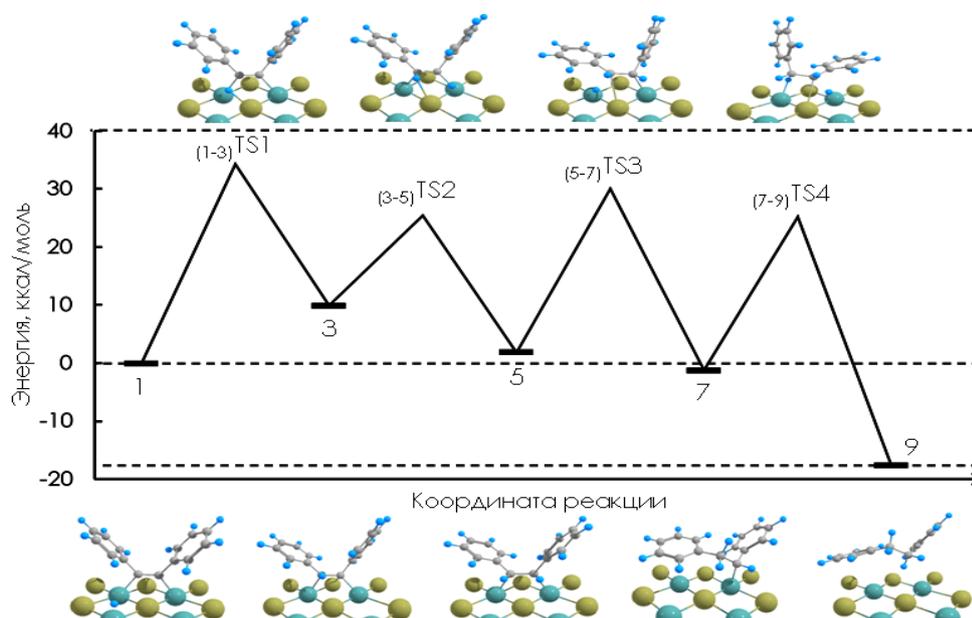


Рисунок 60 показывает вычисленный энергетический профиль для гидрогенизации ДФА на поверхности катализатора Pd–Ag с изолированными атомами Pd, происходящей поэтапным гидрированием активным водородом. ДФА адсорбируется и хорошо активируется на двух соседних сайтах Pd кластеров Pd–Ag–Pd.

Полученные предварительные результаты [497] показывают, что использованный метод позволяет адекватно описать протекание процесса гидрирования на первой стадии (присоединение первых двух атомов водорода, приводящее к образованию соответствующего алкена). На второй стадии гидрирования (трансформация олефин → алкан) наблюдается существенно большее расхождение между результатами теоретического расчета и экспериментальными данными, и это является предметом дальнейших исследований.

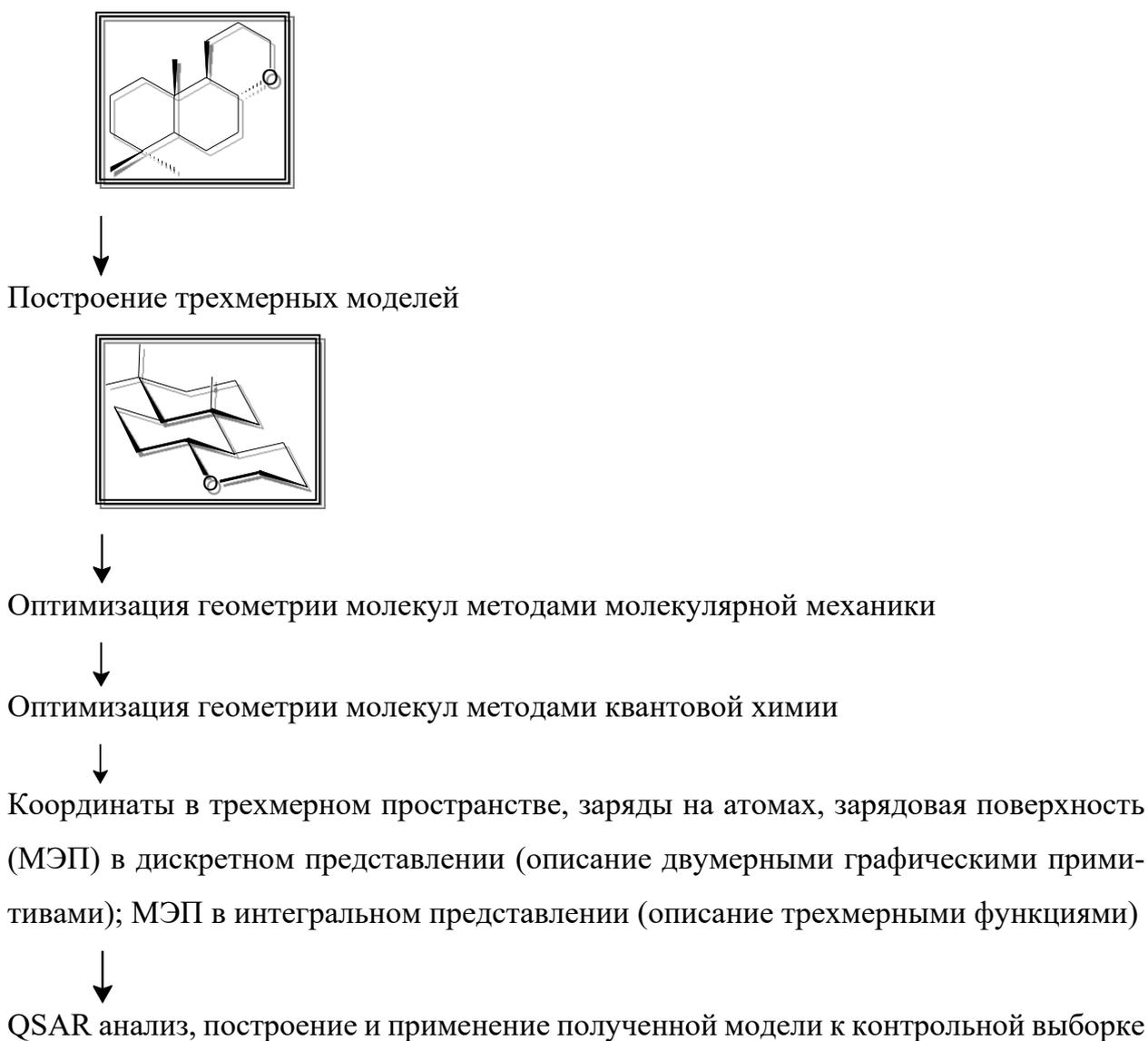
4. Экспериментальная часть. Описания использованных в практическом моделировании типовых методик расчетного эксперимента; схемы вспомогательных синтезов

4.1. Математические методы, примененные в расчетах методами 3D-QSAR

Общая постановка задачи моделирования в рамках современного 3D-QSAR – в работе [400].

4.1.1. Основные этапы построения молекулярных представлений для последующего QSAR-анализа

Схема 21



4.1.2. Определения и постановка задачи расчетов методами QSAR

M-граф (меченый молекулярный граф $G = \{E, V\}$) – это помеченный граф, вершины которого $\{E\}$ интерпретируются как атомы молекулы, а ребра $\{V\}$ – как валентные связи между парами атомов. Метки вершин и ребер (числа или символы) кодируют атомы и связи различной химической природы. В качестве меток вершин могут быть использованы любые характеристики соответствующих атомов (например, символ химического элемента, заряд ядра, поляризуемость, атомный вес, атомный радиус и др.), а в качестве меток ребер – любые характеристики соответствующих связей (кратность, длины, порядки связей, полученные из квантово-химических расчетов, и т.д.) Предполагаем, что различным с точки зрения химии соединениям соответствуют различные *M-графы*.

Множество всевозможных молекулярных графов обозначим $MG = \{G\}$.

В данной работе мы отождествляем химические соединения, молекулы и *M-графы*. В литературе можно найти также задание каждой молекулы одним из способов: молекулярным графом, молекулярной поверхностью или множеством особых точек.

Обучающая выборка $LS = \{(G_i, C_i)\}_{i=1}^N$, $G_i \in MG, C_i \in \{Cl_1, \dots, Cl_H\}$ – совокупность из N химических соединений, где:

- i -е соединение представлено меченым молекулярным графом G_i ;
- i -е соединение отнесено к C_i - одному из H классов активности $\{Cl_1, Cl_2, \dots, Cl_H\}$ (например, «активных», «слабоактивных», «неактивных» веществ) согласно исследуемому свойству.

Дескриптором будем называть какое-либо свойство, численное значение которого может быть вычислено для произвольного молекулярного графа G (произвольного химического соединения).

Алфавитом дескрипторов будем называть множество всех дескрипторов, используемых для анализа обучающей выборки, обозначенных различными символьными метками.

Пусть алфавит дескрипторов состоит из M элементов. *Вектором признаков* молекулярного графа G будем называть вектор $x = (x_1, \dots, x_M) \in \check{Y}^M$, где x_i – значение i -ого дескриптора, вычисленное для G .

МД-матрицей или матрицей «молекула – дескриптор» (матрицей признаков)

для рассматриваемой обучающей выборки $LS = \{(G_i, C_i)\}_{i=1}^N$ будем называть матрицу X размера $N \times M$, в i -ой строке которой стоит вектор признаков $x_i = (x_{i1}, \dots, x_{iM})$ i -ого соединения.

Традиционно задача построения прогнозирующей модели при решении задачи «структура–свойство» разбивается на два достаточно независимых этапа:

Этап описания обучающей выборки. Кратко, этап описания обучающей выборки состоит в выборе представления информации о структуре молекулы, то есть, набора признаков–дескрипторов. Более подробно, в ходе этапа описания обучающей выборки необходимо решить следующие задачи:

- выбрать и зафиксировать алфавит дескрипторов для данной обучающей выборки,
- для каждого молекулярного графа G_i из обучающей выборки $LS = \{(G_i, C_i)\}_{i=1}^N$ вычислить его вектор признаков $x_i = (x_{i1}, \dots, x_{iM})$.

Результатом работы этапа описания, как правило, принято считать построенную по обучающей выборке матрицу «молекула – дескриптор» MD .

Этап поиска функциональной зависимости или этап анализа МД-матрицы. В ходе данного этапа может решаться большое число дополнительных подзадач, таких как: кластерный анализ обучающей выборки, поиск выбросов в обучающей выборке, отбор дескрипторов для прогнозирования и другие. Но ключевым для данного этапа является решение задачи поиска функциональной зависимости f между значениями признаков и значением активности/свойства. *Классифицирующая* функция $f: \check{Y}^M \rightarrow \{Cl_1, Cl_2, \dots, Cl_H\}$ получает в качестве аргумента вектор признаков $x = (x_1, \dots, x_M) \in \check{Y}^M$ произвольного молекулярного графа G и относит соответствующее этому M -графу соединение к одному из классов активности $C \in \{Cl_1, Cl_2, \dots, Cl_H\}$.

Предполагаем, что для классифицирующей функции может быть вычислен некоторый функционал качества, характеризующий её способность предсказывать активность неизвестных соединений. Когда у нас есть только обучающая выборка удобно использовать функционал качества со *скользящим контролем*. В таблицах, приведенных в данной работе, используется именно этот функционал качества.

Теперь общую задачу можно сформулировать так: по обучающей выборке

необходимо построить распознающие модели – то есть набор классифицирующих функций со своими областями определения (имеются в виду ограничения на допустимые соединения) и оценки качества этих функций для согласованного прогнозирования активности новых соединений.

4.1.3. Дескрипторное описание соединений обучающей выборки

Дескрипторы первого уровня

В качестве *дескрипторов* для описания соединений выбран метод выделения линейных фрагментов с введением маркировки вершин молекулярных графов [3].

Регулируемые *параметры описания*:

- 1) длина линейных фрагментов ($k= 2, 3, 4$);
- 2) маркеры, участвующие в описании (d – степень вершины молекулярного графа, b – информация о наличии химических связей, r – положение в кольце).

В соответствии с выбором параметров построено 24 матрицы молекула–дескриптор – 8 вариантов включения маркеров и 3 варианта длины линейных фрагментов. Строкам матрицы соответствуют молекулы выборки, столбцам – дескрипторы. Значение матрицы на пересечении i -ой строки и j -го столбца – количество повторений j -го дескриптора в i -ой молекуле.

Дескрипторы второго уровня

В использованном подходе описания молекул с помощью дескрипторов первого уровня присутствует существенный недостаток – дескрипторы представляют собой линейные фрагменты молекулярных графов, не позволяющие получить какую-либо информацию о пространственной структуре соединения. Для устранения этого недостатка рассматривается тип дескрипторов второго уровня.

Дескриптор второго уровня представляет собой пару дескрипторов предыдущего этапа (дескрипторы первого уровня – линейные фрагменты) с введением расстояния между ними. В качестве меры расстояния между двумя подграфами молекулярного графа выбирается геометрическое или топологическое расстояние между парами вершин подграфов. Таким образом, новый дескриптор выглядит следующим образом:

$\langle Desc_2 \rangle = \langle Desc_1 \rangle + \langle Desc_1 \rangle + \langle расстояние \rangle$, где $Desc_1$ – дескриптор первого уровня, $Desc_2$ – дескриптор второго уровня.

Следует отметить, что, хотя такие дескрипторы и вносят дополнительную информацию о структуре молекул, но описание обучающей выборки значительно усложняется. Так, если количество дескрипторов для одной молекулы M на предыдущем этапе описания $C_{от_1}(M) = N$, то при использовании дескрипторов второго уровня сложность описания $C_{от_2}(M) \approx N^2$.

Построение МД-матриц аналогично описанному выше методу.

Построение функции прогнозирования

На этапе построения функции прогнозирования активности использовались различные подходы к обработке полученных МД-матриц.

Кластеризация и Метод Группового Учета Аргументов (МГУА).

В варианте моделирования МГУА МД-матрицы кластеризовались разными способами, в т.ч. многопараметрическим иерархическим кластерным анализом (параметры: количество кластеров, матрица связности с расстояниями между объектом и кластером). Всего использовалось 53 варианта кластеризации для каждой из сформированных МД-матриц.

Для каждого из построенных кластеров независимо строилась прогнозирующая функция в соответствии с МГУА [6]. В результате для кластеров получены функции прогноза, для каждой из которых известен коэффициент качества прогноза R .

Построение функций прогноза при описании дескрипторами первого уровня

С помощью описания обучающей выборки дескрипторами первого уровня получено 24 варианта описания, в каждом случае выборка разбита на кластеры, и каждому кластеру присвоена некоторая функция прогноза. Показано, что при таком методе построения модели можно добиться разбиения обучающей выборки на несколько крупных кластеров, в которых параметр качества прогнозирующей функции превышает 95 %.

Для обработки соединений с неизвестной активностью описываем новые соединения с помощью дескрипторов, полученных на этапе описания обучающей выборки. Для каждого фиксированного варианта описания определяем принадлежность данного соединения к одному из построенных кластеров. Для этого в каждом кластере методом к-средних выделяется несколько представителей и вычисляется расстояние от нового объекта до представителей. Принадлежность одному из кластеров определяется методом голосования. Далее применяется функция прогноза, соответствующая выбранному кластеру. Применяя к каждому соединению несколько различных моделей, итоговую активность можно определить с помощью анализа полученных результатов.

В результате применения описанного алгоритма построения моделей на обучающей выборке и применения функций прогноза к выборке соединений с неизвестной активностью выделено несколько соединений, на которых получено положительное значение активности, то есть предположительно обладающих активностью и рекомендованных для синтеза.

Нечеткие деревья решений и эволюционный отбор дескрипторов

МД–матрицы, полученные с помощью дескрипторов первого уровня, обрабатывались также с помощью подхода, сочетающего использование для прогноза активности деревьев решений, строящихся локально внутри нечетких кластеров, с эволюционным отбором дескрипторов.

Имея различные описания обучающей выборки в виде 24 МД–матриц, для анализа кластерной структуры применялся алгоритм нечёткой кластеризации (в нашем случае широко известный *c-means fuzzy* – нечеткие к-средних, но в данном подходе возможно использование и любого другого алгоритма нечёткой кластеризации).

Объекты, отнесенные нечеткими методами кластеризации, могут принадлежать разным кластерам с индивидуальными коэффициентами отнесения (сумма коэффициентов =1).

4.2. Программы моделирования (докинг и ВСЭ)

Для подготовки топологий лигандов использовался скрипт ACPYPE; квантово-механическую оптимизацию геометрии лигандов, а также определение зарядов на атомах проводили с помощью пакета программ AmberTools 1.5. Для подготовки трехмерных моделей и эмпирической оптимизации геометрии низкомолекулярных соединений использовался программный пакет ChemDraw Pro. Визуализация 3D-структур осуществлялась программой VMD 1.8.7. Подготовка 3D-структуры лиганда, комплекса лиганд–мишень, включая сольватированные, осуществлялась с помощью Gromacs 4.5.4. Молекулярно-динамические расчёты, визуализация траекторий и расчёты ВСЭ – также при помощи системы Gromacs 4.5.4. Для автоматизированного анализа результатов моделирования структур комплексов использовался специально разработанный пакет программ на языке Perl.

4.2.1. Методики моделирования, использованные в данной работе

[498,499,531]

Методы термодинамического семплирования боковых радикалов аминокислотных остатков белка (TSAR)

В рамках алгоритма термодинамического семплирования боковых радикалов аминокислотных остатков белка (TSAR), белок рассматривается как система независимых подвижных боковых радикалов, закрепленных на жестком полипептидном остове. Боковые радикалы могут иметь состояния, различающиеся протонированием и конформацией.

Сопоставим белку молекулярный граф, построенный следующим образом: каждому узлу графа сопоставляется набор состояний, которые могут принимать боковые функциональные группы. Заданная пара вершин графа соединяется ребром, если

- минимальное расстояние между различными состояниями соответствующих остатков меньше 4 Å
- разница минимальной и максимальной энергий взаимодействия между различными состояниями остатков больше порогового значения 0,1 ккал/моль

(энергетические оценки проводятся при помощи скоринговой функции, описанной ниже).

Поскольку каждому узлу графа сопоставляется набор состояний, которые могут принимать боковые радикалы, каждому ребру графа сопоставляется таблица парных энергий взаимодействия состояний соединяемых ребром вершин.

Полученный граф белка представляет собой конструкцию, известную в математике как Байесовская сеть (Belief Network), в которой вероятность пребывания каждого узла в определенном состоянии зависит от состояний соседних вершин. Аналогия становится полной, если перевести энергии состояний в вероятности в соответствии с известным соотношением.

$$P(A | B) = P(B) \cdot \exp\left(-\frac{E_A + E_{AB}}{RT}\right) / \sum_A \exp\left(-\frac{E_A + E_{AB}}{RT}\right) \quad (1)$$

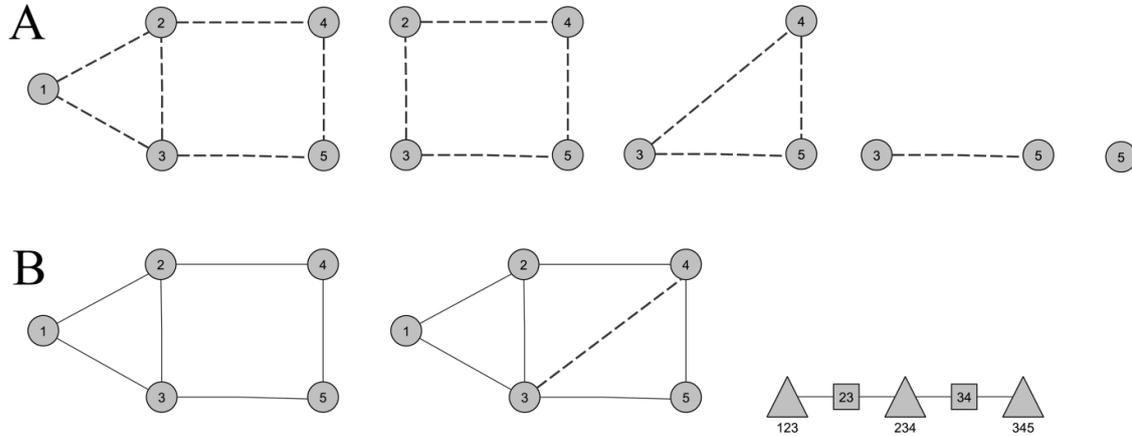
в котором условная вероятность состояния вершины А определяется энергией взаимодействия с соседней вершиной (набором соседних вершин) В, находящейся в одном из своих состояний. В выражении также присутствует внутренняя энергия А, не зависящая от состояний соседей, но влияющая на вероятность. Следует отметить, что в случае боковых радикалов белка внутренняя энергия может определяться не только торсионной составляющей, но и энергией взаимодействия с неподвижным остовом, а также рН-зависимым вкладом в потенциал ионизации.

Общая идея алгоритма TSAR

Нахождение функций распределения состояний индивидуальных вершин согласно теории Байесовских сетей проводится следующим образом:

- Вначале проводится триангуляция исходного графа, из которого далее выделяются клики, и строится дерево обхода клик.
- Затем проводится распространение потенциала по дереву в обе стороны, в ходе которого проводится пересчет функций распределения клик, через которые передается сигнал.
- И наконец, после прохождения сигнала в обе стороны каждой из клик соответствует искомая функция распределения, из которой функция распределения индивидуальных вершин находится простой сверткой.

Рисунок 61. Иллюстрация идеи алгоритма TSAR на модельном графе. Интуитивный маршрут объединения вершин графа и нахождения функций распределения вершин (А). Формализованный подход теории Belief–Network (В)



На Рисунке 61 изображен граф модельной системы. Пусть каждой вершине соответствует набор состояний, нумеруемых переменной с номером соответствующей вершины. Очевидно, что состояние вершины 1 зависит только от состояний вершин 2 и 3. Поэтому для каждой пары состояний вершин 2 и 3 можно обновить эффективный потенциал их взаимодействия следующим образом:

$$E'_{23} = \min_1 (E_1 + E_{12} + E_{23}) \quad (2)$$

Физический смысл обновленного потенциала E_{23} состоит в том, что он фигурирует в функции распределения новой системы, получаемой из исходного графа формальным исключением вершины 1. После исключения вершины 1 граф представляет собой 4-членный цикл. Аналогично можно удалить из полученного графа одну из вершин (например, 2), связанную с двумя другими (3 и 4), при этом возникнет новая эффективная связь между вершинами 3 и 4, которой не было в исходном графе. Физически, новая связь входит в функцию распределения полученного графа (представляющего собой трехчленный цикл) и несет в свернутом виде информацию о распределении остальных вершин (1 и 2). Далее из полученного графа аналогичным образом исключается любая из вершин, например, вершина 3. Результирующий граф состоит из двух вершин, а функция распределения содержит 2 переменные – 4 и 5. Далее, исключая еще одну вершину, например, 4, получаем граф, состоящий из

одной вершины 5. Полученная функция распределения и определяет состояние вершины 5 в исходном графе.

Методика 1. Подготовка трехмерных структур и топологий лигандов

Перед проведением молекулярной динамики необходимо было подготовить полноатомные (содержащие атомы водорода) трёхмерные структуры лигандов, а также топологические файлы *top*, содержащие информацию о типах атомов лиганда, их зарядах, параметрах связей, валентных и двугранных углов лиганда в выбранном силовом поле (OPLSA, (7)).

Сначала с помощью программы ACD ChemSketch 12.0 (4) проводилась предварительная эмпирическая оптимизация трёхмерных структур лигандов. Состояния протонирования ионогенных групп определялись следующим образом: для карбоксильной группы и алифатической аминогруппы всегда задавалась ионизированная форма, для остальных групп проводился предварительный анализ литературных данных по их pK_a . Для заданной таким образом двумерной структуры проводилась трёхмерная оптимизация, и результат экспортировался в *mol*-файл. В отдельных случаях проводилась квантово-химическая оптимизация трехмерной структуры соединений реализованным в программном пакете AmberTools ограниченным методом Хартри–Фока в базисе 6–31G.

Затем с помощью программы АСРУРЕ, использующей программный пакет AmberTools 1.5 для квантово-химических расчётов (для расчёта зарядов применялась схема AM1–ВСС), генерировались файлы оптимизированных трёхмерных структур, а также топологические файлы. Корректность топологических файлов проверялась по тексту, типы атомов, не определённые АСРУРЕ (*opls_x*), маркировались отдельно.

Методика 2. Подготовка трехмерных структур и топологий лигандов

Пример *mdp*-файла, использовавшегося для проведения расчётов методом ВСЭ, приведён ниже.

MDP-файл для NPT расчётной динамики для $\lambda=0.30$.

```
integrator          =          sd
dt                  =          0.002
```

nsteps	1000000
nstcomm	100
nstxout	5000
nstvout	5000
nstfout	0
nstlog	5000
nstenergy	5000
nstxtcout	0
xtc-precision	1000
ns_type	grid
pbc	xyz
rlist	1.0
coulombtype	PME
rcoulomb	1.0
vdw-type	switch
rvdw-switch	0.8
rvdw	0.9
ewald_rtol	1e-06
epsilon_surface	0
optimize_fft	no
tau_t	1.0
ref_t	300
Pcoupl	Parrinello–Rahman
tau_p	0.5
compressibility	4.5e-05
ref_p	1.0
init_lambda	0.30
delta_lambda	0
foreign_lambda	0.25 0.35
sc-alpha	0.5

sc-power	1.0
sc-sigma	0.3
couple-intramol	no
nstdhdl	10

Независящие от шага λ mdp-файлы для минимизации (steep и l-bfgs), изохорической (NVT), изобарической и расчётной (NPT) динамик готовились аналогичным образом. Для минимизации применялись интеграторы steep и l-bfgs, для динамик – интегратор sd. NVT и NPT-динамики проводились с использованием хоростата и баростата Паринелло–Рамана, соответственно. Во всех случаях использовался термостат Берендсена. Mdp-файлы для каждого последующего шага λ генерировались при помощи вспомогательной программы на языке Perl.

Методика 3. Расчет ингибиторов ПАРП1 дискретными методами докинга (Lead Finder).

Докинг лигандов производился в компьютерной системе Lead Finder, версия 1.1.15 [189] с параметрами, выставленными по умолчанию. Полноатомная модель ПАРП1 подготовлена при помощи системы Model Builder, являющейся составной частью дистрибутивного пакета Lead Finder. Исходная структура ПАРП1 взята в Protein Database, PDB ID: 1UK0.

Энергия связывания лигандов были оценены с помощью ΔG -скоринговой функции пакета Lead Finder. Фрагментный скрининг базировался на библиотеке STK, предоставленной VitasM, <http://www.vitasmlab.com/>. Соединения из библиотеки STK были разбиты на фрагменты в соответствии с опубликованной ранее процедурой [384].

Значения IC_{50} были пересчитаны в K_i с использованием уравнения ингибирования $K_i = IC_{50}/(1 + [NAD^+]/K_m)$, где $[NAD^+]$ – концентрация NAD^+ , в экспериментах, а K_m – константа реакции Микаэлиса (по опубликованным данным, $22 K_m = 50 \mu M$).

Методика 4. Расчёт изменения свободной энергии при превращении лигандов Syk-киназы и ПАРП1 в растворе методом ВСЭ

Топология лигандов Syk-киназы и ПАРП1 генерировалась с помощью

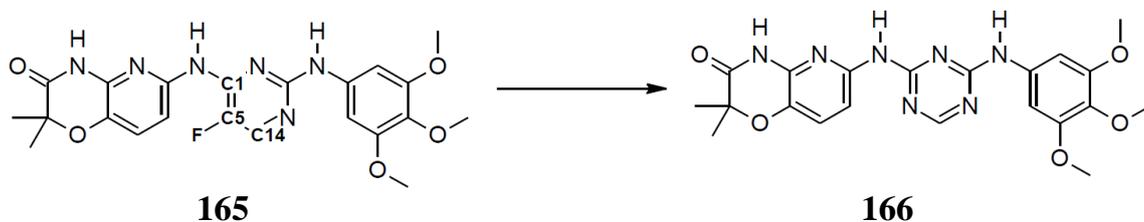
скрипта АСРУРЕ (8). Топологические файлы перевода лигандов друг в друга методом ВСЭ создавались вручную по следующей схеме (перевод лиганда А в лиганд В):

1. отдельно генерировались и проверялись топологические файлы лигандов А и В;
2. в топологический файл лиганда А к соответствующим строкам добавлялись топологические параметры (тип в силовом поле, заряд и масса) изменяемых атомов и атомов, соседних с ними;

Например, для перевода лиганда R406 в МТ-030 в топологическом файле R406 были произведены следующие модификации:

#	type_A	resnum	res name	atom	cgmr	charge_A	mass_A	type_B	charge_B	mass_B
...
4	opls_916	1	LIG	C1	4	-0.534201	12.011	opls_916	-0.761902	12.01100
11	opls_145	1	LIG	C5	11	-0.225700	12.011	opls_313	-0.791000	14.00670
22	opls_145	1	LIG	C14	22	-0.490200	12.011	opls_145	-0.717901	12.01100
25	opls_719	1	LIG	F	25	-0.109900	18.998	MW	-0.000000	18.99840

Данные строки соответствуют замене атома С5 на атом N (изменение OPLS-типа атома и массы), F – на дамми-атом MW, а также изменению заряда на всех атомах, включая ближайших соседей С14 и С1.



Для расчёта энергии перехода лигандов использовался диапазон значений λ от 0 до 1 с шагом 0,025. Расчёт состоял из следующих этапов, идентичных для каждого шага λ :

1. минимизация системы лиганд–вода алгоритмами steep и l-bfgs (максимальная допустимая сила – 10 кДж/моль/нм, количество шагов – 5000, длина шага – 2 фс);
2. NVT–динамика системы лиганд–вода (50000 шагов стохастической динамики, длина шага – 2 фс; общее время траектории – 100 пс);
3. NPT–динамика системы лиганд–вода (50000 шагов стохастической динамики; длина шага – 2 фс, баростат Паринелло–Рамана; общее время

- траектории – 100 пс);
4. финальная динамика системы лиганд–вода (1000000 шагов стохастической динамики; длина шага – 2 фс, баростат Паринелло–Рамана; общее время траектории – 2 нс);
 5. Разница в свободных энергиях соединений определялась программой `g_bar: g_bar -f md*.xvg -o -oi -oh`

Методика 5. Расчёт изменения свободной энергии при переводе лигандов *Syk*-киназы и *PARP1* в белке методом ВСЭ

Топологические файлы перевода лигандов генерировалась аналогично переходам в растворе.

Для расчёта энергии перехода лигандов использовался диапазон значений λ от 0 до 1 с шагом 0,05. Расчёт состоял из следующих этапов, одинаковых для каждого шага λ :

1. Минимизация системы белок–лиганд алгоритмами `steep` и `l-bfgs` (максимальная допустимая сила – 100 кДж/моль/нм, количество шагов – 5000, длина шага – 2 фс);
2. NVT–динамика системы белок-лиганд (50000 шагов стохастической динамики, длина шага – 2 фс; общее время траектории - 100 пс);
3. NPT-динамика системы белок-лиганд (500000 шагов стохастической динамики; длина шага – 2 фс, баростат Паринелло–Рамана; общее время траектории – 1 нс);
4. финальная динамика системы белок–лиганд (2500000 шагов стохастической динамики; длина шага – 2 фс, баростат Паринелло–Рамана, общее время траектории – 5 нс);

Разницу в свободных энергиях соединений, а также ошибку, определённую по блок–среднему из 5 блоков, определяли с помощью команды `g_bar: g_bar -f md*.xvg -o -oi -oh`.

Методика 6. Расчёт энергии конформеров методом ВСЭ [499]

Изменение двугранных углов для осуществления конформационного перехода достигалось с помощью плавного изменения фазы. При этом вид потенциала в

конечной и начальной точке перехода совпадал, однако при промежуточных значениях λ значение фазы плавно менялось. Поскольку константы и совпадают, потенциал при каждом шаге отличается лишь начальной фазой, отличающейся в начальной и конечной точке на 360° . Таким образом, для обеспечения конформационного перехода было необходимо использовать структуру, полученную при предыдущем значении λ , в качестве стартовой для последующего значения λ , так как в противном случае структуры с $\lambda > 0.5$ были бы идентичны структурам с $\lambda < 0.5$. Для этого сначала проводилась последовательная минимизация алгоритмами *steep* и *lbfgs* для каждого шага по λ , а затем NVT и NPT динамика рассчитывалась параллельно.

В некоторых случаях изменение конформации не могло быть достигнуто смещением фазы 4 двугранных углов из-за возникающих внутримолекулярных стерических затруднений (лиганды Syk). В подобных ситуациях одновременно изменялась конформация и другой четверки двугранных углов, что обеспечивало возможность требуемого конформационного перехода. Однако при этом происходило одновременное изменение конформаций двух систем двугранных углов в молекуле. Для того, чтобы изменить конформацию первой части молекулы, сохранив конформацию второй части, было предложено в средней точке перехода ($\lambda = 0.5$) возвращать начальную фазу второго угла к стартовому значению. Для реализации подобного перехода в топологический файл были добавлены следующие строки:

```
#ifndef LAMBDA
#define phi_final -180.00
#define phi_init 180.00
#else
#define phi_init -180.00
#define phi_final 180.00
#endif
```

обеспечивающие смещение фазы вспомогательных двугранных углов в сторону изменения конформации, если переменная LAMBDA инициализирована, и в обратную сторону – если LAMBDA не инициализирована. Параметр LAMBDA, в свою очередь, был инициализирован в каждом mdp-файле с $\lambda > 0.5$:

```
define =-DLAMBDA
```

Для расчёта энергии конформационного перехода использовался следующий набор значений λ : от 0 до 0,45 и от 0,6 до 1,0 – с шагом 0,0125, от 0,45 до 0,6 – с шагом 0,05. Выбор различного шага обусловлен наибольшим изменением значения энергии по краям интервала (0,1) и относительно слабым изменением энергии ближе к 0,5.

Таким образом, расчёт состоял из следующих этапов:

1. последовательная по шагу λ минимизация систем лиганд–вода алгоритмами `steep` и `l-bfgs` (максимальная допустимая сила – 10 кДж/моль/нм, количество шагов – 5000, длина шага – 2 фс);
2. NVT–динамика системы лиганд–вода (10000 шагов стохастической динамики, длина шага – 2 фс; общее время траектории – 20 пс);
3. NPT–динамика системы лиганд–вода (10000 шагов стохастической динамики; длина шага – 2 фс, баростат Паринелло–Рамана; общее время траектории – 20 пс);
4. финальная динамика системы лиганд–вода (250000 шагов стохастической динамики; длина шага – 2 фс, баростат Паринелло–Рамана; общее время траектории – 500 пс);
5. Разницу в энергии конформаций, а также ошибку, определённую по блок-среднему из 5 блоков, определяли с помощью команды `g_bar: g_bar -f md*.xvg -o -oi -oh`

Методика 7. Определение корректности расчётов методом ВСЭ [499]

Корректность всех проведённых расчётов относительной свободной энергии оценивалась по ряду параметров, включающих в себя отклонение для блок-среднего, степень перекрытия фазовых пространств, сходимости свободной энергии от времени, и RMSD атомов белка.

Для определения ошибки расчётов, связанных с недостаточным семплированием системы, рассчитывалась ошибка для блок-среднего по 5 блокам при помощи программы `g_bar: g_bar -f md*.xvg -o -oi -oh`.

RMSD атомов белка рассчитывалось для последовательных кадров траектории с интервалом 500 пс, а также между последним кадром предыдущего значения λ и первым кадром последующего значения λ .

Для расчёта RMSD использовался скрипт на языке Perl, выполняющий следующие шаги:

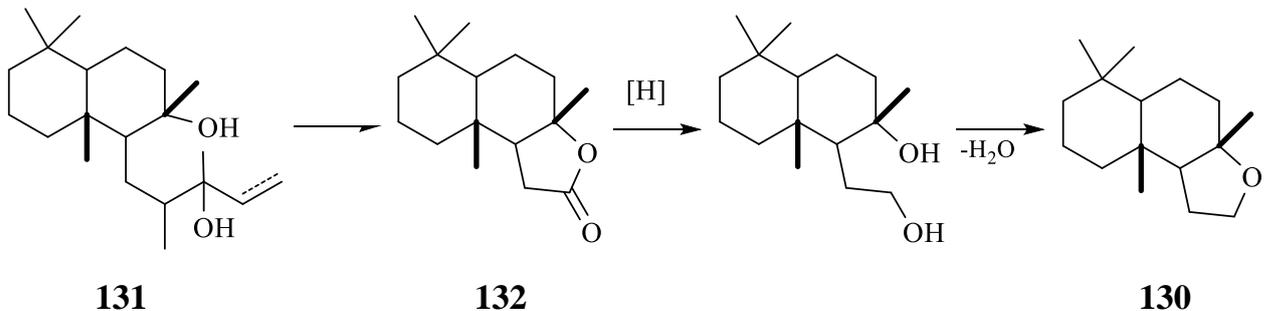
1. вырезание кадров траектории каждые 500 пс с одновременным центрированием по молекуле лиганда (с помощью программы *trjconv*);
2. определение RMSD атомов белка при наложении кадров по полипептидному остову (*g_rms*).

Для определения сходимости свободной энергии одного перехода (изменение λ от 0 до 1, что соответствует модификации кулоновских или Ван-дер-Ваальсовых параметров) использовалась программа *g_bar*. Расчёт проводился по последовательно увеличивающимся (на 5 пс) участкам траектории (5 пс, 10 пс, ...). Полученные данные объединялись в один текстовый файл с помощью скрипта на языке Perl.

4.3. Синтетические методики и схемы.

4.3.1. Синтез веществ с амбровым запахом электрохимическим окислением склареола.

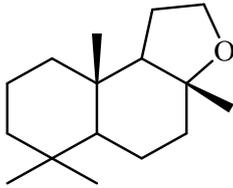
Схема 22



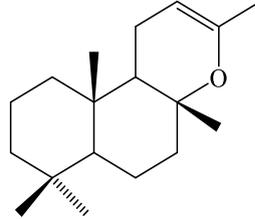
Окисление склареола ведут электролитическим методом в бездиафрагменном электролизере в среде метанола или этанола в присутствии электролита – гидроксида щелочного металла при температуре 60–80°C и пропускании 25–30 F электричества на 1 моль склареола (плотность тока 0,1 А/см²). В качестве анода для электролиза склареола используют платину, графит или другие высокостойкие анодные материалы, в качестве катода – металлы с низким перенапряжением водорода или графита. При токе 1 А и площади поверхности электродов 10 см² время реакции 2,5–15 ч на 1/60 моль склареола. Продукт реакции получают отгонкой растворителя,

последующей промывкой водой и экстракцией углеводородным растворителем (или диэтиловым эфиром, или хлороформом).

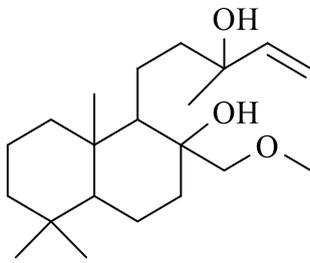
Готовый продукт представляет собой смесь веществ с амбровым запахом (**129**, **130**, **135**) и лишенных запаха (**133**, **134**, склареол), не мешающих парфюмерному использованию смеси и не влияющих на парфюмерную оценку.



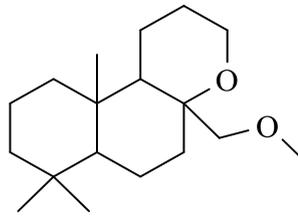
130 Амброксид (10–15%)



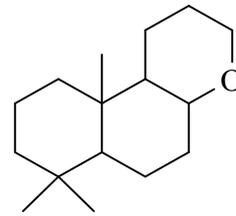
129 Оксид склареола (20–30%)



133



134

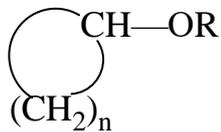


135

Запах полученной смеси был охарактеризован как амбровый, чистый, парфюмерным советом Института душистых веществ Союзпарфюмерпрома.

4.3.2. Синтез циклоалкиловых эфиров и амбролигнана электрокаталитическим присоединением спиртов к двойной связи

Схема 23. Способ получения циклоалкиловых эфиров общей формулы:



$n = 5$, $R\text{—CH}_3$;

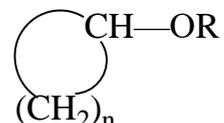
$n = 7$, $R\text{—CH}_3$;

$n = 11$, $R\text{—CH}_3$, C_2H_5 , $\text{CH}_2\text{CH}_2\text{OCH}_3$

путем электролиза циклического олефина в бездиафрагменном электролизере на аноде и катоде в среде соответствующего спирта, в присутствии солей аммония.

В качестве материала катода используют латунь, в качестве соли аммония – тетраэтиламмоний бромистый и процесс проводят при 10–40°C, пропуская 10–14 Ф электричества на 1 моль исходного циклоолефина.

Таблица 24. Получение циклоалкиловых эфиров общей формулы

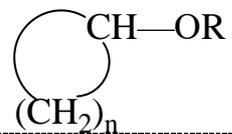


Опыт	n	R	Температура, °C	Растворитель	Количество электричества Ф/моль	Выход эфира, %
1	5	CH ₃	10	CH ₃ OH	10	88
2	7	CH ₃	20	CH ₃ OH	12	86
3	11	CH ₃	30	CH ₃ OH	12	90

Опыт	n	R	Температура, °C	Растворитель	Количество электричества Ф/моль	Выход эфира, %
4	11	C ₂ H ₅	30	C ₂ H ₅ OH	10	85
5	11	CH ₂ CH ₂ OCH ₃	30	CH ₃ OCH ₂ CH ₂ OH	12	86
6	11	CH ₃	40	CH ₃ OH	12	85
7	11	CH ₃	10	CH ₃ OH		92
8	11	CH ₃	10	CH ₃ OH		85

Примечания: 1. Колебания температуры в ходе электролиза ±2 °C.
 2. Катод – везде латунь.
 3. Анод – везде С.
 4. Электролит – везде NH₄Br.

Таблица 25. Получение циклоалкиловых эфиров общей формулы



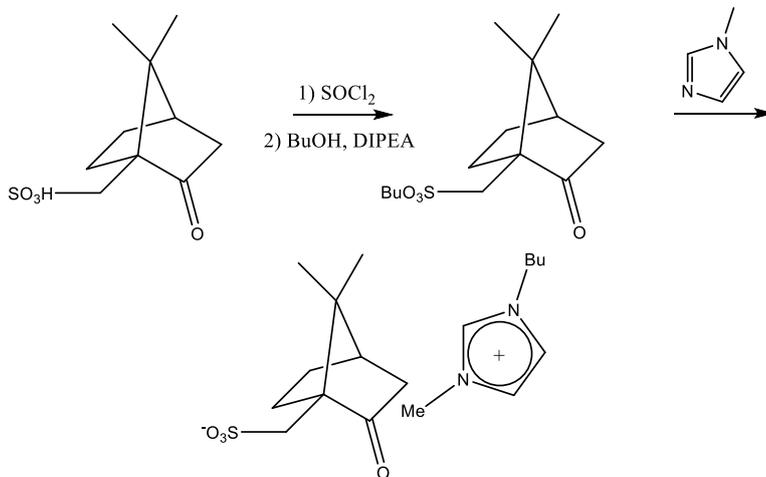
Опыт	n	R	Катод	Температура, °C	Количество электричества Ф/моль	Выход эфира, %
9	7	CH ₃	Латунь	10	9	72
10	11	CH ₃	Латунь	0	10	70
11	7	CH ₃	Pb	30	12	—
12	11	CH ₃	Hg	30	11	15
13	11	CH ₃	Fe	30	13	10
14	11	CH ₃	Zn	30	12	20
15	11	CH ₃	С	30	13	14
16	11	CH ₃	Латунь	50	14	65

- Примечание:
1. Колебания температуры в ходе электролиза ± 2 °С.
 2. Катод – везде латунь.
 3. Анод – везде С.
 4. Электролит – везде NH_4Br .

4.3.3. Моделирование асимметрического синтеза методом ВСЭ, реакция Дильса–Альдера.

Вычислительные методы. Оптимизация геометрии и расчет свободной энергии Гиббса (в вакууме) были проведены с использованием Gaussian09 [438] и методом B3LYP функционала DFT (с 6–31G(d) базой). Совместные вычисления молекулярной механикой и возмущения свободной энергии (MM–FEP) были проведены в пакете Gromacs[500] в силовом поле OPLS–AA.[501] Сольватированная молекула была помещена в куб молекул растворителя со стороной 70Å, и применена утилита Gromacs «genbox». Далее, ВСЭ применен для разграничения влияния растворителя на эту кубическую структуру. Расчеты ВСЭ производились отдельно для Ван-дер-Ваальсовых взаимодействий и кулоновских.

Синтез *Vtm*–CSA.



4.3.4. Вспомогательные синтезы промежуточных соединений и соединений сравнения в синтезе ингибиторов *SyK*-киназы

Схема 24. Синтез 2,4-дихлор-5-фторпиридина

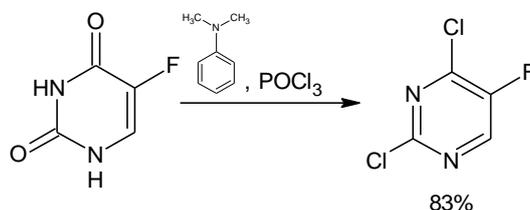


Схема 25. Общая схема синтеза соединения сравнения **165**

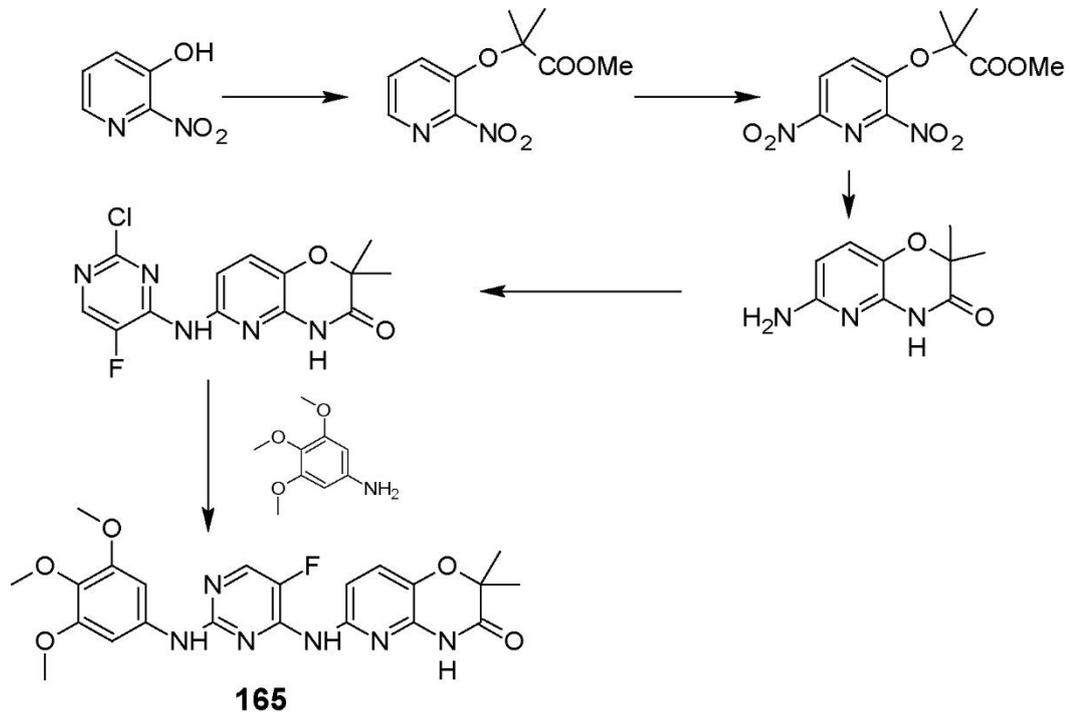


Схема 26. Синтез 6-[(2-хлор-5-фторпиримидин-4-ил)амино]-2,2-диметил-2Н-пирид[3,2-*b*][1,4]оксазин-3(4Н)-она

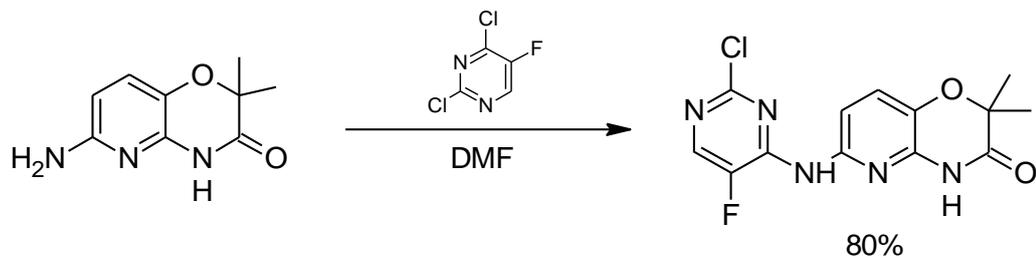


Схема 27. Синтез 6-({5-фтор-2-[(3,4,5-триметоксифенил)амино]пиримидин-4-ил}амино)-2,2-диметил-2Н-пирид[3,2-*b*][1,4]оксазин-3(4Н)-она (**165**)

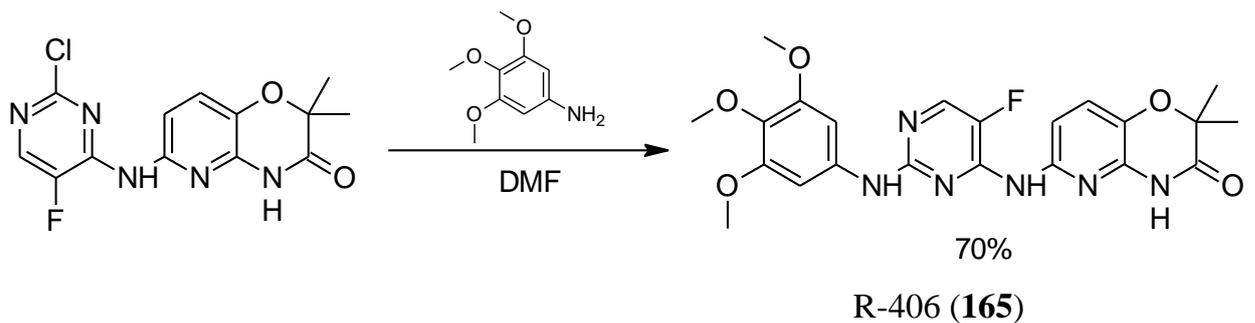


Схема 28. Синтез 6-[(2-хлорпиридин-4-ил)амино]-2,2-диметил-2Н-пирид[3,2-
b][1,4]оксазин-3(4Н)-она

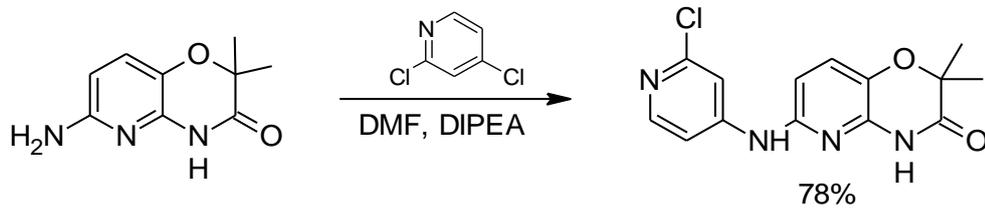


Схема 29. Синтез 2,2-диметил-6-({2-[(3,4,5-триметоксифенил)амино]пиридин-4-
ил}амино)-2Н-пирид[3,2-*b*][1,4]оксазин-3(4Н)-она

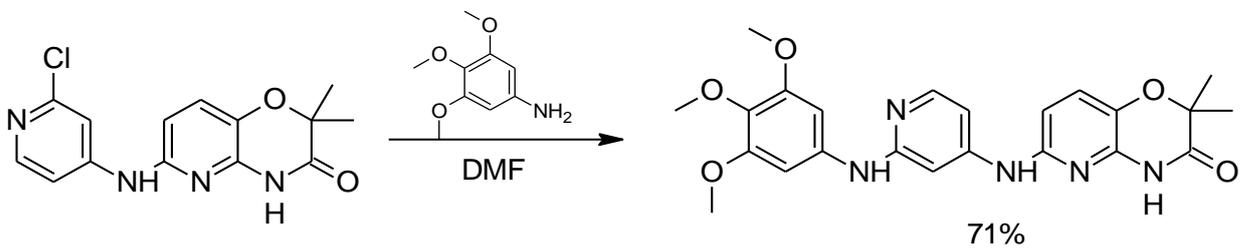


Схема 30. Общая схема синтеза соединения МТ-Syk-02

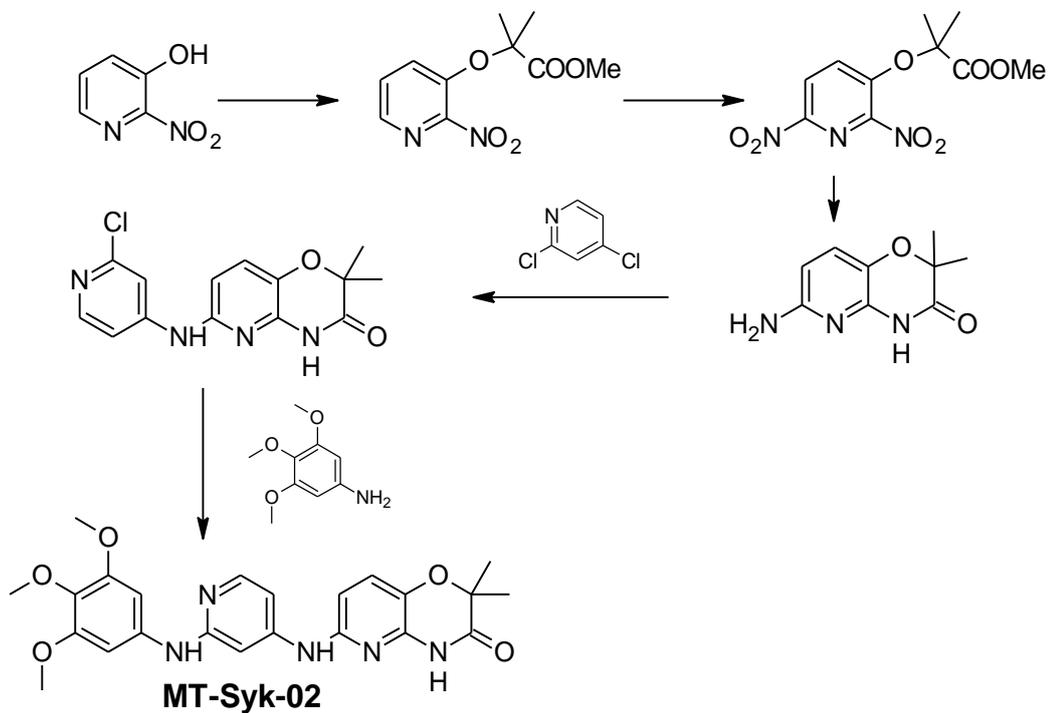


Схема 31. Синтез N-цианохлорформамидина

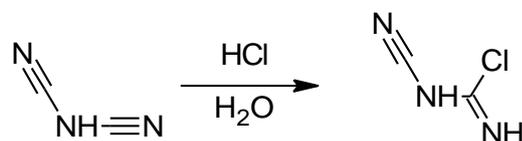


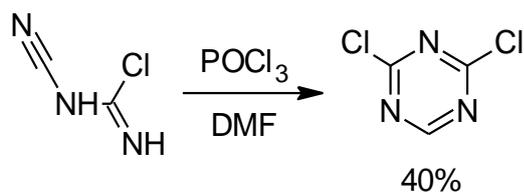
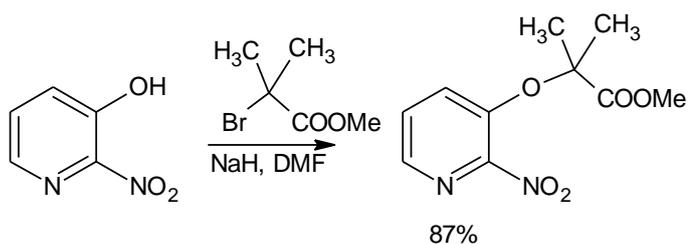
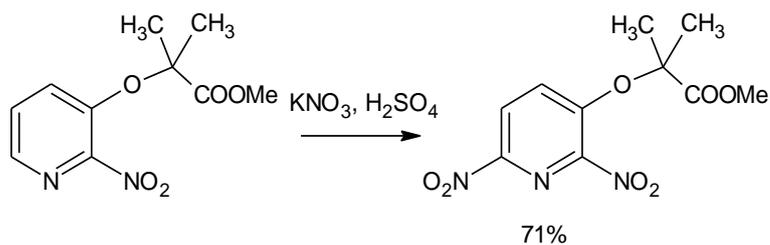
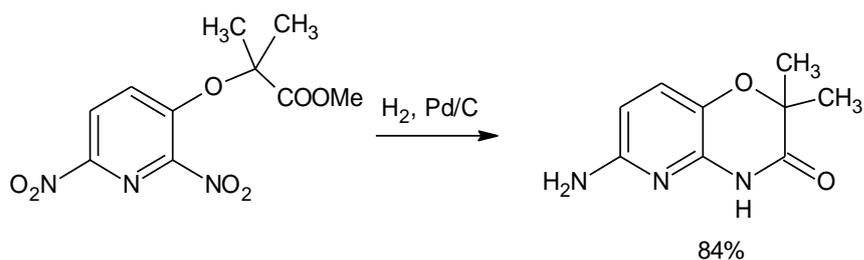
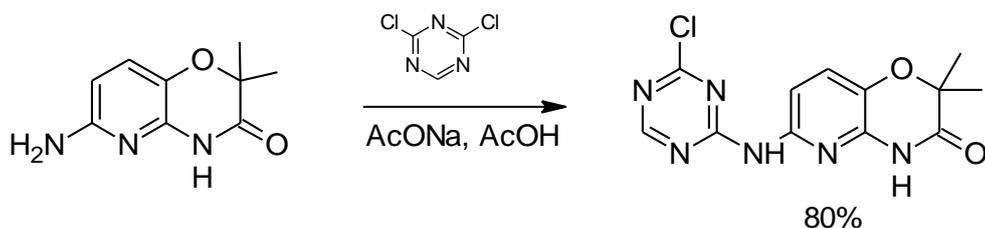
Схема 32. Синтез дихлор-1,3,5-триазина**Схема 33.** Синтез метилового эфира 2-метил-2-[(2-нитропиридин-3-ил)окси]пропановой кислоты**Схема 34.** Синтез метилового эфира 2-метил-2-[(2,6-динитропиридин-3-ил)окси]пропановой кислоты**Схема 35.** Синтез 6-амино-2,2-диметил-2Н-пирид[3,2-*b*][1,4]оксазин-3(4Н)-она**Схема 36.** Синтез 6-[(4-хлор-1,3,5-триазин-2-ил)амино]-2,2-диметил-2Н-пирид[3,2-*b*][1,4]оксазин-3(4Н)-она

Схема 37. Синтез 2,2-диметил-6-({4-[(3,4,5-триметоксифенил)амино]-1,3,5-триазин-2-ил}амино)-2H-пирид[3,2-*b*][1,4]оксазин-3(4H)-она, МТ-Syk-03 (**166**)

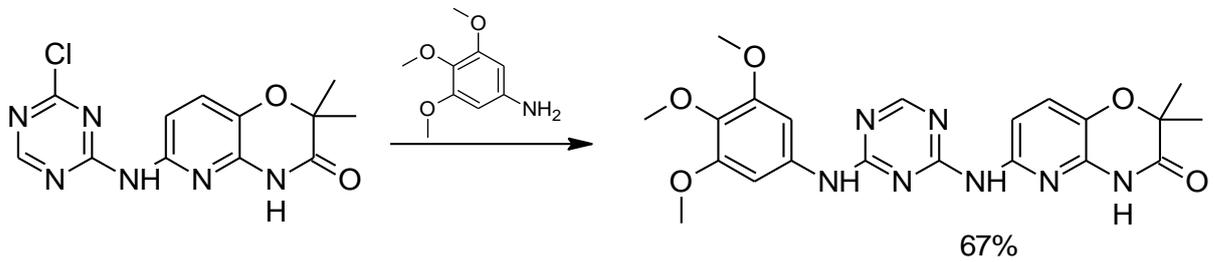


Схема 38. Синтез 2-амино-7H-пирроло[2,3-*d*]пиримидин-4-ола

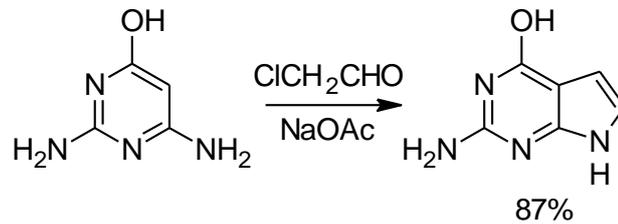


Схема 39. Синтез 2-амино-4-хлор-7H-пирроло[2,3-*d*]пиримидина

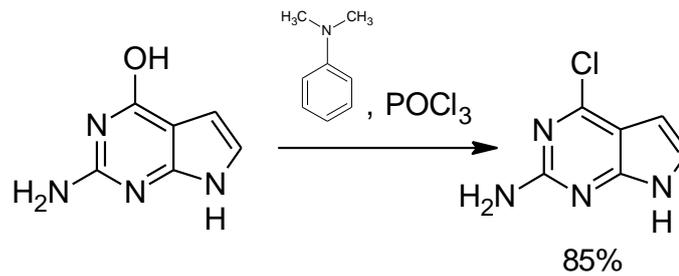


Схема 40. Синтез 2-амино-7H-пирроло[2,3-*d*]пиримидина

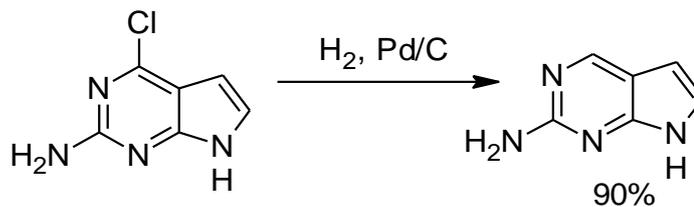


Схема 41. Синтез 2-гидрокси-7H-пирроло[2,3-*d*]пиримидина

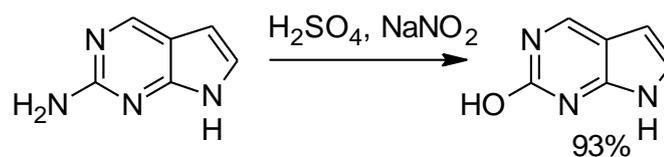


Схема 42. Синтез 2-хлор-7Н-пирроло[2,3-*d*]пиримидина

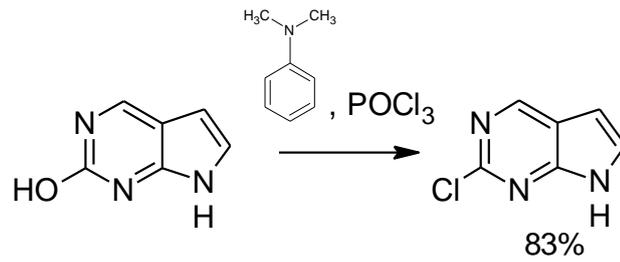
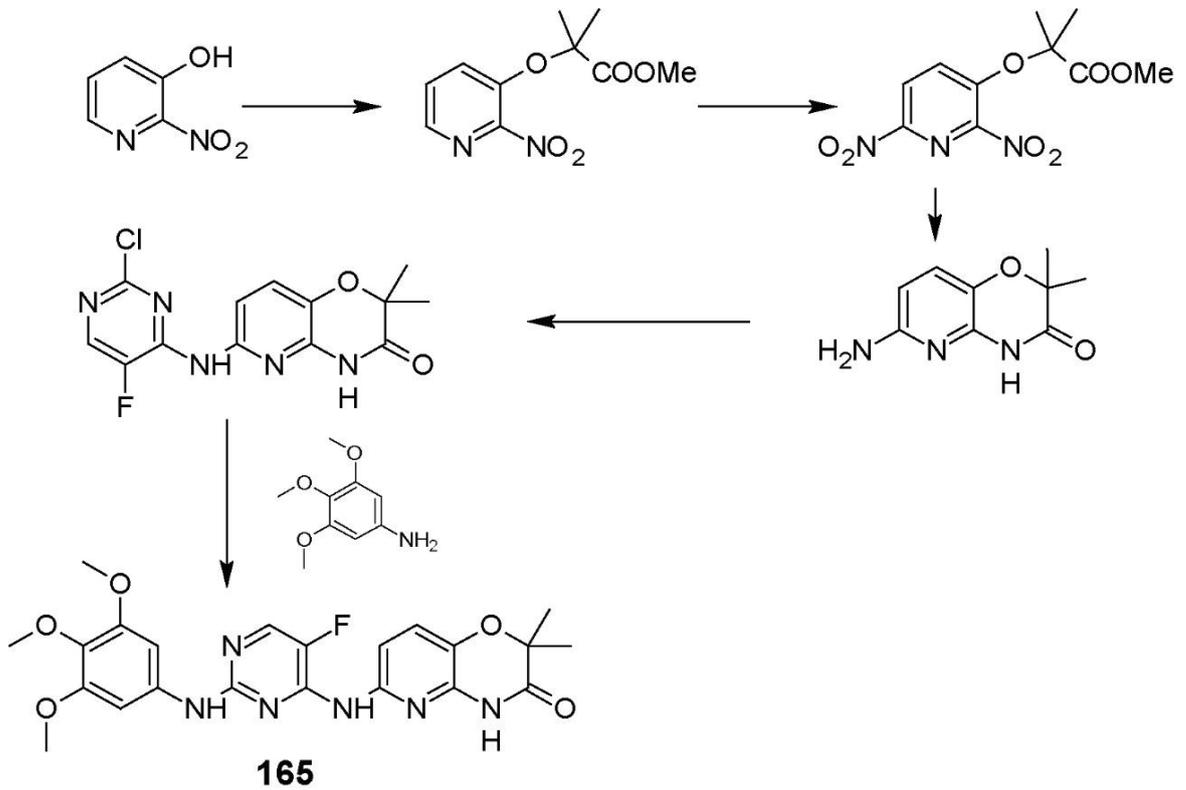


Схема 43. Общая схема синтеза соединения МТ-Syk-06



5. Заключение

Проделанная работа хронологически охватывает и базируется на процессе совершенствования алгоритмов и вычислительных методов в моделировании целевых структур и прогнозировании продуктов направленного синтеза веществ в медицинской химии, гомогенном (включая асимметрический) катализе; и в перспективе затрагивает важную тему моделирования гетерогенных каталитических реакций.

Процесс моделирования активных структур поделен на три базовых методики:

- (1) Статическое моделирование координации лиганд–мишень (единственная либо скорость-определяющая стадия взаимодействия) в отсутствие информации как о структуре мишени, так и о структуре образующегося комплекса.
- (2) Аналогичный расчет для случая, когда такая (экспериментально полученная) информация о мишени известна (докинг). Модификация докинга – динамическая конформационная фильтрация плюс учет действия растворителя.
- (3) Динамическое квантово-химическое моделирование хода реакции с расчетом энергетики переходного комплекса либо переходного состояния.

Результат работы структурирован по этим трем частям процесса моделирования следующим образом:

- 1а) На обучающих выборках соединений с известной активностью построена математическая модель и определена связь структура–запах в амбровом и мускусном запахе. Синтезированы практически важные душистые вещества.
- 1б) На обучающей выборке соединений с известной активностью построена модель, и на ее основе синтезированы новые вещества в ряду бицикломочевин, обладающие нейрологической активностью.
- 1в) Усовершенствованный математический метод обработки обучающей выборки с применением нечеткой логики и элементов машинного обучения позволил построить модель биологически активных гликольбурилов, валидированную последующим синтезом. Показано, что этот метод дает результаты, отличающиеся и дополняющие метод докинга (2); и есть основания считать, что методы классического QSAR и в настоящее время достаточно эффективны в прогнозировании

биологической активности.

2а) Смоделировано статическое взаимодействие лиганд–терапевтическая мишень для ингибиторов ряда киназ. Проведен синтез и биологические испытания полученных соединений и показана 100%-ная валидация методов.

2б) На основе теории возмущения свободной энергии разработан усовершенствованный метод моделирования, проводящий конформационную фильтрацию активных реагирующих субстратов и учитывающий влияние растворителя на комплекс лиганд–мишень. Усовершенствован метод п.2а, получены новые ингибиторы ПАРП1 и Syc-киназы, показавшие активность выше, чем у промышленных образцов.

2в) С использованием метода п. 2б) на модельных объектах асимметрического синтеза с участием асимметрического индуктора осуществлено прогнозирование (подтвержденное синтезом) конфигурации продуктов реакции.

2г) Проведено моделирование переходного состояния реакции Сузуки–Мияуры; расчетами показано, что молекулы воды на последней стадии реакции координируются с Pd с образованием связи Pd–H, а не Pd–O, как принято в каноническом описании механизма.

Новые методы и возможности вычислительной техники позволили перейти к квантово-механическим методам расчета динамики реакций.

3а) На примере модельной реакции Дильса–Альдера, происходящей в живой природе, проведено моделирование переходного состояния и показано, что базовый механизм включает цикл [6+4], что существенно меняет сложившиеся представления о ее катализе природным ферментом SpnF.

3б) Затронут вопрос и предварительно опробованы методы моделирования «single-atom» гетерогенного катализа. Получены результаты, обозначившие перспективы дальнейших исследований.

Суммарные выводы:

1. Построена универсальная система вычислительных методов, обеспечивающих расчет энергии комплекса либо переходного состояния реакции, и позволяющая на этих данных построить модельную структуру для синтеза либо прогнозировать такую модельную структуру как результат направленного синтеза.

2. Показано, что предварительное моделирование веществ с заданной активностью, а также моделирование процессов, приводящих к таким веществам, способно существенно сократить количество «слепых» экспериментов, активность продуктов в которых заранее неизвестна; за счет направленного синтеза смоделированных структур, с высокой вероятностью обладающих такой активностью. Параллельно сокращается ресурсоемкость синтеза и анализа (особенно анализа биологической активности) продуктов.
3. Направленным синтезом получены вещества с заданной активностью, проявившие в валидирующем эксперименте заданные свойства. Этим показана эффективность разработанных расчетных методов в связке структура–активность продуктов органического синтеза.

6. Сокращения и условные обозначения

PDB: Protein Data Bank, база данных по трёхмерным структурам белков.
<http://www.rcsb.org>

PubMed: База данных медицинских и биологических публикаций.
<http://www.ncbi.nlm.nih.gov/pubmed/>

RMSD: Root Mean Square Deviation, масс-взвешенное среднеквадратичное отклонение. Мера правильности наложения двух структур.

TSAR – Thermodynamic Sampling of Amino acid Residues, Термодинамическое Сэмплирование Аминокислотных Фрагментов

ΔG: Свободная энергия Гиббса

IC₅₀: Концентрация ингибитора, подавляющая 50% ферментативной активности

Mol: Распространенный файловый формат хранения информации о двухмерной или трёхмерной структуре молекулы

Sdf: Structure Data Format, распространенный файловый формат хранения информации о двухмерной или трехмерной структуре набора молекул.
<http://www.epa.gov/ncct/dsstox/MoreonSDF.html>

Df: Файловый формат программы Lead Finder для описания структурных фильтров

Tsf: Формат текстового файла со знаком табуляции в качестве разделителя

Pdb: Файловый формат хранения информации о трёхмерной структуре молекулы в формате базы PDB. <http://www.wwpdb.org/docs.html>

Bin: Формат бинарного файла для хранения сеточного представления энергий, используемых программой для докинга Lead Finder.

Log: Формат выходного текстового файла программы для докинга Lead Finder.

BCЭ (FEP): возмущение свободной энергии (free energy perturbation)

КФ: – конформационная фокусировка

ПАРП, ПАРП1 (PARP1): поли(АДФ-рибоза)-полимераза (poly-(ADP-ribose)-polymerase)

Syk: – селезёночная тирозинкиназа (spleen tyrosine kinase)

ITAM – Immunoreceptor Tyrosine–Based Activation Motif, фосфорилированный тирозиновый активирующий агент

МД (MD): – молекулярная динамика

GMDH –Group Method of Data Handling (Групповой метод расчета обработки данных).

ИДО – индол-2,3-диоксигеназа.

ДМФА – диметилформамид.

ВСЭ – возмущение свободной энергии

КФ – конформационная фокусировка

PARP – поли(АДФ-рибоза)-полимераза (poly-(ADP-ribose)-polymerase)

Syk – селезеночная тирозинкиназа (spleen tyrosine kinase)

МД – молекулярная динамика

7. Список терминов

Молекулярная поверхность - Определяется как поверхность объема, являющегося объединение шаров с центром на ядрах с радиусом равным Ван-дер-Ваальсову радиусу атома (табулировано со свободных ресурсов) плюс Ван-дер-Ваальсов радиус воды (определено приблизительно).

Сглаживание поверхности (QSAR): процедура, устраняющая разрывы в первой производной.

Ключевая точка (QSAR): определяется как точка триангуляция, имеющая гравитационный потенциал (сумма слагаемых вида $pot(v) = \sum (i) ((v-r(i))/R(i)) ^ - 2$, где $r(i)$ – координаты i -того атома, $R(i)$ – его же Ван-дер-Ваальсов радиус.) больше или меньше, чем все, соединенные с ней ребрами. Процедура несовершенна и подлежит улучшению.

Дискретизация значения (QSAR): Для каждого значения $v(i)$ найти функцию $W(i) = \sum (j) \exp (- k (\text{mod} (v(j) - v(i)))$). Границы классов проводятся по значениям, которые имеют вес меньше соседей. Метка точек класса - максимальное значение веса в классе. Константа подгоняется вручную так, чтобы было два-три-четыре больших класса и не было маленьких.

Двоичное дерево решений (QSAR): - представление классификатора. В каждом узле классификатора содержится либо решение (да-нет) либо условие (условие на один признак), которым управляется выбор правого или левого поддерева.

Нативные лиганды. Соединения, обладающие сродством к конкретному белку, согласно полученным или имеющимся в литературе экспериментальным данным.

Скоринговая функция: в общем случае – набор и тип параметров, доступных для выбранной модели.

Библиотека (база) химических соединений: набор химических соединений, объединенных по тому или иному признаку (например, по доступности от определенного производителя).

Молекулярный докинг: численный способ нахождения оптимальной структуры комплекса двух молекул, основанный на методах теоретической химии и строения молекул.

Виртуальный скрининг: проведение молекулярного докинга каждого соединения из библиотеки к определенной терапевтической мишени (белку) с целью расчета структуры и прочности (энергии) комплекса белок-соединение.

Терапевтическая мишень: белок, участвующий в том или ином патофизиологическом процессе, и, следовательно, представляющий интерес для направленного лекарственного воздействия.

Лиганд: химическое соединение, участник образования межмолекулярного комплекса.

UniProt: база данных, содержащая различную информацию (название, функции, организм, кодирующая последовательность и т.д.) о белках.
<http://www.uniprot.org/>

8. Список литературы

1. Наточин Ю. В. Физиология, биомедицина, медицина // Успехи физиологических наук. – 2008. – Т. 39, № 2. – С. 8–31.
2. Общая социология: Учебное пособие. Под ред. Эфендиев А. Г. – М.: ИНФРА-М, 2008. . – С. 47.
3. Tang Y., Zhu W., Chen K., Jiang H. New technologies in computer-aided drug design: Toward target identification and new chemical entity discovery // Drug discovery today: Technologies. – 2006. – Т. 3, № 3. – С. 307–13.
4. Скворцова М. И., Станкевич И. В., Палюлин В. А., Зефирова Н. С. Концепция молекулярного подобия и ее использование для прогнозирования свойств химических соединений // Успехи химии. – 2006. – Т. 75, № 11. – С. 1074–1093.
5. Seifert M. H. J., Wolf K., Vitt D. Virtual high-throughput in silico screening // Biosilico. – 2003. – Т. 1, № 4. – С. 143–149.
6. Clark D. E. Computational prediction of ADMET properties: Recent developments and future challenges // Annual reports in computational chemistry / Spellmeyer D. – USA: Elsevier Science, 2005. – С. 133–151.
7. Tropsha A. Variable selection QSAR modeling, model validation, and virtual screening // Annual reports in computational chemistry / Spellmeyer D. C. – USA: Elsevier Science, 2006. – С. 113–126.
8. Раевский О. А. Дескрипторы молекулярной структуры в компьютерном дизайне биологически активных веществ // Успехи химии. – 1999. – Т. 68. – С. 555–576.
9. Mager D. E. Quantitative structure-pharmacokinetic/pharmacodynamic relationships // Advanced Drug Delivery Reviews. – 2006. – Т. 58, № 12–13. – С. 1326–1356.
10. Molecular connectivity in structure-activity analysis. Chemometrics research studies press. / Kier L. B., Hall L. H. – Letchworth, Herefordshire, England: John Wiley and Sons, 1986. – Т. 2: Chemometrics Research Studies Press., – С. 1–101.
11. Kier L. B., Hall L. H. The electrotopological state: structure modeling for QSAR and database analysis // Topological indices and related descriptors in QSAR and QSPAR / Devillers J., Balaban A. T. – Amsterdam, The Netherlands: Gordon and Breach Science Publishers, 1999. – С. 307–336.

12. Hall L. H., Hall L. M. QSAR modeling based on structure-information for properties of interest in human health // SAR and QSAR in Environmental Research – 2005. – Т. 16, № 1–2. – С. 13–41.
13. Modeling the drug-receptor interaction in quantum pharmacology. Molecules in physics, chemistry, and biology. topics in molecular organization and engineering. / Gómez-Jeria J. S.; Под ред. Maruani J. – Dordrecht, The Netherlands: Springer, 1989. С. 215–231.
14. Drug-receptor interaction: Interaction of one or more drugs with one receptor system. Molecular Pharmacology. / Ariens E. J., Simonis A. M., van Rossum J. M.; Под ред. Ariens E. J.: Academic Press 1964. – С. 119–286.
15. Hawkins D. M. The problem of overfitting // Journal of Chemical Information and Computer Sciences –2004. – Т. 44, № 1. – С. 1–12.
16. Hansch C., Maloney P. P., Fujita T., Muir R. M. Correlation of biological activity of phenoxyacetic acids with Hammett substituent constants and partition coefficients // Nature. – 1962. – Т. 194, № 4824. – С. 178–180.
17. Polanski J. Receptor dependent multidimensional QSAR for modeling drug – receptor interactions // Current Medicinal Chemistry. – 2009. – Т. 16, № 25. – С. 3243–3257.
18. Моделирование соотношений "структура–свойства" / Раевский О. А. – Москва: КДУ Добросвет, 2015. – 288 с.
19. Molecular descriptors for chemoinformatics. / Todeschini R., Consonni V.; Под ред. Mannhold R. и др. – Weinheim: Wiley-VCH, 2009. 1257 с.
20. Ivanciuc O. QSAR Comparative study of wiener descriptors for weighted molecular graphs // Journal of Chemical Information and Computer Sciences. – 2000. – Т. 40, № 6. – С. 1412–1422.
21. Moreau G., P. B. Autocorrelation of molecular structures. Application to SAR studies // New Journal of Chemistry –1980. – Т. 4. – С. 757–764.
22. Consonni V., Todeschini R. Structure – activity relationships by autocorrelation descriptors and genetic algorithms // Chemoinformatics and advanced machine learning perspectives: complex computational methods and collaborative techniques. IGI Global. – 2011. – С. 60–93.

23. Rand Randić M. Generalized molecular descriptors // *Journal of Mathematical Chemistry*. – 1991. – T. 7, № 1. – C. 155–168.
24. Cherkasov A., Muratov E. N., Fourches D., Varnek A., Baskin, II, Cronin M., Dearden J., Gramatica P., Martin Y. C., Todeschini R., Consonni V., Kuz'min V. E., Cramer R., Benigni R., Yang C., Rathman J., Terfloth L., Gasteiger J., Richard A., Tropsha A. QSAR modeling: where have you been? Where are you going to? // *Journal of Medicinal Chemistry*. – 2014. – T. 57, № 12. – C. 4977–5010.
25. Marshall G. R., Cramer R. D. Three-dimensional structure-activity relationships // *Trends in Pharmacological Sciences*. – 1988. – T. 9, № 8. – C. 285–289.
26. Cramer III R. D., Milne M. The lattice model: a general paradigm for shaperelated structure/activity correlation // 19th National meeting of the american chemical Society – Washington, DC, USA: American Chemical Society, 1979. – C. COMP 44.
27. Cramer R. D., Patterson D. E., Bunce J. D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins // *Journal of the American Chemical Society* –1988. – T. 110, № 18. – C. 5959–5967.
28. Wold S., Abano C., Dunn III W. J., Esbensen K., Hellberg S., Johansson E., Lindberg W., Sjostrom M. Modelling data tables by principal componentsand PLS: class patterns and quantitative predictive relations // *Analysis*. – 1984 –T. 12, № 10. – C. 477–485.
29. Cramer R. D., Bunce J. D., Patterson D. E., Frank I. E. Crossvalidation, bootstrapping, and partial least squares compared with multiple regression in conventional QSAR studies // *Quantitative Structure-Activity Relationships*. – 1988. – T. 7, № 1. – C. 18–25.
30. Cramer R. D. BC(DEF) parameters. 1. The intrinsic dimensionality of intermolecular interactions in the liquid state // *Journal of the American Chemical Society*. – 1980. – T. 102, № 6. – C. 1837–1849.
31. Marshall G. R., Arimoto R., Ragno R., Head R. D. Predicting affinity: The sina qua non of activity // *Abstracts of Papers, Joint Conference - Chemical Institute of Canada and American Chemical Society*. – 2000. – T. 219. – C. 056–COMP.
32. Head R. D., Smythe M. L., Oprea T. I., Waller C. L., Green S. M., Marshall G. R. VALIDATE: A new method for the receptor-based prediction of binding affinities of novel ligands // *Journal of the American Chemical Society*. – 1996. – T. 118, № 16. – C. 3959–3969.

33. Klebe G., Abraham U., Mietzner T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity // *Journal of Medicinal Chemistry*. – 1994. – T. 37, № 24. – C. 4130–4146.
34. Goodford P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules // *Journal of Medicinal Chemistry*. – 1985. – T. 28, № 7. – C. 849–857.
35. Pastor M., Cruciani G., McLay I., Pickett S., Clementi S. Grid-independent descriptors (grind): A novel class of alignment-independent three-dimensional molecular descriptors // *Journal of Medicinal Chemistry*. – 2000. – T. 43, № 17. – C. 3233–3243.
36. Duran A., Zamora I., Pastor M. Suitability of GRIND-based principal properties for the description of molecular similarity and ligand-based virtual screening // *Journal of Chemical Information and Modeling*. – 2009. – T. 49, № 9. – C. 2129–2138.
37. Low C. M., Vinter J. G. Rationalizing the activities of diverse cholecystokinin 2 receptor antagonists using molecular field points // *Journal of Medicinal Chemistry*. – 2008. – T. 51, № 3. – C. 565–573.
38. Dixon S. L., Smondyrev A. M., Rao S. N. Phase: A novel approach to pharmacophore modeling and 3D database searching // *Chemical Biology & Drug Design* –2006. – T. 67, № 5. – C. 370–372.
39. Ortiz A. R., Pisabarro M. T., Gago F., Wade R. C. Prediction of drug binding affinities by comparative binding energy analysis // *Journal of Medicinal Chemistry*. – 1995. – T. 38, № 14. – C. 2681–2691.
40. Gohlke H., Klebe G. DrugScore meets CoMFA: Adaptation of fields for molecular comparison (AFMoC) or how to tailor knowledge-based pair-potentials to a particular protein // *Journal of Medicinal Chemistry*. – 2002. – T. 45, № 19. – C. 4153–4170.
41. Doweiko A. M. The hypothetical active site lattice. An approach to modelling active sites from data on inhibitor molecules // *Journal of Medicinal Chemistry*. – 1988. – T. 31, № 7. – C. 1396–1406.
42. Simon Z., Badilescu I., Racovitan T. Mapping of dihydrofolate-reductase receptor site by correlation with minimal topological (steric) differences // *Journal of Theoretical Biology*. – 1977. – T. 66, № 3. – C. 485–495.

43. Varela R., Walters W. P., Goldman B. B., Jain A. N. Iterative refinement of a binding pocket model: active computational steering of lead optimization // *Journal of Medicinal Chemistry*. – 2012. – Т. 55, № 20. – С. 8926–8942.
44. de Mello Castanho Amboni R. D., da Silva Junkes B., Yunes R. A., Heinzen V. E. F. Quantitative structure – odor relationships of aliphatic esters using topological indices // *Journal of Agricultural and Food Chemistry*. – 2000. – Т. 48, № 8. – С. 3517–3521.
45. Баскин И. И., Палюлин В. А., Зефирова Н. С. Молекулярное моделирование рецепторов физиологически активных веществ для целей медицинской химии // *Успехи химии*. – 2009 –Т. 78, № 6. – С. 539–557.
46. Johann, Gasteiger J. Risk assessment of chemicals and prediction of metabolism // Erlangen, Germany. Molecular networks. Inspiring chemical discovery. – URL: https://www.epa.gov/sites/production/files/2014-08/documents/epa_riskassessment_short_810.pdf.
47. Tropsha A. Recent trends in quantitative structure-activity relationships. // *Burger's medicinal chemistry and drug discovery*. – John Wiley and Sons, Inc., 2003.
48. Cherkasov A., Muratov E. N., Fourches D., Varnek A., Baskin, II, Cronin M., Dearden J., Gramatica P., Martin Y. C., Todeschini R., Consonni V., Kuz'min V. E., Cramer R., Benigni R., Yang C., Rathman J., Terfloth L., Gasteiger J., Richard A., Tropsha A. QSAR modeling: where have you been? Where are you going to? // *Journal of Medicinal Chemistry*. – 2014. – Т. 57, № 12. – С. 4977–5010.
49. Cramer R. D. The inevitable QSAR renaissance // *Journal of Computer-Aided Molecular Design*. – 2012. – Т. 26, № 1. – С. 35–38.
50. Pirhadi S., Sunseri J., Koes D. R. Open source molecular modeling // *Journal of Molecular Graphics and Modelling*. – 2016. – Т. 69. – С. 127–143
51. Svitan'ko I. V., Kumskov M. I., Zyryanov I. L., Suslov I. A. A method for describing the molecular electrostatic potential in determining structure-activity relationship // *Mendeleev Communications*. – 1994. – Т. 4, № 5. – С. 161–162.
52. Zauhar R. J., Gianti E., Welsh W. J. Fragment-based shape signatures: A new tool for virtual screening and drug discovery // *Journal of Computer-Aided Molecular Design*. – 2013. – Т. 27, № 12. – С. 1009–1036.

53. Weiner P. K., Kollman P. A. AMBER: Assisted model building with energy refinement. A general program for modeling molecules and their interactions // *Journal of Computational Chemistry*. – 1981. – Т. 2, № 3. – С. 287–303.
54. Biegler-König F. W., Bader R. F. W., Tang T.-H. Calculation of the average properties of atoms in molecules. II // *Journal of Computational Chemistry*. – 1982. – Т. 3, № 3. – С. 317–328.
55. P Smith P. H., Derissen J. L., van Duijneveldt F. B. Intermolecular interactions in crystals of carboxylic acids // *Molecular Physics*. – 1979. – Т. 37, № 2. – С. 521–539.
56. Singh U. C., Kollman P. A. An approach to computing electrostatic charges for molecules // *Journal of Computational Chemistry*. – 1984. – Т. 5, № 2. – С. 129–145.
57. Williams D. E. Representation of the molecular electrostatic potential by atomic multipole and bond dipole models // *Journal of Computational Chemistry*. – 1988. – Т. 9, № 7. – С. 745–763.
58. 3D QSAR in drug design: Theory methods and applications. Three-dimensional quantitative structure activity relationships. Под ред. Kubinyi H.: Springer Netherlands, 1994. Three-Dimensional Quantitative Structure Activity Relationships.
59. 1. Exploring QSAR: Fundamentals and applications in chemistry and biology ACS Professional Reference Book /Hansch C., Leo A.; Под ред. Heller S. R. – Washington: American Chemical Society, 1995. ACS Professional Reference Book
60. Advanced computer-assisted techniques in drug discovery. Methods and principles in medicinal chemistry. Под ред. van de Waterbeemd H. – Второе изд. – Weinheim: VCH, 1995. Methods and Principles in Medicinal Chemistry.
61. Goodsell D. S., Olson A. J. Automated docking of substrates to proteins by simulated annealing // *Proteins*. – 1990. – Т. 8, № 3. – С. 195–202.
62. Kuntz I. D. Structure-based strategies for drug design and discovery // *Science*. – 1992. – Т. 257, № 5073. – С. 1078–1082.
63. Witteveen J. G., van der Weerd A. J. A. Structure-odour relationships of some new synthetic sandalwood aroma chemicals: Synthesis and olfactive properties in a series of bicycle[4.4.0]decan-3-ols // *Recueil des Travaux Chimiques des Pays-Bas*. – 2010. – Т. 106, № 1. – С. 29–34.

64. Buchbauer G., Spreitzer H., Swatonek H., Wolschann P. Absolute configuration and odour analysis of the enantiomeric *tert*-Butylbicyclo[4.4.0]decan-3-ols // *Tetrahedron: Asymmetry*. – 1992. – Т. 3, № 2. – С. 197–198.
65. Krotz A., Helmchen G. Total syntheses of sandalwood fragrances: (*Z*)- and (*E*)- β -santalol and their enantiomers, ent- β -santalene // *Tetrahedron: Asymmetry*. – 1990. – Т. 1, № 8. – С. 537–540.
66. Brunke E.-J., Hammerschmidt F.-J., Struwe H. (+)-*epi*- β -Santalol isolierung aus sandelholzöl und partialsynthese aus (+)- α -santalol // *Tetrahedron Letters*. – 1980. – Т. 21, № 25. – С. 2405–2408.
67. Christenson P. A., Secord N., Willis B. J. Identification of trans- β -santalol and *epi-cis*- β -santalol in east indian sandalwood oil // *Phytochemistry*. – 1981. – Т. 20, № 5. – С. 1139–1141.
68. Demole E. Sur la structure des terpénylphénols obtenus en condensant le camphène avec le phénol, et celle de leurs dérivés hydrogénés à odeur boisée // *Helvetica Chimica Acta*. – 1964. – Т. 47, № 1. – С. 319–338.
69. Пат. 4052341 (А) Соединенные Штаты Америки, МПК7 C11B9/00. 3-Methyl-5-(2,2,3-trimethylcyclopent-3-en-1-yl)pentan-2-ol compound and perfume compositions / Naipawer R. E., Easter W. M.; заявитель и патентообладатель Corporation G. – № 681.351; заявл. 29.04.1976; опубл. 04.10.1977, – 11 с.: ил.
70. Buchbauer G., Lebeda P., Wiesinger L., Weiss-Greiler P., Wolschann P. On the odor of the enantiomers of Madrol // *Chirality*. – 1997. – Т. 9, № 4. – С. 380–385.
71. Weyerstahl P. Odor and structure // *Journal fur Praktische Chemie/Chemiker-Zeitung*. – 1994. – Т. 336, № 2. – С. 95–109.
72. Mookherjee B. D., Trenkle R. W., Wilson R. A. // 12th International congress of flavours, fragrances and essential oils / Под ред. Woidich H. B., G. – Vienna, 1992. – С. 234.
73. Пат. 2255119 (A1) Германия, МПК7 C07C41/00; C07C43/13; C07C43/178; C07C67/00; C11B9/00; 2-Substituierte 2,6-Dimethyl-7-hydroxyoctanverbindungen, Verfahren zu Ihrer Herstellung und Ihre Verwendung in Parfuemkompositionen / Jones B. N., Ansari H. R., Jagers B. G., Janes J. F.; заявитель и патентообладатель LTD B. V. A. – № DE19722255119 19721110 заявл. 10.11.1972; опубл. 30.05.1973, – 14 с.: ил.

74. H. R. Ansari A. J. C. Sesquiterpenes in the perfumery industry // *Journal of Cosmetic Science*. – 1974, . – Т. 25. – С. 203–232.
75. Bajgrowicz J. A., Frank I., Fráter G., Hennig M. Synthesis and structure elucidation of a new potent sandalwood-oil substitute // *Helvetica Chimica Acta*. – 1998. – Т. 81, № 5–8. – С. 1349–1358.
76. Buchbauer G., Neumann A., Siebenheigl U., Wei P., Wolschann P. Structure-activity-relationships of someexo-isocamphanyl cyclohexanols. Conformational calculations on sandalwood odour IX // *Monatshefte für Chemie - Chemical Monthly*. – 1994. – Т. 125, № 6–7. – С. 747–752.
77. Buchbauer G., Hayn A., Liedl E., Weiss-Greiler P., Wolschann P. Conformational calculations on sandalwood odour, Part XI. Molecular similarity of sandalwood odour compounds // *Flavour and Fragrance Journal*. – 1997. – Т. 12, № 3. – С. 141–146.
78. Brunke E. J. Woody aroma chemicals // *Dragoco Report*. – 1983. Т. 6. – С. 149.
79. Essential oils. / Brunke E. J., Klein E.; Под ред. Mookherjee B. D., Mussinan C. J. – Wheaton, USA: Allured Publishing, 1981. – С. 83.
80. Dimoglo A. S., Beda A. A., Shvets N. M., Gorbachov M. Y., Kheifits L. A., Aulchenko I. S. Investigation of the relationship between sandalwood odor and chemical-structure–electron–topological approach // *New journal of chemistry*. – 1995. – Т. 19, № 2. – С. 149–154.
81. Buchbauer G., Lebeda P., Spreitzer H., Wolschann P. Structure-odor relationships of sandalwood odorants: Synthesis of (Z)-7-oxa- β -santalol // *Liebigs Annalen*. – 1995. – Т. 1995, № 9. – С. 1693-1696.
82. Molecular electrostatic potentials and chemical reactivity / Politzer, P.; Murray, J. S. // *Reviews in Computational Chemistry* / Под ред. Lipkowitz, K. B.; Boyd, D. B., Lipkowitz K. B., Boyd D. B. – New York: VCH, 1991. – С. 273.
83. Kearsley S. K., Smith G. M. An alternative method for the alignment of molecular structures: Maximizing electrostatic and steric overlap // *Tetrahedron Computer Methodology*. – 1990. – Т. 3, № 6. – С. 615–633.
84. Masek B. B., Merchant A., Matthew J. B. Molecular shape comparison of angiotensin II receptor antagonists // *Journal of Medicinal Chemistry*. – 1993. – Т. 36, № 9. – С. 1230–1238.

85. Bonaccorsi R., Scrocco E., Tomasi J. Molecular SCF calculations for the ground state of some three-membered ring molecules: $(\text{CH}_2)_3$, $(\text{CH}_2)_2\text{NH}$, $(\text{CH}_2)_2\text{NH}_2^+$, $(\text{CH}_2)_2\text{O}$, $(\text{CH}_2)_2\text{S}$, $(\text{CH})_2\text{CH}_2$, and N_2CH_2 // *The Journal of Chemical Physics*. – 1970. – T. 52, № 10. – C. 5270–5284.
86. Cheng L. P., Gu Y. J., Jin Q. QSAR of α -campholenic derivatives with sandalwood odor, and molecular design // *Monatshefte für Chemie – Chemical Monthly*. – 2010. – T. 141, № 9. – C. 953–959.
87. Gavezzotti A. The calculation of molecular volumes and the use of volume analysis in the investigation of structured media and of solid-state organic reactivity // *Journal of the American Chemical Society*. – 1983. – T. 105, № 16. – C. 5220–5225.
88. Atoms in molecules: A quantum theory. / Bader R. F. W. – Oxford: Oxford University Press, 1990. – 458 c.
89. Amoore J. E. Odor theory and odor classification // *Fragrance chemistry. The science of the sense of smell* / Theimer E. T. – N.Y., London: Academic Press, 1982. – C. 27–76.
90. Polak E. H. Multiple profile-multiple receptor site model for vertebrate olfaction // *Journal of Theoretical Biology*. – 1973. – T. 40, № 3. – C. 469–484.
91. Rinaldi A. The scent of life. The exquisite complexity of the sense of smell in animals and humans // *EMBO Rep*. – 2007. – T. 8, № 7. – C. 629–633.
92. Zou Z., Buck L. B. Combinatorial effects of odorant mixes in olfactory cortex // *Science*. – 2006. – T. 311, № 5766. – C. 1477–1481.
93. Sell C. S. On the unpredictability of odor // *Angewandte Chemie International Edition* –2006. – T. 45, № 38. – C. 6254–6261.
94. Dyson G. M. Some aspects of the vibration theory of odor // *Perfumery and essential oil record*. – 1928. – T. 19. – C. 456–459.
95. Turin L. A spectroscopic mechanism for primary olfactory reception // *Chemical Senses*. – 1996. – T. 21, № 6. – C. 773–791.
96. The emperor of scent: A true story of perfume and obsession. / Burr C. – New York, NY, USA: Random House Trade Paperbacks, 2003.
97. Solomon M. History of science: On smell and scientific practice // *Science*. – 2006. – T. 313, № 5788. – C. 763a–764a.

98. Block E., Jang S., Matsunami H., Sekharan S., Dethier B., Ertem M. Z., Gundala S., Pan Y., Li S., Li Z., Lodge S. N., Ozbil M., Jiang H., Penalba S. F., Batista V. S., Zhuang H. Implausibility of the vibrational theory of olfaction // *Proceedings of the National Academy of Sciences of the United States of America*. – 2015. – T. 112, № 21. – C. E2766–E2774.
99. Franco M. I., Turin L., Merzhin A., Skoulakis E. M. Molecular vibration-sensing component in *Drosophila melanogaster* olfaction // *Proceedings of the National Academy of Sciences of the United States of America*. – 2011. – T. 108, № 9. – C. 3797–802.
100. Keller A., Vosshall L. B. A psychophysical test of the vibration theory of olfaction // *Nature Neuroscience*. – 2004. – T. 7, № 4. – C. 337–338.
101. Turin L. A Method for the calculation of odor character from molecular structure // *Journal of Theoretical Biology*. – 2002. – T. 216, № 3. – C. 367–385.
102. Kaupp U. B. Olfactory signalling in vertebrates and insects: differences and commonalities // *Nature Reviews Neuroscience*. – 2010. – T. 11, № 3. – C. 188–200.
103. Luan F., Liu H. T., Wen Y. Y., Zhang X. Y. Classification of the fragrance properties of chemical compounds based on support vector machine and linear discriminant analysis // *Flavour and Fragrance Journal*. – 2008. – T. 23, № 4. – C. 232–238.
104. Rojas C., Duchowicz P. R., Tripaldi P., Diez R. P. QSPR analysis for the retention index of flavors and fragrances on a OV-101 column // *Chemometrics and Intelligent Laboratory Systems*. – 2015. – T. 140. – C. 126–132.
105. Kumskov M. I., Zyryanov I. L., Svitan'ko I. V. A new method for representing spatial electronic structures of molecules in the problem of structure-biological activity relationship // *Pattern Recognition and Image Analysis*. – 1995. – T. 3. – C. 477–484.
106. Mityushev D.F., Kumskov M.I. An integrated software system to predict properties of chemical compounds on a personal computer // *Pattern Recognition and Image Analysis*, – 1996. – T. 6. C. 809–822.
107. Kovatcheva A., Golbraikh A., Oloff S., Feng J., Zheng W., Tropsha A. QSAR modeling of datasets with enantioselective compounds using chirality sensitive molecular descriptors // *SAR QSAR Environmental Research*. – 2005. – T. 16, № 1–2. – C. 93–102.
108. Anonis D. P. Ambergris // *Perfumer & Flavorist*. – 1995. – T. 20, № 1. – C. 7–11.

109. Lederer E. Chemistry and biochemistry of some mammalian secretions and excretions // *Journal of the Chemical Society (Resumed)*. – 1949. – C. 2115–2125.
110. Ohloff G., Naf F., Decorzant R., Thommen W., Sundt E. Synthesis of potential ambradorants: 5,5,9-trimethyldecalyl derivatives // *Helvetica Chimica Acta*. – 1973. – T. 56, № 4. – C. 1414–1448.
111. Winter B. Ring-opened analogues of Ambrox®: Synthesis and structure-odour relationships // *Helvetica Chimica Acta*. – 1989. – T. 72, № 6. – C. 1278–1283.
112. Vial C., Thommen W., Näf F. Structure-activity relationship in ambergris-type woody odorants possessing a hydronaphthalene skeleton // *Helvetica Chimica Acta*. – 1989. – T. 72, № 6. – C. 1390–1399.
113. Gorbachov M. Y., Rossiter, K. J. A New Electronic topological investigation of the relationship between chemical structure and ambergris odour // *Chemical Senses*. – 1999. – T. 24, № 2. – C. 171–178.
114. Bersuker I. B., Dimoglo A. S., Gorbachov M. Y., Koltsa M. N., Vlad P. F. Structural and electronic origin of ambergris odor of cyclic compounds // *Nouveau journal de chimie*. – 1985. – T. 9. – C. 211–218.
115. Dimoglo A. S., Vlad P. F., Shvets N. M., Coltsa M. N. Electronic-topological investigations of the relationship between chemical structure and ambergris odor // *New Journal of Chemistry*. – 1995. – T. 19. – C. 1217–1226.
116. Bersuker I. B., Dimoglo A. S. The electron-topological approach to the QSAR problem // *Reviews in computational chemistry* / ред. Lipkowitz K. B., Boyd D. B. – Hoboken, NJ, USA: John Wiley & Sons, Inc., 1991.
117. Shvets N. Applied program system for the prognosis of the biological activity of chemical compounds: development and use // *Computer Science Journal of Moldova (Kishinev)*. – 1993. – T. 1, № 3. – C. 31–41.
118. Winter B. Ring-opened analogues of Ambrox®: Synthesis and structure-odour relationships // *Helvetica Chimica Acta*. – 1989. – T. 72, № 6. – C. 1278–1283.
119. Winter B. Quantitative structure-activity relationships in drug design // *QSAR in olfaction: Ambergris type odorants* / ред. Alan R. – New York, USA: Liss, Inc., 1989. – C. 401–405

120. Bajgrowicz J. A., Frank I. Camphor-derived amber/woody odorants: 1,7,7-trimethyl-2'-iso-propylspiro[bicyclo[2.2.1]heptane-2,4'-(1,3-dioxanes)] // *Tetrahedron: Asymmetry*. – 2001. – T. 12, № 14. – C. 2049–2057.
121. Bajgrowicz J. A. B., C. Molecular modelling in the design of new odorants: Scope and limitations // 13th International congress of flavours, fragrances and essential oils. – T. 3 – Istanbul: AREP, 1995. – C. 1–15.
122. CATALYST 4.7. Installation guide. / Molecular Simulations Inc. (Accelrys): San Diego, CA, 2002. // URL:http://www.esi.umontreal.ca/accelrys/life/catalyst47/install/47_install.pdf
123. Greene J., Kahn S., Savoj H., Sprague P., Teig S. Chemical function queries for 3D database search // *Journal of Chemical Information and Modeling*. – 1994. – T. 34, № 6. – C. 1297–1308.
124. Bilke S., Mosandl A. Enantioselective analysis of 2-methyl-4-(2,2,3-trimethylcyclopent-3-en-1-yl)-but-2-enol, 2-methyl-4-(2,2,3-trimethylcyclopent-3-en-1-yl)-but-2-enal and α -campholene aldehyde by capillary gas chromatography // *Journal of Separation Science*. – 2001. – T. 24, № 10-11. – C. 819–822.
125. Buchbauer G., Lebeda P., Wiesinger L., Weiss-Greiler P., Wolschann P. On the odor of the enantiomers of Madrol // *Chirality*. – 1997. – T. 9, № 4. – C. 380–385.
126. Principles of biochemistry. / Horton H. R., Moran L. A., Raymond S. Ochs (Author), Rawn D. J., Scrimgeour K. G. – 3 изд. – Englewood Cliffs, NJ: Prentice Hall, 2002.
127. Solms J. V., L.; Egli, R. H. The taste of L- and D-amino acids. // *Experientia*. – 1965. – T. 21. – C. 692–694.
128. Schiffman S. S., Clark T. B., Gagnon J. Influence of chirality of amino acids on the growth of perceived taste intensity with concentration // *Physiology & Behavior*. – 1982. – T. 28, № 3. – C. 457–465.
129. De Camp W. H. The FDA perspective on the development of stereoisomers // *Chirality*. – 1989. – T. 1, № 1. – C. 2–6.
130. Hutt A. J., Tan S. C. Drug chirality and its clinical significance // *Drugs*. – 1996. – T. 52, Supplement 5. – C. 1–12.
131. Wnendt S., Zwingenberger K. Thalidomide's chirality // *Nature*. – 1997. – T. 385, № 6614. – C. 303–304.

132. Kovatcheva A., Golbraikh A., Oloff S., Xiao Y. D., Zheng W., Wolschann P., Buchbauer G., Tropsha A. Combinatorial QSAR of ambergris fragrance compounds // *Journal of Chemical Information and Computer Sciences*. – 2004. – T. 44, № 2. – C. 582–595.
133. Labute P. Binary qsar: A new method for quantitative structure activity relationships // 1999 Pacific Symposium / Под ред. Altman – Singapore: World Scientific Publishing, 1999.
134. Protein and antibody modeling. / MOE: Molecular Operating Environment // Chemical Computing Group. // URL: http://www.chemcomp.com/MOE-Protein_and_Antibody_Modeling.htm.
135. Silverman B. D., Platt D. E. Comparative molecular moment analysis (CoMMA): 3D-QSAR without molecular superposition // *Journal of Medicinal Chemistry*. – 1996. – T. 39, № 11. – C. 2129–2140.
136. Cruciani G., Crivori P., Carrupt P. A., Testa B. Molecular fields in quantitative structure–permeation relationships: the VolSurf approach // *Journal of Molecular Structure: THEOCHEM*. – 2000. – T. 503, № 1–2. – C. 17–30.
137. MolconnZ / MolconnZ-Soft – 2008. // URL: <http://www.edusoft-lc.com/molconn/>
138. Golbraikh A., Bonchev D., Tropsha A. Novel chirality descriptors derived from molecular topology // *Journal of Chemical Information and Computer Sciences*. – 2001. – T. 41, № 1. – C. 147–158.
139. Golbraikh A., Tropsha A. QSAR modeling using chirality descriptors derived from molecular topology // *Journal of Chemical Information and Modeling*. – 2003. – T. 43, № 1. – C. 144–154.
140. Rojas C., Duchowicz P. R., Tripaldi P., Diez R. P. QSPR analysis for the retention index of flavors and fragrances on a OV-101 column // *Chemometrics and Intelligent Laboratory Systems*. – 2015. – T. 140. – C. 126–132.
141. Yan J., Cao D. S., Guo F. Q., Zhang L. X., He M., Huang J. H., Xu Q. S., Liang Y. Z. Comparison of quantitative structure-retention relationship models on four stationary phases with different polarity for a diverse set of flavor compounds // *Journal of Chromatography A*. – 2012. – T. 1223. – C. 118–25.

142. Баскин И.И., Палюлин В.А., Зефирова Н.С. Молекулярное моделирование рецепторов физиологически активных веществ для целей медицинской химии // Успехи химии. – 2003 – Т. 72. – С. 495–511.
143. Warshel A., Sharma P. K., Kato M., Parson W. W. Modeling electrostatic effects in proteins // *Biochimica et Biophysica Acta*. – 2006. – Т. 1764, № 11. – С. 1647–76.
144. Warshel A., Sussman F., King G. Free energy of charges in solvated proteins: microscopic calculations using a reversible charging process // *Biochemistry*. – 2002. – Т. 25, № 26. – С. 8368–8372.
145. Russell S. T., Warshel A. Calculations of electrostatic energies in proteins // *Journal of Molecular Biology*. – 1985. – Т. 185, № 2. – С. 389–404.
146. Warshel A., Russell S. T. Calculations of electrostatic interactions in biological systems and in solutions // *Quarterly Reviews of Biophysics*. – 1984. – Т. 17, № 3. – С. 283–422.
147. Sham Y. Y., Chu Z. T., Warshel A. Consistent calculations of pKa's of ionizable residues in proteins: Semi-microscopic and microscopic approaches // *The Journal of Physical Chemistry B*. – 1997. – Т. 101, № 22. – С. 4458–4472.
148. Gilson M. K., Honig B. H. Energetics of charge-charge interactions in proteins // *Proteins*. – 1988. – Т. 3, № 1. – С. 32–52.
149. Bashford D., Karplus M. pKa's of ionizable groups in proteins: atomic detail from a continuum electrostatic model // *Biochemistry*. – 1990. – Т. 29, № 44. – С. 10219–10225.
150. Antosiewicz J., McCammon J. A., Gilson M. K. Prediction of pH-dependent properties of proteins // *Journal of Molecular Biology*. – 1994. – Т. 238, № 3. – С. 415–436.
151. Bashford D., Case D. A. Generalized born models of macromolecular solvation effects // *Annual Review of Physical Chemistry*. – 2000. – Т. 51. – С. 129–152.
152. Mehler E. L., Guarnieri F. A Self-Consistent, microenvironment modulated screened coulomb potential approximation to calculate pH-dependent electrostatic effects in proteins // *Biophysical Journal*. – 1999. – Т. 77, № 1. – С. 3–22.
153. Lee F. S., Chu Z. T., Warshel A. Microscopic and semimicroscopic calculations of electrostatic energies in proteins by the POLARIS and ENZYMIK programs // *Journal of Computational Chemistry*. – 1993. – Т. 14, № 2. – С. 161–185.

154. Ji C., Mei Y., Zhang J. Z. Developing polarized protein-specific charges for protein dynamics: MD free energy calculation of pKa shifts for Asp26/Asp20 in thioredoxin // *Biophysical Journal*. – 2008. – T. 95, № 3. – C. 1080–1088.
155. Mongan J., Case D. A., McCammon J. A. Constant pH molecular dynamics in generalized Born implicit solvent // *Journal of Computational Chemistry* –2004. – T. 25, № 16. – C. 2038–2048.
156. Messer B. M., Roca M., Chu Z. T., Vicatos S., Kilshtain A. V., Warshel A. Multiscale simulations of protein landscapes: using coarse-grained models as reference potentials to full explicit models // *Proteins*. – 2010. – T. 78, № 5. – C. 1212–1227.
157. You T. J., Bashford D. Conformation and hydrogen ion titration of proteins: A continuum electrostatic model with conformational flexibility // *Biophysical Journal*. – 1995. – T. 69, № 5. – C. 1721–1733.
158. Kieseritzky G., Knapp E. W. Optimizing pKa computation in proteins with pH adapted conformations // *Proteins*. – 2008. – T. 71, № 3. – C. 1335–48.
159. Georgescu R. E., Alexov E. G., Gunner M. R. Combining conformational flexibility and continuum electrostatics for calculating pKas in proteins // *Biophysical Journal*. – 2002. – T. 83, № 4. – C. 1731–1748.
160. Song Y., Mao J., Gunner M. R. MCCE2: improving protein pKa calculations with extensive side chain rotamer sampling // *Journal of Computational Chemistry*. – 2009. – T. 30, № 14. – C. 2231–47.
161. Koehl P., Delarue M. Application of a self-consistent mean field theory to predict protein side-chains conformation and estimate their conformational entropy // *Journal of Molecular Biology*. – 1994. – T. 239, № 2. – C. 249–275.
162. Barth P., Alber T., Harbury P. B. Accurate, conformation-dependent predictions of solvent effects on protein ionization constants // *Proceedings of the National Academy of Sciences of the United States of America*. – 2007. – T. 104, № 12. – C. 4898–4903.
163. Huang C., Darwiche A. Inference in belief networks: A procedural guide // *International Journal of Approximate Reasoning*. – 1996. – T. 15, № 3. – C. 225–263.
164. Moitessier N., Englebienne P., Lee D., Lawandi J., Corbeil C. R. Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go // *British Journal of Pharmacology* –2008. – T. 153 Suppl 1. – C. S7–S26.

165. Totrov M. M., Abagyan R. A. Flexible protein–ligand docking by global energy optimization in internal coordinates // *Proteins: Structure, Function, and Bioinformatics*. – 1997. – T. 29, № Supplement S1. – C. 215–220.
166. Tietze S., Apostolakis J. GlamDock: development and validation of a new docking tool on several thousand protein-ligand complexes // *Journal of Chemical Information and Modeling*. – 2007. – T. 47, № 4. – C. 1657–1672.
167. Morris G. M., Goodsell D. S., Halliday R. S., Huey R., Hart W. E., Belew R. K., Olson A. J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function // *Journal of Computational Chemistry*. – 1998. – T. 19, № 14. – C. 1639–1662.
168. Huey R., Morris G. M., Olson A. J., Goodsell D. S. A semiempirical free energy force field with charge-based desolvation // *Journal of Computational Chemistry* –2007. – T. 28, № 6. – C. 1145–1152.
169. Jones G., Willett P., Glen R. C., Leach A. R., Taylor R. Development and validation of a genetic algorithm for flexible docking // *Journal of Molecular Biology*. – 1997. – T. 267, № 3. – C. 727–748.
170. Th Thomsen R., Christensen M. H. MolDock: a new technique for high-accuracy molecular docking // *Journal of Medicinal Chemistry*. – 2006. – T. 49, № 11. – C. 3315–3321.
171. Oshiro C. M., Kuntz I. D., Dixon J. S. Flexible ligand docking using a genetic algorithm // *Journal of Computer-Aided Molecular Design*. – 1995. – T. 9, № 2. – C. 113–130.
172. Rarey, M.; Kramer, B.; Lengauer, T. Multiple automatic base selection: Protein-ligand docking based on incremental construction without manual intervention. // *Journal of Computer-Aided Molecular Design*. – 1997. – T. 11, № 4. – C. 369–384.
173. Kramer B., Rarey M., Lengauer T. Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking // *Proteins: Structure, Function, and Genetics*. – 1999. – T. 37, № 2. – C. 228–241.
174. Jain A. N. Surflex: Fully automatic flexible molecular docking using a molecular similarity-based search engine // *Journal of Medicinal Chemistry*. – 2003. – T. 46, № 4. – C. 499–511.

175. Zsoldos Z., Reid D., Simon A., Sadjad B. S., Johnson A. P. eHiTS: An innovative approach to the docking and scoring function problems // *Current Protein & Peptide Science*. – 2006. – T. 7, № 5. – C. 421–435.
176. Zsoldos Z., Reid D., Simon A., Sadjad S. B., Johnson A. P. eHiTS: a new fast, exhaustive flexible ligand docking system // *Journal of Molecular Graphics and Modelling*. – 2007. – T. 26, № 1. – C. 198–212.
177. McMartin C., Bohacek R. S. QXP: powerful, rapid computer algorithms for structure-based drug design // *Journal of Computer-Aided Molecular Design*. – 1997. – T. 11, № 4. – C. 333–344.
178. Venkatachalam C. M., Jiang X., Oldfield T., Waldman M. LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites // *Journal of Molecular Graphics and Modelling*. – 2003. – T. 21, № 4. – C. 289–307.
179. Friesner R. A., Banks J. L., Murphy R. B., Halgren T. A., Klicic J. J., Mainz D. T., Repasky M. P., Knoll E. H., Shelley M., Perry J. K., Shaw D. E., Francis P., Shenkin P. S. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy // *Journal of Medicinal Chemistry*. – 2004. – T. 47, № 7. – C. 1739–1749.
180. Cole J. C., Murray C. W., Nissink J. W., Taylor R. D., Taylor R. Comparing protein-ligand docking programs is difficult // *Proteins*. – 2005. – T. 60, № 3. – C. 325–32.
181. Nissink J. W., Murray C., Hartshorn M., Verdonk M. L., Cole J. C., Taylor R. A new test set for validating predictions of protein-ligand interaction // *Proteins*. – 2002. – T. 49, № 4. – C. 457–471.
182. Hartshorn M. J., Verdonk M. L., Chessari G., Brewerton S. C., Mooij W. T., Mortenson P. N., Murray C. W. Diverse, high-quality test set for the validation of protein-ligand docking performance // *Journal of Medicinal Chemistry*. – 2007. – T. 50, № 4. – C. 726–741.
183. Jain A. N., Nicholls A. Recommendations for evaluation of computational methods // *Journal of Computer-Aided Molecular Design*. – 2008. – T. 22, № 3–4. – C. 133–139.
184. Kellenberger E., Rodrigo J., Muller P., Rognan D. Comparative evaluation of eight docking tools for docking and virtual screening accuracy // *Proteins*. – 2004. – T. 57, № 2. – C. 225–242.

185. Kontoyianni M., McClellan L. M., Sokol G. S. Evaluation of docking performance: comparative data on docking algorithms // *Journal of Medicinal Chemistry*. – 2004. – T. 47, № 3. – C. 558–565.
186. Warren G. L., Andrews C. W., Capelli A. M., Clarke B., LaLonde J., Lambert M. H., Lindvall M., Nevins N., Semus S. F., Senger S., Tedesco G., Wall I. D., Woolven J. M., Peishoff C. E., Head M. S. A critical assessment of docking programs and scoring functions // *Journal of Medicinal Chemistry*. – 2006. – T. 49, № 20. – C. 5912–5931.
187. Graves A. P., Brenk R., Shoichet B. K. Decoys for docking // *Journal of Medicinal Chemistry*. – 2005. – T. 48, № 11. – C. 3714–3728.
188. Huang N., Shoichet B. K., Irwin J. J. Benchmarking sets for molecular docking // *Journal of Medicinal Chemistry*. – 2006. – T. 49, № 23. – C. 6789–801.
189. Stroganov O. V., Novikov F. N., Stroylov V. S., Kulkov V., Chilov G. G. Lead finder: an approach to improve accuracy of protein-ligand docking, binding energy estimation, and virtual screening // *Journal of Chemical Information and Modeling*. – 2008. – T. 48, № 12. – C. 2371–2385.
190. Jacobson M. P., Pincus D. L., Rapp C. S., Day T. J., Honig B., Shaw D. E., Friesner R. A. A hierarchical approach to all-atom protein loop prediction // *Proteins*. – 2004. – T. 55, № 2. – C. 351–367.
191. Jacobson M. P., Friesner R. A., Xiang Z., Honig B. On the role of the crystal environment in determining protein side-chain conformations // *Journal of Molecular Biology*. – 2002. – T. 320, № 3. – C. 597–608.
192. Spassov V. Z., Yan L., Flook P. K. The dominant role of side-chain backbone interactions in structural realization of amino acid code. ChiRotor: a side-chain prediction algorithm based on side-chain backbone interactions // *Protein Science*. – 2007. – T. 16, № 3. – C. 494–506.
193. Smith R. E., Lovell S. C., Burke D. F., Montalvao R. W., Blundell T. L. Andante: reducing side-chain rotamer search space during comparative modeling using environment-specific substitution probabilities // *Bioinformatics*. – 2007. – T. 23, № 9. – C. 1099–105.

194. Abagyan R., Totrov M. Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins // *Journal of Molecular Biology*. – 1994. – T. 235, № 3. – C. 983–1002.
195. Rossi K. A., Nayeem A., Weigelt C. A., Krystek S. R., Jr. Closing the side-chain gap in protein loop modeling // *Journal of Computer-Aided Molecular Design*. – 2009. – T. 23, № 7. – C. 411–418.
196. Rossi K. A., Weigelt C. A., Nayeem A., Krystek S. R., Jr. Loopholes and missing links in protein modeling // *Protein Science*. – 2007. – T. 16, № 9. – C. 1999–2012.
197. Canutescu A. A., Shelenkov A. A., Dunbrack R. L., Jr. A graph-theory algorithm for rapid protein side-chain prediction // *Protein Science*. – 2003. – T. 12, № 9. – C. 2001–2014.
198. Wang Q., Canutescu A. A., Dunbrack R. L., Jr. SCWRL and MolIDE: computer programs for side-chain conformation prediction and homology modeling // *Nature Protocols*. – 2008. – T. 3, № 12. – C. 1832–1847.
199. Fong P. C., Boss D. S., Yap T. A., Tutt A., Wu P., Mergui-Roelvink M., Mortimer P., Swaisland H., Lau A., O'Connor M. J., Ashworth A., Carmichael J., Kaye S. B., Schellens J. H., de Bono J. S. Inhibition of *poly*(ADP-ribose) polymerase in tumors from BRCA mutation carriers // *The New England Journal of Medicine*. – 2009. – T. 361, № 2. – C. 123–134.
200. Liu X., Shi Y., Maag D. X., Palma J. P., Patterson M. J., Ellis P. A., Surber B. W., Ready D. B., Soni N. B., Ladror U. S., Xu A. J., Iyer R., Harlan J. E., Solomon L. R., Donawho C. K., Penning T. D., Johnson E. F., Shoemaker A. R. Iniparib nonselectively modifies cysteine-containing proteins in tumor cells and is not a bona fide PARP inhibitor // *Clinical Cancer Research*. – 2012. – T. 18, № 2. – C. 510–523.
201. Patel A. G., De Lorenzo S. B., Flatten K. S., Poirier G. G., Kaufmann S. H. Failure of iniparib to inhibit *poly*(ADP-Ribose) polymerase in vitro // *Clinical Cancer Research*. – 2012. – T. 18, № 6. – C. 1655–1662.

202. Mian A. A., Rafiei A., Haberbosch I., Zeifman A., Titov I., Stroylov V., Metodieva A., Stroganov O., Novikov F., Brill B., Chilov G., Hoelzer D., Ottmann O. G., Ruthardt M. PF-114, a potent and selective inhibitor of native and mutated BCR/ABL is active against Philadelphia chromosome-positive (Ph+) leukemias harboring the T315I mutation // *Leukemia*. – 2015. – T. 29, № 5. – C. 1104–1114.
203. Yanagi S., Inatome R., Takano T., Yamamura H. SYK expression and novel function in a wide variety of tissues // *Biochemical and Biophysical Research Communications* – 2001. – T. 288, № 3. – C. 495–498.
204. Riccaboni M., Bianchi I., Petrillo P. Spleen tyrosine kinases: biology, therapeutic targets and drugs // *Drug Discovery Today* – 2010. – T. 15, № 13–14. – C. 517–530.
205. Lee D. M., Weinblatt M. E. Rheumatoid arthritis // *The Lancet*. – 2001. – T. 358, № 9285. – C. 903–911.
206. Choy E. H., Panayi G. S. Cytokine pathways and joint inflammation in rheumatoid arthritis // *The New England Journal of Medicine*. – 2001. – T. 344, № 12. – C. 907–916.
207. Rossi A. B., Herlaar E., Braselmann S., Huynh S., Taylor V., Frances R., Issakani S. D., Argade A., Singh R., Payan D. G., Masuda E. S. Identification of the SYK kinase inhibitor R112 by a human mast cell screen // *Journal of Allergy and Clinical Immunology*. – 2006. – T. 118, № 3. – C. 749–755.
208. Young R. M., Hardy I. R., Clarke R. L., Lundy N., Pine P., Turner B. C., Potter T. A., Refaeli Y. Mouse models of non-Hodgkin lymphoma reveal SYK as an important therapeutic target // *Blood*. – 2009. – T. 113, № 11. – C. 2508–2516.
209. Xie H. Z., Li L. L., Ren J. X., Zou J., Yang L., Wei Y. Q., Yang S. Y. Pharmacophore modeling study based on known spleen tyrosine kinase inhibitors together with virtual screening for identifying novel inhibitors // *Bioorganic & Medicinal Chemistry Letters*. – 2009. – T. 19, № 7. – C. 1944–1949.
210. Rüegg U. T., Gillian B. Staurosporine, K-252 and UCN-01: potent but nonspecific inhibitors of protein kinases // *Trends in Pharmacological Sciences*. – 1989. – T. 10, № 6. – C. 218–220.
211. Atwell S., Adams J. M., Badger J., Buchanan M. D., Feil I. K., Froning K. J., Gao X., Hendle J., Keegan K., Leon B. C., Muller-Dieckmann H. J., Nienaber V. L., Noland B. W., Post K., Rajashankar K. R., Ramos A., Russell M., Burley S. K., Buchanan S. G. A novel

mode of Gleevec binding is revealed by the structure of spleen tyrosine kinase // *The Journal of Biological Chemistry*. – 2004. – Т. 279, № 53. – С. 55827–55832.

212. Ashikawa K., Majumdar S., Banerjee S., Bharti A. C., Shishodia S., Aggarwal B. B. Piceatannol inhibits TNF-induced NF- κ B activation and NF- κ B-mediated gene expression through suppression of I κ B kinase and p65 phosphorylation // *The Journal of Immunology*. – 2002. – Т. 169, № 11. – С. 6490–6497.

213. Tabata H., Matsuoka T., Endo F., Nishimura Y., Matsushita S. Ligation of HLA-DR molecules on B cells induces enhanced expression of IgM heavy chain genes in association with Syk activation // *The Journal of Biological Chemistry*. – 2000. – Т. 275, № 45. – С. 34998–5005.

214. Пат. 2017246167 (A1) Соединенные Штаты Америки, МПК⁷ А61К31/505; А61К31/506; А61К31/519; А61К31/5377; А61P1/00; А61P11/06; А61P17/00; А61P27/02; А61P27/14; А61P27/16; А61P29/00; А61P37/08; А61P43/00; C07D239/48; C07D401/12; C07D401/14; C07D403/12; C07D403/14; C07D405/12; C07D405/14; C07D409/12; C07D409/14; C07D413/12; C07D413/14; C07D417/12; C07D417/14. 2,4-Pyrimidinediamine compounds and their uses / Singh R., Argade A., Payan D., Molineaux S., Holland S. J., Clough J., Keim H., Bhamidipati S., Sylvain C., Li H., Rossii A. B.; заявитель и патентообладатель Rigel Pharmaceuticals Inc. – № US201715443145; заявл. 27.02.2017; опубл. 31.08.2017, – 269 с.: ил.

215. Пат. WO 2004014382 (A1) Соединенные Штаты Америки, МПК⁷ А61К31/506; А61К31/519; А61P37/00; (IPC1-7): А61К31/506; А61К31/519. Methods of treating or preventing autoimmune diseases with 2,4-pyrimidinediamine compounds / Singh R., Argade A., Payan D. G., Clough J., Keim H., Sylvain C., Li H., Bhamidipati S.; заявитель и патентообладатель Rigel Pharmaceuticals Inc. – № WO 2003US24087; заявл. 29.07.2003; опубл. 19.02.2004, – 811 с.: ил.

216. Пат. WO 2005012294 (A1) Соединенные Штаты Америки, МПК⁷ А61К31/505; А61К31/506; А61К31/538; А61К31/5415; А61К31/542; А61P37/00; C07D403/12; C07D405/12; C07D405/14; C07D413/02; C07D413/12; C07D413/14; C07D417/02; C07D417/12; C07D417/14; C07D498/02; C07D498/04; C07D498/06; (IPC1-7): А61К31/505; А61P37/00; C07D413/12; C07D413/14; C07D417/12. 2,4-Pyrimidinediamine compounds for use in the treatment or prevention of autoimmune diseases / Singh

- R., Ankush A., Li H., Bhamidipati S., Carroll D., Sylvain C., Clough J., Keim H.; заявитель и патентообладатель – № WO 2004US24920; заявл. 30.07.2004; опубл. 10.02.2005, – 169 с.: ил.
217. Пат. WO 2006068770 (A1) Соединенные Штаты Америки, МПК⁷ А61К31/5386; А61Р37/00; С07D413/12; С07D413/14. Spiro-2,4-pyrimidinediamine compounds and their uses / Cooper R., Singh R., Clough G.; заявитель и патентообладатель Rigel Pharmaceuticals Inc. – № WO2005US42712; заявл. 23.11.2005; опубл. – 96 с.: ил.
218. Пат. 8314093 (B2) Соединенные Штаты Америки, МПК⁷ А61К31/506; А61К31/538; А61К31/5415; А61Р29/00; А61Р37/00; С12N5/00. 2,4-Pyrimidinediamine compounds for treating or preventing autoimmune diseases / Clough J. W., Bhamidipati S., Singh R., Masuda E., Zhao H.; заявитель и патентообладатель Rigel Pharmaceuticals Inc. – № 2010152172 (A1); заявл. 17.06.2010 опубл. 20.11.2012, – 29 с.: ил.
219. Braselmann S., Taylor V., Zhao H., Wang S., Sylvain C., Baluom M., Qu K., Herlaar E., Lau A., Young C., Wong B. R., Lovell S., Sun T., Park G., Argade A., Jurcevic S., Pine P., Singh R., Grossbard E. B., Payan D. G., Masuda E. S. R406, an orally available spleen tyrosine kinase inhibitor blocks fc receptor signaling and reduces immune complex-mediated inflammation // *Journal of Pharmacology and Experimental Therapeutics*. – 2006. – Т. 319, № 3. – С. 998–1008.
220. Bajpai M. Fostamatinib, a Syk inhibitor prodrug for the treatment of inflammatory diseases. // *IDrugs: the investigational drugs journal*. – 2009. – Т. 12, № 3. – С. 174–185.
221. Пат. WO 2006078846 (A1) Соединенные Штаты Америки, МПК⁷ А61К31/5383; А61К31/675; А61Р37/00; С07D498/04; С07F9/6561. Prodrugs of 2,4-pyrimidinediamine compounds and their uses / Singh R., Bhamidipati S., Masuda E.; заявитель и патентообладатель Rigel Pharmaceuticals Inc. – № WO 2006US01945; заявл. 19.01.2006; опубл. 27.07.2006, – 110 с.: ил.
222. Пат. WO 2007124221 (A1) Соединенные Штаты Америки, МПК⁷ А61К31/505; А61Р35/02. Methods of treating cell proliferative disorders by using pyrimidinediamine compounds / Masuda E., Payan D. G., Grossbard E. B.; заявитель и патентообладатель Rigel pharmaceuticals inc. – № WO 2007US64511; заявл. 21.03.2007 опубл. 01.11.2007, – 88 с.: ил.

223. Пат. WO 2009031011 (A2) Великобритания, МПК⁷ А61К31/538; А61Р11/06; С07D413/12. Salt form / Taylor S., Colin J.; заявитель и патентообладатель Pfizer Ltd. – № WO 2008IB02288 заявл. 27.08.2008; опубл. 12.03.2009, – 34 с.: ил.

224. Yamamoto N., Takeshita K., Shichijo M., Kokubo T., Sato M., Nakashima K., Ishimori M., Nagai H., Li Y. F., Yura T., Bacon K. B. The orally available spleen tyrosine kinase inhibitor 2-[7-(3,4-dimethoxyphenyl)-imidazo[1,2-c]pyrimidin-5-ylamino]nicotinamide dihydrochloride (BAY 61-3606) blocks antigen-induced airway inflammation in rodents // *Journal of Pharmacology and Experimental Therapeutics*. – 2003. – Т. 306, № 3. – С. 1174–1181.

225. Пат. WO 0183485 (A1) Япония, МПК⁷ А61К31/505; А61К31/519; А61К31/5377; А61К31/541; А61Р1/04; А61Р11/02; А61Р11/06; А61Р17/00; А61Р17/04; А61Р17/06; А61Р19/02; А61Р21/04; А61Р25/28; А61Р27/02; А61Р27/14; А61Р29/00; А61Р3/04; А61Р35/00; А61Р37/02; А61Р37/06; А61Р37/08; А61Р43/00; А61Р7/02; А61Р7/06; С07D487/04; С07D519/00; (IPC1-7): А61К31/505; А61Р11/06; А61Р35/00; А61Р37/08; А61Р7/02; С07D487/04. Imidazopyrimidine derivatives and triazolopyrimidine derivatives / Yura T., Concepcion A. B., Han G., Hiraoka M., Katsumata H., Kawamura N., Kokubo T., Komura H., Li Y., Lowinger T. B., Mogi M., Yamamoto N., Yoshida N.; заявитель и патентообладатель Bayer AG. – № WO 2001EP04357; заявл. 17.04.2001; опубл. 08.11.2001, – 205 с.: ил.

226. Пат. WO 03057695 (A1) Соединенные Штаты Америки, МПК⁷ А61Р29/00; С07D471/04; (IPC1-7): А61К31/4375; А61Р29/00; С07D471/04. 1,6 Naphthyridines useful as inhibitors of SYK kinase / Cywin C. L., Jakes S. E., Heider J., Bobko M. A., Des J. R. L., Player M., Rinker J., Winters M., Zhao B.-P.; заявитель и патентообладатель Boehringer Ingelheim Pharma. – № WO 2002US38375; заявл. 03.12.2002; опубл. 17.07.2003, – 86 с.: ил.

227. Cywin C. L., Zhao B.-P., McNeil D. W., Hrapchak M., Prokopowicz A. S., Goldberg D. R., Morwick T. M., Gao A., Jakes S., Kashem M., Magolda R. L., Soll R. M., Player M. R., Bobko M. A., Rinker J., DesJarlais R. L., Winters M. P. Discovery and SAR of novel Naphthyridines as potent inhibitors of spleen tyrosine kinase (SYK) // *Bioorganic & Medicinal Chemistry Letters*. – 2003. – Т. 13, № 8. – С. 1415–1418.

228. Lai J. Y. Q., Cox P. J., Patel R., Sadiq S., Aldous D. J., Thurairatnam S., Smith K., Wheeler D., Jagpal S., Parveen S., Fenton G., Harrison T. K. P., McCarthy C., Bamborough P. Potent small molecule inhibitors of spleen tyrosine kinase (SYK) // *Bioorganic & Medicinal Chemistry Letters*. – 2003. – Т. 13, № 18. – С. 3111–3114.
229. Hisamichi H., Naito R., Toyoshima A., Kawano N., Ichikawa A., Orita A., Orita M., Hamada N., Takeuchi M., Ohta M., Tsukamoto S. Synthetic studies on novel Syk inhibitors. Part 1: Synthesis and structure-activity relationships of pyrimidine-5-carboxamide derivatives // *Bioorganic & Medicinal Chemistry* –2005. – Т. 13, № 16. – С. 4936–4951.
230. Hirabayashi A., Mukaiyama H., Kobayashi H., Shiohara H., Nakayama S., Ozawa M., Miyazawa K., Misawa K., Ohnota H., Isaji M. A novel Syk family kinase inhibitor: design, synthesis, and structure-activity relationship of 1,2,4-triazolo[4,3-c]pyrimidine and 1,2,4-triazolo[1,5-c]pyrimidine derivatives // *Bioorganic & Medicinal Chemistry Letters*. – 2008. – Т. 16, № 15. – С. 7347–7357.
231. Hirabayashi A., Mukaiyama H., Kobayashi H., Shiohara H., Nakayama S., Ozawa M., Tsuji E., Miyazawa K., Misawa K., Ohnota H., Isaji M. Structure-activity relationship studies of imidazo[1,2-c]pyrimidine derivatives as potent and orally effective Syk family kinases inhibitors // *Bioorg Med Chem*. – 2008. – Т. 16, № 20. – С. 9247–60.
232. Пат. 2004203748 (А) Япония, МПК⁷ А61К31/519; А61К31/5377; А61Р11/02; А61Р11/06; А61Р17/00; А61Р27/02; А61Р37/06; А61Р37/08; А61Р43/00; С07D487/04. New imidazo[1,2-c]pyrimidine derivative, medicinal composition comprising the same and application thereof / Hirabayashi A., Mukoyama H., Shiobara H., Kobayashi H., Terao Y., Miyazawa T., Misawa K., Onoda H.; заявитель и патентообладатель Pharmaceutical К. – заявл. 24.12.2002; опубл. 22.07.2004, – 117 с.: ил.
233. Farmer L. J., Bemis G., Britt S. D., Cochran J., Connors M., Harrington E. M., Hoock T., Markland W., Nanthakumar S., Taslimi P., Ter Haar E., Wang J., Zhaveri D., Salituro F. G. Discovery and SAR of novel 4-thiazolyl-2-phenylaminopyrimidines as potent inhibitors of spleen tyrosine kinase (SYK) // *Bioorganic & Medicinal Chemistry Letters* –2008. – Т. 18, № 23. – С. 6231–6235.

234. Пат. WO 02096905 (A1) Соединенные Штаты Америки, МПК⁷ А61К31/427; А61К31/506; А61Р3/10; С07D405/06; С07D405/14; С07D417/04; С07D417/14; С07D487/10; С07D521/00; С07F7/18; С07F9/6561; (IPC1-7): А61К31/427; А61Р3/10; С07D417/04. Thiazole compounds useful as inhibitors of protein kinases / Cochran J., Nanthakumar S., Harrington E., Wang J.; заявитель и патентообладатель Vertex Pharma. – № WO 2002US16352; заявл. 23.05.2002; опубл. 05.12.2002, – 47 с.: ил.
235. Shichijo M., Yamamoto N., Tsujishita H., Kimata M., Nagai H., Kokubo T. Inhibition of Syk Activity and Degranulation of Human Mast Cells by Flavonoids // Biological & Pharmaceutical Bulletin. – 2003. – Т. 26, № 12. – С. 1685–1690.
236. Moriya K., Rivera J., Odom S., Sakuma Y., Muramoto K., Yoshiuchi T., Miyamoto M., Yamada K. ER-27319, an acridone-related compound, inhibits release of antigen-induced allergic mediators from mast cells by selective inhibition of Fc ϵ receptor I-mediated activation of Syk // Proceedings of the National Academy of Sciences of the United States of America. – 1997. – Т. 94, № 23. – С. 12539–12544.
237. Пат. WO 0109134 (A1) Великобритания, МПК⁷ А61К31/52; А61К31/522; А61К31/5377; А61Р11/00; А61Р11/02; А61Р17/00; А61Р17/02; А61Р17/06; А61Р17/14; А61Р29/00; А61Р37/08; А61Р43/00; А61Р9/14; С07D473/16; С07D473/18; С07D473/24; (IPC1-7): А61К31/52; С07D473/16; С07D473/18; С07D473/24. Purine derivatives inhibitors of tyrosine protein kinase SYK / Collingwood S. P., Hayler J., Le G. D. M., Mattes H., Menear K. A., Walker C. V., Cockcroft X.-L.; заявитель и патентообладатель Novartis AG. – № WO 2000EP07311; заявл. 28.07.2000; опубл. 08.02.2001, – 70 с.: ил.
238. Пат. WO 2005026158 (A1) Китайская Народная Республика, МПК⁷ А61К31/506; С07D239/48; С07D403/12; С07D405/12; С07D405/14; С07D413/14; С07D417/14; (IPC1-7): А61К31/506; С07D239/48; С07D403/12; С07D405/12; С07D405/14; С07D413/14; С07D417/14. 2,4 Di (hetero) -arylamino-pyrimidine derivatives as ZAP-70 and/or SYK inhibitors / Baenteli R., Bernhard M. C., Buehlmayer P., Cooke N. G., Duthaler R., Hinterding K., Thoma G., Van E. M., Von M. A., Walliser L., Zenke G.; заявитель и патентообладатель Novartis AG. – № WO 2004EP10348; заявл. 15.09.2004; опубл. 24.03.2005, –: ил.

239. Пат. WO 2007107469 (A1) Соединенные Штаты Америки, МПК⁷ А61К31/502; А61Р11/06; А61Р19/02; А61Р25/00. Methods of inhibiting BTK and SYK protein kinases / Goldstein D. M., Rueth M.; заявитель и патентообладатель Hoffmann La Roche. – № WO 2007EP52267; заявл. 12.03.2007 опубл. 27.09.2007, – 130 с.: ил.
240. Пат. WO 2007042299 (A1) Великобритания, МПК⁷ А61К31/519; А61Р37/08; C07D487/04. Pyrrolopyrimidine derivatives as SYK inhibitors / Ancliff R. A., Atkinson F. L., Barker M. D., Vox P. C., Daniel C., Gore P. M., Guntrip S. B., Hasegawa M., Inglis G. G. A., Kano K., Miyazaki Y., Patel V. K., Ritchie T. J., Swanson S., Walker A. L., Wellaway C. R., Michael W.; заявитель и патентообладатель Glaxo Group Ltd. – № WO 2006EP09870; заявл. 11.10.2006 опубл. 19.04.2007, – 354 с.: ил.
241. Friedberg J. W., Sharman J., Sweetenham J., Johnston P. B., Vose J. M., Lacasce A., Schaefer-Cuttillo J., De Vos S., Sinha R., Leonard J. P., Cripe L. D., Gregory S. A., Sterba M. P., Lowe A. M., Levy R., Shipp M. A. Inhibition of SYK with fostamatinib disodium has significant clinical activity in non-Hodgkin lymphoma and chronic lymphocytic leukemia // *Blood*. – 2010. – Т. 115, № 13. – С. 2578–2585.
242. Passamonti S., Terdoslavich M., Franca R., Vanzo A., Tramer F., Braidot E., Petrusa E., Vianello A. Bioavailability of Flavonoids: A review of their membrane transport and the function of bilitranslocase in animal and plant organisms // *Current Drug Metabolism*. – 2009. – Т. 10, № 4. – С. 369–394.
243. Weinblatt M. E., Kavanaugh A., Genovese M. C., Musser T. K., Grossbard E. B., Magilavy D. B. An oral spleen tyrosine kinase (SYK) inhibitor for rheumatoid arthritis // *The New England Journal of Medicine*. – 2010. – Т. 363, № 14. – С. 1303–1312.
244. Jorgensen W. L. Free energy calculations: a breakthrough for modeling organic chemistry in solution // *Accounts of Chemical Research*. – 2002. – Т. 22, № 5. – С. 184–189.
245. Kollman P. Free energy calculations: Applications to chemical and biochemical phenomena // *Chemical Reviews*. – 1993. – Т. 93, № 7. – С. 2395–2417.
246. Åqvist J., Medina C., Samuelsson J.-E. A new method for predicting binding affinity in computer-aided drug design // *Protein Engineering, Design and Selection*. – 1994. – Т. 7, № 3. – С. 385–391.

247. Srinivasan J., Cheatham I. T. E., Cieplak P., Kollman P. A., Case D. A. Continuum solvent studies of the stability of DNA, rna, and phosphoramidate–DNA helices // *Journal of the American Chemical Society*. – 1998. – T. 120, № 37. – C. 9401–9409.
248. Novikov F. N., Zeifman A. A., Stroganov O. V., Stroylov V. S., Kulkov V., Chilov G. G. CSAR scoring challenge reveals the need for new concepts in estimating protein-ligand binding affinity // *Journal of Chemical Information and Modeling*. – 2011. – T. 51, № 9. – C. 2090–2096.
249. Zwanzig R. W. High-temperature equation of state by a perturbation method. i. non-polar gases // *The Journal of Chemical Physics*. – 1954. – T. 22, № 8. – C. 1420–1426.
250. Brandsdal B. O., Österberg F., Almlöf M., Feierberg I., Luzhkov V. B., Åqvist J. Free energy calculations and ligand binding // *Advances in Protein Chemistry*. – 2003. – T. 66. – C. 123–158.
251. Jorgensen W. L., Ruiz-Caro J., Tirado-Rives J., Basavapathruni A., Anderson K. S., Hamilton A. D. Computer-aided design of non-nucleoside inhibitors of HIV-1 reverse transcriptase // *Bioorganic & Medicinal Chemistry Letters*. – 2006. – T. 16, № 3. – C. 663–667.
252. Zeevaart J. G., Wang L., Thakur V. V., Leung C. S., Tirado-Rives J., Bailey C. M., Domaal R. A., Anderson K. S., Jorgensen W. L. Optimization of azoles as anti-human immunodeficiency virus agents guided by free-energy calculations // *Journal of the American Chemical Society*. – 2008. – T. 130, № 29. – C. 9492–9499.
253. Mouchlis V. D., Mavromoustakos T. M., Kokotos G. Design of new secreted phospholipase A2 inhibitors based on docking calculations by modifying the pharmacophore segments of the FPL67047XX inhibitor // *Journal of Computer-Aided Molecular Design*. – 2010. – T. 24, № 2. – C. 107–115.
254. Luccarelli J., Michel J., Tirado-Rives J., Jorgensen W. L. Effects of water placement on predictions of binding affinities for P38ALPHA MAP kinase inhibitors // *Journal of Chemical Theory and Computation* – 2010. – T. 6, № 12. – C. 3850–3856.
255. Palma P. N., Bonifacio M. J., Loureiro A. I., Soares-da-Silva P. Computation of the binding affinities of catechol-O-methyltransferase inhibitors: Multisubstate relative free energy calculations // *Journal of Computational Chemistry* – 2012. – T. 33, № 9. – C. 970–986.

256. Lawrenz M., Wereszczynski J., Amaro R., Walker R., Roitberg A., McCammon J. A. Impact of calcium on N1 influenza neuraminidase dynamics and binding free energy // *Proteins*. – 2010. – T. 78, № 11. – C. 2523–2532.
257. Ruiz F., Grigera J. Free energy perturbation calculations on glucosidase-inhibitor complexes // *Medicinal Chemistry*. – 2005. – T. 1, № 5. – C. 455–460.
258. Tirado-Rives J., Jorgensen W. L. Contribution of conformer focusing to the uncertainty in predicting free energies for protein-ligand binding // *Journal of Medicinal Chemistry*. – 2006. – T. 49, № 20. – C. 5880–5884.
259. Perola E., Charifson P. S. Conformational analysis of drug-like molecules bound to proteins: an extensive study of ligand reorganization upon binding // *Journal of Medicinal Chemistry*. – 2004. – T. 47, № 10. – C. 2499–2510.
260. Li H., Fajer M., Yang W. Simulated scaling method for localized enhanced sampling and simultaneous "alchemical" free energy simulations: a general method for molecular mechanical, quantum mechanical, and quantum mechanical/molecular mechanical simulations // *The Journal of Chemical Physics*. – 2007. – T. 126, № 2. – C. 024106.
261. Min D., Li H., Li G., Bitetti-Putzer R., Yang W. Synergistic approach to improve "alchemical" free energy calculation in rugged energy surface // *The Journal of Chemical Physics*. – 2007. – T. 126, № 14. – C. 144109.
262. Jiang W., Roux B. Free energy perturbation hamiltonian replica-exchange molecular dynamics (FEP/H-REMD) for absolute ligand binding free energy calculations // *Journal of Chemical Theory and Computation* – 2010. – T. 6, № 9. – C. 2559–2565.
263. Wang L., Berne B. J., Friesner R. A. On achieving high accuracy and reliability in the calculation of relative protein-ligand binding affinities // *Proceedings of the National Academy of Sciences of the United States of America*. – 2012. – T. 109, № 6. – C. 1937–1942.
264. Michel J., Essex J. W. Hit identification and binding mode predictions by rigorous free energy simulations // *Journal of Medicinal Chemistry*. – 2008. – T. 51, № 21. – C. 6654–6664.
265. Chaudry U. A., Popelier P. L. A. Ester hydrolysis rate constant prediction from quantum topological molecular similarity descriptors // *The Journal of Physical Chemistry A*. – 2003. – T. 107, № 22. – C. 4578–4582.

266. Roy K., Kar S., R.N. D. Statistical Methods in QSAR/QSPR // A Primer on QSAR/QSPR modeling. – Springer, Cham, 2015. – C. 37–59.
267. Estrada E., Matamala A. R. Generalized topological indices. Modeling gas-phase rate coefficients of atmospheric relevance // Journal of Chemical Information and Modeling. – 2007. – T. 47, № 3. – C. 794–804.
268. Struebing H., Ganase Z., Karamertzanis P. G., Siougkrou E., Haycock P., Piccione P. M., Armstrong A., Galindo A., Adjiman C. S. Computer-aided molecular design of solvents for accelerated reaction kinetics // Nature Chemistry. – 2013. – T. 5, № 11. – C. 952–957.
269. Stanescu I., Achenie L. E. K. A theoretical study of solvent effects on Kolbe–Schmitt reaction kinetics // Chemical Engineering Science. – 2006. – T. 61, № 18. – C. 6199–6212.
270. Dev V. A., Chemmangattuvalappil N. G., Eden M. R. Designing reactants and products with properties dependent on both structures // Computer aided chemical engineering. – 2015. – T. 37. – C. 1445–1450.
271. Nandi S., Monesi A., Drgan V., Merzel F., Novic M. Quantitative structure-activation barrier relationship modeling for Diels-Alder ligations utilizing quantum chemical structural descriptors // Chemistry Central Journal. – 2013. – T. 7, № 1. – C. 171.
272. Zhou T., McBride K., Zhang X., Qi Z., Sundmacher K. Integrated solvent and process design exemplified for a Diels-Alder reaction // AIChE Journal. – 2015. – T. 61, № 1. – C. 147–158.
273. Evans D. A., Johnson J. S. Diels–Alder reactions // Comprehensive asymmetric catalysis I–III / Под ред. Jacobsen E. N. и др. – Berlin: Springer-Verlag Berlin Heidelberg. – C. 1178–1235.
274. Rideout D. C., Breslow R. Hydrophobic acceleration of Diels-Alder reactions // Journal of the American Chemical Society. – 1980. – T. 102, № 26. – C. 7816–7817.
275. The molecular modeling workbook for organic chemistry. / Hehre W. J., Shusterman A. J., Nelson J. E. – USA: Wavefunction Inc., 1998. – 307 с.
276. Cramer C. J. Essentials of computational chemistry: Theories and models. – Бромпое изд. – Wiley, 2004. – 618 с.
277. Anslyn E. V., Dougherty D. A. Modern physical organic chemistry // Book Modern physical organic chemistry. – Herndon, VA: University Science, 2005. – C. 1104.

278. Jayamurugan G., Gisselbrecht J. P., Boudon C., Schoenebeck F., Schweizer W. B., Bernet B., Diederich F. Expanding the chemical space for push-pull chromophores by non-concerted [2+2] and [4+2] cycloadditions: access to a highly functionalised 6,6-dicyanopentafulvene with an intense, low-energy charge-transfer band // *Chemical Communication (Cambridge)*. – 2011. – T. 47, № 15. – C. 4520–4522.
279. Yamada M., Schweizer W. B., Schoenebeck F., Diederich F. Unprecedented thermal rearrangement of push-pull-chromophore-[60]fullerene conjugates: formation of chiral 1,2,9,12-tetrakis-adducts // *Chemical Communication (Cambridge)*. – 2010. – T. 46, № 29. – C. 5334–5336.
280. Chiu M., Tchitchanov B. H., Zimmerli D., Sanhueza I. A., Schoenebeck F., Trapp N., Schweizer W. B., Diederich F. Strain-accelerated formation of chiral, optically active buta-1,3-dienes // *Angewandte Chemie International Edition* –2015. – T. 54, № 1. – C. 349–354.
281. Zhou L., Li Z., Zou Y., Wang Q., Sanhueza I. A., Schoenebeck F., Goeke A. Tandem nucleophilic addition/oxy-2-azonia-Cope rearrangement for the formation of homoallylic amides and lactams: total synthesis and structural verification of motuporamine G // *Journal of the American Chemical Society*. – 2012. – T. 134, № 49. – C. 20009–20012.
282. Zou Y., Ding C., Zhou L., Li Z., Wang Q., Schoenebeck F., Goeke A. Tandem cross-dimerisation/oxonia-cope reaction of carbonyl compounds to homoallylic esters and lactones // *Angewandte Chemie International Edition* –2012. – T. 51, № 23. – C. 5647–51.
283. Diethelm S., Schoenebeck F., Carreira E. M. Mechanistic insight into the spirocyclopropane isoxazolidine ring contraction // *Organic Letters* –2014. – T. 16, № 3. – C. 960–963.
284. Hendriks C. M., Penning T. M., Zang T., Wiemuth D., Grunder S., Sanhueza I. A., Schoenebeck F., Bolm C. Pentafluorosulfanyl-containing flufenamic acid analogs: Syntheses, properties and biological activities // *Bioorganic & Medicinal Chemistry Letters*. – 2015. – T. 25, № 20. – C. 4437–4440.
285. Sperger T., Sanhueza I. A., Schoenebeck F. Computation and Experiment: A Powerful Combination to Understand and Predict Reactivities // *Accounts of Chemical Research*. – 2016. – T. 49, № 6. – C. 1311–1319.

286. Becke A. D. Density-functional thermochemistry. III. The role of exact exchange // *The Journal of Chemical Physics*. – 1993. – T. 98, № 7. – C. 5648–5652.
287. Zhao Y., Truhlar D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals // *Theoretical Chemistry Accounts*. – 2007. – T. 120, № 1–3. – C. 215–241.
288. Barone V., Cossi M. Quantum calculation of molecular energies and energy gradients in solution by a conductor solvent model // *The Journal of Physical Chemistry A*. – 1998. – T. 102, № 11. – C. 1995–2001.
289. Cossi M., Rega N., Scalmani G., Barone V. Energies, structures, and electronic properties of molecules in solution with the C-PCM solvation model // *Journal of Computational Chemistry*. – 2003. – T. 24, № 6. – C. 669–681.
290. Marenich A. V., Cramer C. J., Truhlar D. G. Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions // *The Journal of Physical Chemistry B*. – 2009. – T. 113, № 18. – C. 6378–6396.
291. Fage C. D., Isiorho E. A., Liu Y., Wagner D. T., Liu H. W., Keatinge-Clay A. T. The structure of SpnF, a standalone enzyme that catalyzes [4 + 2] cycloaddition // *Nature Chemical Biology*. – 2015. – T. 11, № 4. – C. 256–258.
292. Kim H. J., Rusczycky M. W., Liu H. W. Current developments and challenges in the search for a naturally selected Diels-Alderase // *Current Opinion in Chemical Biology*. – 2012. – T. 16, № 1–2. – C. 124–131.
293. Kirst H. A. The spinosyn family of insecticides: Realizing the potential of natural products research // *The Journal of Antibiotics (Tokyo)*. – 2010. – T. 63, № 3. – C. 101–111.
294. Preiswerk N., Beck T., Schulz J. D., Milovnik P., Mayer C., Siegel J. B., Baker D., Hilvert D. Impact of scaffold rigidity on the design and evolution of an artificial Diels-Alderase // *Proceedings of the National Academy of Sciences of the United States of America*. – 2014. – T. 111, № 22. – C. 8013–8018.

295. Siegel J. B., Zanghellini A., Lovick H. M., Kiss G., Lambert A. R., St Clair J. L., Gallaher J. L., Hilvert D., Gelb M. H., Stoddard B. L., Houk K. N., Michael F. E., Baker D. Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction // *Science*. – 2010. – T. 329, № 5989. – C. 309–313.
296. Hilvert D. Design of protein catalysts // *Annual Review of Biochemistry* –2013. – T. 82. – C. 447–470.
297. Kiss G., Celebi-Olcum N., Moretti R., Baker D., Houk K. N. Computational enzyme design // *Angewandte Chemie International Edition* –2013. – T. 52, № 22. – C. 5700–5725.
298. Yano B. L. Spinosad insecticide: subchronic and chronic toxicity and lack of carcinogenicity in Fischer 344 rats // *Toxicological Sciences*. – 2002. – T. 65, № 2. – C. 288–298.
299. Kim H. J., Ruzsyczky M. W., Choi S. H., Liu Y. N., Liu H. W. Enzyme-catalysed [4+2] cycloaddition is a key step in the biosynthesis of spinosyn A // *Nature*. – 2011. – T. 473, № 7345. – C. 109–112.
300. Gordeev E. G., Ananikov V. P. Computational study of a model system of enzyme-mediated [4+2] cycloaddition reaction // *PLoS One*. – 2015. – T. 10, № 4. – C. e0119984.
301. Hess B. A., Jr., Smentek L. Concerted, highly asynchronous, enzyme-catalyzed [4 + 2] cycloaddition in the biosynthesis of spinosyn A; computational evidence // *Organic & Biomolecular Chemistry*. – 2012. – T. 10, № 37. – C. 7503–7509.
302. Patel A., Chen Z., Yang Z., Gutierrez O., Liu H. W., Houk K. N., Singleton D. A. Dynamically complex [6+4] and [4+2] cycloadditions in the biosynthesis of Spinosyn A // *Journal of the American Chemical Society*. – 2016. – T. 138, № 11. – C. 3631–3634.
303. Reed A. E., Weinstock R. B., Weinhold F. Natural population analysis // *The Journal of Chemical Physics*. – 1985. – T. 83, № 2. – C. 735–746.
304. Medvedev M. G., Zeifman A. A., Novikov F. N., Bushmarinov I. S., Stroganov O. V., Titov I. Y., Chilov G. G., Svitanko I. V. Quantifying possible routes for SpnF-catalyzed formal Diels-Alder cycloaddition // *Journal of the American Chemical Society*. – 2017. – T. 139, № 11. – C. 3942–3945.
305. Baskin I.I., Stankevich M.I., Devdariani R.O., Zefirov N.S. A program set for finding structure-property correlations using topological indices // *Journal of Structural Chemistry*. – 1989 – T. 30, № 6, – C. 995–997.

306. de Mello Castanho Amboni R. D., da Silva Junkes B., Yunes R. A., Heinzen V. E. F. Quantitative structure–odor relationships of aliphatic esters using topological indices // *Journal of Agricultural and Food Chemistry*. – 2000. – Т. 48, № 8. – С. 3517–3521.
307. Станкевич М. И., Станкевич И. В., Зефирова Н. С. Топологические индексы в органической химии // *Успехи химии*. – 1988. – Т. 57, № 3. – С. 337–366.
308. Gorbachev M. Y., Dimoglo A. S., Bersuker I. B., Vlad P. F., Koltsa M. N. Role of electronic and structural factors in the origin of ambergris-type odors in cyclic compounds // *Theoretical and Experimental Chemistry*. – 1986. – Т. 22, № 3. – С. 340–344.
309. Водородная связь. / Пиментел Д., Мак-Клеллан О. – М.: "Мир", 1964.
310. Логико-комбинаторные методы конструирования лекарств. / Розенблит А. Б., Голендер В. Е. – Рига: "Знание", 1983.
311. DesJarlais R. L., Sheridan R. P., Dixon J. S., Kuntz I. D., Venkataraghavan R. Docking flexible ligands to macromolecular receptors by molecular shape // *Journal of Medicinal Chemistry*. – 1986. – Т. 29, № 11. – С. 2149–2153.
312. Colloc'h N., Mornon J.-P. A new tool for the qualitative and quantitative analysis of protein surfaces using B-spline and density of surface neighborhood // *Journal of Molecular Graphics*. – 1990. – Т. 8, № 3. – С. 133–140.
313. Chothia C., Janin J. Principles of protein–protein recognition // *Nature*. – 1975. – Т. 256, № 5520. – С. 705–708.
314. Wodak S. J., Janin J. Computer analysis of protein–protein interaction // *Journal of Molecular Biology*. – 1978. – Т. 124, № 2. – С. 323–342.
315. Kuntz I. D., Blaney J. M., Oatley S. J., Langridge R., Ferrin T. E. A geometric approach to macromolecule–ligand interactions // *Journal of Molecular Biology*. – 1982. – Т. 161, № 2. – С. 269–288.
316. Hadkin E. E., Richards W. G. Molecular similarity based on electrostatic potential and electric fields // *International journal of quantum chemistry. Quantum biology symposium: Proceedings of the international symposium on quantum biology and quantum pharmacology*. – 1987. – Т. 14. – С. 105–110.

317. Sjoberg P., Murray J. S., Brinck T., Evans P., Politzer P. The use of the electrostatic potential at the molecular surface in recognition interactions: Dibenzo-p-dioxins and related systems // *Journal of Molecular Graphics*. – 1990. – Т. 8, № 2. – С. 81–85.
318. Richard A. M. Quantitative comparison of molecular electrostatic potentials for structure–activity studies // *Journal of Computational Chemistry*. – 1991. – Т. 12, № 8. – С. 959–969.
319. Crippen G. M. Distance geometry approach to rationalizing binding data // *Journal of Medicinal Chemistry*. – 1979. – Т. 22, № 8. – С. 988–997.
320. Кумсков М. И., Пономарева Л. А., Смоленский Е. А., Митюшев Д. Ф., Зефирова Н. С. Метод автоматического формирования структурных дескрипторов органических соединений для количественных корреляций "структура–свойство" // *Известия Академии наук. Серия химическая*. – 1994. – Т. 8. – С. 1391–1393.
321. Кумсков М. И., Смоленский Е. А., Пономарева Л. А., Митюшев Д. Ф., Зефирова Н. С. Системы структурных дескрипторов для решения задач "структура–свойство" // *Доклады АН*. – 1994. – Т. 336, № 1. – С. 64–68.
322. Sanner M. F., Olson A. J., Spehner J.-C. Fast and robust computation of molecular surfaces // *SCG '95. The eleventh annual symposium on Computational geometry – Vancouver, British Columbia, Canada, 05–07 June 1995* – С. 406–407
323. Connolly M. L. Shape complementarity at the hemoglobin alpha 1 beta 1 subunit interface // *Biopolymers*. – 1986. – Т. 25, № 7. – С. 1229–1247
324. Korostelev A. P., Promakhina I. M. Efficiency of learning information employment in local algorithms of pattern recognition // *Pattern recognition and image analysis (advances in mathematical theory and applications)* – 1996. – Т. 8, № 4. – С. 497–500.
325. Душистые вещества и другие продукты для парфюмерии. / Хейфиц Л. А., Дашунин В. М. – Москва: Химия, 1994. – 256 с.
326. А. С. 1426073 СССР, МПК³ С 11 В 9/00. Способ получения душистого вещества с амбровым запахом / Свитанько И. В., Ситнова Л. М., Скворцова А. Б., Бельфер А. Г., Никитина М. А., Новиков Н. А., Элинсон М. Н., Махова И. В., Никишин Г. И.; заявитель и патентообладатель Всесоюзный научно-исследовательский институт синтетических и натуральных душистых веществ и Институт органической химии АН СССР. – опубл. 11.02.87 – с.: ил.

327. А. С. 1648092 СССР, МПК³ С 25 В 3/10, С 07 С 43/184. Способ получения циклоалкиловых эфиров / Элинсон М. Н., Махова И. В., Свитанько И. В., Никишин Г. И.; заявитель и патентообладатель Институт органической химии им. Н.Д. Зелинского. – опубл. 18.04.89, – с.: ил.
328. Svitan'ko I. V., Kumskov M. I., Zyryanov I. L., Suslov I. A. A Method for describing the molecular electrostatic potential in determining structure–activity relationship // *Mendeleev Communications*. – 1994. – Т. 4, № 5. – С. 161–162.
329. Svitan'ko I. V., Zyryanov I. L., Kumskov M. I., Khmel'nitskii L. I., Suvorova L. I., Kravchenko A. N., Markova T. B., Lebedev O. V., Orekhova G. A., Belova S. V. The surface molecular potential method for calculating the structure–activity relationship for psychotropic compounds // *Mendeleev Communications*. – 1995. – Т. 5, № 2. – С. 49–51.
330. Triller A., Boulden E. A., Churchill A., Hatt H., Englund J., Spehr M., Sell C. S. Odorant-receptor interactions and odor percept: a chemical perspective // *Chemistry & Biodiversity*. – 2008. – Т. 5, № 6. – С. 862–86.
331. Лужков В. Б., Богданов Г. Н. Квантовохимические расчеты в изучении противоопухолевых соединений // *Успехи химии*. – 1986. – Т. 55, № 1. – С. 3–28.
332. Станкевич М. И., Станкевич И. В., Зефиоров Н. С. Топологические индексы в органической химии // *Успехи химии*. – 1988. – Т. 57, № 3. – С. 337–366.
333. Martin Y. C. 3D database searching in drug design // *Journal of Medicinal Chemistry*. – 1992. – Т. 35, № 12. – С. 2145–2154.
334. Crippen G. M. Distance geometry approach to rationalizing binding data // *Journal of Medicinal Chemistry*. – 1979. – Т. 22, № 8. – С. 988–997.
335. Crippen G. M. Quantitative structure-activity relationships by distance geometry: systematic analysis of dihydrofolate reductase inhibitors // *Journal of Medicinal Chemistry*. – 1980. – Т. 23, № 6. – С. 599–606.
336. Richards F. M. Areas, volumes, packing and protein structure // *Annual Review of Biophysics and Bioengineering*. – 1977. – Т. 6. – С. 151–176.
337. Пономарева Л. А., Кумсков М. И., Зефиоров Н. С. Формирование моделей ККСА на основе большого числа структурных дескрипторов. (Метод группового учета аргументов) // Тезисы докл. 1-ой Всесоюзной конф. по теоретич. органической химии - ВАТОХ – Волгоград, Россия, 1991. – С. 545.

338. Ересько В. А., Епишина Л. В., Лебедев О. В., Повстяной М. В., Хмельницкий Л. И., Новиков С. С. Исследование в области химии бициклических мочевинов. Сообщение 5. Синтез 2,4,6,8- тетраазабицикло[3.3.0]октанон-3-тионов-7. // Изв. АН СССР. Сер. хим. – 1980. – № 7. – С. 1594–1597.
339. Svitank'o I. V., Kumskov M. I., Zyryanov I. L., Suslov I. A. A Method for Describing the Molecular Electrostatic Potential in Determining Structure–Activity Relationship // *Mendeleev Communications*. – 1994. – Т. 4, № 5. – С. 161–162.
340. Кумсков М. И., Пономарева Л. А., Смоленский Е. А., Митюшев Д. Ф., Зефирова Н. С. Метод автоматического формирования структурных дескрипторов органических соединений для количественных корреляций "структура–свойство" // *Известия Академии наук. Серия химическая*. – 1994. – Т. 8. – С. 1391–1393.
341. Молекулярная механика. / Буркерт У., Н. Э. – М.: "Мир", 1986. – 364 с.
342. Stewart J. J. P. Optimization of parameters for semiempirical methods II. Applications // *Journal of Computational Chemistry*. – 1989. – Т. 10, № 2. – С. 221–264.
343. Rakitina T. V., Zeifman A. A., Titov I. Y., Svitanko I. V., Lipkin A. V., Stroylov V. S., Stroganov O. V., Novikov F. N., Chilov G. G. Efficacy of novel Syk-Kinase Inhibitors MT-Syk-03 and MT-Syk-322 in Cellular Models of Autoimmunity and Cancer // *Mendeleev Communications*. – 2012. – Т. 22, № 6. – С. 287–289.
344. Zeifman A. A., Titov I. Y., Svitanko I. V., Rakitina T. V., Lipkin A. V., Stroylov V. S., Stroganov O. V., Novikov F. N., Chilov G. G. Rational design and synthesis of novel Syk-kinase inhibitors // *Mendeleev Communications*. – 2012. – Т. 22, № 2. – С. 73–74.
345. Sukhorukov A. Y., Andryushkevich S. O., Chilov G. G., Zeifman A. A., Svitanko I. V., Ioffe S. L. The first synthesis and molecular docking studies of diastereomerically pure substituted 4-amino-7-hydroxyheptanoic acids // *Mendeleev Communications*. – 2011. – Т. 21, № 4. – С. 183–185.
346. Romashov L. V., Zeifman A. A., Zakharenko A. L., Novikov F. N., Stroilov V. S., Stroganov O. V., Chilov G. G., Khodyreva S. N., Lavrik O. I., Titov I. Y., Svitank'o I. V. Rational design and synthesis of new PARP1 inhibitors // *Mendeleev Communications*. – 2012. – Т. 22, № 1. – С. 15–17.

347. Prokhorov E. I., Ponomareva L. A., Permyakov E. A., Kumskov M. I. Fuzzy classification and fast rules for refusal in the QSAR problem // *Pattern Recognition and Image Analysis*. – 2011. – Т. 21, № 3. – С. 542–544.
348. Svitanko I. V., Devetyarov D. A., Tcheboukov D. E., Dolmat M. S., Zakharov A. M., Grigor'eva S. S., Chichua V. T., Ponomareva L. A., Kumskov M. I. QSAR modeling on the basis of 3D descriptors representing the electrostatic molecular surface (ambergris fragrances) // *Mendeleev Communications*. – 2007. – Т. 17, № 2. – С. 90–91.
349. Gazieva G., Poluboyarov P., Popov L., Kolotyrkina N., Kravchenko A., Makhova N. A Novel synthesis of thioglycolurils by ring contraction of 5,7-dialkyl-3-thioxoperhydroimidazo[4,5-e]-1,2,4-triazin-6-ones // *Synthesis*. – 2012. – Т. 44, № 21. – С. 3366–3370.
350. Лекарственные средства. / Машковский М. Д.; Под ред. Дивов С. Б. – 14-е изд. – М.: ООО «Издательство Новая волна», 202. – 540 с.
351. Smith R. D., Dunbar J. B., Jr., Ung P. M., Esposito E. X., Yang C. Y., Wang S., Carlson H. A. CSAR benchmark exercise of 2010: Combined evaluation across all submitted scoring functions // *Journal of Chemical Information and Modeling*. – 2011. – Т. 51, № 9. – С. 2115–2131.
352. d'Amours D., Desnoyers S., d'Silva I., Poirier G. G. Poly(ADP-ribosyl)ation reactions in the regulation of nuclear functions // *Biochemical Journal*. – 1999. – Т. 342, № 2. – С. 249–268.
353. Hassa P. O., Haenni S. S., Elser M., Hottiger M. O. Nuclear ADP-ribosylation reactions in mammalian cells: where are we today and where are we going? // *Microbiology and Molecular Biology Reviews*. – 2006. – Т. 70, № 3. – С. 789–829.
354. Hassa P. O. The diverse biological roles of mammalian PARPS, a small but powerful family of poly-ADP-ribose polymerases // *Frontiers in Bioscience*. – 2008. – Т. 13, № 13. – С. 3046.
355. Pacher P., Szabo C. Role of poly(ADP-ribose) polymerase 1 (PARP-1) in cardiovascular diseases: the therapeutic potential of PARP inhibitors // *Cardiovascular Drug Reviews* – 2007. – Т. 25, № 3. – С. 235–260.

356. Calvert H., Azzariti A. The clinical development of inhibitors of poly(ADP-ribose) polymerase // *Annals of Oncology*. – 2011. – T. 22 Suppl 1. – C. i53–i59.
357. Novikov F. N., Chilov G. G. Molecular docking: Theoretical background, practical applications and perspectives // *Mendeleev Communications*. – 2009. – T. 19, № 5. – C. 237–242.
358. Stroylov V. S., Rakitina T. V., Novikov F. N., Stroganov O. V., Chilova G. G., Lipkin A. V. Novel fragment-like inhibitors of EphA2 obtained by experimental screening and modelling // *Mendeleev Communications*. – 2010. – T. 20, № 5. – C. 263–265.
359. Novikov F. N., Stroylov V. S., Stroganov O. V., Kulkov V., Chilov G. G. Developing novel approaches to improve binding energy estimation and virtual screening: a PARP case study // *Journal of Molecular Modeling* – 2009. – T. 15, № 11. – C. 1337–1347.
360. Luccarelli J., Michel J., Tirado-Rives J., Jorgensen W. L. Effects of water placement on predictions of binding affinities for p38alpha MAP kinase inhibitors // *Journal of Chemical Theory and Computation*. – 2010. – T. 6, № 12. – C. 3850–3856.
361. Gewald K., Schinke E., Böttcher H. Heterocyclen aus CH-aciden Nitrilen, VIII. 2-Amino-thiophene aus methylenaktiven Nitrilen, Carbonylverbindungen und Schwefel // *Chemische Berichte*. – 1966. – T. 99, № 1. – C. 94–100.
362. Baker B. R., Schaub R. E., Joseph J. P., McEvoy F. J., Williams J. H. An antimalarial alkaloid from hydrangea. XV. Synthesis of 5-, 6-, 7-, and 8-derivatives with two identical substituents // *The Journal of Organic Chemistry*. – 1952. – T. 17, № 1. – C. 149–156.
363. Kanwar S., Sharma S. D. Thienopyrimidines as hetryl moiety in 2-azetidiones: Synthesis of 4-hetryl-2-azetidiones // *Indian Journal of Chemistry. Section B*. – 2005. – T. 44B, № 11. – C. 2367–2371.
364. Shishoo C. J., Jain K. S., Rathod I. S., Thakkar B. J., Brahmhatt S. B., Gandhi T. P., Bangaru R., Goyal R. K. // *Arzneimittel Forschung. Drug Research*. – 1996. – T. 46. – C. 273.
365. Ameen M. A., Karsten S., Liebscher J. A convenient method for constructing novel tetrahydropyrido[4',3':4,5]thieno[2,3-d]-pyrimidinones-carbohydrate and amino acid conjugates via copper(I)-catalyzed alkyne-azide 'Click Chemistry' // *Tetrahedron*. – 2010. – T. 66, № 12. – C. 2141–2147.

366. Пат. WO 03103575 (A2) Соединенные Штаты Америки, МПК7 А61К31/519; А61Р1/04; А61Р17/02; А61Р17/06; А61Р19/02; А61Р29/00; А61Р35/00; А61Р35/02; А61Р37/00; А61Р37/06; А61Р41/00; А61Р43/00; А61Р9/00; А61Р9/10; В42Д15/00; С07Д471/04; С07Д475/02; С07Д475/04; С07Д487/04; (IPC1-7): А61К. Compounds, compositions, and methods / Dhanak D., Knight S. D., Lu P.-P., Morgans D. J. J., Yao B.; заявитель и патентообладатель Cytokinetics Inc. – № WO 2003US16500; заявл. 22.05.2003; опубл. 18.12.2003, – 89 с.: ил.
367. Banasik M., Komura H., Shimoyama M., Ueda K. Specific inhibitors of poly(ADP-ribose) synthetase and mono(ADP-ribosyl)transferase. // *Journal of Biological Chemistry*. – 1992. – Т. 267. – С. 1569–1575.
368. White A. W., Almassy R., Calvert A. H., Curtin N. J., Griffin R. J., Hostomsky Z., Maegley K., Newell D. R., Srinivasan S., Golding B. T. Resistance-modifying agents. 9.1 synthesis and biological properties of benzimidazole inhibitors of the DNA repair enzyme *poly(ADP-ribose) polymerase* // *Journal of Medicinal Chemistry*. – 2000. – Т. 43, № 22. – С. 4084–4097.
369. Donawho C. K., Luo Y., Luo Y., Penning T. D., Bauch J. L., Bouska J. J., Bontcheva-Diaz V. D., Cox B. F., DeWeese T. L., Dillehay L. E., Ferguson D. C., Ghoreishi-Haack N. S., Grimm D. R., Guan R., Han E. K., Holley-Shanks R. R., Hristov B., Idler K. B., Jarvis K., Johnson E. F., Kleinberg L. R., Klinghofer V., Lasko L. M., Liu X., Marsh K. C., McGonigal T. P., Meulbroek J. A., Olson A. M., Palma J. P., Rodriguez L. E., Shi Y., Stavropoulos J. A., Tsurutani A. C., Zhu G. D., Rosenberg S. H., Giranda V. L., Frost D. J. ABT-888, an orally active poly(ADP-ribose) polymerase inhibitor that potentiates DNA-damaging agents in preclinical tumor models // *Clinical Cancer Research*. – 2007. – Т. 13, № 9. – С. 2728–2737.
370. Rossi A. B., Herlaar E., Braselmann S., Huynh S., Taylor V., Frances R., Issakani S. D., Argade A., Singh R., Payan D. G., Masuda E. S. Identification of the SYK kinase inhibitor R112 by a human mast cell screen // *Journal of Allergy and Clinical Immunology*. – 2006. – Т. 118, № 3. – С. 749–755.
371. Ruzza P., Biondi B., Calderan A. Therapeutic prospect of Syk inhibitors // *Expert Opinion on Therapeutic Patents*. – 2009. – Т. 19, № 10. – С. 1361–1376.

372. Weinblatt M. E., Kavanaugh A., Burgos-Vargas R., Dikranian A. H., Medrano-Ramirez G., Morales-Torres J. L., Murphy F. T., Musser T. K., Straniero N., Vicente-Gonzales A. V., Grossbard E. Treatment of rheumatoid arthritis with a Syk kinase inhibitor: a twelve-week, randomized, placebo-controlled trial // *Arthritis & Rheumatology* –2008. – Т. 58, № 11. – С. 3309–3318.
373. Weinblatt M. E., Kavanaugh A., Genovese M. C., Musser T. K., Grossbard E. B., Magilavy D. B. An oral spleen tyrosine kinase (SYK) inhibitor for rheumatoid arthritis // *The New England Journal of Medicine*. – 2010. – Т. 363, № 14. – С. 1303–1312.
374. Weinblatt M. E., Kavanaugh A., Genovese M. C., Musser T. K., Grossbard E. B., Magilavy D. B. An oral spleen tyrosine kinase (SYK) inhibitor for rheumatoid arthritis // *The New England Journal of Medicine*. – 2010. – Т. 363, № 14. – С. 1303–1312.
375. Lee D. M., Weinblatt M. E. Rheumatoid arthritis // *The Lancet*. – 2001. – Т. 358, № 9285. – С. 903–911.
376. Young R. M., Hardy I. R., Clarke R. L., Lundy N., Pine P., Turner B. C., Potter T. A., Refaeli Y. Mouse models of non-Hodgkin lymphoma reveal SYK as an important therapeutic target // *Blood*. – 2009. – Т. 113, № 11. – С. 2508–2516.
377. Stahura F., Bajorath J. Virtual screening methods that complement HTS // *Combinatorial Chemistry & High Throughput Screening*. – 2004. – Т. 7, № 4. – С. 259–269.
378. Schneider G., Böhm H.-J. Virtual screening and fast automated docking methods // *Drug Discovery Today*. – 2002. – Т. 7, № 1. – С. 64–70.
379. Пат. WO 2008064274 (A1) Соединенные Штаты Америки, МПК⁷ А61К31/5383; А61К31/675; А61P37/00; C07D498/04; C07F9/6561. Prodrug salts of 2,4-pyrimidinediamine compounds and their uses / Bhamidipati S., Singh R., Sun T., Masuda E.; заявитель и патентообладатель Rigel Pharmaceuticals Inc. – № WO 2007US85313; заявл. 20.11.2007; опубл. 29.05.2008, – 139 с.: ил.
380. Novikov F. N., Stroylov V. S., Stroganov O. V., Chilov G. G. Improving performance of docking-based virtual screening by structural filtration // *Journal of Molecular Modeling*. – 2010. – Т. 16, № 7. – С. 1223–1230.
381. Riccaboni M., Bianchi I., Petrillo P. Spleen tyrosine kinases: biology, therapeutic targets and drugs // *Drug Discovery Today* – 2010. – Т. 15, № 13–14. – С. 517–530.

382. Braselmann S., Taylor V., Zhao H., Wang S., Sylvain C., Baluom M., Qu K., Herlaar E., Lau A., Young C., Wong B. R., Lovell S., Sun T., Park G., Argade A., Jurcevic S., Pine P., Singh R., Grossbard E. B., Payan D. G., Masuda E. S. R406, an orally available spleen tyrosine kinase inhibitor blocks fc receptor signaling and reduces immune complex-mediated inflammation // *Journal of Pharmacology and Experimental Therapeutics*. – 2006. – Т. 319, № 3. – С. 998–1008.
383. Novikov F. N., Stroylov V. S., Zeifman A. A., Stroganov O. V., Kulkov V., Chilov G. G. Lead Finder docking and virtual screening evaluation with Astex and DUD test sets // *Journal of Computer-Aided Molecular Design*. – 2012. – Т. 26, № 6. – С. 725–735.
384. Zeifman A. A., Stroylov V. S., Novikov F. N., Stroganov O. V., Zakharenko A. L., Khodyreva S. N., Lavrik O. I., Chilov G. G. Hit clustering can improve virtual fragment screening: CDK2 and PARP1 case studies // *Journal of Molecular Modeling*. – 2012. – Т. 18, № 6. – С. 2553–2566.
385. Нефедов О. М., Трепалин С. В., Королева Л. М., Бессонов Ю. Е. Быстрый поиск точных химических структур в больших базах данных с использованием InChI Key кодировки структур // *Научно-техническая информация. Серия 2: Информационные процессы и системы* –2013. – № 12. – С. 27–33.
386. Zeifman A. A., Novikov F. N., Stroylov V. S., Stroganov O. V., Chilov G. G., Skoblov A. Y., Miroshnikov A. I., Skoblov Y. S. 2,3-Dihydroxy-quinoxaline induces ATPase activity of Herpes Simplex Virus thymidine kinase // *FEBS Letters*. – 2014. – Т. 588, № 3. – С. 509–511.
387. Stroganov O. V., Novikov F. N., Zeifman A. A., Stroylov V. S., Chilov G. G. TSAR, a new graph-theoretical approach to computational modeling of protein side-chain flexibility: modeling of ionization properties of proteins // *Proteins*. – 2011. – Т. 79, № 9. – С. 2693–2710.
388. Zeifman A. A., Stroylov V. V., Novikov F. N., Stroganov O. V., Kulkov V., Chilov G. G. Alchemical FEP calculations of ligand conformer focusing in explicit solvent // *Journal of Chemical Theory and Computation*. – 2013. – Т. 9, № 2. – С. 1093–1102.
389. Bliss C. I. The toxicity of poisons applied jointly // *Annals of Applied Biology*. – 1939. – Т. 26, № 3. – С. 585–615.

390. Beuming, et al.; Thermodynamic analysis of water molecules at the surface of proteins and applications to binding site prediction and characterization; *Proteins: Structure, Function, and Bioinformatics*; 2012; 80; 3; 871-883
391. Krivov, et al.; Improved prediction of protein side-chain conformations with SCWRL4; *Proteins*; 2009; 77; 4; 778-95
392. Amin, Anwar, Kamel, Kassem, Syam and Elseginy; Synthesis, cytotoxic evaluation and molecular docking study of novel quinazoline derivatives as PARP-1 inhibitors; *Acta Pol Pharm*; 2013; 70; 5; 833-49
393. Новиков, Федор Николаевич. Алгоритм расчета конформационно-зависимых свойств белков для моделирования их координации с химическими соединениями : дис. ... канд. хим. наук : 02.00.04, 02.00.15: защищена 17.05.2011 : / Новиков, Федор Николаевич – М., 2011. – 225 с. – Библиогр.: с. 218–225.
394. Журавлёв Ю. И. Об алгебраическом подходе к решению задач распознавания или классификации // *Проблемы кибернетики*. – 1978. – Т. 33. – С. 5–68.
395. Воронцов К. В., Каневский Д. Ю. Козволюционный метод обучения алгоритмических композиций // *Таврический вестник информатики и математики*. – 2005. № 2. – С. 51–66.
396. Коллективные правила распознавания. / Растринин Л. А., Эренштейн Р. Х. – М.: Энергия, 1981.
397. Jacobs R. A., Jordan M. I., Nowlan S. J., Hinton G. E. Adaptive mixtures of local experts // *Neural Computation*. – 1991. – Т. 3, № 1. – С. 79-87.
398. Применение кернел-методов в задаче «структура – свойство». Прогнозирование свойств и биологической активности химических соединений. / А.В. Б. – Саарбрюккен, Германия: LAP Lambert Academic Publishing, 2012. – 96 с.
399. Прохоров Е. И., Беккер А. В., Перевозников А. В., Свитанько И. В., Захаренко А. Л., Суханова М. В., Кумсков М. И. Приложения метода эволюционного отбора дескрипторов в математическом моделировании зависимости биологической активности соединения от его структуры –Изд-во МАКС Пресс Москва, 2012. – С. 3–24.
400. Perevoznikov A. V., Shestov A. M., Permyakov E. A., Kumskov M. I. A way to increase the prediction quality for the large set of molecular graphs by using the k-NN classifier // *Pattern Recognition and Image Analysis*. – 2011. – Т. 21, № 3. – С. 545-548.

401. Прохоров Е. И. Нейронные сети для построения ограничений допустимости в задаче «структура – свойство» // *Нейрокомпьютеры: Разработка, применение* –2012. – Т. 10. – С. 46–56.
402. Tan, M., Chao, L., Pan, S. Yang L. Combined QSAR and molecule docking studies on predicting P-glycoprotein inhibitors // *Journal of Computer-Aided Molecular Design*. – 2013. – Т. 27, № 12. – С. 1067–1073
403. Blaser H.-U. Industrial asymmetric catalysis: Approaches and results // *Rendiconti Lincei*. – 2007. – Т. 18, № 4. – С. 281–304.
404. Hawkins J. M., Watson T. J. Asymmetric catalysis in the pharmaceutical industry // *Angewandte Chemie International Edition in English*. – 2004. – Т. 43, № 25. – С. 3224–3228.
405. Ananikov V. P., Khokhlova E. A., Egorov M. P., Sakharov A. M., Zlotin S. G., Kucherov A. V., Kustov L. M., Gening M. L., Nifantiev N. E. Organic and hybrid molecular systems // *Mendeleev Communications*. – 2015. – Т. 25, № 2. – С. 75–82.
406. North M., Villuendas P. A chiral solvent effect in asymmetric organocatalysis // *Organic Letters*. – 2010. – Т. 12, № 10. – С. 2378–2381.
407. Kochetkov S. V., Kucherenko A. S., Zlotin S. G. Asymmetric aldol reactions in ketone/ketone systems catalyzed by ionic liquid-supported C_2 -symmetrical organocatalyst // *Mendeleev Communications*. – 2015. – Т. 25, № 3. – С. 168–170.
408. Schulz P. S., Müller N., Bösmann A., Wasserscheid P. Effective chirality transfer in ionic liquids through ion-pairing effects // *Angewandte Chemie*. – 2007. – Т. 119, № 8. – С. 1315–1317.
409. Wagner V., Schulz P. S., Wasserscheid P. Asymmetric hydrogenation catalysis via ion-pairing in chiral ionic liquids // *Journal of Molecular Liquids*. – 2014. – Т. 192. – С. 177–184.
410. Glorius F., Gnas Y. Chiral auxiliaries – principles and recent applications // *Synthesis*. – 2006. – Т. 2006, № 12. – С. 1899–1930.
411. Miyaura N., Yamada K., Suzuki A. A new stereospecific cross-coupling by the palladium-catalyzed reaction of 1-alkenylboranes with 1-alkenyl or 1-alkynyl halides // *Tetrahedron Letters*. – 1979. – Т. 20, № 36. – С. 3437–3440.

412. Azarian D., Dua S. S., Eaborn C., Walton D. R. M. Reactions of organic halides with $R^3 MMR^3$ compounds ($M = Si, Ge, Sn$) in the presence of tetrakis(triarylphosphine)palladium // *Journal of Organometallic Chemistry*. – 1976. – T. 117, № 3. – C. C55–C57.
413. Milstein D., Stille J. K. A general, selective, and facile method for ketone synthesis from acid chlorides and organotin compounds catalyzed by palladium // *Journal of the American Chemical Society*. – 1978. – T. 100, № 11. – C. 3636–3638.
414. Cordovilla C., Bartolomé C., Martínez-Ilarduya J. M., Espinet P. The stille reaction, 38 years later // *ACS Catalysis*. – 2015. – T. 5, № 5. – C. 3040–3053.
415. Kashin A. S., Ananikov V. P. Catalytic C–C and C-heteroatom bond formation reactions: in situ generated or preformed catalysts? Complicated mechanistic picture behind well-known experimental procedures // *The Journal of Organic Chemistry* – 2013. – T. 78, № 22. – C. 11117–11125.
416. Braga A. A. C., Ujaque G., Maseras F. A DFT study of the full catalytic cycle of the Suzuki–Miyaura cross-coupling on a model system // *Organometallics*. – 2006. – T. 25, № 15. – C. 3647–3658.
417. de Courcy B., Derat E., Piquemal J. P. Bridging organometallics and quantum chemical topology: Understanding electronic relocalisation during palladium-catalyzed reductive elimination // *Journal of Computational Chemistry*. – 2015. – T. 36, № 15. – C. 1167–1175.
418. Ananikov V. P., Musaev D. G., Morokuma K. Real size of ligands, reactants and catalysts: Studies of structure, reactivity and selectivity by ONIOM and other hybrid computational approaches // *Journal of Molecular Catalysis A: Chemical*. – 2010. – T. 324, № 1-2. – C. 104–119.
419. Fomine S., Vargas S. M., Tlenkopatchev M. A. Molecular modeling of ruthenium alkylidene mediated olefin metathesis reactions. DFT study of reaction pathways // *Organometallics*. – 2003. – T. 22, № 1. – C. 93–99.
420. Kozuch S., Martin J. M. L. What makes for a bad catalytic cycle? A theoretical study on the suzuki–miyaura reaction within the energetic span model // *ACS Catalysis*. – 2011. – T. 1, № 4. – C. 246–253.

421. Zhang S. L., Huang L., Sun L. J. The mechanism, electronic and ligand effects for reductive elimination from arylPd(II) trifluoromethyl complexes: a systematic DFT study // Dalton Trans. – 2015. – T. 44, № 10. – C. 4613–22.
422. Ananikov V. P., Musaev D. G., Morokuma K. Theoretical Insight into the C–C Coupling Reactions of the Vinyl, Phenyl, Ethynyl, and Methyl Complexes of Palladium and Platinum // Organometallics. – 2005. – T. 24, № 4. – C. 715–723.
423. Proutiere F., Schoenebeck F. Solvent effect on palladium-catalyzed cross-coupling reactions and implications on the active catalytic species // Angewandte Chemie International Edition in English. – 2011. – T. 50, № 35. – C. 8192–8195.
424. Svensson M., Humbel S., Froese R. D. J., Matsubara T., Sieber S., Morokuma K. ONIOM: A multilayered integrated MO+MM method for geometry optimizations and single point energy predictions. A test for Diels–Alder reactions and Pt(P(*t*-Bu)₃)₂ + H₂ Oxidative Addition // The Journal of Physical Chemistry. – 1996. – T. 100, № 50. – C. 19357–19363.
425. Chung L. W., Sameera W. M., Ramozzi R., Page A. J., Hatanaka M., Petrova G. P., Harris T. V., Li X., Ke Z., Liu F., Li H. B., Ding L., Morokuma K. The ONIOM method and its applications // Chem Rev. – 2015. – T. 115, № 12. – C. 5678–5796.
426. Vreven T., Mennucci B., da Silva C. O., Morokuma K., Tomasi J. The ONIOM-PCM method: Combining the hybrid molecular orbital method and the polarizable continuum model for solvation. Application to the geometry and properties of a merocyanine in solution // The Journal of Chemical Physics. – 2001. – T. 115, № 1. – C. 62–72.
427. Gouda H., Kuntz I. D., Case D. A., Kollman P. A. Free energy calculations for theophylline binding to an RNA aptamer: Comparison of MM-PBSA and thermodynamic integration methods // Biopolymers. – 2003. – T. 68, № 1. – C. 16–34.
428. Jiao D., Zhang J., Duke R. E., Li G., Schnieders M. J., Ren P. Trypsin-ligand binding free energies from explicit and implicit solvent simulations with polarizable potential // Journal of Computational Chemistry. – 2009. – T. 30, № 11. – C. 1701–1711.
429. Hallman K., Frolander A., Wondimagegn T., Svensson M., Moberg C. OH–Pd(0) interaction as a stabilizing factor in palladium-catalyzed allylic alkylations // Proceedings of the National Academy of Sciences of the United States of America. – 2004. – T. 101, № 15. – C. 5400–5404.

430. Acevedo O., Jorgensen W. L. Advances in quantum and molecular mechanical (QM/MM) simulations for organic and enzymatic reactions // *Accounts of Chemical Research*. – 2010. – T. 43, № 1. – C. 142–151.
431. Jorgensen W. L., Tirado-Rives J. Potential energy functions for atomic-level simulations of water and organic and biomolecular systems // *Proceedings of the National Academy of Sciences of the United States of America*. – 2005. – T. 102, № 19. – C. 6665–6670.
432. Chandrasekhar J., Jorgensen W. L. Energy profile for a nonconcerted SN2 reaction in solution // *Journal of the American Chemical Society*. – 1985. – T. 107, № 10. – C. 2974–2975.
433. Chandrasekhar J., Smith S. F., Jorgensen W. L. Theoretical examination of the SN2 reaction involving chloride ion and methyl chloride in the gas phase and aqueous solution // *Journal of the American Chemical Society*. – 1985. – T. 107, № 1. – C. 154–163.
434. Rod T. H., Ryde U. Accurate QM/MM free energy calculations of enzyme reactions: Methylation by catechol O-methyltransferase // *Journal of Chemical Theory and Computation*. – 2005. – T. 1, № 6. – C. 1240–1251.
435. Miyaura N., Suzuki A. Palladium-catalyzed cross-coupling reactions of organoboron compounds // *Chemical Reviews*. – 1995. – T. 95, № 7. – C. 2457–2483.
436. Marenich A. V., Cramer C. J., Truhlar D. G. Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions // *The Journal of Physical Chemistry B*. – 2009. – T. 113, № 18. – C. 6378–6396.
437. Ananikov V. P., Musaev D. G., Morokuma K. Critical effect of phosphane ligands on the mechanism of carbon–carbon bond formation involving palladium(II) complexes: A theoretical investigation of reductive elimination from square-planar and T-shaped species // *European Journal of Inorganic Chemistry*. – 2007. – T. 2007, № 34. – C. 5390–5399.
438. Frisch M. J., Trucks G. W., Schlegel H. B., Scuseria G. E., Robb M. A., Cheeseman J. R., Scalmani G., Barone V., Mennucci B., Petersson G. A., Nakatsuji H., Caricato M., Li X., Hratchian H. P., Izmaylov A. F., Bloino J., Zheng J. L., Sonnenberg M., Hada M., Ehara M., Toyota K., Fukuda R., Hasegawa J., Ishida M., Nakajima T., Honda Y., Kitao O., Nakai H., Vreven T., Montgomery J. A. J., Peralta J. E., Ogliaro F., Bearpark M., Heyd J. J., Brothers E., Kudin K. N., Staroverov V. N., Kobayashi R., Normand J., Raghavachari

- K., Rendell A., Burant J. C., Iyengar S. S., Tomasi J., Cossi M., Rega N., Millam J. M., Klene M., Knox J. E., Cross J. B., Bakken V., Adamo C., Jaramillo J., Gomperts R., Stratmann R. E., Yazyev O., Austin A. J., Cammi R., Pomelli C., Ochterski J. W., Martin R. L., Morokuma K., Zakrzewski V. G., Voth G. A., Salvador P., Dannenberg J. J., Dapprich S., Daniels A. D., Farkas Ö., Foresman J. B., Ortiz J. V., Cioslowski J., Fox D. J. 438 09, Revision D.01. – 2009. – URL: <http://gaussian.com/g09citation/>.
439. MacKerell A. D., Banavali N., Foloppe N. Development and current status of the CHARMM force field for nucleic acids // *Biopolymers*. – 2000. – T. 56, № 4. – C. 257–265.
440. Humphrey W., Dalke A., Schulten K. VMD: visual molecular dynamics. // *Journal of Molecular Graphics*. – 1996. – T. 14, № (33–38). – C. 27–38.
441. Heinz H., Vaia R. A., Farmer B. L., Naik R. R. Accurate simulation of surfaces and interfaces of face-centered cubic metals using 12–6 and 9–6 lennard-jones potentials // *The Journal of Physical Chemistry C*. – 2008. – T. 112, № 44. – C. 17281–17290.
442. Jorgensen W. L., Chandrasekhar J., Madura J. D., Impey R. W., Klein M. L. Comparison of simple potential functions for simulating liquid water // *The Journal of Chemical Physics*. – 1983. – T. 79, № 2. – C. 926–935.
443. Pronk S., Pall S., Schulz R., Larsson P., Bjelkmar P., Apostolov R., Shirts M. R., Smith J. C., Kasson P. M., van der Spoel D., Hess B., Lindahl E. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit // *Bioinformatics*. – 2013. – T. 29, № 7. – C. 845–854.
444. Wu D., Kofke D. A. Phase-space overlap measures. I. Fail-safe bias detection in free energies calculated by molecular simulation // *The Journal of Chemical Physics*. – 2005. – T. 123, № 5. – C. 054103.
445. Hallman K., Frolander A., Wondimagegn T., Svensson M., Moberg C. OH-Pd(0) interaction as a stabilizing factor in palladium-catalyzed allylic alkylations // *Proceedings of the National Academy of Sciences of the United States of America*. – 2004. – T. 101, № 15. – C. 5400–5404.
446. Braga D., Grepioni F., Tedesco E., Biradha K., Desiraju G. R. Hydrogen bonding in organometallic crystals. 6. X–H---M Hydrogen bonds and M---(H–X) pseudo-agostic bonds // *Organometallics*. – 1997. – T. 16, № 9. – C. 1846–1856.

447. Brammer L., Zhao D., Ladipo F. T., Braddock-Wilking J. Hydrogen bonds involving transition metal centres – a brief review // *Acta Crystallographica Section B Structural Science*. – 1995. – T. 51, № 4. – C. 632–640.
448. Sunoj R. B., Anand M. Microsolvated transition state models for improved insight into chemical properties and reaction mechanisms // *Physical Chemistry Chemical Physics*. – 2012. – T. 14, № 37. – C. 12715–12736.
449. Wang B., Cao Z. Mechanism of acid-catalyzed hydrolysis of formamide from cluster-continuum model calculations: concerted versus stepwise pathway // *The Journal of Physical Chemistry A* – 2010. – T. 114, № 49. – C. 12918–12927.
450. Jacquemin D., Preat J., Wathelet V., Perpète E. A. Substitution and chemical environment effects on the absorption spectrum of indigo // *J Chem Phys*. – 2006. – T. 124, № 7. – C. 74104.
451. Böes E. S., Livotto P. R., Stassen H. Solvation of monovalent anions in acetonitrile and N,N-dimethylformamide: Parameterization of the IEF-PCM model // *Chemical Physics*. – 2006. – T. 331, № 1. – C. 142–158.
452. Activation of dioxygen at transition metal centers: Fundamental investigations of the heavy atom isotope effect and its applications. / Lanci M. P. – Baltimore, MD Johns Hopkins University, 2009.
453. Fundamentals of statistical and thermal physics. / Reif F. – Нью-Йорк, США: McGraw-Hill, 1965. – 651 с.
454. Voleská I., Akola J., Jóvári P., Gutwirth J., Wágner T., Vasileiadis T., Yannopoulos S. N., Jones R. O. Structure, electronic, and vibrational properties of glassy Ga₁₁Ge₁₁Te₇₈: Experimentally constrained density functional study // *Physical Review B*. – 2012. – T. 86, № 9. – C. 094108.
455. Medvedev M. G., Bushmarinov I. S., Lyssenko K. A. Z-effect reversal in carboxylic acid associates // *Chemical Communications (Cambridge)*. – 2016. – T. 52, № 39. – C. 6593–6596.
456. Gerasimchuk V. V., Kucherenko A. S., Fakhrutdinov A. N., Medvedev M. G., Neylyubina Y. V., Zlotin S. G. Towards sustainable amino acid derived organocatalysts for asymmetric syn-aldol reactions // *European Journal of Organic Chemistry*. – 2017. – T. 2017, № 17. – C. 2540–2544.

457. Harris S. A., Kendon V. M. Quantum-assisted biomolecular modelling // *Philosophical Transactions of the Royal Society A*. – 2010. – T. 368, № 1924. – C. 3581–3592.
458. van Duin A. C. T., Dasgupta S., Lorant F., Goddard W. A. ReaxFF: A reactive force field for hydrocarbons // *The Journal of Physical Chemistry A*. – 2001. – T. 105, № 41. – C. 9396–9409.
459. Truhlar D. G., Garrett B. C. Variational transition state theory // *Annual Review of Physical Chemistry*. – 1984. – T. 35, № 1. – C. 159-189.
460. New trends in Kramers' reaction rate theory. Understanding chemical reactivity. Под ред. Talkner P., Hänggi P. – Dordrecht, Netherlands: Springer, 1995. – Т. 11: Understanding chemical reactivity. – 251 с.
461. Seeman J. I. The Curtin-Hammett principle and the Winstein-Holness equation: new definition and recent extensions to classical concepts // *Journal of Chemical Education*. – 1986. – T. 63, № 1. – C. 42.
462. Zefirov N. S. Stereochemical studies—XXI // *Tetrahedron*. – 1977. – T. 33, № 20. – C. 2719–2722.
463. Seeman J. I., Farone W. A. Analytical solution to the Curtin-Hammett/Winstein-Holness kinetic system // *The Journal of Organic Chemistry*. – 1978. – T. 43, № 10. – C. 1854–1864.
464. O'Boyle N. M., Banck M., James C. A., Morley C., Vandermeersch T., Hutchison G. R. Open Babel: An open chemical toolbox // *Journal of Cheminformatics*. – 2011. – T. 3. – C. 33.
465. RDKit: Open-source cheminformatics software // *RDKit: Open-source cheminformatics and Machine Learning, 2017*. URL: <http://www.rdkit.org>.
466. Schrödinger: Maestro // Schrödinger, LLC, New York, 2016. URL: <https://www.schrodinger.com/>
467. Patra N., Ioannidis E. I., Kulik H. J. Computational investigation of the interplay of substrate positioning and reactivity in catechol *O*-methyltransferase // *PLoS One*. – 2016. – T. 11, № 8. – C. e0161868.
468. Jindal G., Warshel A. Exploring the dependence of QM/MM calculations of enzyme catalysis on the size of the QM region // *Journal of Physical Chemistry B*. – 2016. – T. 120, № 37. – C. 9913-9921.

469. Medvedev M. G., Bushmarinov I. S., Sun J., Perdew J. P., Lyssenko K. A. Density functional theory is straying from the path toward the exact functional // *Science*. – 2017. – T. 355, № 6320. – C. 49–52.
470. Vasilevskaya T., Khrenova M. G., Nemukhin A. V., Thiel W. Reaction mechanism of matrix metalloproteinases with a catalytically active zinc ion studied by the QM(DFTB)/MM simulations // *Mendeleev Communications*. – 2016. – T. 26, № 3. – C. 209–211.
471. *Methods and applications of cycloaddition reactions in organic syntheses*. Под ред. Nishiwaki N. – Hoboken, New Jersey: John Wiley & Sons, 2013. – 672 с.
472. Preiswerk N., Beck T., Schulz J. D., Milovnik P., Mayer C., Siegel J. B., Baker D., Hilvert D. Impact of scaffold rigidity on the design and evolution of an artificial Diels-Alderase // *Proceedings of the National Academy of Sciences of the United States of America*. – 2014. – T. 111, № 22. – C. 8013–8018.
473. Guimaraes C. R., Udier-Blagovic M., Jorgensen W. L. Macrophomate synthase: QM/MM simulations address the Diels-Alder versus Michael-Aldol reaction mechanism // *Journal of the American Chemical Society*. – 2005. – T. 127, № 10. – C. 3577–3588.
474. Klas K., Tsukamoto S., Sherman D. H., Williams R. M. Natural Diels-Alderase: Elusive and Irresistible // *Journal of Organic Chemistry*. – 2015. – T. 80, № 23. – C. 11672–11685.
475. Minami A., Oikawa H. Recent advances of Diels-Alderase involved in natural product biosynthesis // *The Journal of Antibiotics (Tokyo)*. – 2016. – T. 69, № 7. – C. 500–506.
476. Byrne M. J., Lees N. R., Han L. C., van der Kamp M. W., Mulholland A. J., Stach J. E., Willis C. L., Race P. R. The catalytic mechanism of a natural diels-alderase revealed in molecular detail // *Journal of the American Chemical Society*. – 2016. – T. 138, № 19. – C. 6095–6098.
477. Li L., Yu P., Tang M. C., Zou Y., Gao S. S., Hung Y. S., Zhao M., Watanabe K., Houk K. N., Tang Y. Biochemical characterization of a eukaryotic decalin-forming Diels-Alderase // *Journal of the American Chemical Society*. – 2016. – T. 138, № 49. – C. 15837–15840.

478. Yano B. L. Spinosad insecticide: Subchronic and chronic toxicity and lack of carcinogenicity in Fischer 344 Rats // *Toxicological Sciences*. – 2002. – Т. 65, № 2. – С. 288–298.
479. Gordeev E. G., Ananikov V. P. Computational study of a model system of enzyme-mediated [4+2] cycloaddition reaction // *PLoS One*. – 2015. – Т. 10, № 4. – С. e0119984.
480. Kim H. J., Rusczycky M. W., Choi S. H., Liu Y. N., Liu H. W. Enzyme-catalysed [4+2] cycloaddition is a key step in the biosynthesis of spinosyn A // *Nature*. – 2011. – Т. 473, № 7345. – С. 109–112.
481. Hess B. A., Jr., Smentek L. Concerted, highly asynchronous, enzyme-catalyzed [4+2] cycloaddition in the biosynthesis of spinosyn A; computational evidence // *Organic & Biomolecular Chemistry*. – 2012. – Т. 10, № 37. – С. 7503–7509.
482. Patel A., Chen Z., Yang Z., Gutierrez O., Liu H. W., Houk K. N., Singleton D. A. Dynamically complex [6+4] and [4+2] cycloadditions in the biosynthesis of spinosyn A // *Journal of the American Chemical Society*. – 2016. – Т. 138, № 11. – С. 3631–3634.
483. Kim N. H. Mechanistic Investigations of SpnF- and SpnL-Catalyzed Cyclizations in the Biosynthesis of Spinosyn A : дисс PhD, защищена: декабрь 2013. / Kim Nam Ho, The University of Texas at Austin. – Остин, США. – 357 с. – Библиогр.: с. 335–356.
484. Black K., Liu P., Xu L., Doubleday C., Houk K. N. Dynamics, transition states, and timing of bond formation in Diels-Alder reactions // *Proceedings of the National Academy of Sciences of the United States of America*. – 2012. – Т. 109, № 32. – С. 12860–12865.
485. Rogers D. W., Matsunaga N., Zavitsas A. A. Stabilization energies of extensively conjugated propargylic radicals // *Journal of Organic Chemistry*. – 2006. – Т. 71, № 6. – С. 2214–2219.
486. Bergeler M., Simm G. N., Proppe J., Reiher M. Heuristics-Guided Exploration of Reaction Mechanisms // *Journal of Chemical Theory and Computation*. – 2015. – Т. 11, № 12. – С. 5712–5722.
487. Adamo C., Barone V. Toward reliable density functional methods without adjustable parameters: The PBE0 model // *The Journal of Chemical Physics*. – 1999. – Т. 110, № 13. – С. 6158–6170.
488. Grimme S., Antony J., Ehrlich S., Krieg H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu // *Journal of Chemical Physics*. – 2010. – Т. 132, № 15. – С. 154104.

489. Watts K. S., Dalal P., Tebben A. J., Cheney D. L., Shelley J. C. Macrocyclic conformational sampling with MacroModel // *Journal of Chemical Information and Modeling*. – 2014. – Т. 54, № 10. – С. 2680–2696.
490. Rogers D. W., Matsunaga N., Zavitsas A. A. Stabilization energies of extensively conjugated propargylic radicals // *Journal of Organic Chemistry*. – 2006. – Т. 71, № 6. – С. 2214–2219.
491. McGill R., Tukey J. W., Larsen W. A. Variations of Box Plots // *The American Statistician* –1978. – Т. 32, № 1. – С. 12–16.
492. Atoms in molecules: A quantum theory. / Bader R. F. W. – Oxford: Oxford University Press, 1990. . – 458 с.
493. Warshel A., Sharma P. K., Kato M., Xiang Y., Liu H., Olsson M. H. Electrostatic basis for enzyme catalysis // *Chemical Reviews*. – 2006. – Т. 106, № 8. – С. 3210–3235.
494. Aragonés A. C., Haworth N. L., Darwish N., Ciampi S., Bloomfield N. J., Wallace G. G., Díez-Pérez I., Coote M. L. Electrostatic catalysis of a Diels-Alder reaction // *Nature*. – 2016. – Т. 531, № 7592. – С. 88–91.
495. Medvedev M. G., Panova M. V., Chilov G. G., Bushmarinov I. S., Novikov F. N., Stroganov O. V., Zeifman A. A., Svitanko I. V. Exhaustive conformational search for transition states: the case of catechol O-methyltransferase active site // *Mendeleev Communications*. – 2017. – Т. 27, № 3. – С. 224–227.
496. Wannere C. S., Paul A., Herges R., Houk K. N., Schaefer H. F., 3rd, von Rague Schleyer P. The existence of secondary orbital interactions // *Journal of Computational Chemistry*. – 2007. – Т. 28, № 1. – С. 344–361.
497. Medvedev M. G., Novikov F. N., Bushmarinov I. S., Zeifman A. A., Polkovnichenko M. S., Stroganov O. V., Chilov G. G., Lyssenko K. A., Svitanko I. V. Charge redistribution in the SpnF-catalyzed Diels–Alder reaction // *Mendeleev Communications*. – 2017. – Т. 27, № 5. – С. 500–502.
498. Стройлов, В. С. Моделирование координации биологически активных соединений с терапевтическими мишенями: дис. ... канд. хим. наук: 02.00.15, 02.00.04: защищена 21.12.2010 / Стройлов Виктор Сергеевич – М., 2010. – 190 с. – Библиогр.: с. 178–190. – 004616463

499. Зейфман, А. А. Моделирование конформационного равновесия и сольватации при связывании органических соединений с биологическими мишенями: дис. ... канд. хим. наук: 02.00.03: защищена 11.06.2013 / Зейфман, Алексей Александрович– М., 2013. – 129 с. – Библиогр.: с. 123–129. – 005059448
500. S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. Van Der Spoel, B. Hess and E. Lindahl, *Bioinformatics*, 2013, 29, 845.
501. L. Jorgensen, D. S. Maxwell and J. Tirado-Rives, *J. Am. Chem. Soc.*, 1996, 118, 11225.