

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

На правах рукописи

Кривошапов Николай Владиславович

**Конформационная подвижность и стереоэлектронные эффекты
в полициклических молекулах**

1.4.4 – Физическая химия

ДИССЕРТАЦИЯ

на соискание учёной степени

кандидата физико-математических наук

Научный руководитель:
к.ф.-м.н. Медведев М. Г.

Москва – 2026

Оглавление

Введение и общая характеристика работы	4
Раздел 1. Методы конформационного поиска.....	14
1.1. Обзор методов конформационного поиска	14
1.1.1. Методы глобального перебора.....	15
1.1.2. Методы движения по поверхности потенциальной энергии	28
1.1.3. Методы аугментации ансамблей.....	41
1.1.4. Методы машинного обучения.....	46
1.1.5. Гибридные методы (пайплайны) конформационного поиска.....	49
1.2. Тестирование методов конформационного поиска.....	51
1.2.1. Метрики конформационного подобия	52
1.2.2. Подходы к построению тестовых задач	54
Раздел 2. Стереoeлектронные эффекты в каркасных органических пероксидах	58
2.1. Моделирование реакции β,γ' -трикетонов с пероксидом водорода	58
2.1.1. Используемые методы.....	60
2.1.2. Результаты квантово-химического моделирования	62
2.1.3. Типовые стереoeлектронные эффекты и способы их моделирования ...	69
2.1.4. Анализ орбитальных взаимодействий.....	76
2.2. Добавление конкурирующего N-нуклеофила. Моделирование реакции β,δ' -трикетонов, пероксида водорода и аммиака.....	82
2.2.1. Используемые методы.....	83
2.2.2. Результаты квантово-химического моделирования	84
2.2.3. Анализ орбитальных взаимодействий.....	90
2.3. Аналог обратного α -эффекта с атомами азота. Моделирование реакции β,δ' -трикетонов с пероксидом водорода и гидразидом	94
2.3.1. Используемые методы.....	95
2.3.2. Результаты моделирования и анализ орбитальных взаимодействий	96

2.4. Другие проявления стереоэлектронных эффектов. Моделирование реакции δ -дикетонов, пероксида водорода и аммиака.....	99
2.4.1. Используемые методы.....	100
2.4.2. Результаты квантово-химического моделирования	101
2.4.3. Анализ орбитальных взаимодействий.....	103
Раздел 3. Конформационная подвижность циклических молекул.....	110
3.1. Механическая модель подвижности молекул. Прямая кинематика.....	111
3.2. Степени свободы циклов. Постановка задачи обратной кинематики.....	115
3.3. Решение обратной кинематики одиночных циклов.....	118
3.4. Обобщение обратной кинематики на полициклические системы	120
3.5. Метод конформационного поиска MCR и его тестирование	125
Заключение	132
Публикации автора по теме диссертации	134
Список используемых сокращений.....	136
Список литературы.....	138
Приложение 1. Результаты квантово-химических расчётов относительных энергий предполагаемых интермедиатов вдоль нескольких предложенных путей сборки пероксидов A1b и A2b (см. подраздел 2.1).....	154
Приложение 2. Результаты квантово-химических расчётов относительных энергий предполагаемых интермедиатов и ПС вдоль нескольких предложенных путей сборки азаозонидов D1a и D2a (см. подраздел 2.4)	155

Введение и общая характеристика работы

Актуальность темы и степень её разработанности. Циклические молекулы представляют интерес для поиска новых лекарственных препаратов. Например, среди недавно одобренных FDA соединений можно выделить несколько макроциклических молекул: циклический пептид Мотиксафортид [1] и макроцикл Репотректиниб [2], использующиеся при лечении онкологических заболеваний, а также противогрибковый циклический пептид Резафунгин [3]. В то же время полициклические каркасы из малых циклов также представляют фармакологический интерес, в частности, к ним относятся многие лекарственные органические пероксиды. Ярким примером является природный противомаларийный препарат Артемизинин, открытие которого было удостоено Нобелевской премии в 2015 году, и который продолжает применяться по сей день. Наряду с этим активно развиваются и другие направления по созданию синтетических пероксидов с противомаларийной [4,5], противораковой [6,7], фунгицидной [5,8], противовирусной [9] и другими видами активности.

Точность прогнозов виртуального скрининга биологически активных молекул, а также квантово-химического моделирования реакций с их участием требует рассмотрения репрезентативного набора их конформеров. Для циклических систем, содержащих атомы азота и кислорода, построение конформационного ансамбля затрудняется тем, что эти атомы зачастую вступают в стереоэлектронные взаимодействия, усложняющие конформационное пространство молекулы и делающие его чувствительным к небольшим изменениям в молекуле или даже межмолекулярных взаимодействиях [10]. Это приводит к необходимости конформационного анализа таких систем перед их квантово-химическим моделированием.

Конденсация поликетонов с пероксидом водорода, являющаяся удобным методом одностадийного получения полициклических пероксидов из доступных субстратов, зачастую приводит к образованию непредсказуемых продуктов, поскольку каскадная сборка может протекать по множеству альтернативных путей

со смешанным термодинамическим и кинетическим контролем. Поэтому для направленного синтеза таких молекул важной задачей является выявление ключевых факторов, определяющих конечный продукт сборки каркасных пероксидов из поликетонов и пероксида водорода. Так, Gomes и соавторы обнаружили бициклический пероксид, искаженная конформация которого ослабляет делокализацию $n_o \rightarrow \sigma_{C-O}^*$, вследствие чего образуется другой продукт реакции, в котором аналогичная делокализация не ослаблена [11]. На основании квантово-химических расчётов Juaristi и соавторы предположили, что пероксидная группа является более слабым стабилизатором карбокатионов по сравнению с одиночным атомом кислорода вследствие своей пониженной π -донорной способности, вызванной обратным α -эффектом [12]. Виль и соавторы обнаружили реальный пример реакции, которая не достигает термодинамического продукта вследствие этого эффекта [13]. Однако, для трициклических и азотсодержащих каркасных пероксидов влияние делокализации $n \rightarrow \sigma^*$ и обратного α -эффекта на продукт сборки остаётся неизученным и его изучение является **актуальной задачей**. Стоит отметить, что в случае конденсированных систем из малых циклов, к которым относятся многие каркасные пероксиды, конформационный поиск является критически важной, но несложной задачей, поскольку такие молекулы обладают ограниченной подвижностью и их структуры могут быть сгенерированы методами дистанционной геометрии [14].

В случае же конформационно гибких макроциклических молекул многие существующие методы конформационного поиска оказываются неэффективны из-за большей размерности конформационного пространства и сложности перебора значений торсионных углов с сохранением замкнутости цикла в получаемых геометриях. Перспективным направлением к построению конформаций циклов является обратная кинематика (ОК). В этом подходе определяется набор степеней свободы полициклической системы, то есть независимых торсионных углов каждого цикла, при этом количество степеней свободы всегда меньше, чем количество связей в цикле. Остальные торсионные углы вычисляются путём решения системы алгебраических уравнений, так как они определяются из необходимости замыкания цикла с определёнными длинами связей и валентными углами. Применение ОК уже

было предложено ранее [15,16], однако универсальный подход к анализу топологии произвольной полициклической молекулы с определением степеней свободы и их последующим перебором для формирования конформационного ансамбля ранее отсутствовал. Следовательно, разработка универсального подхода для автоматического анализа топологии произвольной полициклической молекулы и генерации её конформаций на основе ОК является **актуальной задачей**.

Таким образом, изучение роли стереоэлектронных эффектов, направляющих конденсацию поликетонов до конформационно жёстких полициклических пероксидов, и разработка методов глобального конформационного поиска для конформационно подвижных полициклических молекул являются **актуальными** нерешенными задачами.

Цель работы состоит в изучении конформационной подвижности полициклических органических пероксидов и её влияния на процесс их сборки, а также развитии подхода обратной кинематики для конформационного поиска циклических и полициклических молекул.

Для достижения поставленной цели необходимо было решить следующие **задачи**:

1. Применить существующие методы конформационного поиска полициклических молекул для решения практической задачи квантово-химического моделирования сборки каркасных пероксидов из β,γ' -трикетонов, β,δ' -трикетонов и δ -дикетонов под действием пероксида водорода и иных нуклеофилов (аммиака, гидразидов) и воспроизведения ключевых экспериментальных наблюдений, таких как преимущественный продукт, характерное время реакции и конформация из рентгеноструктурного анализа.
2. Изучить влияние стереоэлектронных эффектов (отрицательной гиперконъюгации и обратного α -эффекта) на конформацию полициклических каркасов, их термодинамическую стабильность, химические свойства и кинетические барьеры их сборки.

3. Разработать метод построения конформаций конденсированных циклов путём последовательного решения обратной кинематики для каждого отдельного цикла и на его основе реализовать алгоритм определения независимых торсионных углов (степеней свободы) полициклических молекул и генерации их конформаций по заданным значениям этих углов.
4. Реализовать алгоритм конформационного поиска на основе случайного перебора кинематически независимых торсионных углов, применимый к молекулам, содержащим полициклические и макроциклические фрагменты.
5. Протестировать разработанный алгоритм конформационного поиска и оценить его эффективность по сравнению с существующими методами.

Объектами исследования выступали реакции сборки каркасных полициклических пероксидов из поликетонов (β,γ' -трикетон, β,δ' -трикетон и δ -дикетон) с пероксидом водорода и иными нуклеофилами (аммиак, гидразиды), а также 150 макроциклических молекул лекарственно-подобных органических веществ, составляющих тестовый набор MacroModel, который использовался для тестирования разработанного нами метода конформационного поиска.

Предметом исследования являлись стереоэлектронные эффекты и их роль в механизме сборки каркасных полициклических пероксидов, а также методы конформационного поиска и получаемые с их помощью ансамбли низкоэнергетических конформеров исследуемых молекул.

Научная новизна

Результаты, перечисленные ниже, были получены впервые:

1. Построены квантово-химические модели реакций сборки полициклических органических пероксидов из β,γ' -трикетонов, β,δ' -трикетонов и δ -дикетонов под действием пероксида водорода и N-нуклеофилов. Построенные модели согласуются с экспериментальными наблюдениями, а результаты конформационного анализа продуктов подтверждены результатами

рентгеноструктурного анализа. На основе построенных моделей определены скорость-лимитирующие стадии, определяющие продукты этих реакций.

2. Изучено влияние делокализации неподелённых пар электронов гетероатомов полициклических пероксидов на их конформацию, термодинамическую стабильность и химические свойства. Показано, что селективность сборки полициклических пероксидов из поликетонов, пероксида водорода и N-нуклеофилов во многих случаях кинетически контролируется обратным α -эффектом, то есть пониженной стабилизацией карбокатионов пероксидной группой.
3. Предложен и реализован универсальный алгоритм построения конформаций полициклических молекул по их степеням свободы, основанный на оригинальном подходе через последовательное решение обратной кинематики для каждого отдельного цикла. Реализация данного алгоритма находится в открытом доступе <https://pypi.org/project/ringo-ik> [17].
4. Изучены скорости генерации низкоэнергетических конформаций, полнота ансамблей и точность воспроизведения экспериментальных структур в методе конформационного поиска полициклических молекул на основе обратной кинематики.

Теоретическая и практическая значимость

1. Показано, что активация орбитальных взаимодействий с участием неподелённых пар электронов гетероатомов является основным фактором термодинамического контроля сборки каркасных пероксидов. Установлено, что реакции, протекающие через бимолекулярную нуклеофильную атаку по пероксикарбениевому иону, как правило, кинетически блокируются обратным α -эффектом. Практическая значимость заключается в возможности использования этих принципов для планирования синтеза органических пероксидов и реакций с их участием.
2. Разработан эффективный метод глобального сэмплирования конформаций полициклических молекул методом обратной кинематики, что позволяет повысить эффективность их конформационного анализа. Практическая значимость заключается в возможности предсказания конформаций молекул

потенциальных лекарственных веществ, включая макроциклические органические пероксиды, что необходимо для дальнейшего моделирования их связывания с белками, что открывает путь к более эффективному виртуальному скринингу лекарственных веществ.

Методология и методы исследования

Все квантово-химические расчёты реакций сборки органических пероксидов были проведены в программном пакете Gaussian16 A.03 [18] на уровне теории PBE0-D3BJ [19,20] / 6-311++G(d,p) [21,22] с моделями неявной сольватации SMD [23] или CPCM [24]. В расчётах свободной энергии использовалась квазигармоническая поправка, рассчитываемая с помощью кода GoodVibes [25]. Для конформационного поиска использовалась собственная имплементация алгоритма ringmaker Monte-Carlo (RMMC), ETKDG [14], имплементированный в пакете RDKit [26], а также пайплайн iMTD-GC в пакете CREST 2.12 [27]. Для локализации переходных состояний (ПС) и их конформационного поиска применялся подход [28], заключающийся в предварительной оптимизации геометрии ПС в минимум энергии при фиксированных длинах разрывающихся и образующихся связей на их ожидаемой длине в ПС (2 Å для связей между атомами второго периода) с последующей оптимизацией в седловую точку – данный подход упрощает локализацию ПС. Для всех рассмотренных геометрий проводилась проверка числа мнимых колебательных мод (0 – для интермедиатов, 1 – для ПС).

При анализе результатов квантово-химического моделирования в данной работе используется понятие энергетического барьера. Под ним понимается наибольшая разность свободных энергий между каким-либо ПС и любым из интермедиатов, предшествующих этому ПС по ходу реакции. Это основано на выводе предшествующих работ [29,30] о том, что в многостадийных реакционных системах наблюдаемая скорость не всегда определяется энергией активации ΔG^\ddagger одной элементарной стадии, поскольку несколько последовательных стадий могут создать более высокий эффективный барьер. Поэтому энергетический барьер оказывается

информативным даже в случаях, когда ни одна отдельная стадия не является скоростью-лимитирующей.

Разработанные алгоритмы обратной кинематики и конформационного поиска полициклических молекул были реализованы в виде библиотеки для Python, имплементированной на языке C++ [17]. Для выделения независимых циклов в полициклических фрагментах молекул в алгоритме используется метод вычисления базиса циклов минимального веса FAST-MCB [31]. Для построения конформаций одиночных циклов по заданным независимым торсионным углам используется алгоритм Triaxial Loop Closure (TLC) [15]. Сборка конформаций из геометрий отдельных циклов и нециклических фрагментов производится с помощью матриц стандартного представления специальной евклидовой группы $SE(3)$.

Для оценки эффективности нового метода конформационного поиска мы сравнили его с рядом популярных методов: (1) алгоритм ETKDG из библиотеки RDKit; (2) метод MTD в библиотеке XTB на основе молекулярной динамики с локальным подъёмом потенциала; (3) CREST – алгоритм, объединяющий MTD, классическую молекулярную динамику и генетические операции; (4) метод торсионного перебора Монте-Карло в программе MacroModel (MCMM); (5) конформационное сэмплирование макроциклов в MacroModel (MMBS). В расчётах MTD и CREST использовался уровень теории GFN-FF [32] / ALPB(H₂O) [33], а в MacroModel применялось силовое поле OPLS-3 с неявным учётом сольватации водой. Перед дальнейшим анализом все конформационные ансамбли оптимизированы на едином уровне теории MMFF94 [34]. В качестве тестового набора рассмотрены 150 лекарственных и лекарственно-подобных макроциклических соединений, использовавшихся для тестирования MMBS [35].

Для проверки глобальности конформационного поиска разработанным нами алгоритмом мы построили двумерные диаграммы конформационных ансамблей каждой молекулы, где точки соответствуют кластерам конформаций макроцикла; кластер считается найденным, если хотя бы одна конформация из ансамбля принадлежит этому кластеру. Построение диаграмм требует способа измерения сходства между парами конформеров. Вместо обычного полноатомного

среднеквадратичного отклонения (RMSD) была использована метрика циклического RMSD (CRMSD). Единственное отличие заключается в том, что CRMSD вычисляется только для циклических атомов, а при наличии нескольких циклических фрагментов CRMSD рассчитывается независимо для каждого фрагмента, и в качестве окончательного значения CRMSD выбирается максимальное из вычисленных. Построение ансамблевой диаграммы для каждой молекулы включало следующие шаги: (1) расчёт матрицы CRMSD для объединенного ансамбля по всем методам после удаления конформеров с энергией > 15 ккал/моль от минимального, (2) кластеризация на основе полученной матрицы методом DBSCAN [36], (3) расчёт двумерных координат кластеров из матриц CRMSD путём понижения размерности с помощью алгоритма t-SNE [37] (параметр perplexity – 50, скорость обучения – 10, число итераций – 2000).

Энергии орбитальных взаимодействий, свободные энергии реакций ΔG_r и активации ΔG^\ddagger в тексте и на рисунках приведены в ккал/моль.

Работа выполнена с использованием оборудования центра коллективного пользования «Комплекс моделирования и обработки данных исследовательских установок мега-класса» НИЦ «Курчатовский институт», <http://ckp.nrcki.ru/>.

Основные положения, выносимые на защиту

1. Стереoeлектронные эффекты с участием неподелённых пар электронов гетероатомов, контролирующие направление реакций сборки полициклических пероксидов из поликетонов и пероксида водорода.
2. Подход обратной кинематики для построения конформаций полициклических молекул из заданного набора значений их независимых торсионных углов, сводящий задачу обратной кинематики полицикла к решению последовательности задач обратной кинематики для одиночных циклов.
3. Алгоритм глобального конформационного поиска полициклических молекул, который обеспечивает высокую скорость генерации низкоэнергетических конформаций, полноту ансамблей и точность воспроизведения

экспериментальных структур, превосходя по одному или нескольким этим показателям существующие методы (CREST, ETKDG, MacroModel).

Достоверность полученных результатов обеспечивается использованием современных методов теории функционала плотности и локализации орбиталей для квантово-химического моделирования сборки каркасных пероксидов и их анализа, а также детальным сравнением разработанного алгоритма конформационного поиска с существующими передовыми методами на тестовом наборе лекарственно-подобных соединений.

Апробация работы

Основные результаты диссертационной работы были представлены на международных и всероссийских конференциях:

1. XIII International Conference on Chemistry for Young Scientists «Mendeleev 2024». – г. Санкт-Петербург, 2–6 сентября 2024 г. – устный доклад.
2. XXVII Всероссийская конференция молодых ученых-химиков. – г. Нижний Новгород, Новгородская обл., 16–18 апреля 2024 г. – устный доклад.
3. V. A. Fock Meeting on Theoretical, Quantum and Computational Chemistry. – г. Великий Новгород, 4–8 сентября 2023 г. – устный доклад.
4. Всероссийская научная конференция посвященная 65-летию со дня основания НИОХ СО РАН «Современные проблемы органической химии». – г. Новосибирск, 26–30 июня 2023 г. – флеш-доклад.
5. Всероссийская научная конференция «Марковниковские чтения: органическая химия от Марковникова до наших дней». – п. Лоо, Краснодарский край, 16–21 сентября 2022 г. – устный доклад.
6. Международная конференция «18-th Session of the V. A. Fock Meeting on Quantum and Computational Chemistry». – г. Великий Новгород, Новгородский государственный университет им. Ярослава Мудрого, 22–26 августа 2022 г. – устный доклад.

7. Всероссийская молодежная научная школа-конференция «Актуальные проблемы органической химии». – п. Шерегеш, Кемеровская обл., 20–26 марта 2022 г. – устный доклад.

Личный вклад. В работах по моделированию сборки каркасных пероксидов автор лично предложил квантово-химические модели исследуемых реакций, спланировал и провёл необходимые расчёты, а также участвовал в интерпретации и анализе теоретических и экспериментальных данных и в написании статей. В обзорных работах по стереоэлектронным эффектам автор участвовал в изучении литературы, написании текста, проведении квантово-химических расчётов и подготовке иллюстраций. В работах по обратной кинематике автор лично определил, предложил и реализовал алгоритм конформационного поиска, провёл тестирование, проанализировал полученные данные и написал статьи.

Структура и объем работы. Диссертация состоит из введения, 3 разделов, заключения, списка публикаций автора по теме диссертации, списка используемых сокращений и списка литературы. Работа изложена на 155 страницах, содержит 50 рисунков, 202 библиографические ссылки и 2 приложения.

Раздел 1. Методы конформационного поиска

1.1. Обзор методов конформационного поиска

Задача конформационного поиска заключается в выявлении низкоэнергетических конформеров¹ молекул – информация об их геометрии и энергии принципиально важна для моделирования физических и химических свойств вещества. С математической точки зрения конформационный поиск сводится к задаче нахождения глобального минимума (глобальной оптимизации) и низколежащих локальных минимумов. В данной постановке потенциальная энергия молекулы выступает в качестве целевой функции (при фиксированной топологии молекулярного графа), а переменными являются внутренние координаты молекулы или декартовы координаты атомов.

Методы конформационного поиска могут исследовать поверхность потенциальной энергии (ППЭ) исходя либо из некоторого стартового конформера, либо непосредственно из молекулярного графа²; в последнем случае генерация исходных конформаций становится отдельной задачей. В отличие от локальной минимизации,³ для которой направление спуска определяется антиградиентом $-\nabla V(\vec{r})$, глобальный поиск минимумов оказывается существенно сложнее. Это связано с тем, что для достижения далёких от начальной структуры минимумов необходимы как нисходящие, так и восходящие перемещения по ППЭ при том, что

¹ Здесь и далее в рамках адиабатического приближения под «конформацией» понимается любое пространственное расположение атомов (набор ядерных координат), отвечающее заданной топологии и стереохимии. Термин «конформер» используется для обозначения конформации, которая соответствует локальному минимуму на ППЭ.

² Здесь и далее под «молекулярным графом» понимается граф связности молекулы с атрибутами вершин (химический элемент, заряд) и рёбер (кратность связи).

³ Под локальной минимизацией понимается стандартная для квантово-химического моделирования процедура поиска минимума ППЭ. В простейшем случае она реализуется градиентным спуском, а в популярных программных пакетах используются квази-ньютоновские методы в избыточных внутренних координатах. Выражения «минимизация энергии» и «локальная оптимизация геометрии» используются ниже в том же значении.

наиболее перспективные повышающие энергию движения априори неизвестны. Тем не менее к настоящему времени разработаны различные подходы к глобальному поиску низкоэнергетических конформеров, которые, согласно нашему опыту, можно классифицировать по трём группам:

1. Методы глобального перебора – случайный, систематический или эмпирический выбор начальных точек на ППЭ с последующей локальной оптимизацией в ближайший минимум.
2. Методы движения по поверхности потенциальной энергии – динамика или случайное блуждание, максимизирующее вероятность посещения низкоэнергетических областей; алгоритмы этой группы пересекаются с методами построения выборок из термодинамических ансамблей молекул.
3. Методы аугментации ансамблей – анализ набора уже найденных конформеров и их энергий для выявления потенциально отсутствующих низкоэнергетических локальных минимумов с последующей генерацией новых перспективных геометрий и их локальной оптимизацией.

Методы первой группы основаны на взаимно независимых, некоррелированных итерациях перебора и запусках локальной оптимизации. В методах второй группы каждая последующая итерация определяется состоянием системы, полученным на предыдущем шаге (в методах с адаптивным смещающим потенциалом – несколькими предыдущими состояниями). Методы третьей группы опираются на глобальный анализ всех полученных на текущий момент конформеров и их энергий с целью предсказания новых низкоэнергетических структур. Ниже представлен обзор методов конформационного поиска, организованный согласно данной классификации, обсуждение методов на основе машинного обучения выделено в отдельный подраздел 1.1.4.

1.1.1. Методы глобального перебора

Данный подход заключается в случайном или систематическом переборе стартовых конформаций, распределённых по ППЭ. Из каждой такой конформации с использованием методов локальной оптимизации выполняется спуск в ближайший локальный минимум, после чего дублирующиеся структуры исключаются.

Ключевым элементом методов перебора является алгоритм генерации стартовых конформаций, обеспечивающий быстрое получение точек на ППЭ, находящихся преимущественно в окрестности низкоэнергетических минимумов. Для задачи конформационного поиска органических молекул разработано множество алгоритмов перебора:

1. Методы дистанционной геометрии – получение случайных конформаций путём оптимизации декартовых координат атомов для удовлетворения геометрических ограничений.
2. Методы в приближении жёсткого ротатора – перебор торсионных углов, отвечающих за конформационную подвижность молекулы, при фиксированных длинах связей, валентных углах и жёстких торсионных углах (как правило, соответствующих кратным связям).
3. Эмпирические методы – генерация конформаций на основе геометрий фрагментов, извлечённых из структурных баз данных, и статистически наиболее вероятных значений торсионных углов.

Ниже приведён обзор методов конформационного перебора, согласно данной классификации. Многие методы на основе машинного обучения также можно отнести к методам перебора, однако они будут рассмотрены отдельно, в подразделе 1.1.4.

1.1.1.1. Дистанционная геометрия и родственные методы

Если решать задачу перебора конформаций в декартовых координатах атомов, то необходимо обеспечить корректность топологии и валентных углов. Очевидным подходом является выбор случайных положений атомов в объёме куба или шара достаточного размера с последующей оптимизацией для создания необходимой топологии и локальной корректировки геометрии. Ярким примером этого подхода является метод Stochastic Proximity Embedding (SPE) [38,39], который получает конформации из случайно распределённых в объёме куба атомов путём минимизации ошибок в межатомных расстояниях (для фиксации длин связей, валентных углов и абсолютных значений торсионных углов) и ориентированных

объёмах (для контроля стереоцентров и знака торсионных углов). Однако, для крупных систем такой метод неэффективен, так как требует длительной оптимизации для сборки каждой новой конформации из случайно расположенных атомов. Для частичного решения этой проблемы был предложен метод Self-Organizing Superimposition (SOS) [40], который повышает скорость и вероятность успеха оптимизации стартовой геометрии за счёт использования библиотеки структур распространённых фрагментов (например, бензольных колец).

Как оказалось, вероятность успешной оптимизации случайной геометрии можно дополнительно повысить, если на этапе её генерации учесть геометрические ограничения (длины связей, валентные углы и абсолютные значения торсионных углов), которые эквивалентно выражаются через межатомные расстояния (Рисунок 1). В такой формулировке для этого можно использовать методы дистанционной геометрии (ДГ), раздела геометрии, занимающегося безкоординатным описанием конечных наборов точек, используя только информацию об их расстояниях друг от друга. В частности, методы ДГ позволяют проверить возможность реализации набора точек в трёхмерном пространстве по полной матрице парных расстояний и получить такую реализацию алгоритмически.

В методах конформационного поиска, основанных на ДГ, межатомные расстояния, не зафиксированные длинами связей, валентными углами или жёсткими торсионными углами, используют как источник случайности для генерации конформаций. Для этого задают две матрицы $N \times N$ (где N – число атомов) минимально и максимально допустимых парных расстояний, затем генерируют случайные дистанционные матрицы с элементами в этих интервалах и преобразуют их в трёхмерные координаты алгоритмами ДГ. Например, часто используется алгоритм [41], вычисляющий координаты атомов по трём собственным векторам матрицы Грама, что даёт точное решение для корректных дистанционных матриц и наилучшее приближение по среднеквадратичной ошибке, если точного трёхмерного вложения не существует. Интересно, что более точные ограничения для матриц минимальных и максимальных расстояний можно получить из данных спектроскопии ядерного эффекта Оверхаузера, которая экспериментально выявляет

пространственные контакты между атомами. Комбинация описанного подхода ДГ с такими экспериментальными данными успешно применяется для построения трёхмерных структур белков [42], предсказание структуры которых из первых принципов является крайне сложной задачей.



Рисунок 1. Принцип конформационного перебора методами дистанционной геометрии.

Было предложено множество алгоритмов конформационного поиска на основе ДГ [14,43–47]. Наиболее распространённым методом из этого класса является ETKDG (Experimental-Torsion + basic Knowledge + Distance Geometry) [14,45], доступный в библиотеке RDKit. ETKDG является примером того, как экспериментальные данные о предпочтительных торсионных углах можно включать в процесс семплирования методами ДГ; в нём также используется механизм контроля «вытянутости» получаемых конформаций циклов за счёт учёта парных расстояний между точками модельного эллипса с заранее заданным эксцентриситетом. По результатам сравнения алгоритмов конформационного поиска в работе [48], алгоритм ETKDG был выявлен лидером по воспроизведению конформаций лигандов, реализующихся в комплексах с белками. Также ДГ используется в коммерческом программном пакете OMEGA [44].

Таким образом, подход ДГ отлично подходит для конформационного перебора, когда многие межатомные расстояния известны наперёд, в частности, эмпирически известные торсионные углы в методе ETKDG или экспериментальные данные о

межатомных контактах при построении структуры белков. Однако в случае конформационного поиска гибких молекул без дополнительных экспериментальных данных, межатомные расстояния становятся координатами для перебора.

Декартовы координаты и межатомные расстояния являются неудачными способами параметризации конформационного пространства молекулы, потому что на них наложена система геометрических ограничений. Так как случайный выбор координат или межатомных расстояний приводит к сильно искажённым структурам, то требуется сравнительно долгая оптимизация для корректировки длин связей, валентных углов и расстояний между несвязанными атомами (замедление скорости семплирования ETKDG обсуждено в подразделе 3.5). Кроме того, систематический конформационный перебор затруднителен при использовании декартовых координат ядер или межатомных расстояний, так как их избыточно много даже для малых молекул ($3N - 6$ и $(3N - 6)(3N - 7)/2$, соответственно), и их значения не ограничены сверху (декартовы координаты также не ограничены снизу).

1.1.1.2. Приближение жёсткого ротатора. Перебор во внутренних координатах.

Во многих химических системах априори можно выделить степени свободы (СС), определяющие конформационную подвижность, то есть изменения этих СС достаточно для перехода по ППЭ в окрестность любого конформера с последующим его нахождением через локальную минимизацию.

Геометрия молекул задаётся тремя типами СС: длинами связей, валентными и торсионными углами; при этом конформационную подвижность обуславливают преимущественно торсионные углы. Как показано на Рисунке 2, потенциальные кривые, соответствующие длинам связей и валентным углам, имеют только один минимум, тогда как для торсионных углов типично существование нескольких минимумов. Кроме того, жёсткость (то есть кривизна потенциала вблизи минимума) резко возрастает в ряду: торсионные углы < валентные углы < длины связей (см. [49] с. 170–175). Это приводит к тому, что длины связей практически постоянны по конформационному набору, валентные углы варьируются в пределах порядка $\pm 10^\circ$, а торсионные углы, обладающие невысокой амплитудой потенциала, могут в

принципе принимать любые значения – для них имеет место лишь большая плотность распределения конформеров вблизи минимумов потенциала.

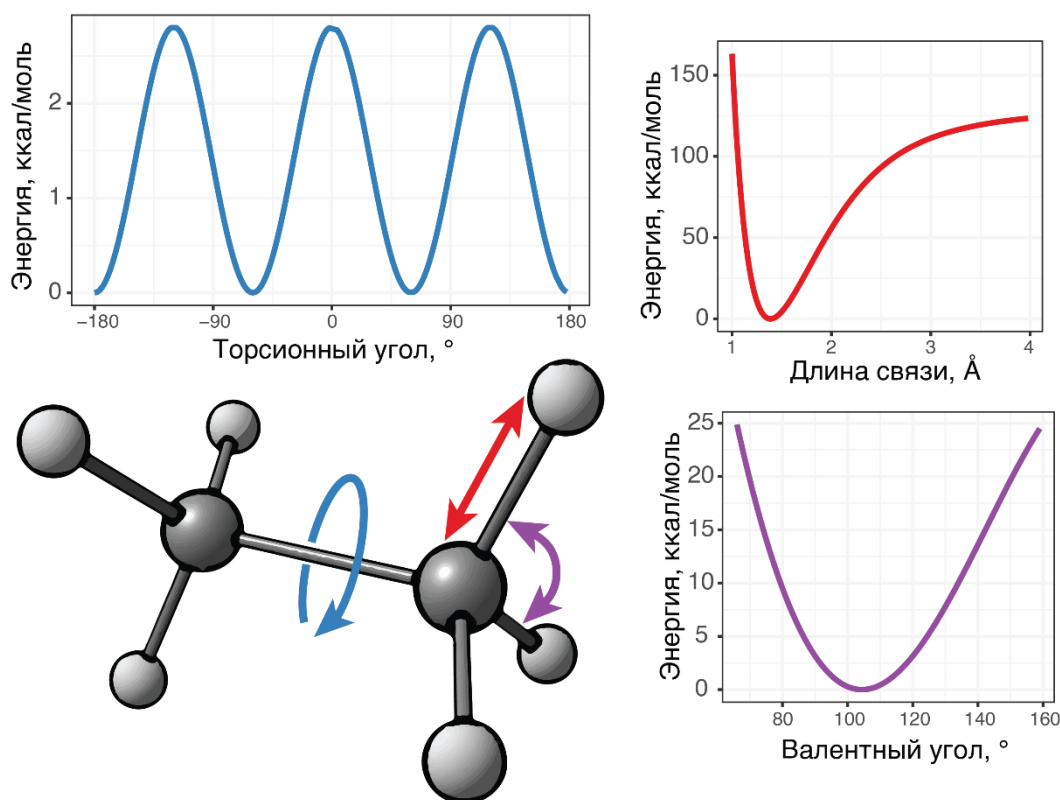


Рисунок 2. Характерные потенциальные кривые для длин связей, валентных и торсионных углов.

Подход к исследованию ППЭ путём изменения торсионных углов при фиксированных длинах связей, валентных углах и торсионных углах кратных связей называют приближением жёсткого ротатора. Другими словами, это является моделью химического конструктора – точечные атомы соединены между собой нерастяжимыми и негнуцимыми связями. Глобальный перебор конфигураций такого конструктора для интересующей молекулы можно реализовать алгоритмически, что и используется в ряде методов конформационного поиска – как случайного, так и систематического.

Подход к случайному перебору торсионных углов с последующей локальной оптимизацией был предложен в 1989 году и назван методом Монте-Карло для поиска многих минимумов (Monte Carlo conformational search for Multiple Minima, MCMM) [50]. С тех пор метод MCMM активно используется для перебора конформаций органических молекул и в качестве базового метода сравнения с новыми методами конформационного поиска. MCMM реализован, например, в программном пакете Schrödinger MacroModel [51] и в библиотеке OpenBabel [52].

Систематический перебор торсионных углов основан на построении сетки перебора по значениям торсионных углов, соответствующих заторможенным конфигурациям (см. положения локальных минимумов на Рисунке 2). Систематический перебор реализован, например, в библиотеке OpenBabel, однако этот подход применим только для малых молекул вследствие «экспоненциального взрыва» числа точек сетки перебора с увеличением количества подвижных торсионных углов. Поэтому был предложен ряд подходов к повышению эффективности систематического перебора. Например, методы SUMM [53] и Frog [54] используют переменное разрешение сетки для автоматического перехода к грубому перебору с большим шагом и меньшим количеством торсионных углов, если размеры молекулы не позволяют провести детальное семплирование.

Не все точки сетки перебора соответствуют стерически допустимым конформациям, поэтому была предложена стратегия рекурсивного построения конформаций для раннего исключения неперспективных подпространств сетки; например, если последовательность соседних торсионных углов делает фрагмент молекулы стерически затруднённым, то перебор остальных торсионных углов уже не требуется. Данный рекурсивный подход реализован, например, в методах CAESAR [55] и MULTIC [56]. Кроме этого, порядок обхода сетки тоже может иметь значение – при простейшем последовательном обходе и досрочном завершении поиска до рассмотрения всех точек ППЭ будет покрыта неравномерно. Поэтому применяются стратегии обхода сетки в псевдослучайном порядке, основанные, например, на методе латинского гиперкуба и регистрах сдвига. Это обеспечивает эффект случайного перебора на ранних этапах поиска, при этом к концу процедуры

вся сетка оказывается равномерно покрытой, и каждая точка посещена однократно. Подобные принципы используются в методах SUMM [53] и Confab [57].

У приближения жёсткого ротатора имеется ряд преимуществ, благодаря которым система координат торсионных углов используется в конформационном переборе чаще, чем декартовы координаты атомов: (1) меньшее количество степеней свободы для перебора и компактность пространства перебора (конфигурационное пространство одного торсионного угла гомеоморфно окружности, а значит конфигурационное пространство жёсткого ротатора из N торсионных углов гомеоморфно N -мерному тору), (2) система координат торсионных углов по своему построению учитывает ограничения на длины связей, валентные углы и жёсткие торсионные углы, поэтому проверка допустимости конформации сводится только к выявлению нежелательных стерических контактов.

Если молекула содержит циклические фрагменты, использование торсионных углов в качестве независимых переменных перестаёт быть тривиальной задачей. В этом случае произвольный выбор торсионных углов не гарантирует сохранения заданных длин связей и валентных углов внутри цикла. Следовательно, торсионные степени свободы для последовательности из $(N - 3)$ связей в цикле, состоящем из N атомов, не являются полностью независимыми; их изменения должны согласовываться с условием замкнутости цикла. Простейший подход к решению этой проблемы, известный как метод ringmaker [58], заключается в следующем: одна из связей в каждом цикле временно разрывается, оставляя только остовное дерево молекулярного графа; далее осуществляется случайный или систематический перебор торсионных углов для полученной ациклической структуры и для каждой сгенерированной геометрии проверяются условия на длины связей и валентные углы исходного цикла (Рисунок 3). Прошедшие проверку конформации подвергаются дальнейшей локальной оптимизации. Случайный перебор торсионных углов отсеиванием полученных геометрий по условию достаточной замкнутости циклов называется методом ringmaker Monte-Carlo (RMMC). Подход ringmaker также используется, например, в систематическом алгоритме MULTIC [56] и методе ConfBuster [59] на основе генетического алгоритма и систематического перебора,

однако в зависимости от порогов отбора конформаций, он приводит или к сильно искажённым конформациям, или к очень неэффективному перебору с долей выбросов на стадии отсеивания близкой к 100%.

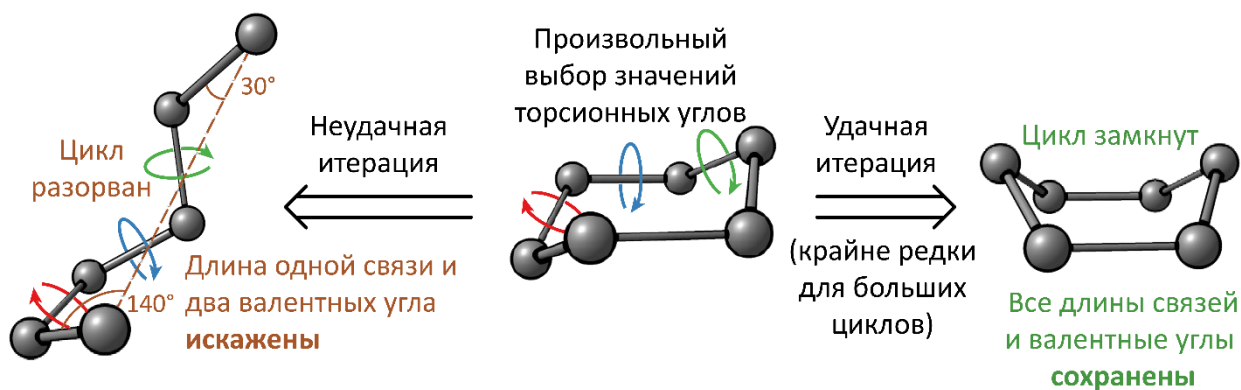


Рисунок 3. Иллюстрация подхода ringmaker построения конформаций циклических молекул через перебор торсионных углов с проверкой условий замыкания цикла.

Подход малоэффективен, так как большинство итераций приводят к сильно искажённым геометриям.

Для повышения эффективности конформационного перебора для циклических соединений был предложен ряд подходов: (1) корректировка торсионных углов для минимизации искажений цикла, реализованная в методе SUMM [53]; (2) генерация конформаций циклов методом дистанционной геометрии, выполняемая отдельно от основного перебора торсионных углов, что реализовано в методах Frog2 [60] и в CORINA [61] для генерации макроциков; (3) использование предварительно рассчитанных библиотек конформаций (поли)циклических фрагментов; (4) подход обратной кинематики (ОК), основанный на точном решении системы алгебраических уравнений, описывающих условия замкнутости цикла. В первом случае дополнительные шаги оптимизации снижают общую эффективность перебора. Второй подход наследует недостатки методов дистанционной геометрии, а именно сравнительно низкую эффективность и невозможность проведения систематического поиска. Методы, использующие структурные библиотеки, будут

рассмотрены в следующем подразделе, тогда как обсуждению подходов на основе ОК посвящён Раздел 3 данной диссертации.

1.1.1.3. Методы эмпирического перебора

Практически все методы конформационного поиска, независимо от их положения в классификации, опираются на те или иные эвристические предположения (элементы химической интуиции), позволяющие уменьшить вычислительную сложность задачи по сравнению с глобальной минимизацией энергии как «чёрного ящика». К числу таких эвристик можно отнести:

1) предположение о постоянстве молекулярного графа и стандартных значениях длин связей, валентных углов и торсионных углов кратных связей;

2) существование типовых торсионных углов для связей между атомами с заданной гибридизацией, например, для связи $sp^3 - sp^3$ характерны значения углов $\{\pm 60^\circ, 180^\circ\}$;

3) представление о том, что конформационные переходы происходят вдоль направлений низкочастотных колебательных мод и не требуют преодоления высоких энергетических барьеров, таких как барьеры разрыва химических связей;

4) предположение о существовании определённых геометрических мотивов (например, конфигураций циклов, межатомных контактов и т. п.), наличие которых делает конформеры более энергетически выгодными.

Первая эвристика наиболее важна для методов систематического и случайного перебора, а вторая эвристика используется при построении сеток при систематическом переборе. Третья и четвёртая эвристики лежат в основе конформационного поиска путём движения системы по ППЭ и аугментации ансамбля, соответственно (см. подразделы 1.1.2 и 1.1.3).

Особую группу составляют эмпирические алгоритмы – они основаны не на небольшом наборе встроенных в алгоритм эвристик, а на больших массивах структурных данных, заранее подготовленных на основе предварительных расчётов

или анализа баз экспериментальных данных. В качестве первичных источников экспериментальных структур наибольшую популярность имеют геометрии лигандов из базы данных белков (Protein Data Bank, PDB) [62] и конформации молекул в кристаллах из Кембриджской структурной базы данных (Cambridge Structural Database, CSD) [63]. Поскольку нельзя полагаться на то, что конформации интересующей молекулы имеются в структурных базах данных, применяется подход с выделением структурных фрагментов внутри молекулы. Если каждый фрагмент представлен в базах, то для каждого фрагмента можно построить модель условного распределения конформаций при фиксированном химическом окружении, которое встречается в целевой молекуле $P(q_{\text{фраг}} | \text{хим. окружение})$. Структуры фрагментов, $q_{\text{фраг}}$, представляются через координаты атомов, либо через торсионные углы (наборы характерных значений или гистограммы). Для построения конформационного ансамбля делается семплирование из условных распределений фрагментов с последующей сборкой полных конформаций – это также позволяет оценить частоту встречаемости конформаций из условных вероятностей конформаций фрагментов. Каждому фрагменту сопоставляются условия, при которых данный фрагмент считается релевантным для рассматриваемого химического окружения. Степень абстракции таких условий варьируется в зависимости от задач, на которые ориентирован метод конформационного поиска.

В середине спектра абстрактности лежат правила точного соответствия молекулярному подграфу с учётом типов атомов, кратностей связей и конфигураций стереоцентров. Такие наборы структур и правил часто называют библиотеками фрагментов. Характерным примером является метод VCL::Conf [64], база данных которого содержит около $2 \cdot 10^5$ фрагментов, для каждого из которых хранится одна или несколько конформаций, наиболее часто наблюдаемых в экспериментах. Работа с меньшим количеством более универсальных правил возможна с помощью структурных шаблонов, то есть подграфов с *частично* определёнными химическими данными. Наиболее распространённым языком для описания таких правил является SMARTS (SMILES Arbitrary Target Specification) [65], который позволяет независимо указывать или не указывать химический элемент, ароматичность/алифатичность, количество заместителей на атоме, кратность связи и т. п. Правила на основе

SMARTS используются в методах Conformer [66] и CONFORGE [67] для реализации систематического перебора ациклических торсионных углов, а также в алгоритме ETKDG [14,45] из библиотеки RDKit – для учёта экспериментальных торсионных предпочтений при построении матриц расстояний. С другой стороны, специализированные методы конформационного поиска могут использовать библиотеки с очень строгими правилами, например алгоритмы перебора боковых белковых цепей MC-SCE [68] и Rosetta [69] выполняют перебор из распределений торсионных углов, которые табулированы не только по аминокислоте, но и по величине двух ближайших торсионных углов остова (т.н. остов-зависимые библиотеки ротамеров [70]).

Стоит отметить, что библиотеки фрагментов и шаблонов, построенные на основе экспериментальных и расчётных структур, отлично подходят для конформационного поиска молекул с малыми циклами, потому что такие фрагменты имеют конечный набор конформаций, широко представленных в структурных базах. Не удивительно, что библиотеки циклических фрагментов или шаблонов очень часто включены в методы конформационного поиска, например, ConfGen [71], CONCORD [72], Conformer [66], CONAN [73], BCL::Conf [64], OMEGA [74], CONFORGE [67], SOS [40] и CORINA [61]. Ситуация существенно изменяется при переходе к макроциклам (около 9 и более атомов) – построение универсальных библиотек для данного класса соединений затруднено ввиду их высокой конформационной гибкости и отсутствия набора типовых геометрий, в то же время циклическое строение делает торсионные углы нестандартными и сильно коррелированными, что затрудняет перебор из априорных распределений. Показательно, что многие методы, успешно использующие библиотеки малых циклов, при работе с макроциклами и гибкими полициклическими системами обращаются к альтернативным подходам. Так, методы CORINA [61], OMEGA [74] и CONFORGE [67] применяют алгоритмы дистанционной геометрии (подраздел 1.1.1.1), а Conformer [66] пользуется подходом ringmaker [58] (подраздел 1.1.1.2). Интересно обратить внимание на эмпирический «суперструктурный» подход генерации структур полициклов в методе CORINA (Рисунок 4), который использует конформации малых полициклов для построения конформаций макрополициклов. Он включает следующие этапы: (1)

определение атомов в точках ветвления полициклического каркаса; (2) замена линейных фрагментов между этими атомами на фиктивные связи с получением суррогатного полицикла, содержащего меньшее число атомов; (3) генерация конформации суррогатного полицикла с использованием библиотеки фрагментов; (4) восстановление удалённых атомов через подстановку и оптимизацию геометрических ограничений.

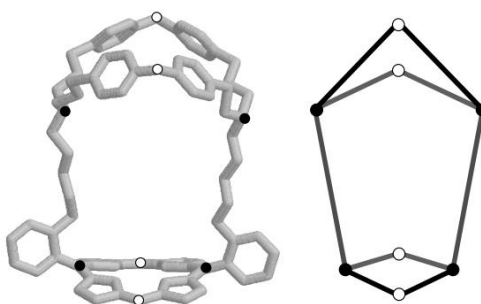


Рисунок 4. Принцип генерации приблизительных конформаций макрополициклов в методе CORINA.

Наконец, так как эмпирические методы по своей природе строят выборки из распределений структур фрагментов, то они все должны каким-то образом решать задачу сборки единой конформации из найденных геометрий фрагментов. Обычно эта задача делегируется другому алгоритму конформационного поиска, например, Conformer [66], OMEGA [74], CONFORGE [67] и ConfGen [71] выполняют систематический перебор вариантов сборки из найденных геометрий фрагментов и типовых значений торсионных углов. Метод ETKDG использует эмпирические данные для дальнейшей сборки конформаций алгоритмом дистанционной геометрии, а метод BCL::Conf находит низкоэнергетические комбинации структур фрагментов методом Метрополиса (см. подраздел 1.1.2.2) с использованием оценочной функции, учитывающей априорные вероятности геометрий фрагментов и стерическую затруднённость собранной конформации.

Предлагались и подходы к сборке конформаций, полностью основанные на эмпирических данных. Ярким примером является метод CONAN [73], в котором молекула разбивается на попарно пересекающиеся фрагменты, что позволяет делать сборку путём парного соединения только совместимых геометрий фрагментов. Однако данный подход не исключает возникновения стерических наложений фрагментов. Метод BCL::Conf также использует парные соединения фрагментов для получения первичной конформации, но для устранения геометрических конфликтов он использует вышеупомянутый алгоритм Метрополиса. Встречаются и приёмы ускоренной сборки конформаций из фрагментов без дополнительного поиска и перебора с использованием дополнительных эвристик, например, алгоритм CORINA выбирает ациклические торсионные углы таким образом, чтобы максимизировать расстояние между началом и концом самой длинной цепи в молекуле, а метод CONCORD добавляет фрагменты один за другим и на каждом этапе оптимизирует новый торсионный угол для минимизации стерической загруженности.

Эмпирические методы конформационного поиска показывают высокую эффективность генерации конформеров, структурно близких к экспериментально наблюдаемым геометриям. Так, в работе [75] показано, что эмпирический метод дистанционной геометрии ETKDG, реализованный в пакете RDKit, является одним из лидеров среди алгоритмов конформационного поиска с открытым исходным кодом по частоте воспроизведения экспериментальных структур. А согласно тестам в [76], эмпирический метод OMEGA превосходит другие коммерческие программы конформационного поиска. Тем не менее, существующие эмпирические подходы не подходят для создания универсальных методов конформационного поиска макроциклов по причинам ограниченной подвижности их торсионных углов и недоступности наборов типовых конфигураций.

1.1.2. Методы движения по поверхности потенциальной энергии

В рассмотренных выше методах перебора каждый следующий шаг не зависит от результатов расчёта энергии и минимизации на предыдущем шаге. В свою очередь, методы построения термодинамических ансамблей получают траекторию движения

по ППЭ путём отсчёта новой точки от предыдущей по определённому правилу, другими словами, путём моделирования динамической системы. Чтобы такая процедура позволяла преодолевать энергетические барьеры конформационных переходов, но при этом преимущественно находила низкоэнергетические конформеры, используются две стратегии перемещения по ППЭ (Рисунок 5): (1) короткие детерминированные шаги, соответствующие численному интегрированию уравнений молекулярной динамики (МД), (2) более крупные случайные шаги с критерием принятия шага, благоприятствующим перемещению в более энергетически выгодные области, но допускающим повышение энергии для преодоления кинетических барьеров (подраздел 1.1.2.2). В методах из первой группы кратковременное повышение энергии в ходе МД возможно благодаря переносу энергии между кинетической и потенциальной формами, а предпочтение низкоэнергетических областей возникает вследствие законов статистической механики. В методах из второй группы, которые называются методами Монте-Карло с марковскими цепями, возможность преодоления барьеров контролируется критерием принятия шага, из которых наиболее распространённым является критерий Метрополиса. Согласно этому критерию, все понижающие потенциальную энергию шаги принимаются, а повышающие энергию шаги принимаются случайным образом с вероятностью $\exp(-\Delta V/T)$, где ΔV – это положительная разница между энергиями новой и предшествующей конформацией, а T – приведённая температура моделирования. Методы Монте-Карло с марковскими цепями, использующие критерий Метрополиса, называются методами Монте-Карло Метрополиса (МКМ, Metropolis Monte Carlo).

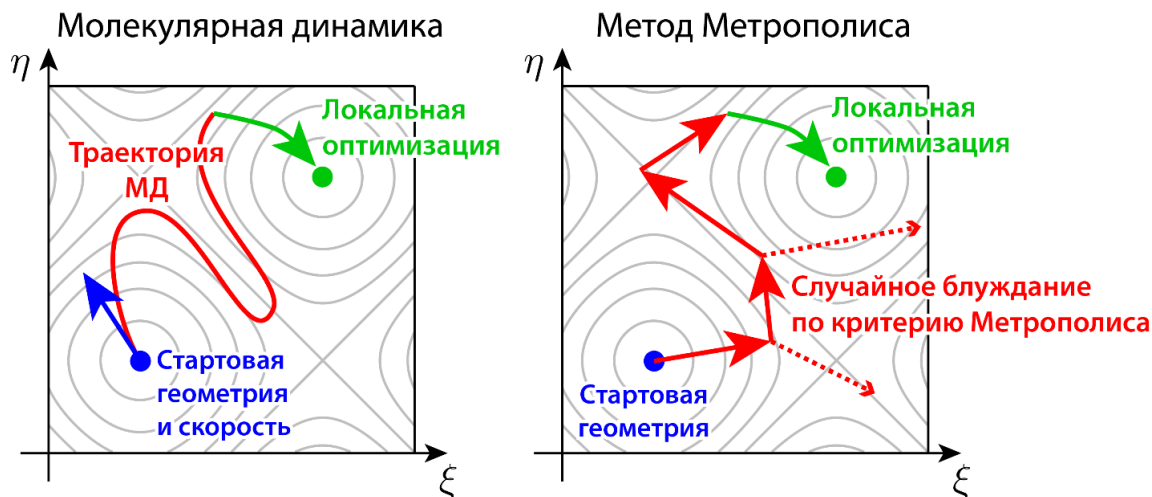


Рисунок 5. Сравнение движения по ППЭ методами молекулярной динамики и Метрополиса.

Несмотря на принципиальные отличия подходов МД и МКМ, они оба в пределе бесконечно долгого моделирования генерируют репрезентативную выборку из канонического ансамбля исследуемой системы при температуре термостата в случае МД и температуре T критерия Метрополиса в случае МКМ, соответственно. Более того, как МД, так и МКМ имеют проблему кинетической блокировки (kinetic trapping): если достижение глобального минимума требует преодоления высокого энергетического барьера ΔV , то МД будет пересекать такой барьер с экспоненциально малой частотой $\exp(-\Delta V/kT)$ вследствие необходимости накопления достаточной кинетической энергии для перехода через седловую область ППЭ, а МКМ – принимать соответствующие повышающие энергию шаги с той же экспоненциально малой вероятностью, определяемой критерием Метрополиса. Как будет показано ниже, проблема кинетической блокировки решается применением подходов ускоренного семплирования, которые, как оказывается, имеют много общего между ускоренными вариантами МКМ и МД (Таблица 1).

Таблица 1. Сравнение наиболее распространённых методов конформационного поиска через движение по ППЭ.

Метод конф. поиска	Тип шага	Подходы к ускорению				
		Избирательное направление	Отжиг	Сглаживание потенциала	История в потенциале	Обмен репликами
Классическая МД	МД					
Смягчение скоростей		+				
МД с отжигом			+			
МД с выталкиванием				+		
Local elevation ¹					+	
Basin hopping	МКМ					
Low mode search		+				
Симуляция отжига			+			
Табу-поиск					+	
T-REMC				+		+
T-REMD	МД + МКМ		+			+
H-REMD				+		+

МД – молекулярная динамика; МКМ – метод Монте-Карло Метрополиса;

T-REMC – метод Монте-Карло с марковскими цепями с температурным обменом репликами;

T-REMD/H-REMD – МД с температурным/гамильтоновым обменом репликами;

¹ – к этой группе относится популярный метод MTD, используемый в пакете CREST.

1.1.2.1. Методы на основе молекулярной динамики

Родоначальником этой группы методов является, собственно, молекулярная динамика (МД). Для вычисления потенциальной энергии в данной группе методов обычно используется приближение молекулярной механики. Расчёт начинается в некотором локальном минимуме ППЭ со случайными скоростями на атомах, которые отмасштабированы в соответствии с выбранной температурой. Последующая эволюция системы во времени (траектория движения на ППЭ) определяется путём численного интегрирования уравнений классической механики с типичным шагом интегрирования порядка $1 \text{ фс} = 10^{-15} \text{ с}$. Соответствие получаемого ансамбля выбранной температуре обеспечивается использованием термостата – алгоритма, корректирующего скорости атомов для поддержания требуемой среднеквадратичной кинетической энергии. На завершающем этапе строится конформационный набор путём отбора структур из накопленной траектории, их оптимизации и удаления дубликатов. МД широко применяется для релаксации и исследования

низкоэнергетических минимумов для биологических макромолекул, в том числе белков [77,78].

Однако практическое применение МД для моделирования конформационной релаксации при целевой температуре ограничено: время, необходимое для преодоления энергетических барьеров, зачастую оказывается недостижимым даже в приближении молекулярной механики. В промежутках между конформационными переходами система длительное время находится в окрестности одного и того же локального минимума, что неэффективно расходует вычислительные ресурсы. Одним из часто используемых способов ускорения структурных переходов является отжиг – МД при повышенной температуре. Увеличение среднеквадратичной кинетической энергии атомов облегчает преодоление потенциальных барьеров, что делает отжиг часто используемым методом релаксации напряжённых структур [79]. Однако, отжиг имеет существенные недостатки. Во-первых, часть кинетической энергии переходит в колебание связей и валентных углов, что делает структуры в среднем более искажёнными. Во-вторых, повышение температуры нивелирует статистические предпочтения низкоэнергетических областей ППЭ (по распределению Больцмана), снижая вероятность получения энергетически выгодных конформеров при последующей оптимизации точек траектории. Проблема пребывания системы в искажённом состоянии решается плавным понижением температуры в ходе отжига или применением алгоритма SHAKE [80], который позволяет фиксировать степени свободы молекулы, не участвующие в конформационных переходах, например, длины ковалентных связей [81]. Однако, проблема сглаживания энергетических предпочтений остаётся.

Другим параметром, который можно настроить для улучшения эффективности конформационного поиска с помощью МД, являются начальные скорости атомов. Это реализуют алгоритмы МД со смягчением скоростей. Основываясь на предположении, что конформационные переходы происходят вдоль направлений низкой кривизны ППЭ, начальные скорости выбираются таким образом, чтобы минимизировать вклад высокочастотных колебательных мод. Простейшая реализация смягчения скоростей потребовала бы явного вычисления собственных

значений и векторов Гессиана (матрицы вторых производных потенциальной энергии), что было бы очень ресурсозатратно для больших систем. Поэтому на практике используются приближённые методы, основанные на генерации случайных скоростей с последующим подавлением высокочастотных компонент. Это достигается либо через минимизацию вращательных сил, действующих на пару точек на ППЭ (Рисунок 6) [82], либо с применением спектрального фильтра Чебышева, построенного на фиксированном числе производных градиента энергии вдоль случайных направлений, что реализовано в алгоритме LowModeMD [83]. Принципиальных отличий между этими подходами нет; первый был предложен в области предсказания структур кристаллов и атомных кластеров, а второй используется в моделировании биохимических систем [83,84]. Интересно, что первый подход аналогичен методу «димера» [85] – алгоритму поиска переходных состояний, не требующему вычисления вторых производных. Этот метод также работает в удвоенном конфигурационном пространстве (одновременно оптимизируются две близко расположенные точки на ППЭ) и выполняет поиск направления наименьшей кривизны путём минимизации ортогональных сил.

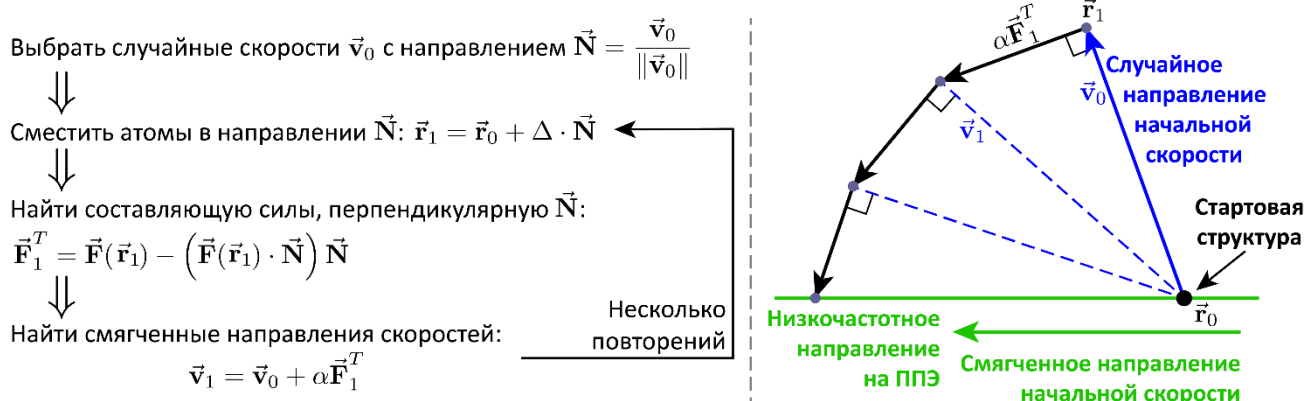


Рисунок 6. Схема алгоритма смягчения скоростей (слева). Итеративное перемещение пробной точки по ППЭ для расчёта смягченного направления скоростей (справа). \vec{F}_1^T – вращательная сила, \vec{v}_0 – случайные скорости, \vec{v}_1 – смягчённые скорости.

Еще один элемент алгоритма МД, допускающий модификацию для ускорения конформационных переходов – это, собственно, ППЭ. Модификации потенциалов обычно направлены на снижение отталкивания ядер при минимальном влиянии на положения и энергии минимумов ППЭ. Поскольку отдельные слагаемые силовых полей, как правило, имеют физическое обоснование, стандартным решением является снижение вкладов дальних взаимодействий – кулоновских и ван-дер-ваальсовых. Важно отметить, что потенциалы обоих типов взаимодействий обладают сингулярностью при нулевом межатомном расстоянии, $\lim_{r \rightarrow 0} V(r) = +\infty$. Соответственно, подходы к снижению этих потенциалов можно разделить на две группы: (1) с сохранением сингулярности и (2) с её устранением, то есть энергетический вклад от полного наложения ядер делается конечным. Потенциалы второго типа называются мягкими (soft-core potentials). Возможность контроля «мягкости» этих потенциалов (Рисунок 7) позволяет выполнять МД с выталкиванием (push-off MD), то есть постепенным понижением мягкости потенциала, что используется для удаления стерического напряжения в сгенерированных атомистических моделях, например полимеров [86,87].

Применительно к системам малого и среднего размера широкое распространение получил подход, основанный не на снижении стерического отталкивания по всей ППЭ, а на «вытеснении» системы из ранее посещённых областей путём добавления смещающего потенциала V_{bias} , который модифицирует геометрию ППЭ только в уже посещённых областях. Данный метод известен как молекулярная динамика с локальным подъёмом потенциала (local elevation, LE). Ключевыми элементами LE являются: (1) выбор геометрической метрики для оценки подобия конформаций, (2) выбор аналитической формы вкладов в V_{bias} . Форма V_{bias} должна обеспечивать возможность быстрого вычисления как самого потенциала, так и его вклада в силы ($-\nabla V_{bias}$). Стандартным выбором является гауссова форма потенциала:

$$V_{bias}(\mathbf{s}, \mathbf{s}_{ref}) = h \exp\left(-\alpha \cdot d(\mathbf{s}, \mathbf{s}_{ref})^2\right), \quad (1)$$

где \mathbf{s} и \mathbf{s}_{ref} – это текущая геометрия и геометрия, от которой производится отталкивание, $d(\mathbf{s}, \mathbf{s}_{ref})$ – это коллективная переменная, то есть метрика подобия двух структур, а h и α – это эмпирические параметры, отвечающие за дальность и максимальную энергию одного вклада.

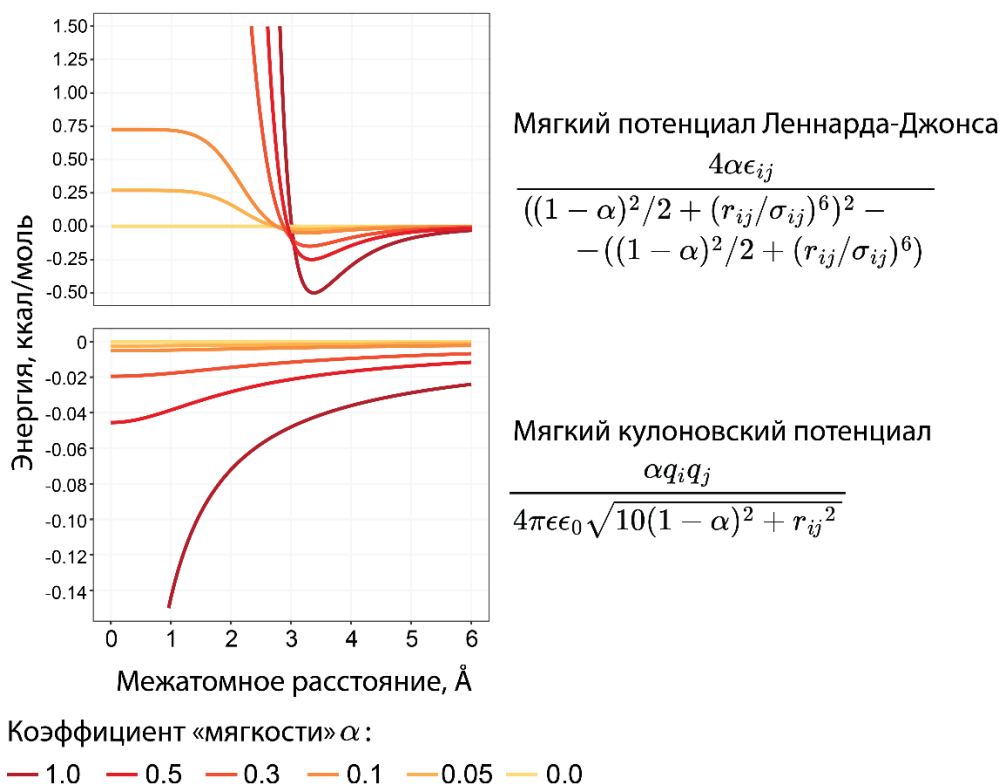


Рисунок 7. Мягкие потенциалы, используемые в пакете LAMMPS. Для контроля мягкости используется коэффициент $\alpha \in [0,1]$. Граничные значения соответствуют стандартному «жесткому» потенциалу при $\alpha = 1$ и константно нулевому – при $\alpha = 0$. При $\alpha < 1$ сингулярность в нуле отсутствует. Графики адаптированы из [87].

В качестве метрики предлагалось использование торсионных углов и межатомных расстояний [88], однако наибольшую популярность для конформационного поиска молекул получил алгоритм LE с использованием среднеквадратичных отклонений координат атомов (root-mean-square deviation, RMSD, см. подраздел 1.2.1), реализованный в пакете ХТВ [89] и используемый в популярном методе конформационного поиска CREST [27]. Стоит отметить, что

разработчики назвали данный метод «метадинамикой» (MTD), подчёркивая его родство с одноимённым методом построения модельных поверхностей свободной энергии, однако сам метод MTD в ХТВ и CREST рассчитан именно на семплирование ППЭ и, формально говоря, метадинамикой не является. Структуры сравнения для V_{bias} отбираются из траектории MTD через фиксированные интервалы времени (каждую 1 пс), что обеспечивает отталкивание от ранее посещённых областей ППЭ (Рисунок 8).

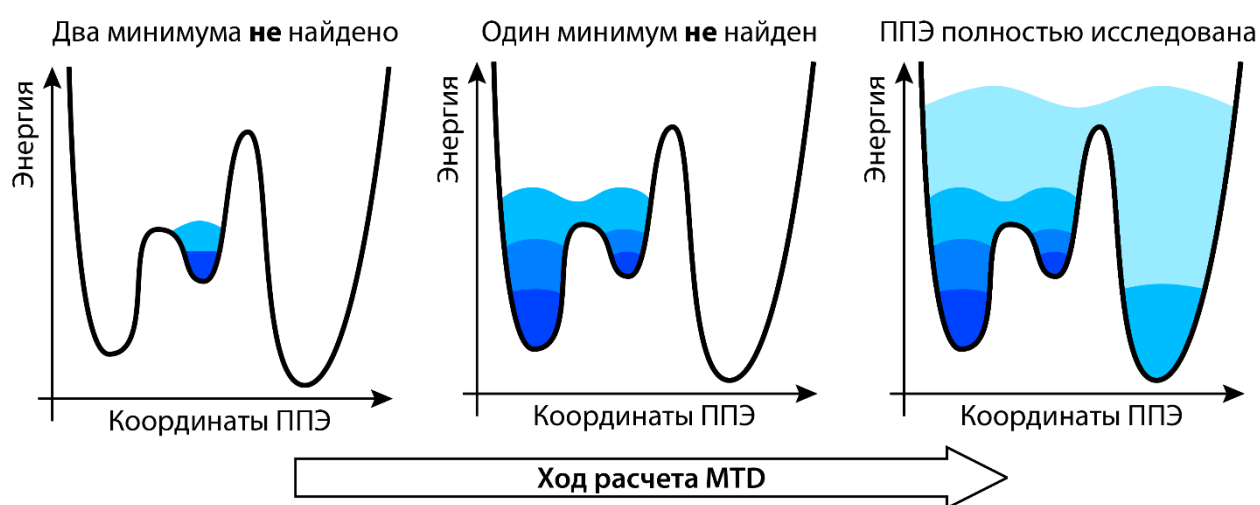


Рисунок 8. Иллюстрация принципа локального поднятия потенциала, в частности метода MTD, используемого в пакетах ХТВ и CREST. Чёрная кривая показывает профиль исследуемой ППЭ, а цветом показан смещающий потенциал, который накапливается в уже исследованных областях.

При всех достоинствах методов МД они исследуют ППЭ путём совершения множества мелких шагов, что требует значительного числа вычислений градиентов. Альтернативный подход на основе крупных случайных шагов обсуждён в следующем подразделе.

1.1.2.2. Методы случайных блужданий на основе критерия Метрополиса

Методы Монте-Карло Метрополиса (МКМ) потенциально позволяют более эффективно исследовать ППЭ по сравнению с методами МД, поскольку они не ограничены в длине шагов и способны перемещаться на большие расстояния, рассматривая существенно меньше точек. Однако, и МД, и МКМ могут сталкиваться со сходными трудностями из-за кинетической блокировки (см. начало подраздела 1.1.2). В МД это проявляется, когда значительная часть вычислительного времени расходуется на интегрирование траектории, остающейся вблизи одного или нескольких минимумов. А в МКМ вследствие кинетической блокировки генерируются неперспективные случайные шаги, которые отклоняются по критерию Метрополиса. Однако, в отличие от детерминированной динамики МД, стохастическая постановка задачи в МКМ предоставляет большую свободу в выборе шагов. Поэтому значительная часть усилий по разработке таких методов направлена на создание операторов перемещения, которые с большей вероятностью направляли бы систему в энергетически перспективные области ППЭ.

Так как для химических систем локальная оптимизация геометрии выполняется с относительно небольшими вычислительными затратами, Wales и Doye заметили, что эффективность алгоритма МКМ можно повысить, выполняя минимизацию энергии после каждого пробного шага. Был предложен алгоритм перескоков между бассейнами (basin hopping, ВН) [90], где под бассейном понимается область ППЭ, все точки которой при локальной оптимизации сходятся к одному и тому же минимуму. Алгоритм ВН состоит из двух чередующихся шагов: (1) случайный шаг, (2) локальная оптимизация полученной геометрии, приводящая систему к минимуму соответствующего бассейна. Критерий Метрополиса при этом применяется к энергии после оптимизации. Как замечают авторы, случайные шаги фактически выполняются не по исходной ППЭ, а по кусочно-постоянной функции, которая сопоставляет каждой конформации энергию минимума её бассейна. Wales и Doye применили данный метод для поиска глобальных минимумов кластеров Леннарда–Джонса, где использовались два типа случайных шагов: (1) случайное смещение всех атомов кластера на небольшую величину; (2) перенос атома с наименее выгодными

взаимодействиями с остальными атомами в случайную точку на сфере того же радиуса относительно центра масс кластера.

Аналогично алгоритмам глобального перебора для молекул, которые оказываются более эффективны в координатах торсионных углов (см. подраздел 1.1.1), алгоритмы МКМ также часто реализуют случайные шаги посредством изменения торсионных углов. Например, подход ВН применялся для конформационного поиска малого пептида [Met⁵]-энкефалина, где шаги заключались в случайном изменении одного произвольного подвижного торсионного угла [91]. Аналогично, при моделировании жидкого гексана предлагалось использовать МКМ без этапа минимизации [92]. Шаги включали трансляцию и поворот одной случайно выбранной молекулы как целого, небольшое изменение всех её торсионных углов, а также случайное изменение объёма ячейки моделирования.

Методы эмпирического перебора для крупных систем зачастую используют МКМ для поиска выгодных комбинаций экспериментальных геометрий фрагментов. Как отмечалось в подразделе 1.1.1.3, в таких подходах критерий Метрополиса может также учитывать частоты встречаемости в структурных базах данных. Таким образом, предпочтительность конформации определяется не только её энергией, но и распространённостью составляющих её геометрий фрагментов. Этот принцип реализован в вышеупомянутом методе BCL::Conf [64], а также в специализированных методах построения боковых цепей пептидов, например OPUS-Rota [93].

Многие приёмы, использующиеся для ускорения семплирования методами МД, имеют близкие аналоги среди методов МКМ. Так, подходу смягчения скоростей соответствует аналогичный подход поиска вдоль низких частот колебаний или их произвольной линейной комбинации (low-mode search). Его эффективность была показана как для конформационного поиска малых молекул [94,95], так и для крупных белковых систем [96,97].

Вариант МКМ с применением отжига называется методом симуляции отжига – его суть состоит в постепенном понижении значения температуры в критерии Метрополиса, таким образом плавно переносится баланс с глобального

исследования ППЭ на локальную оптимизацию. SA предлагался, например, для конформационного поиска малых молекул [98,99], пептидов [100,101] и кластеров Леннарда-Джонса [102].

Аналогично подходу LE, существуют варианты МКМ, добавляющие зависимость критерия Метрополиса от истории расчёта. Так, для конформационного поиска малых молекул предлагался табу-поиск [103] – в отличие от LE, табу-поиск не использует смещающего потенциала, а вместо этого случайные шаги в уже исследованные области ППЭ просто запрещаются. Кроме того, для конформационного поиска пептидов и белков предлагался МКМ с адаптирующимся смещающим потенциалом (Monte Carlo Minimization with an Adaptive Bias) [104,105]. Такой потенциал позволяет по ходу расчёта моделировать вероятностные распределения для торсионных углов на основе ранее принятых низкоэнергетических структур, включая парные корреляции между торсионными углами.

Уникальный принцип генерации случайных шагов реализован в методе конформационного поиска GOAT [106]. Каждый шаг включает выбор случайного направления, подъём по ППЭ вдоль антиградиента энергии, отражённого относительно выбранного направления, и последующую минимизацию. Интересно, что подход, основанный на отражении градиента, ранее применялся в алгоритмах поиска переходных состояний, например, в методе Newton trajectory scan [107], а также в универсальных методах поиска седловых точек, таких как gentlest ascent dynamics [108]. Помимо этого, в GOAT реализованы автоматическая настройка температуры и механизм памяти, аналогичный табу-поиску, который предотвращает повторное посещение минимумов.

Случайные шаги по критерию Метрополиса играют ключевую роль в методах обмена репликами (replica-exchange, RE). Отличительной особенностью этих методов является работа в расширенном конфигурационном пространстве, представляющим собой декартовую степень Q^n , где Q – это конформационное пространство и $n \in \mathbb{N}$. Иными словами, на каждом шаге траектории одновременно эволюционируют n точек на ППЭ, которые называются репликами. Для каждой из n

позиций в Q^n , называемых «мирами», задаются различные параметры МД или МКМ. Так, в МД с обменом репликами по температуре (T-REMД) в мирах устанавливается разная температура ($T_i < T_{i+1}$), что задает различную степень отжига при временной эволюции. Связь между мирами обеспечивается через шаги обмена реплик между соседними мирами по критерию Метрополиса [109,110]:

$$P(i \leftrightarrow j) = \min(1, e^{\Delta}),$$

$$\Delta = (\beta_i - \beta_j) (V(q_i) - V(q_j)), \quad j = i + 1, \quad \beta = \frac{1}{kT} \quad (2)$$

При этом вместе с координатами атомов обмениваются и их скорости с соответствующим температурным масштабированием. Таким образом, подход REMD объединяет шаги МД и МКМ так, что невыгодные конформации с высокой энергией в результате обменов оказываются в мирах с высокой температурой, где они способны быстрее преодолевать конформационные изменения и понижать свою энергию. Выгодные конформации с низкой энергией, напротив, перемещаются в миры с низкой температурой и релаксируют к ближайшему минимуму (Рисунок 9). Аналог T-REMД, использующий шаги МКМ вместо МД, известен как параллельный отжиг (parallel tempering или T-REMC).

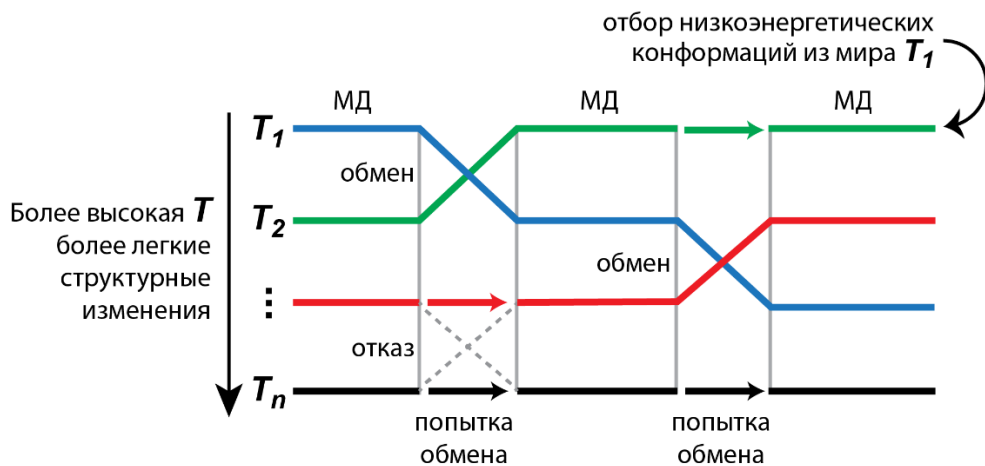


Рисунок 9. Суть метода обмена репликами T-REMД.

Было замечено, что для эффективного обмена между мирами в T-REMD необходимо хорошее перекрытие их энергетических распределений. Это создаёт ограничение: количество миров, требуемое для охвата одного и того же температурного диапазона, оказывается пропорционально квадратному корню из числа степеней свободы системы [111]. Вследствие этого расчёт необходимого числа реплик становится слишком ресурсозатратным для больших систем. Для преодоления этой проблемы был предложен метод гамильтонова обмена репликами (Hamiltonian REMD, H-REMD). В этом подходе все миры имеют одинаковую температуру, но различаются параметрами потенциальной энергии. Например, существуют реализации H-REMD, в которых варьировались следующие параметры: заряды атомов и параметры потенциала Леннарда-Джонса [112]; вклад энергии, отвечающий за гидрофобность, максимальные энергии водородных связей и вандер-ваальсовых взаимодействий [111]; масштабирующие множители для вкладов смещающих потенциалов торсионных углов [113]; а также специальные параметры, регулирующие «мягкость» атомов [87,114,115].

Таким образом, методы МД и МКМ предоставляют широкий инструментарий для исследования низкоэнергетических областей ППЭ. Полное устранение кинетической блокировки недостижимо, однако существует ряд подходов, направленных на ускорение семплирования, – от имитации отжига до адаптивных смещающих потенциалов. С точки зрения работы с циклическими молекулами, МД не имеет никаких дополнительных ограничений. В то же время многие методы МКМ сталкиваются с принципиальными трудностями, поскольку один из наиболее эффективных типов шагов для молекул – варьирование торсионных углов – для циклов оказывается неприменимым.

1.1.3. Методы аугментации ансамблей

Методами аугментации конформационных ансамблей мы называем алгоритмы, на каждой итерации которых делается попытка предсказания новой перспективной конформации по набору геометрий и энергий уже известных конформеров [116].

В качестве простейшего подхода к поиску более глубокого минимума, упущенного в уже известном ансамбле, можно предложить применение локальных сдвигов по типу методов МКМ к наиболее низкому из известных минимумов. Так, например, метод LLMOD [117], реализованный в пакете MacroModel, применяет сдвиги вдоль низкочастотных мод. Кроме этого, описанные выше методы с использованием смещающих потенциалов, а именно local elevation и МКМ с адаптивным потенциалом, тоже отчасти являются методами аугментации ансамбля, так как они направляют динамику или блуждание в сторону малоисследованных и более энергетически перспективных областей ППЭ, соответственно. Однако, эти методы остаются главным образом методами совершения одного шага из одной предшествующей геометрии.

В данном подразделе рассмотрены более сложные подходы к аугментации: эволюционные алгоритмы и Байесовская оптимизация.

1.1.3.1. Эволюционные алгоритмы

Эволюционные алгоритмы являются классическим методом аугментации ансамблей, которые работают на «популяции» конформаций и итеративно выполняют над ней операции, нацеленные на использование перспективных характеристик имеющихся конформаций для поиска более низкоэнергетических. Для этого задаётся набор эволюционных операторов, которые случайным образом генерируют новую структуру из одной или двух уже имеющихся. Эволюционные операторы, использующие одну структуру, часто называют «мутациями» (например, сдвиг одного атома), а на двух структурах – «скрещиваниями» (например, сборка геометрии из структурных частей родительских конформеров). Продолжая аналогию с эволюцией в природе, многие методы используют «хромосомное» представление конформаций в виде строк чисел, что позволяет использовать стандартные операторы мутации – случайное изменение одного или нескольких значений в «хромосоме», и стандартное скрещивание – разрез родительских векторов в одном и том же месте и сшивание частей.

Для задач конформационного поиска атомных кластеров [118–120] многие эволюционные алгоритмы используют в качестве мутаций случайные сдвиги одиночных атомов или перестановки пар атомов разных химических элементов. В качестве оператора скрещивания часто используется операция разреза и соединения (cut-and-splice) – обе родительские структуры разрезаются на два фрагмента вдоль случайных плоскостей, а затем дочерние структуры собираются из полученных частей после корректировки стехиометрии.

Конформации молекул, как правило, представляются в виде строк значений их подвижных торсионных углов; например, этот подход в комбинации со стандартными операторами скрещивания и мутации использовался в методе Fafoom [121] и алгоритме Sun для предсказания структур белков и пептидов [122]. Более того, если делается эволюционный поиск для системы из нескольких молекул, то в строку добавляются также координаты центров масс и повороты составляющих молекул, например, это используется в алгоритме конформационного поиска переходных состояний QMTSDock [123] и алгоритме предсказания структур молекулярных кристаллов [124]. Для эволюции конформаций молекул предлагались и нестандартные координаты, например был предложен бинарный генетический алгоритм [125], который кодирует конформации малых циклов через направление отклонения атомов от плоскости кольца и способен контролировать эти отклонения через алгоритм *corner flapping* [126], а обобщение этого подхода используется для мутации конформаций циклов в алгоритме CONFLEX [127].

Наибольшую популярность эволюционные алгоритмы получили в области предсказания структуры кристаллов. Современные методы, например USPEX (Universal Structure Predictor: Evolutionary Xtallography), воспроизводят экспериментальные структуры с очень высокой долей успеха [128,129]. Как и для атомных кластеров, для периодических систем широко распространено скрещивание методом cut-and-splice, а для мутации структур используются обмены атомов разных элементов, а также сдвиги атомов и изменения параметров ячейки. Предлагались также и эволюционные операции, аналогичные шагам в методах МД и МКМ. Например USPEX [129] и XtalOpt [130] реализуют шаги вдоль низкочастотных

колебательных мод, при этом USPEX также учитывает близость новых структур с уже найденными, для чего используется аналог гауссова смещающего потенциала в методах LE.

В начале каждой итерации эволюционного алгоритма необходимо делать выбор родительских конформаций для выполнения над ними эволюционных операций. Такой выбор делается случайным образом, по аналогии с естественным отбором, используя оценки «приспособленности» конформации, которые варьируются в зависимости от желаемого соотношения между локальной минимизацией энергии и глобальным исследованием ППЭ. Приспособленность может быть просто больцмановским весом конформации в уже известном наборе, а может также учитывать исследованность области ППЭ, которой эта геометрия принадлежит – второй вариант реализован в алгоритме GAtor [131].

Однако, эволюционные алгоритмы не являются идеальным подходом к аугментации ансамблей. В частности, при кодировании конформаций циклических или разветвлённых молекул значениями торсионных углов, топология отрезка у «хромосом» никак не отражает молекулярной топологии, соответственно, скрещивание путём разрезания и сшивания «хромосом» не даёт преимущества перед обменом значений торсионных углов в случайных положениях. Кроме того, стабилизирующие внутримолекулярные контакты не могут быть явно указаны через строки торсионных углов. Поэтому представляет интерес построение статистической модели ППЭ на лету, что и делают методы из следующего подраздела.

1.1.3.2. Байесовская оптимизация

Байесовская оптимизация (БО) – это метод оптимизации функций, вычисление которых является вычислительно затратным. Основная идея БО заключается в построении суррогатной модели целевой функции на основе уже известных значений в некоторых точках. Такая модель позволяет предсказывать как ожидаемое значение функции в произвольной точке, так и дисперсию: вблизи известных точек дисперсия мала, а в удалённых областях – высока. Ключевыми компонентами

любого алгоритма БО являются структура суррогатной модели и функция выбора (acquisition function, AF), определяющая целесообразность следующего вычисления целевой функции в той или иной точке.

Наиболее распространённой суррогатной моделью для конформационного поиска малых молекул являются гауссовы процессы, в частности она используется в методах BOA [132], BOKEI [133], BOSS [134] и BOCSER [116]. Во всех указанных методах конформации представлены в координатах их торсионных углов. В качестве AF применялись как стандартные expected improvement (EI), ориентированная на локальный поиск минимума, и lower confidence bound (LCB), нацеленная на глобальное исследование ППЭ [132], так и специализированная функция explorational variance minimizer (EVM) [116], предназначенная для глобального поиска в заданном энергетическом диапазоне.

$$\begin{aligned}
 EI(x) &= \mathbb{E}[\max(I(x), 0)] = \mathbb{E}[\max(f(x^*) - f(x), 0)] \\
 LCB(x) &= \mu(x) - \beta\sigma(x) \\
 EVM(x) &= V[\max(f(x^*) - f(x) + E_{thr}, 0)]
 \end{aligned}
 \tag{3}$$

Здесь x^* – наилучшая точка, обнаруженная на данный момент, $\mu(x)$ – предсказанное среднее значение в точке x , $\sigma(x)$ – предсказанное стандартное отклонение в точке x , β – коэффициент, управляющий соотношением между исследованием и минимизацией, а E_{thr} – верхняя граница энергетического диапазона, относительно минимума.

Одним из ключевых направлений развития методов БО в конформационном анализе является физическое информирование [135], то есть разработка способов учёта априорной информации о структуре ППЭ с целью повышения точности модели при ограниченных данных. Так, в методе BOKEI [133] учитываются парные корреляции торсионных углов, полученные из предварительных двумерных сканирований модельных соединений, однако эта информация включена только в AF, но не в саму суррогатную модель. Этот недостаток преодолён в методе BOCSER [116], где априорные знания о потенциалах торсионных углов (без учёта парных взаимодействий) закладываются непосредственно в функцию ядра на основе

предварительных одномерных сканирований. Кроме того, во всех перечисленных методах используется периодическая функция ядра, что позволяет учитывать цикличность зависимости энергии от торсионных углов молекулы.

Однако, гауссовы процессы становятся слишком вычислительно затратными с увеличением числа структур в ансамбле (временная сложность составляет $O(N^3)$, [136]). Для решения этой проблемы было предложено использовать другую суррогатную модель – древесно-структурированный оценщик Парзена (Tree-structured Parzen Estimator, TPE). В отличие от гауссовских процессов, которые моделируют целевую функцию напрямую, вычисляя ожидаемое значение и неопределённость, подход TPE разделяет наблюдения на «удачные» и «неудачные» и для них строит модели распределений. Временная сложность при этом понижается до $O(N \cdot \log N)$. В работе [137] TPE применяется для конформационного поиска малых молекул с разбиением по энергии удачных и неудачных конформаций в соотношении 1:9. Было показано, что подход TPE требует меньше расчётов энергии, чем методы ВН и эволюционные алгоритмы [137].

Существенным ограничением рассмотренных выше методов БО является их неспособность работать с циклическими молекулами. Это наследуется от используемой в них системы координат торсионных углов. В рамках такой системы координат возможно либо полное исключение циклических торсионных углов из рассмотрения, либо применение стратегии ringmaker (см. подраздел 1.1.1.2), то есть разрыв одной связи в каждом цикле и использование торсионных углов оставшихся связей в качестве координат для оптимизации. Второй подход не является удовлетворительным, так как он не учитывает условий замыкания цикла, вследствие чего подавляющее большинство геометрий оказываются искажёнными (см. Рисунок 3).

1.1.4. Методы машинного обучения

Методы конформационного поиска на основе машинного обучения (machine learning, ML) продолжают идею методов эмпирического перебора, описанных в подразделе 1.1.1.3. В обоих подходах генерация конформаций полагается на данные,

извлечённые из экспериментальных и расчётных баз: PDB [62], CSD [63], GEOM [138]. Различие заключается в способе кодирования и использования этих данных. В классических эмпирических методах на этапе генерации конформации используются библиотеки фрагментов и шаблонов, а также таблицы экспериментальных распределений торсионных углов. Конформации получаются путём комбинаторной сборки по конкретно заданным правилам из заранее подготовленных частей.

В методах ML, напротив, эмпирическая информация кодируется неявно в параметрах (весах) глубоких нейронных сетей. Нейронная сеть – это дифференцируемое отображение между пространствами высокой размерности. Такие сети обладают свойством универсальных аппроксиматоров, то есть способностью приближать разнообразные функции [139]. Более того, введение в функцию скрытых аргументов со случайными значениями позволяет перейти от аппроксимации детерминированных функций к моделированию вероятностных распределений. В частности, машинно-обучаемые методы конформационного поиска моделируют распределение $P(\text{конформация} | \text{граф})$, которое могло бы порождать корректные структуры в обучающих данных. Таким образом, разработка методов на основе ML включает в себя не только сбор структурных данных о конформациях молекул, но ещё и этап обучения – оптимизацию этих многочисленных параметров для минимизации ошибки модели. Генерация же новой конформации в зависимости от типа модели ML выполняется либо прямым вычислением обученной функции, либо численным интегрированием обученного динамического процесса (диффузии или потока), который, в идеальном случае, эволюционирует шумовое распределение на ППЭ в равновесное больцмановское распределение конформаций.

Методы прямого вычисления являются методами конформационного перебора, но с использованием ML. Как обсуждалось в подразделе 1.1.1, конформации можно предсказывать во внутренних координатах, в виде матриц межатомных расстояний или напрямую в виде декартовых координат атомов. Подход прямого предсказания внутренних координат реализован в методах GeoMol [140] и COSMIC [141]. Предсказание межатомных расстояний предлагалось в методах ConfVAE [142] и

GraphDG [47], после чего конформация строилась с помощью алгоритма дистанционной геометрии. Наконец, прямое предсказание декартовых координат реализовано в методах CVGAE [143] и DMCG [144].

Динамические методы ML не дают готовой конформации за одно вычисление. Вместо этого нейронная сеть задаёт закон эволюции системы во времени, аналогично тому, как уравнения молекулярной динамики и эвристические шаги в МКМ задают траекторию исследования ППЭ (см. подраздел 1.1.2). Популярность получили генеративные модели ML, моделирующие процесс диффузии, которые в процессе обучения принимают молекулярные геометрии с примесью шума и учатся удалять этот шум. Это напоминает шаги имитации отжига, однако ключевое отличие в том, что операторы перехода обучаются и уже не основываются на критерии Метрополиса. Было предложено множество методов на основе диффузии декартовых координат (GeoDiff [145], MCF [146] и другие), а также координат торсионных углов (Torsional Diffusion [147] и RINGER [148]).

Другим подходом динамической генерации являются машинно-обученные непрерывные нормализующие потоки. В отличие от стохастической диффузии, эти методы обучают детерминированное уравнение движения системы в виде векторного поля $dz/dt = v_\theta(z, t)$, которое трансформирует простое априорное распределение (обычно многомерное Гауссово распределение) в сложное распределение равновесных конформаций. Например, метод конформационного поиска ConfFlow [149] основан на поле скоростей над декартовыми координатами атомов, тогда как CGCF [150] использует нормализующие потоки для генерации матриц межатомных расстояний, а конформации затем восстанавливаются методом дистанционной геометрии.

Наконец, предлагались подходы генерации на основе ланжевеновской динамики, которая совмещает движение вдоль машинно-обученного градиента $\nabla_x \log P(x)$ с добавлением случайных отклонений. Эта динамика моделирует явление диффузии в потенциальном поле – система притягивается к областям, заселённым конформациями из обучающей выборки, а стохастическое слагаемое помогает с преодолением барьеров для лучшего семплирования распределения. Как правило,

методы конформационного поиска из данной группы генерируют конформации в виде матриц межатомных расстояний, например методы ConfGF [151], SDEGen [152] и DGSM [153].

Как и в случае с методами Байесовской оптимизации, одним из ключевых направлений развития методов конформационного поиска на основе ML является физическое информирование [135], то есть разработка способов учёта априорной информации о структуре ППЭ в архитектуре модели или в процессе её обучения. Это можно видеть на примере используемых представлений конформаций. Частым выбором являются матрицы межатомных расстояний, так как они представляют геометрию $SE(3)$ -инвариантным образом: тогда как разные наборы декартовых координат атомов, различающиеся на трансляцию или поворот, соответствуют сути одной и той же геометрии, межатомные расстояния такой неопределённости не имеют. Также есть примеры [140,147] использования торсионных углов в качестве системы координат, которые, очевидно, тоже инвариантны относительно изометрии. Более того, даже методы ML, которые всё-таки генерируют конформации сразу в представлении декартовых координат, на самом деле используют G -эквивариантную архитектуру нейронных сетей, где в качестве G встречаются $SE(3)$, $E(3)$ и $O(3)$, а также группы перестановок атомов одинаковых химических элементов [154].

Применение методов ML в конформационном поиске не ограничивается генерацией структур. Алгоритмы на основе ML используются и на других этапах построения конформационных ансамблей, в частности, для проведения расчётов МД, оптимизации геометрии и ранжирования конформеров, а также для построения атомистических моделей системы и расчёта состояний протонирования [155–157].

1.1.5. Гибридные методы (пайплайны) конформационного поиска

Программные пакеты конформационного поиска, предназначенные для использования в химических и фармацевтических исследованиях и претендующие на достаточно полное покрытие низкоэнергетических областей ППЭ, крайне редко ограничиваются каким-либо одним из описанных выше подходов. Ни один отдельно взятый метод перебора, динамики или аугментации не способен одновременно

удовлетворить всем требованиям: обеспечить и высокую скорость семплирования, и точность учёта энергии; достичь и глобального охвата ППЭ, и локального уточнения в окрестностях минимумов. Вместо этого подавляющее большинство программ реализованы в виде композиции нескольких подходов конформационного поиска, – пайплайнов, где различные алгоритмы выстроены в последовательность, и каждый из них вносит свои сильные стороны в итоговую эффективность.

Одним из популярных пайплайнов глобального конформационного поиска является iMTD-GC, реализованный в пакете CREST [27,158]. В основе алгоритма лежит метод MTD – молекулярная динамика с локальным повышением потенциала на основе среднеквадратичных отклонений координат атомов (RMSD) от ранее посещённых точек ППЭ (см. подраздел 1.1.2.1). Ключевым этапом пайплайна является запуск 14 параллельных расчётов MTD с различными параметрами смещающего потенциала. Такое разнообразие параметров обеспечивает качественное семплирование для всех типов молекул. Продолжительность этих расчётов MTD фиксированная и определяется автоматически из показателя гибкости молекулы, который вычисляется до запуска параллельных MTD. Из полученных траекторий MTD отбираются точки с шагом 1 пс (совпадают с точками смещающих потенциалов) и подвергаются трёхэтапному отсеиванию с оптимизацией геометрии: на каждом следующем этапе усиливаются критерии сходимости оптимизаций и уменьшается энергетическое окно для отсеивания. После второго этапа проводятся короткие классические расчёты МД при нескольких температурах для шести самых низкоэнергетических конформеров. Это делается для локального анализа ППЭ вблизи наиболее выгодных геометрий. По завершению отсеивания на конформерах выполняются генетические операции, что дополнительно аугментирует ансамбль. Для новых конформаций выполняется двухэтапная процедура отсеивания и оптимизации. Пайплайн является итеративным, т. е. если найден новый более низкий по энергии конформер, вся процедура (начиная с MTD) перезапускается с использованием этого конформера в качестве входного.

Ещё одним из примером гибридного метода является MMBS (MacroModel's macrosycle baseline search), разработанный для конформационного поиска

макроциклов [35]. MMBS последовательно применяет алгоритм симуляции отжига (см. подраздел 1.1.2.2) и LLMOD [117]. Для этих алгоритмов делается максимум по 5000 повторений и полученные структуры оптимизируются. Каждый цикл симуляции отжига включает краткую высокотемпературную часть при 1000 К для преодоления торсионных барьеров, затем охлаждение до 300 К для релаксации. После 5000 циклов симуляции отжига MMBS выполняет до 5000 итераций LLMOD – простого алгоритма аугментации ансамбля путём сдвигов из известного глобального минимума вдоль низкочастотных мод с последующей оптимизацией. В завершение делается отсеивание дубликатов по RMSD и энергетическому окну в 10 ккал/моль.

1.2. Тестирование методов конформационного поиска

Большое разнообразие описанных выше методов построения конформационных ансамблей путём перебора, динамики, случайных блужданий и аугментации, а также композиций из сразу нескольких подходов приводит к необходимости валидации, тестирования и сравнения методов конформационного поиска. Для любой программы компьютерного моделирования цель тестирования состоит в прогнозировании её эффективности применительно к системам, на которых программа еще не тестировалась. Эти прогнозы используются исследователями при выборе метода для решения конкретной задачи, а также разработчиками для определения направлений дальнейшего совершенствования инструмента.

Для методов конформационного поиска двумя ключевыми подходами к тестированию являются оценка разнообразия сгенерированных конформеров и охвата конформационного пространства в окрестностях экспериментально наблюдаемых структур [159]. Обе эти задачи полагаются на существование метода численной оценки подобия конформаций.

1.2.1. Метрики конформационного подобия

В общем случае, метрикой на некотором пространстве M называется функция $D: M \times M \rightarrow \mathbb{R}_{\geq 0}$. В случае моделирования ППЭ, M представляет собой конформационное пространство. Метрики обобщают понятие расстояния, поэтому к ним выдвигается ряд требований:

- 1) неотрицательность,
- 2) тождественность неразличимых точек (расстояние нулевое только между одинаковыми конформациями),
- 3) симметричность $D(a, b) = D(b, a)$,
- 4) неравенство треугольника $D(a, c) \leq D(a, b) + D(b, c)$.

Например, если конформационное пространство представлено трёхмерными координатами атомов, \mathbb{R}^{3n} , то естественно на нём ввести метрику среднеквадратичного отклонения (RMSD, root-mean-square deviation):

$$RMSD(v, w) = \sqrt{\frac{1}{n} \sum_{i=1}^n \|v_i - w_i\|^2}, \quad (4)$$

где v и w – матрицы координат ядер для сравниваемых конформаций. Однако \mathbb{R}^{3n} является избыточным представлением конформационного пространства, за исключением случаев сравнения конформаций в фиксированной системе отсчёта (например, конфигураций лигандов в докинге). Более физически обоснованная модель конформационного пространства, $\mathbb{R}^{3n}/SE(3)$, получается отождествлением геометрий, переходящих друг в друга посредством поворотов и параллельных переносов. В таком случае расчёт RMSD сводится к минимизации

$$RMSD([v], [w]) = \min_{T \in SE(3)} RMSD(v, Tw), \quad (5)$$

то есть к расчёту отклонения при оптимальном наложении геометрий, где $[v]$ и $[w]$ – это классы эквивалентности матриц ядерных координат v и w с точностью до трансляции и поворота. Минимизация RMSD обычно выполняется путём трансляции конформаций в общий центр масс с последующей оптимизацией их

относительного поворота в смысле метода наименьших квадратов с использованием сингулярного разложения. Данный подход известен как алгоритм Кабша [160]; также существуют альтернативные методы на основе кватернионов [161–165]. Существование алгоритмов эффективной оптимизации поворота даёт RMSD вычислительное преимущество перед метриками ℓ^p при $p \neq 2$ и ℓ^∞ . Натуральное число p контролирует баланс между усреднением отклонений и приоритизацией больших отклонений в метрике; в пределе ℓ^∞ является метрикой сравнения конформаций по наибольшему отклонению положений атомов.

Молекулярный граф может обладать элементами симметрии (нетривиальными автоморфизмами), то есть перестановками атомов, не нарушающими его топологию. Для их учёта при расчёте RMSD достаточно перебрать все перестановки из группы автоморфизмов графа и выбрать наименьшее из полученных значений, что реализовано, например, в [166,167]. Существуют и другие подходы: например, численная оптимизация поворота с учётом симметрии графа реализована в G-RMSD [168]. Для задач молекулярного докинга учёт симметрии графа в фиксированной системе координат (без оптимизации по трансляции и повороту) реализован с помощью венгерского алгоритма в методах DOCK [169] и DockRMSD [170].

RMSD повсеместно используется в вычислительной химии при поиске дублирующихся конформаций (например, для отсеивания дубликатов при конформационном поиске) и для численной оценки геометрической близости различных геометрий (например, при тестировании методов конформационного поиска или в качестве компоненты смещающего потенциала MTD). Вместе с тем RMSD не является универсальной метрикой для всех задач сравнения конформаций: усреднение не позволяет детектировать сильные локальные отличия на фоне множества небольших отклонений в крупных молекулах. Так, известно, что RMSD неинформативен для оценки подобия белков, поэтому в этой области применяются специализированные метрики, такие как Contact Mode Score и TM-score [171].

При поиске дубликатов в больших конформационных ансамблях использование RMSD затруднительно, так как требует вычислительных затрат для перебора всех автоморфизмов графа. Для решения этой проблемы предлагался метод TFD (Torsion

Fingerprint Deviation) [172] для приближенного сравнения конформаций через их торсионные углы. Этот метод учитывает симметрически эквивалентные повороты, однако не способен учитывать более сложные перестановки атомов с участием нескольких торсионных углов. Также для избежания расчёта RMSD с учётом автоморфизмов графа в пайплайнах программы CREST дубликаты отсеивают приближённым способом на основе разности энергий конформаций и разностей их вращательных постоянных.

Несмотря на вычислительные затраты учёта симметрии и неинформативность при сравнении особо крупных молекул, RMSD остаётся наиболее распространённой и универсальной метрикой конформационного подобия. В частности, RMSD и его варианты, рассматривающие отдельные атомы (только тяжёлые атомы или только атомы одного циклического фрагмента), использовались при тестировании передовых методов конформационного поиска, таких как CREST, ETKDG, MMBS и GOAT.

1.2.2. Подходы к построению тестовых задач

Для валидации новых методов конформационного поиска и их сравнения с альтернативными алгоритмами, которые могут быть основаны на совершенно иных принципах, используется подход сквозного тестирования. Это предполагает моделирование практического сценария работы алгоритма: запуск на выборке релевантных молекул и оценку полученных результатов по некоторым метрикам, либо сопоставление с заранее известными результатами. Наиболее интересными задачами для такого типа тестирования являются оценка точности воспроизведения экспериментально наблюдаемых структур и оценка покрытия низкоэнергетических областей ППЭ.

Для составления репрезентативной тестовой выборки молекул с подходящим элементным составом, молекулярной массой, наличием или отсутствием макроциклов, наиболее распространённым подходом является поиск таких молекул в структурных базах экспериментальных данных – PDB и CSD. Например, так поступили авторы методов ETKDG и MMBS; в случае ориентированного на

макроциклы метода MMBS авторы составили тестовый набор из 150 макроциклических лекарственно-подобных молекул из PDB и CSD. С другой стороны, тестирование алгоритма CREST включало валидацию конформационных ансамблей через воспроизведение экспериментальных спектров ^1H ЯМР и структур, детектированных методом микроволновой спектроскопии. Тестирование метода GOAT проводилось на составленном авторами наборе сложных для исследования ППЭ систем: переходных состояний, кластеров металлов и молекул воды, металлокомплексов, циклических пептидов, а также нереалистичных гипотетических молекул.

Методика сравнения сгенерированных ансамблей с экспериментальным конформером стандартна и заключается в расчёте минимального RMSD по ансамблю. Однако отсутствует общепринятая методика оценки размера и разнообразия ансамблей: разные авторы применяют различные критерии. Для тестирования CREST использовалась конформационная энтропия S_{conf} : если относительные энергии конформеров ΔE отсчитываются от одного и того же уровня, то S_{conf} строго возрастает с увеличением числа найденных низкоэнергетических конформеров.

$$S_{conf} = R \sum_{i=1}^N p_i \log p_i, \quad p_i = \frac{\exp(-\Delta E_i/RT)}{\sum_{j=1}^N \exp(-\Delta E_j/RT)}, \quad (6)$$

где p_i – равновесная заселённость i -го конформера в полученном ансамбле.

Однако, такая оценка учитывает только количество конформеров и их энергию, но никак не учитывает их геометрическую схожесть. Напротив, при тестировании алгоритма ETKDG в качестве метрики разнообразия ансамбля использовались медианные значения RMSD и TFD между конформерами.

Для гибридных методов также используются модульные тесты, позволяющие убедиться, что ключевые элементы метода работают в соответствии с ожиданиями. При тестировании метода CREST отдельное внимание было уделено точности воспроизведения относительных энергий конформеров новыми полуэмпирическими методами GFNn-xTB, используемыми в CREST. Для этого проводилось сравнение

ошибок в энергиях, полученных методами GFNn-хТВ, со стандартными полуэмпирическими методами и методами теории функционала плотности, откуда авторы сделали вывод, что предложенный ими метод GFN2-хТВ является лидером по точности среди рассмотренных полуэмпирических подходов. Похожим образом для тестирования эмпирического метода дистанционной геометрии ETKDG, реализованного в RDKit, была проведена проверка того, что используемые в методе эмпирические потенциалы торсионных углов действительно «информируют» метод дистанционной геометрии и позволяют ему генерировать конформационные наборы с распределением торсионных углов, совпадающим с желаемыми экспериментальными распределениями. Кроме того, существуют примеры абляционного тестирования методов конформационного поиска, то есть проверки роли некоторого элемента алгоритма путём сравнения эффективности метода с ним и без него. Например, при тестировании ETKDG авторы удостоверились, что добавление потенциала, обеспечивающего планарность ароматических колец, как и ожидалось, снижает разнообразие конформаций (поскольку исключает нереалистичные геометрии), но улучшает воспроизведение экспериментальных конформаций.

Таким образом, в литературном обзоре в Разделе 1 показано существование разнообразных методов построения ансамблей конформеров молекул – от простого перебора торсионных углов до гибридных методов, совмещающих метадинамику с генетическим алгоритмом. Однако, способность многих этих методов к глобальному исследованию конформационного пространства полагается на возможность свободного варьирования торсионных углов молекулы, что создаёт серьёзные ограничения при работе с циклическими и полициклическими молекулами.

В случае конформационно жёстких или малоподвижных полициклических каркасов исследование небольшого конформационного пространства возможно методами, которые не основаны на варьировании торсионных углов (ETKDG) или которые фильтруют получаемые конформации по условию замыкания цикла (RMMC). В Разделе 2 данной работы мы применили эти методы для моделирования

каркасных полициклических пероксидов. Полученные нами геометрии согласовывались со структурами из рентгеноструктурного анализа. На основе полученных низкоэнергетических конформеров мы провели квантово-химическое моделирование реакций сборки каркасных пероксидов, что позволило объяснить ряд экспериментальных наблюдений.

В Разделе 3 данной работы мы предложили универсальное решение задачи перебора торсионных углов циклических молекул на основе подхода обратной кинематики (ОК). Приведённый выше обзор подходов к тестированию методов конформационного поиска помог нам составить процедуру валидации разработанного нами алгоритма случайного перебора циклических торсионных углов с использованием ОК.

Раздел 2. Стереoeлектронные эффекты в каркасных органических пероксидах

Эта часть диссертационной работы посвящена квантово-химическому моделированию четырёх реакций сборки каркасных полициклических пероксидов (Рисунок 10). Органические полициклические пероксиды являются уникальным классом соединений, в которых конформация молекулы определяется как условием цикличности, так и множеством стереoeлектронных эффектов. Поэтому в данном разделе отдельное внимание уделено изучению роли стереoeлектронных эффектов в управлении каскадом сборки продуктов.

Основные результаты данной части диссертационной работы опубликованы в четырёх исследовательских статьях [173–176]; кроме того, результаты систематизации методов анализа стереoeлектронных эффектов и их роли в химии кислородсодержащих соединений, проведённой при участии автора диссертации, представлены в двух обзорах [177,178].

2.1. Моделирование реакции β,γ' -трикетонов с пероксидом водорода

В 2019 году наши коллеги успешно реализовали синтез трициклических пероксидов из β,γ' -трикетонов и H_2O_2 в среде ацетонитрила с добавлением *para*-толуолсульфокислоты. Было экспериментально установлено, что количество пероксидных групп в конечном продукте зависит от количества H_2O_2 , а также от природы заместителя в исходном β,γ' -трикетоне (**ASa-d**, Рисунок 11). Для метильного и аллильного заместителей введение 1 экв. H_2O_2 приводит к образованию трициклического монопероксида **A1**, тогда как использование избытка (3 экв.) H_2O_2 способствует получению бис-пероксида **A2**. В то же время для трикетона, содержащего *para*-нитробензильный фрагмент, реакция останавливается на стадии монопероксида **A1**, а в случае отсутствия заместителя образуется исключительно бис-пероксид **A2**. При этом трис-пероксид **A3** не удавалось синтезировать ни для одной из исследованных комбинаций субстратов и условий.

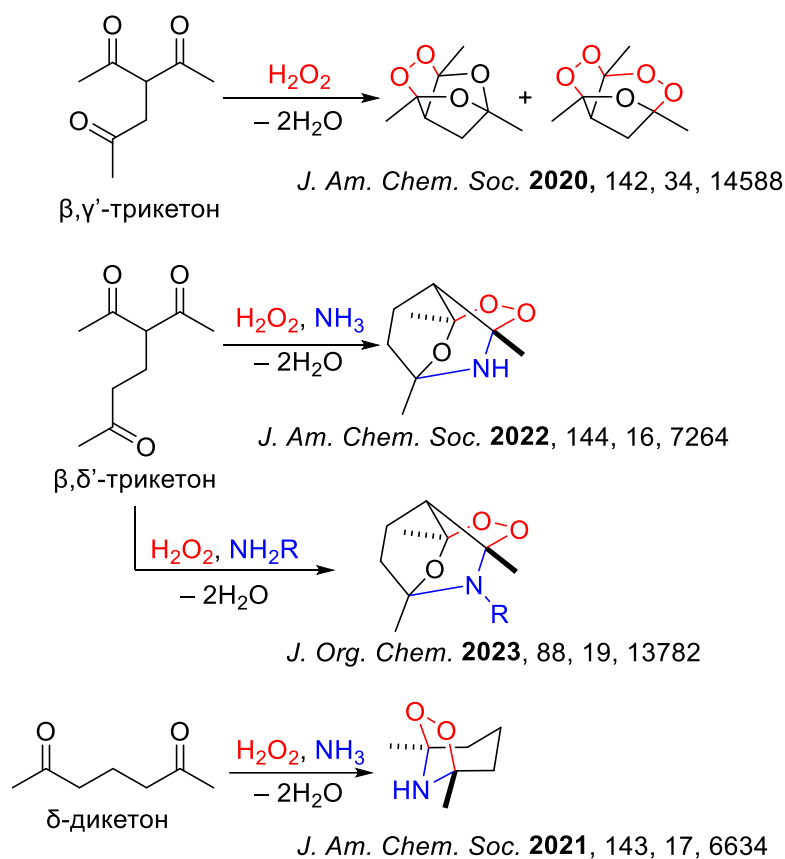
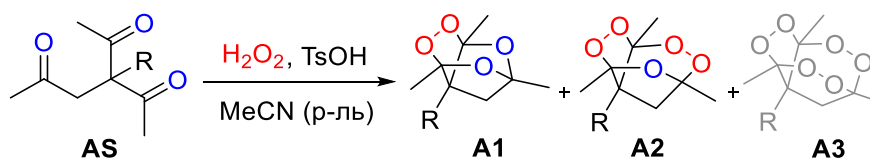


Рисунок 10. Исследуемые реакции получения каркасных пероксидов из ди- и трикетонов. Схемы реакций приведены в порядке их обсуждения тексте. Для наглядности некоторые заместители на этой схеме опущены.



Заместитель в AS	A1	A2	A3
R = ξ -H ASa, без заместителя	✗	✓	✗
R = ξ -CH ₃ ASb, метил	✓	✓	✗
R = ξ -CH ₂ -CH=CH ₂ ASc, аллил	✓	✓	✗
R = ξ -CH ₂ -C ₆ H ₄ -NO ₂ ASd, пара-нитробензил	✓	✗	✗

J. Am. Chem. Soc., **2020**, 142, 34, 14588

Рисунок 11. Схема и экспериментальная селективность исследуемой реакции.

В рамках диссертационного исследования нами было проведено квантово-химическое моделирование для объяснения селективности этой реакции в зависимости от заместителя и количества H_2O_2 .

2.1.1. Используемые методы

Моделирование выполнялись в Gaussian16 A.03 [18] на уровне теории PBE0-D3BJ [19,20] / 6-311++G(d,p) [21,22] / SMD(MeCN) [23]. Выбор функционала PBE0 обусловлен его надёжностью при моделировании органических соединений, а также способностью точно воспроизводить электронные плотности атомных и молекулярных систем [179]. Для всех расчётов свободной энергии применялась квазигармоническая поправка, рассчитываемая в библиотеке GoodVibes [25]. Учёт трансляционной энтропии ацетонитрила выполнен с использованием функции *media* в GoodVibes, которая включает поправку на свободный объём и молярную концентрацию ацетонитрила. Все термодинамические поправки рассчитаны для температуры 0 °C.

Благодаря небольшому размеру интермедиатов и переходных состояний (ПС) в исследуемой реакции, для них был проведён полный конформационный анализ. С этой целью была сделана собственная имплементация алгоритма ringmaker Monte-Carlo (см. подраздел 1.1.1.2 литературного обзора). Поиск начинается из подготовленной вручную стартовой конформации, после чего осуществляется случайный перебор торсионных углов с отсеиванием структур по критериям связности (расстояние между связанными атомами не более 2 Å и не менее 0.95 Å) и стерическим ограничениям (расстояние между несвязанными атомами не менее 2 Å). Отобранные структуры подвергались оптимизации на уровне теории PM6 [180] / SMD(MeCN), после чего отбиралось 10 конформеров с наименьшей энергией. Эти структуры затем оптимизировались на финальном уровне теории PBE0-D3BJ / 6-311++G(d,p) / SMD(MeCN). Для дальнейших расчётов использовался конформер с минимальной свободной энергией при условии проверки числа мнимых частот (0 – для интермедиатов, 1 – для ПС).

Анализ стереоэлектронных эффектов выполнен с помощью программы NBO3.1, встроенной в Gaussian16, а также через моделирование термодинамики изодесмических реакций на том же уровне теории PBE0-D3BJ / 6-311++G(d,p) / SMD(MeCN).

Согласно литературным данным [181,182], *para*-толуолсульфокислота (TsOH) в растворе ацетонитрила диссоциирует неполностью ($pK_a(TsOH) = 8.45$). Это означает, что равновесная концентрация сольватированного протона в растворе TsOH значительно ниже, чем в случае полной диссоциации. Для корректного учёта этого эффекта при расчётах относительных свободных энергий протонированных и нейтральных интермедиатов, к вычисленным относительным энергиям катионных интермедиатов была добавлена величина 10.56 ккал/моль, рассчитанная исходя из константы кислотности TsOH:

$$\begin{aligned} TsOH + MeCN &\rightarrow TsO^- + H^+ \cdot MeCN, \Delta G_{diss}(TsOH) \\ \Delta G_{diss}(TsOH) &= -RT \ln(10^{-pK_a(TsOH)}) \\ &= -8.314 \cdot 273.15 \cdot \ln(10^{-8.45}) = 10.56 \text{ ккал/моль} \end{aligned} \quad (7)$$

Кинетическое интегрирование превращения **A1**→**A2** выполнено алгоритмом LSODE [183], реализованном в библиотеке SciPy [184]. Система используемых кинетических уравнений приведена на Рисунке 12. Константы скорости k получены из свободных энергий активации с помощью уравнения Эйринга с коэффициентом переноса κ , принятым за 1:

$$k = \kappa \frac{k_b T}{h} \exp\left(-\frac{\Delta G^\ddagger}{RT}\right) \quad (8)$$

Энергии активации ΔG^\ddagger стадий 3, 4 и 6 (Рисунок 12) получены из квантово-химически рассчитанных свободных энергий молекул и ПС. Для стадий 2, 5 и 7 нам не удалось локализовать ПС, поэтому, так как для них ожидается очень небольшая ΔG^\ddagger , мы выбрали соответствующие константы скорости так, чтобы в направлении с $\Delta G_r < 0$ энергия активации была равна 6 ккал/моль, а ΔG^\ddagger в обратную сторону были рассчитаны на основе квантово-химических свободных энергий реакций ΔG_r . Начальные концентрации (Рисунок 12) оценены из экспериментальной методики

синтеза бис-пероксидов [173], в частности, сравнительно большая начальная концентрация воды обусловлена использованием 30% раствора H_2O_2 .

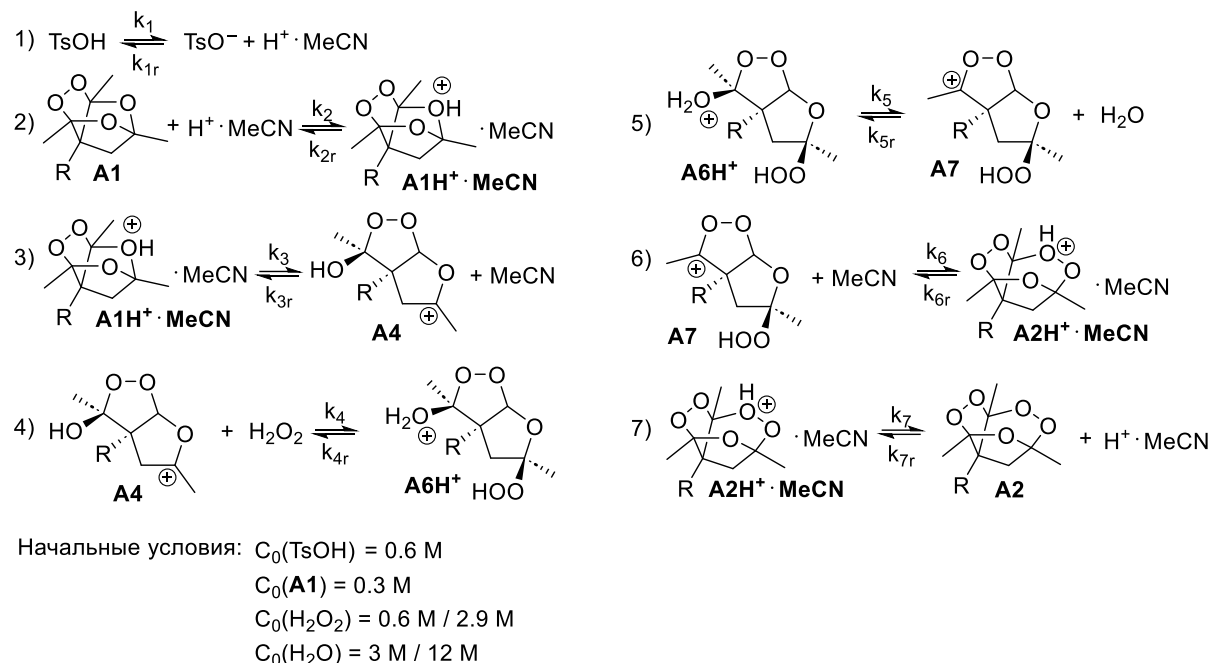


Рисунок 12. Система уравнений, используемая в кинетическом интегрировании для превращения **A1**→**A2**.

2.1.2. Результаты квантово-химического моделирования

Построение квантово-химической модели реакции было начато с расчёта термодинамического соотношения продуктов **A1**–**A3** методом стационарных точек на уровне теории, приведённом выше, и сопоставления полученных соотношений с экспериментальными наблюдениями. Для этого были рассмотрены четыре исходных трикетона: **ASa** без заместителя, а также **ASb**, **ASc** и **ASd** с метильным, аллильным и *para*-нитробензильным заместителем, соответственно (Рисунок 11).

Для расчёта термодинамических соотношений каркасных пероксидов в первую очередь необходимо получить их трёхмерные структуры. Конформационный поиск методом RMMC показал, что для полициклических каркасов **A1** и **A2** существует только один возможный конформер. Сравнение полученных нами геометрий **A1b**,

A2b и **A1d** с данными рентгеноструктурного анализа монокристаллов⁴ (Рисунок 13) показало практически полное совпадение расчётных и экспериментальных геометрий каркасов. Некоторое различие (около 16°) наблюдается лишь для торсионного угла *para*-нитробензильного заместителя в **A1d**, что, вероятно, обусловлено влиянием эффектов кристаллической упаковки, которые не учтены в нашей модели, где мы моделировали геометрии продуктов в растворе. Для соединений **A3**, которые не удалось получить экспериментально, конформационный поиск показал наличие двух отличающихся по энергии конформеров полициклического каркаса (их строение обсуждено в подразделе 2.1.4, см. Рисунок 25).

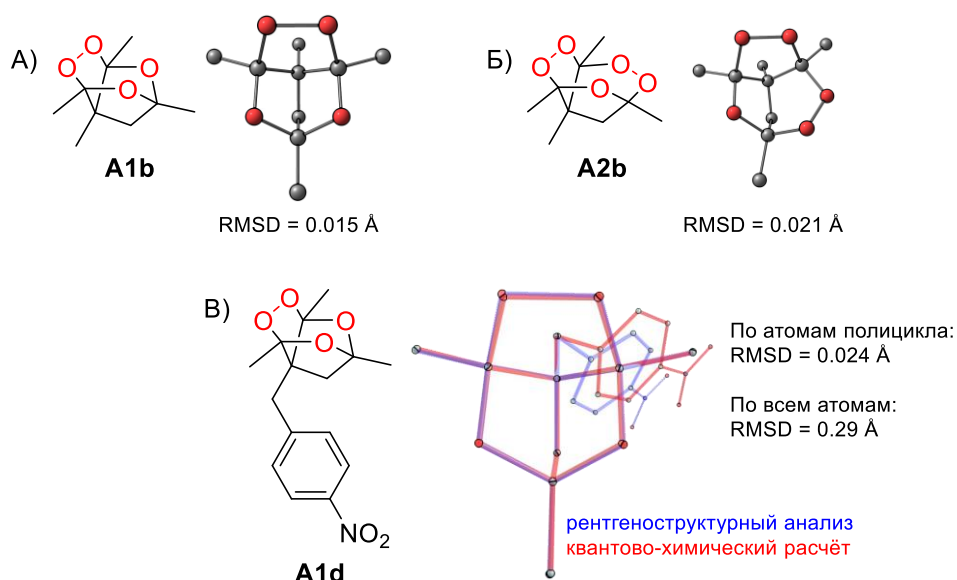


Рисунок 13. Сопоставление конформеров молекул **A1b**, **A2b** и **A1d**, построенных методом РММС с последующей квантово-химической оптимизацией, с экспериментальными конформациями из рентгеноструктурного анализа (РСА) монокристалла. Показаны только неводородные атомы, поскольку РСА не позволяет точно определить длины связей с участием водорода.

⁴ Рентгеновские исследования проведены в Институте элементоорганических соединений им. А. Н. Несмеянова Российской академии наук с помощью дифрактометра Bruker APEX II.

Для конформеров с наименьшей свободной энергией были рассчитаны изменения свободной энергии ΔG_r при взаимопревращениях **AS** \leftrightarrow **A1** \leftrightarrow **A2** \leftrightarrow **A3**. Из анализа полученных значений ΔG_r (Рисунок 14) можно заключить, что сборка монопероксида **A1** и бис-пероксида **A2** термодинамически выгодна для всех рассмотренных трикетонов. Однако, энергетический выигрыш снижается при замещении последнего атома кислорода на пероксидную группу и переход **A2** \rightarrow **A3** уже оказывается термодинамически нейтральным или даже невыгодным для реакций с замещёнными трикетонами (**ASb-d**). Полученная термодинамическая модель объясняет первый парадокс в селективности данной реакции, а именно невозможность экспериментального получения **A1a**: значение ΔG_r последующего превращения **A1a** \rightarrow **A2a** оказывается ниже, чем ΔG_r для собственно образования **A1a** из **ASa**. Это делает молекулу **A1a** склонной к диспропорционированию даже при недостатке H_2O_2 , так как в равновесии половина эквивалента **A2a** более выгодна, чем один эквивалент **A1a**. Аналогичные расчёты для молекул с метильным и аллильным заместителями показывают, что **A1b** и **A1c** не имеют склонности к диспропорционированию, что согласуется с экспериментальными наблюдениями.

Однако, термодинамическая модель не отвечает на остальные вопросы. Так, она предсказывает, что при достижении термодинамического равновесия должны преобладать молекулы **A3a**, **A3c**, **A2d**, а **A3b** должен образовываться в детектируемых количествах, однако ничто из этого не согласуется с экспериментальными наблюдениями (Рисунок 11). Соответственно, дальнейшее моделирование было направлено на учёт энергий интермедиатов и ПС в модели реакции.

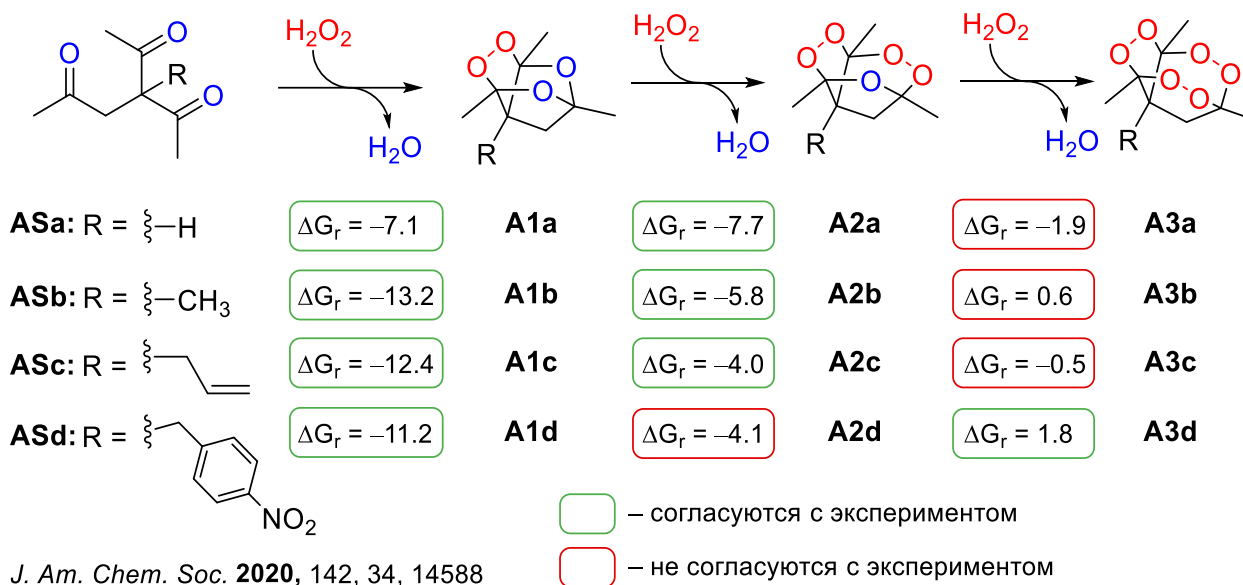
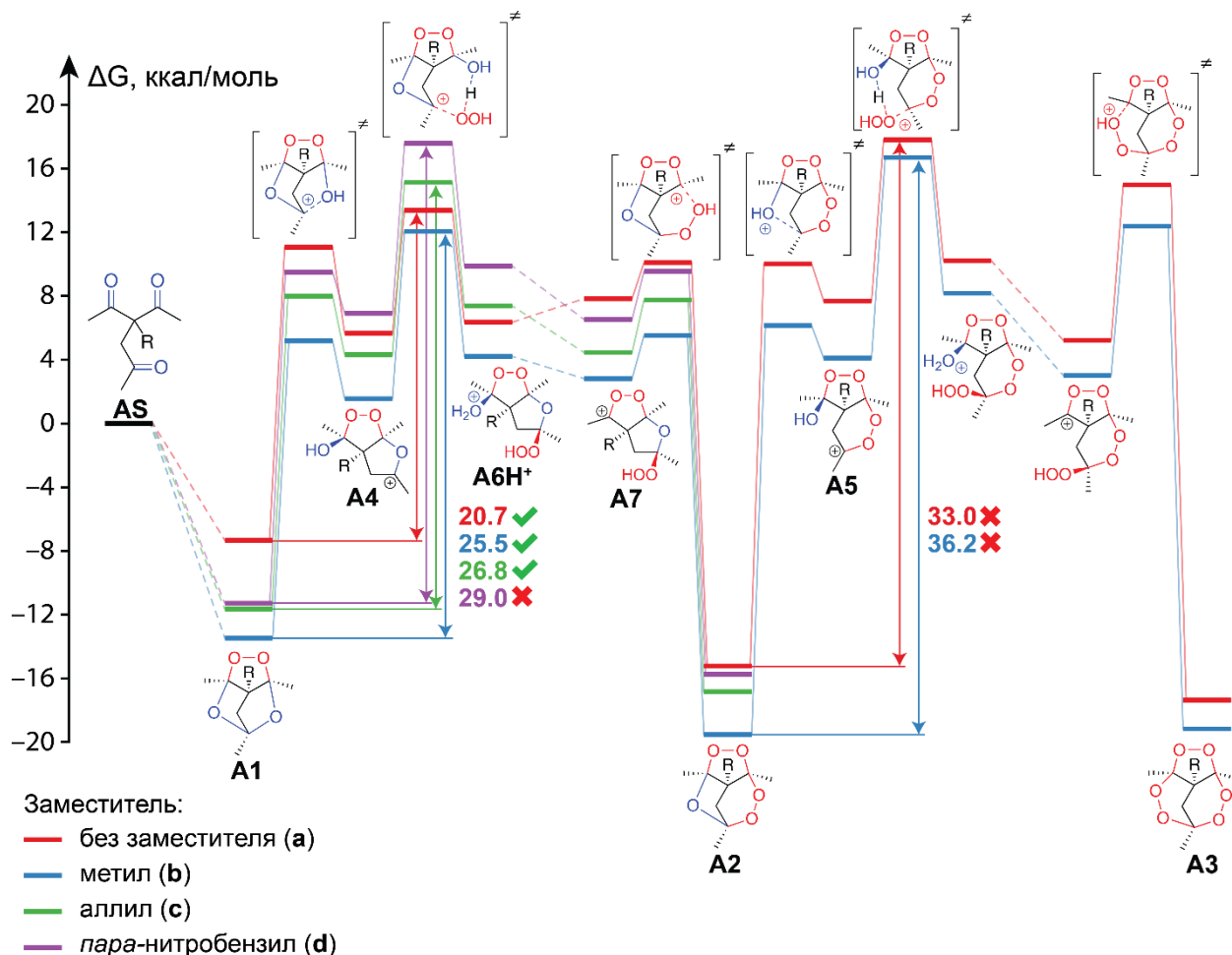


Рисунок 14. Результат термодинамического моделирования реакции для четырёх заместителей.

В первую очередь, из общих химических соображений нами было предложено несколько путей сборки молекул **A1b** и **A2b** из **AS** и затем были рассчитаны энергии всех предположенных интермедиатов (см. Приложение 1). Среди предложенных путей реакции выделяется путь через катион **A4** (Рисунок 15), который оказывается на ~10 ккал/моль выгоднее аналогичных интермедиатов в других предложенных путях сборки. На основе энергий возможных интермедиатов нами было решено сократить моделирование кинетики всей реакции до перегруппировки **A1**→**A4**→**A2**, так как предположительно именно она влияет на селективность. Это можно обосновать постулатом Хэммонда, который предполагает, что при $\Delta G_r < 0$ энергии ПС замыкания катиона в каркасный пероксид коррелируют с энергиями исходных катионов.

Наконец, сокращение механизма до приемлемого количества элементарных стадий позволило нам локализовать и рассчитать энергии ПС этих стадий. Моделирование превращения **A1**→**A2** было проведено для всех четырёх заместителей, а для превращения **A2**→**A3** были рассмотрены только случаи с метильным заместителем (**A2b**→**A3b**) и без заместителя (**A2a**→**A3a**).



J. Am. Chem. Soc., 2020, 142, 34, 14588

Рисунок 15. Расчётный термодинамический профиль для превращений $A1 \rightarrow A2 \rightarrow A3$ с указанием контролирующих эффектов. Пунктиром показаны стадии, которые были исключены из моделирования, а также стадии отщепления воды, для которых не удалось локализовать ПС.

Сравнение энергетических барьеров превращений, то есть разниц энергий между самым высоким по энергии ПС на пути до продукта и самым стабильным предшествующим ему состоянием системы (см. «Методология и методы исследования»), показывает отличное согласие с экспериментальными данными. Несмотря на термодинамическую предпочтительность бис-пероксида $A2d$, барьер его образования, действительно, оказывается более высоким (29 ккал/моль) по

сравнению с барьерами экспериментально наблюдаемых превращений (от 21 до 27 ккал/моль). Закономерности в энергетических барьерах также согласуются с экспериментальным наблюдением, что образование **A2c** происходит существенно медленнее, чем **A2a** и **A2b**. В случае сборки трис-пероксида барьеры оказываются ещё выше: 33 ккал/моль для **A3a** и 36 ккал/моль для **A3b**, что полностью согласуется с их отсутствием среди продуктов реакций.

Далее мы оценили зависимость скорости сборки пероксидов от высоты барьеров на примере перегруппировки **A1**→**A2**. На основе полученных нами выше свободных энергий было проведено численное интегрирование системы кинетических уравнений (см. подраздел 2.1.1) для четырёх заместителей **A1a–d** при двух начальных концентрациях H_2O_2 . Такое моделирование позволяет получить временную зависимость концентраций каркасных пероксидов **A1–2** и интермедиатов их перегруппировки и сопоставить их с экспериментальными наблюдениями.

Полученные результаты показали, что при 0 °C перегруппировка для монопероксида без заместителя (**A1a**) происходит очень быстро – менее чем за минуту при обеих начальных концентрациях H_2O_2 (Рисунок 16). Перегруппировка **A1b**→**A2b** с метильным заместителем оказывается на два порядка медленнее, что согласуется с повышенным на 4.8 ккал/моль барьером. Тем не менее, через 2 часа (экспериментальная продолжительность синтеза **A2b** из трикетона **ASb**) бис-пероксид **A2b** является доминирующим продуктом. Дальнейшее повышение барьера на 1.2 ккал/моль при переходе к аллильному заместителю делает перегруппировку **A1c**→**A2c** ещё более медленной. В этом случае бис-пероксид **A2c** доминирует только в избытке H_2O_2 , что полностью соответствует экспериментальным данным. Ещё большее увеличение барьера на 2.2 ккал/моль при переходе к паранитробензильному заместителю кинетически блокирует перегруппировку **A1d**→**A2d**, делая бис-пероксид **A2d** недостижимым несмотря на отрицательную ΔG_r перегруппировки. Этот результат также согласуется с экспериментальными наблюдениями и подтверждает значимость описанных закономерностей в энергетических барьерах.

Таким образом, нам удалось построить квантово-химическую модель реакции β,γ' -трикетонов с пероксидом водорода в кислой среде, которая воспроизводит все ключевые экспериментальные наблюдения. На следующем этапе работы выполнен детальный анализ этой модели с целью установления причин наблюдаемой селективности реакций (см. подраздел 2.1.4). Поскольку в проведённом анализе значительная роль отводится стереоэлектронным эффектам, в следующем подразделе вводится понятие данных эффектов и кратко обсуждаются основные подходы к их моделированию.

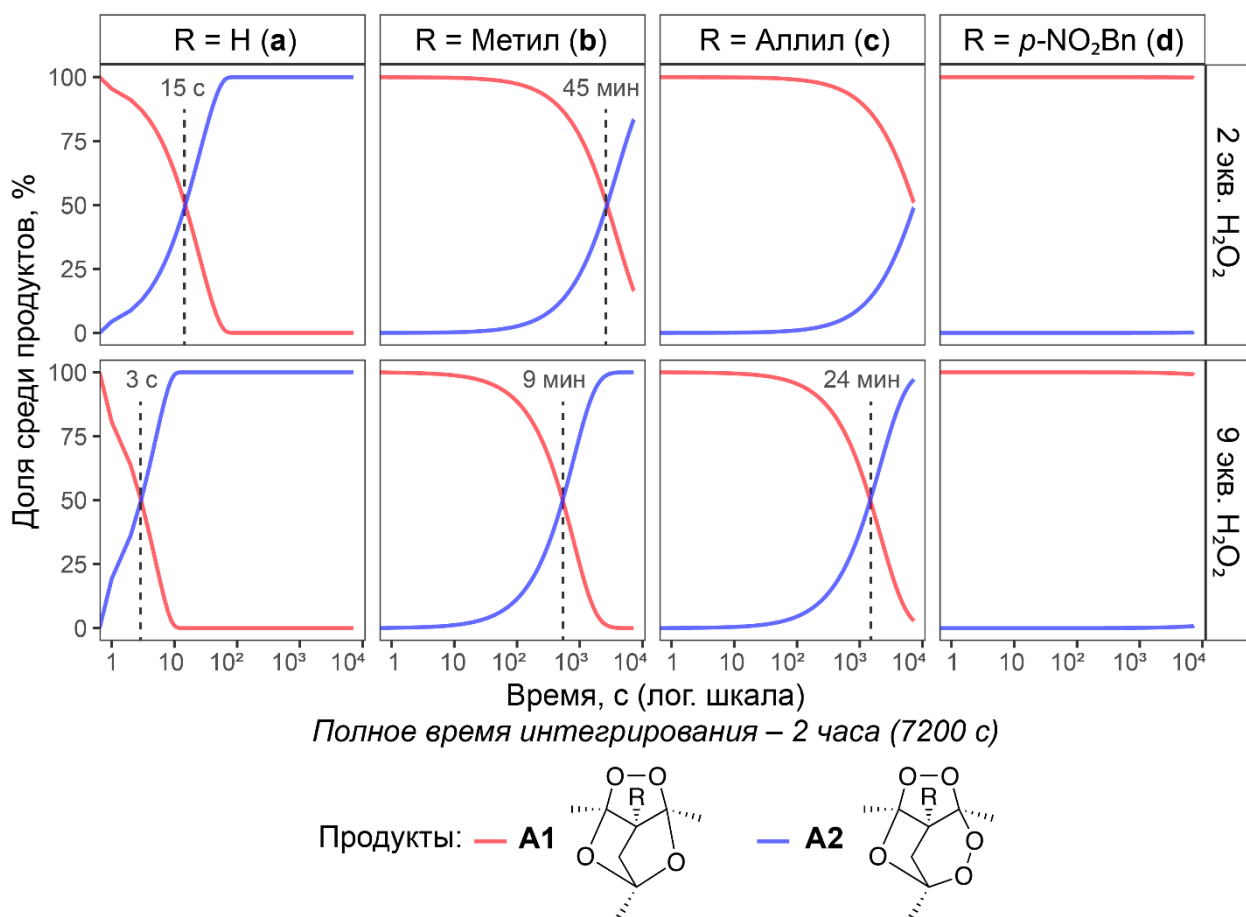


Рисунок 16. Зависимости долей продуктов реакции β,γ' -трикетонов с пероксидом водорода в кислой среде от времени, полученные из кинетического интегрирования. Временная ось приведена в логарифмических координатах.

2.1.3. Типовые стереоэлектронные эффекты и способы их моделирования

Квантово-химическое моделирование электронной структуры молекулы и расчёт её энергии – это непростая вычислительная задача. В теории функционала плотности она требует построения базиса из атомоцентрированных орбиталей, вычисления матрицы их перекрывания и самосогласованного решения нелинейной задачи Кона-Шэма: на каждой итерации рассчитывают эффективный гамильтониан и его собственные векторы, ортонормированные по матрице перекрывания.

Помимо самих результатов квантово-химических расчётов, ценность представляет также извлечение из них достаточно простых интерпретируемых выводов и правил. На сегодняшнем этапе развития химии опытные исследователи часто способны судить об относительной стабильности молекул, предпочтительности их конформаций или сравнительной реакционной способности фрагментов, не прибегая к теоретическим расчётам. В основе этой способности лежит интуитивное понимание устройства атомных орбиталей и тех линейных комбинаций, которые они дают в типичных молекулярных фрагментах. Зная электронное строение отдельных фрагментов, химик может предсказать поведение граничных орбиталей при их взаимодействии. Для этого существует, например, анализ резонансных структур через «стрелочное» смещение электронов [185]; для анализа сложных частных случаев существуют правила Хюккеля, Вудворда-Хофмана и другие. Наконец, когда требуется качественно оценить орбитальное взаимодействие, руководствуются простым принципом: стабилизация тем выше, чем меньше разница энергий орбиталей и чем лучше их перекрывание.

Усиление «химической интуиции» для некоторого класса систем наступает тогда, когда удаётся обнаружить новую, неочевидную структуру граничных орбиталей во фрагменте либо выявить неожиданную закономерность во взаимодействии фрагментов с известной по отдельности электронной структурой.

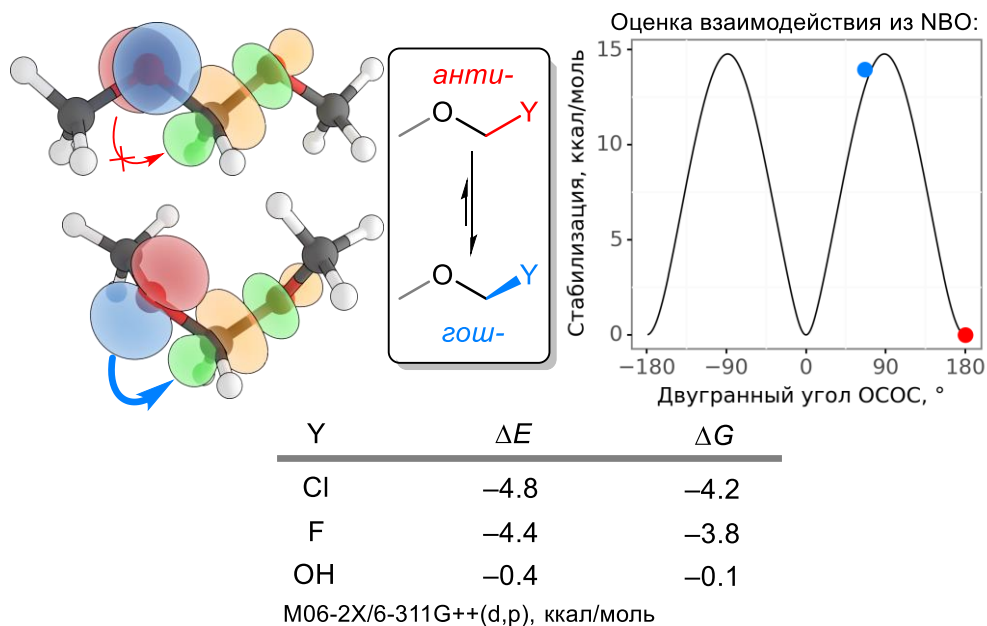
Стартовая точка для формирования интуиции об орбитальных взаимодействиях – это донорно-акцепторные взаимодействия. Их часто называют также стереоэлектронными эффектами (СЭ), что подчёркивает зависимость взаимодействия как от природы орбиталей, так и от их взаимной пространственной

ориентации. Последняя может контролироваться конформацией молекулы или, в случае каркасных структур, самим строением каркаса. Суть эффекта проста: занятая и вакантная орбитали смешиваются с образованием двух новых орбиталей, распределённых по атомам взаимодействующих фрагментов, причём новая занятая орбиталь оказывается ниже по энергии, а новая свободная – выше, чем до смешения. Иными словами, благодаря стереоэлектронному эффекту происходит стабилизация молекулы за счёт частичной делокализации электронов с донора на акцептор (см. ниже анализ в рамках теории возмущений второго порядка).

Один из ключевых стереоэлектронных эффектов – это гиперконъюгация⁵ $n_O \rightarrow \sigma_{C-O}^*$ в ацеталах и кеталах, то есть продуктах присоединения O-нуклеофилов к кетонам (Рисунок 17). Её хорошо известное проявление – повышенное предпочтение аксиальной ориентации гидроксильной группы в пиранозных и фуранозных формах углеводов, известное как аномерный эффект [188].

В более широком случае, даже в ациклических молекулах с фрагментами $-X-CH_2-Y-$, в которых два гетероатома связаны с одним атомом углерода, наблюдаются особые конформационные предпочтения, так как для максимальной стабилизации от гиперконъюгации необходимо, чтобы орбиталь неподелённой пары n_X была расположена в одной плоскости со связью $C-Y$ для максимального взаимодействия с орбиталью σ_{C-Y}^* (Рисунок 17). Обобщая этот принцип на ещё более широкий круг донорных орбиталей, можно увидеть, что все элементарные кислородсодержащие функциональные группы включают множество конъюгаций, то есть делокализаций пар кислорода на σ^* или π^* орбитали. На рисунке 18 проиллюстрированы наиболее распространённые случаи орбитальных взаимодействий с участием атомов кислорода в элементарных органических группах.

⁵Согласно правилам ИЮПАК, взаимодействия вида $n_X \rightarrow \sigma_{C-Y}^*$ принято называть «отрицательной гиперконъюгацией» [186] или «делокализацией n- σ^* » [187]. Поскольку в данной работе классическая (положительная) гиперконъюгация не рассматривается, для взаимодействия вида $n_X \rightarrow \sigma_{C-Y}^*$ для краткости используется термин «гиперконъюгация».



Chem. Soc. Rev., **2021**, 50, 10212

Рисунок 17. Гиперконъюгация $n_O \rightarrow \sigma_{C-Y}^*$, её конформационный эффект, визуализация локализованных орбиталей NBO (см. ниже) и зависимость энергии взаимодействия NBO молекулы диметоксиметана от торсионного угла. В таблице приведены энергии конформационных переходов *анти* \rightarrow *гош*.

Повышение или понижение энергии донорной орбитали прямо сказывается на энергетическом выигрыше от стереоэлектронного эффекта. Здесь стоит выделить два известных явления: смешение орбиталей и электроотрицательность (Рисунок 19, сверху). По отдельности они не меняют энергию системы непосредственно, однако, изменяя энергии занятых орбиталей, способны влиять на стабилизацию косвенно. Смещение пары занятых орбиталей повышает энергию граничной занятой орбитали фрагмента, делая его более сильным донором. Электроотрицательность – это свойство химического окружения; для упрощения анализа его можно моделировать как повышение электростатического потенциала, в котором находится орбиталь, то есть просто как понижение её энергии, ослабляющее донорные свойства.

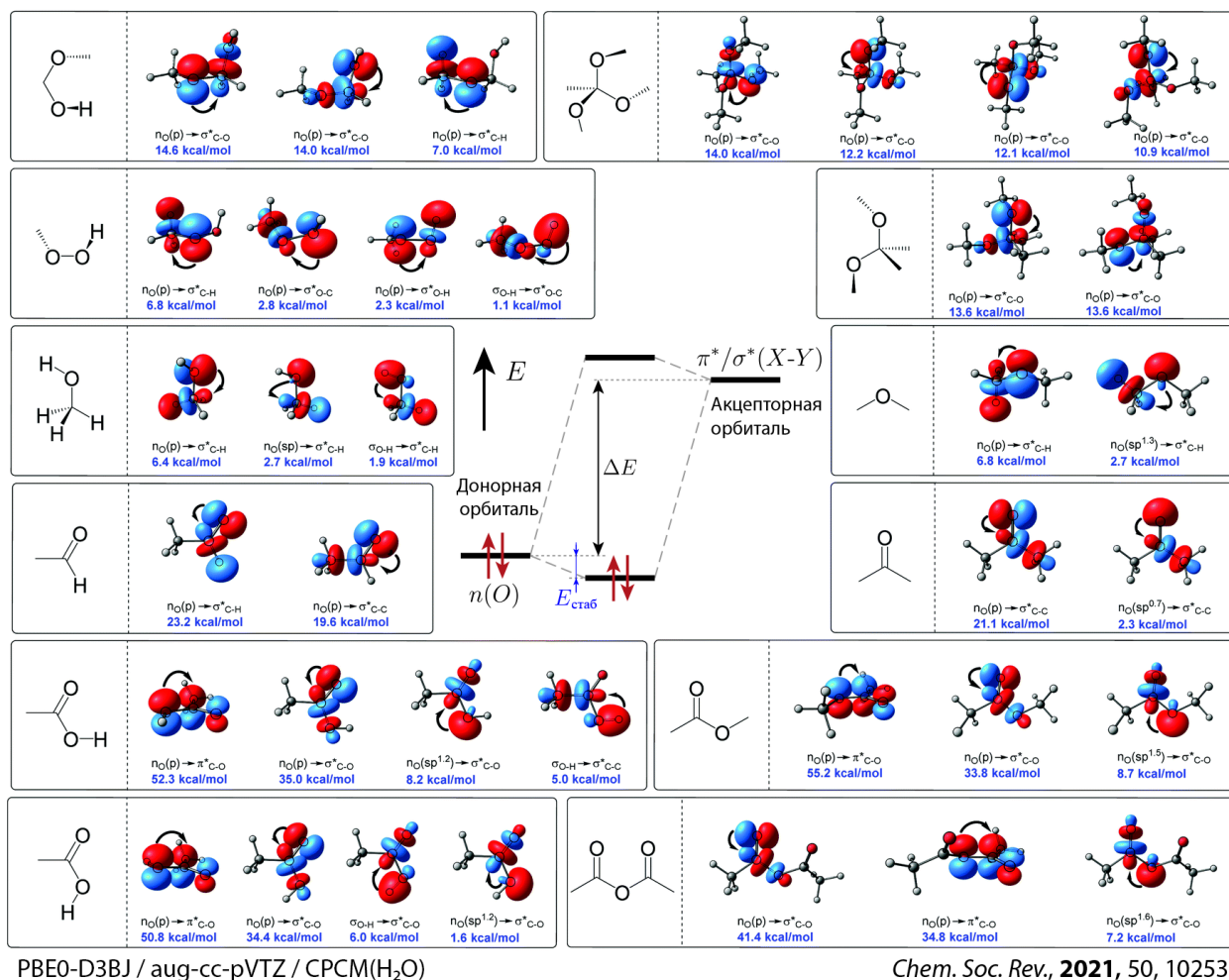


Рисунок 18. Выборка ключевых орбитальных взаимодействий в основных органических кислородсодержащих функциональных группах. Использована иллюстрация из публикации автора диссертации [178].

Стереэлектронный эффект даёт повышенную стабилизацию в сочетании со смещением орбиталей и пониженную – когда донор находится в электроотрицательном окружении (Рисунок 19, снизу). Первое явление известно как α -эффект, второе – как обратный α -эффект (ОАЭ). К α -эффекту часто прибегают для объяснения повышения реакционной способности нуклеофилов при добавлении соседнего гетероатома рядом с нуклеофильным гетероатомом, например пероксид-анион (OOH^-) более реакционноспособен, чем гидроксид-анион (OH^-) вследствие α -эффекта [12]. Однако И. В. Алабугин с соавторами [12] предположили, что для нейтральной пероксидной группы (не аниона) при внутримолекулярных

взаимодействиях может проявляться ОАЭ, то есть она выступает как более слабый донор по сравнению с одиночным атомом кислорода из-за «стягивания» на себя электронной пары кислорода-донора соседним с ним электроотрицательным атомом кислорода. Позднее ему совместно с коллегами из ИОХ РАН удалось найти реальную реакцию (перегруппировка Байера-Виллигера), неожиданная остановка которой была вызвана дестабилизацией переходного состояния именно за счёт ОАЭ [189].

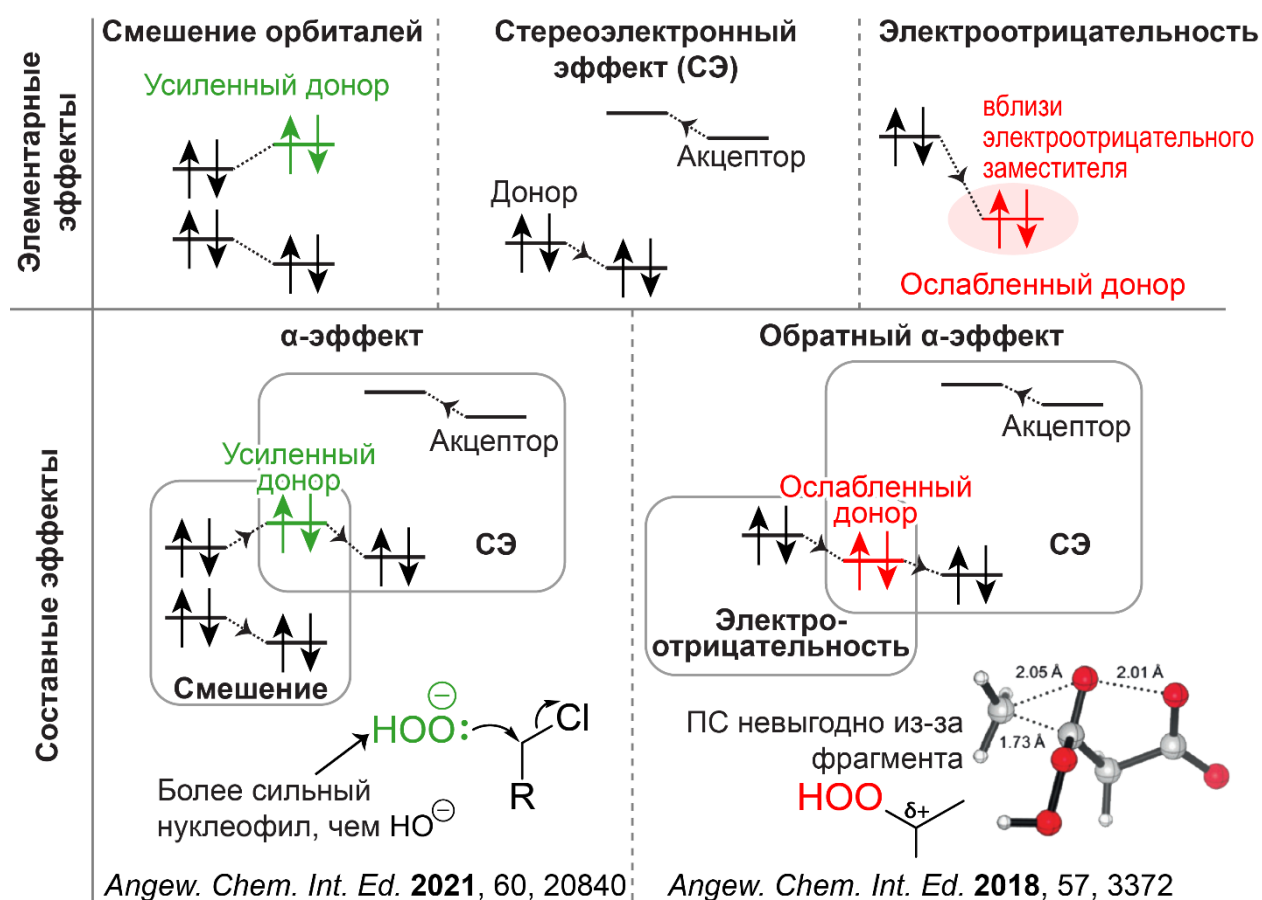


Рисунок 19. Основные типы орбитальных взаимодействий, проявляющиеся, в частности, для каркасных пероксидов.

В связи с высокой предсказательной и объяснительной силой стереoeлектронных эффектов существует задача оценки их энергетического вклада в стабилизацию молекулы из квантово-химических расчётов. Наиболее близкий к химической интуиции подход – это анализ локализованных орбиталей, то есть

построение альтернативного базиса, который соответствует хорошему «первому приближению» волновой функции резонансной структуры без учёта взаимодействий второго порядка. В наиболее популярном методе натуральных связевых орбиталей (Natural Bond Orbitals, NBO) [190] локализованные орбитали классифицируются следующим образом:

- внутренние электроны;
- неподелённые пары и вакантные орбитали;
- связывающие и разрыхляющие орбитали двухцентровых связей;
- связывающие, несвязывающие и разрыхляющие орбитали трёхцентровых связей;
- остаточные одноцентровые функции, называемые ридберговскими орбиталями.

Ключевой этап анализа NBO, связывающий квантово-химическую модель с химической интуицией, – это оценка стабилизации от парных орбитальных взаимодействий. Для этого матрицу Фока переводят в базис NBO и для интересующих пар локализованных орбиталей рассчитывают энергетическую поправку по теории возмущений второго порядка:

$$E_{\text{стаб}} = n_{\alpha} \cdot \frac{F_{ij}^2}{\Delta E}, \quad (9)$$

где n_{α} – электронная заселённость донорной орбитали, ΔE – разность энергий донорной и акцепторной орбиталей (то есть разность соответствующих диагональных элементов матрицы Фока, Рисунок 20), а F_{ij} – недиагональный элемент матрицы Фока.

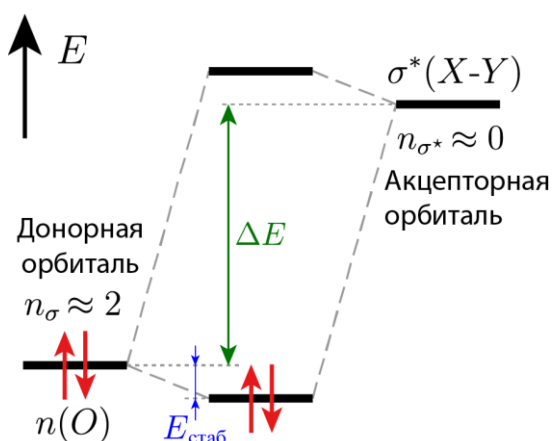


Рисунок 20. Орбитальная диаграмма донорно-акцепторного взаимодействия.

Иногда энергий парной делокализации недостаточно, чтобы проанализировать взаимодействие между интересующими молекулярными фрагментами. В таких случаях зачастую применяются подходы на основе блочно-локализованной волновой функции, которые решают ограниченную задачу самосогласованного поля с запретом на делокализацию определённых орбиталей за пределы отведённых для них фрагментов.

Ещё более «крупномасштабный» анализ возможен при полном отказе от рассмотрения орбиталей и переходе к моделированию изменению энтальпии (ΔH_r) для специально сконструированных гипотетических превращений, позволяющих изолировать энергетический эффект сборки или разборки интересующего молекулярного фрагмента. Например, изодесмические реакции, построенные на принципах материального баланса и сохранения числа связей каждого типа, применяются для оценки энергии межфрагментных взаимодействий [191] или напряжения каркасов [192] (Рисунок 21).

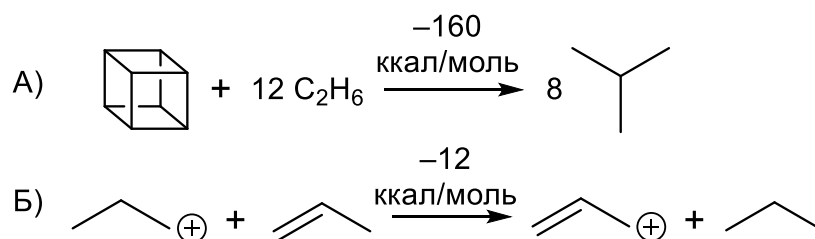


Рисунок 21. Примеры изодесмических реакций, которые могут использоваться для количественной оценки (А) напряжённости кубана, (Б) конъюгации в аллильном катионе. Показаны значения ΔH_r из [191,192].

Наивысший «масштаб» для анализа взаимодействий между фрагментами и проверки гипотез, усиливающих химическую интуицию, представляет хемоинформативный подход. В рамках данного подхода фокус смещается с конкретных геометрий на оценку предпочтительности тех или иных структурных мотивов через анализ частотности графовых или геометрических фрагментов в структурных базах данных, либо на основе энергетических или геометрических корреляций и отклонений от них. Так, через отклонения от корреляции между энергиями протонирования и барьерами нуклеофильной атаки, было предложено определение прямого и обратного α -эффектов [12]. Как другой пример, анализ распределений торсионных углов в кристаллографической базе данных позволил обнаружить склонность к обращению Z-эффекта карбоксильных групп в присутствии акцепторов водородных связей [10].

2.1.4. Анализ орбитальных взаимодействий

Построенная выше квантово-химическая модель (см. подраздел 2.1.2) включает anomalously высокую энергию катиона **A5** и ПС присоединения H_2O_2 к нему, что приводит к кинетической блокировке превращения **A2**→**A3**. Чтобы объяснить эту аномалию, мы более детально изучили электронную структуру катиона **A5**. Его структурный мотив $\text{O} - \text{O} - \text{C}^+$, который далее будет назван пероксикарбениевым катионом (ПКК), включает орбитальную π -конъюгацию, где донор представлен атомом кислорода пероксидной группы. Следуя подразделу 2.1.3 и Рисунку 19,

второй пероксидный кислород может играть двоякую роль в данной конъюгации: будучи заместителем с неподелённой парой, он может её усиливать посредством α -эффекта (через смешение занятых p -орбиталей), однако, являясь электроотрицательным заместителем, он также может ослаблять эту конъюгацию через обратный α -эффект (ОАЭ). На вопрос о том, какой из двух эффектов побеждает в изучаемой нами системе, свидетельства из расчётных данных указывают, что преобладает дестабилизирующий ОАЭ (справа внизу на Рисунке 19). Нами был проведён более детальный анализ катионов и ПС для подтверждения наличия ОАЭ и его количественной оценки.

Во-первых, ОАЭ оценили путём прямого сравнения энергий конъюгации $n_o \rightarrow p_c$, рассчитанных методом NBO, в катионе **A5a** и аналогичном катионе, где второй атомом кислорода заменен на углерод. Полученные оценки согласуются с гипотезой об ОАЭ – конъюгация в ПКК и ПС его реакции с H_2O_2 ослаблена на 14 и 11 ккал/моль, соответственно (Рисунок 22А).

Более крупномасштабное сравнение энергии катионов, стабилизированных одним атомом кислорода и пероксидной группой, было проведено через расчёт изодесмических реакций. Были рассмотрены гипотетические реакции замены одного атома кислорода на пероксидную группу в α -положении к катионному центру в трёх системах (Рисунок 22Б): (1) простой ациклический катион, (2) катионный интермедиат **A4b**, (3) ПС присоединения H_2O_2 к **A4b**. Полученные значения ΔG_r равны 9, 16 и 17 ккал/моль, соответственно, что показывает нежелательность соседства пероксидной группы и катионного центра. Примечательно, что меньшая ΔG_r для ациклического катиона указывает на усиление ОАЭ в конформационно жёстких системах. Это логично: электроотрицательность действует вне зависимости от торсионного угла связи O–O, тогда как эффективное смешение p -орбиталей требует их хорошего перекрывания, что достигается только при некоторых значениях торсионного угла.

Таким образом, результаты анализа NBO и изодесмических реакций подтверждают гипотезу об ОАЭ в ПКК (Рисунок 22В), что объясняет повышенный

энергетический барьер превращения **A2**→**A3** и невозможность экспериментального получения трис-пероксида **A3** в исследуемой реакции.

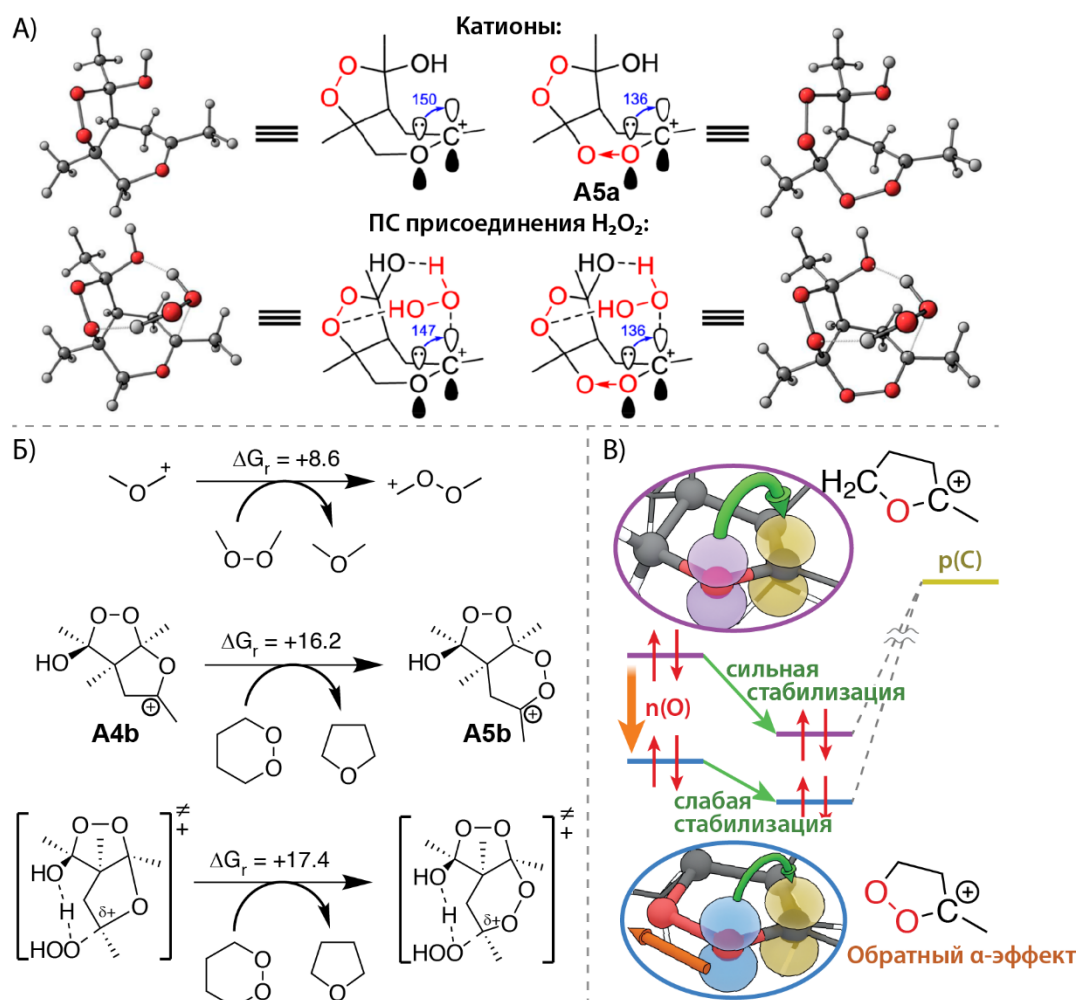


Рисунок 22. Количественная оценка обратного α -эффекта: (А) – через разницу энергий конъюгации $n_o \rightarrow p_c$ в ПКК **A5a** и катионе, стабилизированном одним атомом кислорода, (Б) – через замену фрагментов при катионном центре через изодесмические реакции; (В) – обратный α -эффект как ключевой эффект, дестабилизирующий пероксикарбениевые катионы.

Согласно рассчитанным выше значениям ΔG_r для трёх стадий реакции (см. Рисунок 14), превращения **A1**→**A2** оказываются систематически более энергетически выгодными, чем финальные стадии **A2**→**A3**, для которых ΔG_r оказывается вблизи нуля. Это удивительно, поскольку обе стадии включают лишь

замещение кислорода на пероксидную группу. Для решения этой загадки мы проанализировали стабилизацию молекул **A1–A3** гиперконъюгацией $n_O \rightarrow \sigma_{C-O}^*$. В случае конденсации трикетонов возможно до шести таких орбитальных взаимодействий (Рисунок 23).

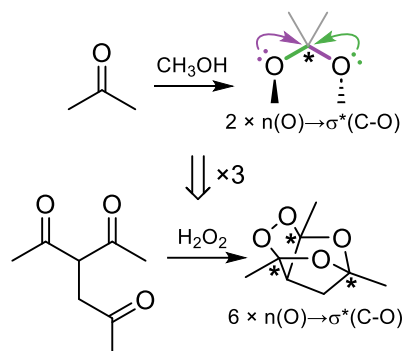


Рисунок 23. Каркасные пероксиды, получаемые из трикетонов, могут стабилизироваться максимум шестью взаимодействиями $n_O \rightarrow \sigma_{C-O}^*$.

Используя анализ NBO, мы обнаружили, что в молекуле трис-пероксида имеются две ослабленные гиперконъюгации (Рисунок 24). При одинаковом химическом окружении стереоэлектронная стабилизация определяется перекрыванием орбиталей, которое зависит от торсионного угла связей C–O. Сопоставление энергий орбитальных взаимодействий из NBO с соответствующими торсионными углами подтвердило, что ослабление гиперконъюгаций в **A3b** вызвано именно геометрией полициклического каркаса трис-пероксида. Следовательно, повышенная ΔG_r превращения **A2**→**A3** обусловлена совместным действием стереоэлектронных и конформационных факторов.

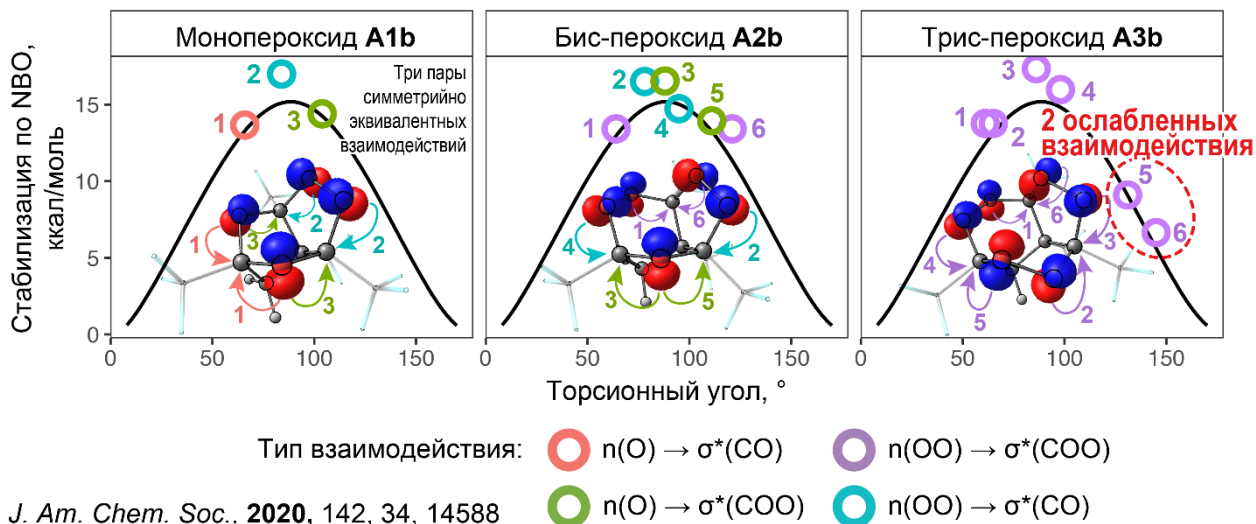
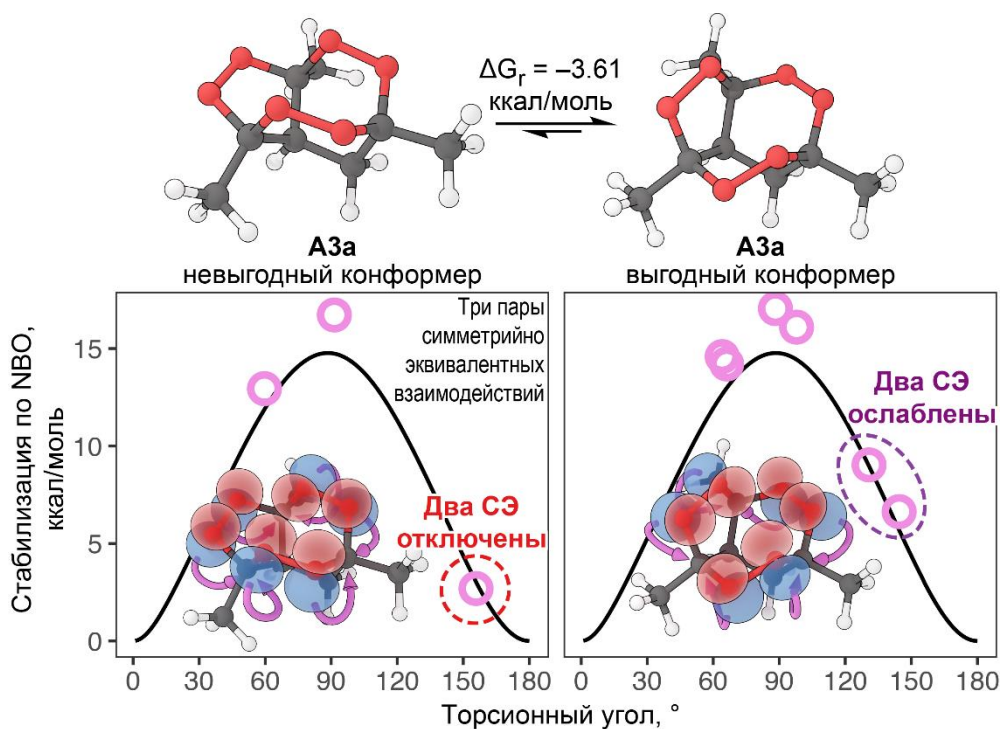


Рисунок 24. Энергии ключевых орбитальных взаимодействий в каркасных пероксидах **A1b–A3b** с указанием торсионных углов связей O–C, определяющих относительную пространственную ориентацию взаимодействующих орбиталей. Черная кривая показывает зависимость энергии гиперконъюгации $n_{\text{O}} \rightarrow \sigma_{\text{C-O}}^*$ от торсионного угла в диметоксиметане $\text{CH}_3\text{OCH}_2\text{OCH}_3$.

Конформационный анализ каркасных пероксидов показал, что существует только один возможный конформер для каркасов моно- и бис-пероксидов **A1** и **A2**. Однако, для трис-пероксидов **A3** были найдены два возможных конформера (см. пример для **A3a** на Рисунке 25). Интересно, что шестичленные циклы в высокоэнергетическом конформере имеют геометрию кресла, а в низкоэнергетическом – геометрию искажённой ванны. Для объяснения этого неожиданного факта, что более энергетически выгодный конформер визуальнее является более искажённым, мы опять использовали анализ NBO. Мы обнаружили, что в высокоэнергетическом конформере почти не реализуются два взаимодействия $n_{\text{O}} \rightarrow \sigma_{\text{C-O}}^*$ (стабилизация 2.7 ккал/моль от каждого, Рисунок 25). Те же самые стереоэлектронные эффекты в низкоэнергетической структуре, хоть и ослаблены, дают значительно бóльшую стабилизацию (9.0 и 6.7 ккал/моль).

Стоит отметить, что проведённый выше анализ и найденные зависимости энергии пероксидного каркаса от стабилизации гиперконъюгацией обладают

значительной новизной – предшествующие исследования ограничивались бициклическими каркасами, бис-пероксидами и озонидами [11].



J. Am. Chem. Soc., **2020**, 142, 34, 14588

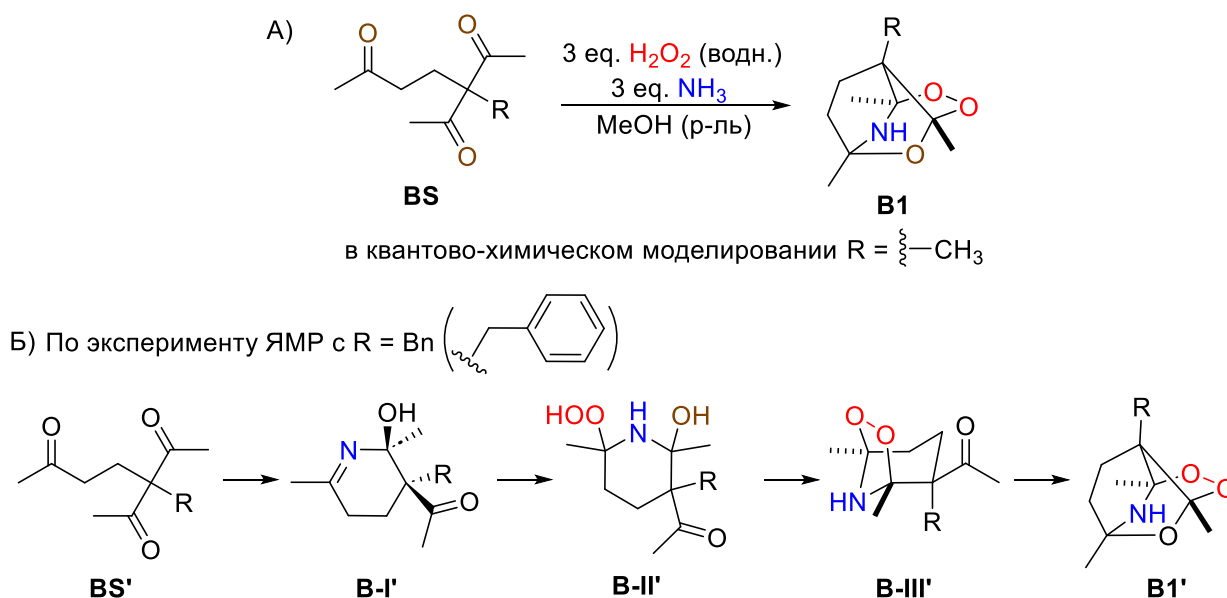
Рисунок 25. Контроль конформационных предпочтений трис-пероксида **A3a** стереоэлектронными эффектами. Кольцами на графике показаны энергии стабилизации из NBO от ключевых орбитальных взаимодействий в конформациях **A3a** с указанием торсионных углов связей O–C, определяющих относительную пространственную ориентацию взаимодействующих орбиталей. Черная кривая показывает зависимость энергии гиперконъюгации $n_O \rightarrow \sigma_{C-O}^*$ от торсионного угла в диметоксиметане $\text{CH}_3\text{OCH}_2\text{OCH}_3$.

Описанные в данном подразделе результаты моделирования опубликованы в *Journal of the American Chemical Society* [173].

2.2. Добавление конкурирующего N-нуклеофила. Моделирование реакции β,δ' -трикетонов, пероксида водорода и аммиака.

Установив ряд электронных эффектов, контролирующих сборку каркасных пероксидов из β,γ' -трикетона и H_2O_2 , мы обратили внимание на схожую реакцию с участием аммиака в качестве дополнительного нуклеофила.

Нашими коллегами был разработан синтез новых трициклических аминпероксидов **V1** из β,δ' -трикетонов **BS**, аммиака и пероксида водорода в среде метанола (Рисунок 26А). Замечательно, что несмотря на возможность разных комбинаций мостиковых групп ($-\text{O}-$, $-\text{O}-\text{O}-$ и $-\text{N}(\text{H})-$) в трициклическом каркасе, в данной системе селективно образуется только трициклический аминпероксид **V1**, содержащий по одному фрагменту каждого типа. С помощью спектроскопии ЯМР удалось зафиксировать три интермедиата сборки продукта из трикетона с бензильным заместителем (**BS'**, Рисунок 26Б): моноциклический имин **V-I'**, амингидропероксид **V-II'**, и бициклический кето-азаозонид **V-III'**, что существенно упростило последующее моделирование.



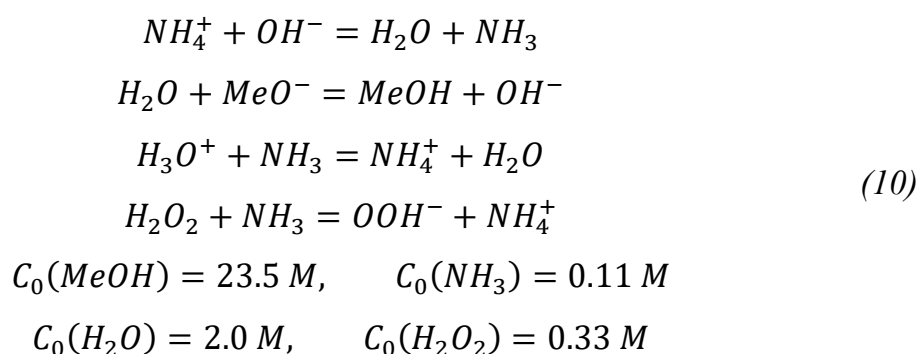
J. Am. Chem. Soc., 2022, 144, 16, 7264

Рисунок 26. (А) – Схема исследуемой реакции β,δ' -трикетонов с пероксидом водорода и аммиаком. (Б) – Интермедиаты процесса сборки трициклического аминпероксида из β,δ' -трикетона **V1'** с бензильным заместителем, обнаруженные с помощью спектроскопии ЯМР.

В рамках диссертационного исследования было проведено квантово-химическое моделирование этой реакции для объяснения наблюдаемой селективности образования аминпероксидов **В1**.

2.2.1. Используемые методы

Моделирование выполнено на уровне теории PBE0-D3BJ [19,20] / 6-311++G(d,p) [21,22] / SMD(MeOH) [23] в Gaussian16 A.03 [18]. Для всех расчётов свободной энергии применялась квазигармоническая поправка, рассчитываемая в библиотеке GoodVibes [25]. Все термодинамические поправки рассчитаны для температуры 0 °С. При расчётах трансляционной энтропии в программе GoodVibes были учтены равновесные концентрации NH_4^+ , NH_3 , H_2O , H_2O_2 и MeOH на основе результатов моделирования следующего кислотно-основного равновесия:



Начальные концентрации были оценены из экспериментальной методики проведения реакции [175]. Для реакций (10) значения ΔG_r были получены из квантово-химических расчётов с учётом трёх явных молекул метанола, связанных с каждой реагирующей молекулой. Численное интегрирование кинетических уравнений с небольшими ΔG^\ddagger (6 ккал/моль в направлении с $\Delta G_r < 0$), согласованными с квантово-химическими величинами ΔG_r , было использовано для нахождения точки равновесия:

$$\begin{aligned}
 [\text{MeOH}] &= 23.5 \text{ M}, & [\text{MeO}^-] &= 2.7 \cdot 10^{-4} \text{ M} \\
 [\text{H}_2\text{O}] &= 2.0 \text{ M}, & [\text{OH}^-] &= 8.78 \cdot 10^{-6} \text{ M} \\
 [\text{H}_3\text{O}^+] &= 3.7 \cdot 10^{-11} \text{ M} \\
 [\text{NH}_3] &= 1.6 \cdot 10^{-2} \text{ M}, & [\text{NH}_4^+] &= 9.3 \cdot 10^{-2} \text{ M}
 \end{aligned} \tag{11}$$

$$[H_2O_2] = 0.24 \text{ M}, \quad [OOH^-] = 9.3 \cdot 10^{-2} \text{ M}$$

Эти концентрации использовались для расчёта термодинамических поправок к энергиям NH_3 , H_2O , H_2O_2 и $MeOH$ (без явных молекул метанола), а также для оценки энергий протонирования интермедиатов. Предполагалось, что NH_4^+ является наилучшим источником протонов в системе, поэтому разница энергий $G(NH_4^+) - G(NH_3)$ использовалась для вычисления относительных энергий положительно заряженных интермедиатов.

Конформационный поиск всех рассмотренных молекул и ПС был проведён путём генерации набора из ~100 конформаций методом ETKDG [14], имплементированном в пакете RDKit [26] и осуществляющего глобальный конформационный перебор как циклических, так и нециклических фрагментов молекулы (см. подраздел 1.1.1.1). Затем для каждой структуры из сгенерированного набора были проведены запуски пайплайна iMTD-GC, имплементированного в пакете CREST [27]. Конформации, отличающиеся друг от друга по величине RMSD ($RMSD_{non-H} > 0.3 \text{ \AA}$), были отобраны и оптимизированы на целевом уровне теории.

Несмотря на то, что эксперименты проводились для субстрата **BS'** с бензильным заместителем, моделирование реакции решено проводить для **BS** с метильным заместителем, так как это упрощает квантово-химические расчёты и конформационный поиск путём снижения количества атомов и степеней свободы системы. Данное решение было подкреплено тем, что низкая стерическая загруженность молекулы позволяет бензильному заместителю развернуться так, чтобы отталкиваться от других заместителей не больше, чем метил на его месте, а отсутствие других ароматических фрагментов исключает его участие в π -стекинге, что могло бы повлиять на конформацию молекулы.

2.2.2. Результаты квантово-химического моделирования

Аналогично описанному в подразделе 2.1.2, квантово-химическое моделирование сначала было выполнено с применением термодинамической модели для сравнения энергий образования **V1**, полученного экспериментально, и гипотетически возможных продуктов **V2–V9**.

Для расчёта ΔG_r сборки каркасных пероксидов в первую очередь необходимо было получить их низкоэнергетические конформеры. Конформационный анализ молекул **V1–V9** методом ETKDG показал, что полициклические каркасы **V4–V6** и **V9** имеют один единственный конформер. Неподвижным полициклическим каркасом обладают также молекулы **V1–V3** и **V8**, содержащие мостики $-N(H)-$, тем не менее для них было найдено несколько конформеров (два – для **V1**, **V3** и **V8**, и четыре – для **V2**), обусловленных двумя направлениями пирамидализации атомов азота (Рисунок 27А); более подробно подвижность связи N–N обсуждена в подразделе 2.4.3. Для каркаса трис-пероксида **V7** были найдены два конформера, различающихся конформацией одного из семичленных циклов и свободной энергией (на 1.9 ккал/моль). Сравнение с результатами конформационного анализа из подраздела 2.1 показывает, что конформационная подвижность **A1**, **A2** и **A3** схожа с их аналогами с дополнительным звеном $-CH_2-$ в углеродном мостике – **V4**, **V9** и **V7**, соответственно. Каркасы **A1 / V4** и **A2 / V9** имеют только один конформер, тогда как **A3 / V7** – два. Различие наблюдается лишь между высокоэнергетическими конформерами **A3** и **V7**: у **A3** шестичленные циклы принимают конформацию кресла, создавая плоскость симметрии, тогда как у **V7** один из семичленных циклов находится в конформации кресла, а другой – в конформации ванны.

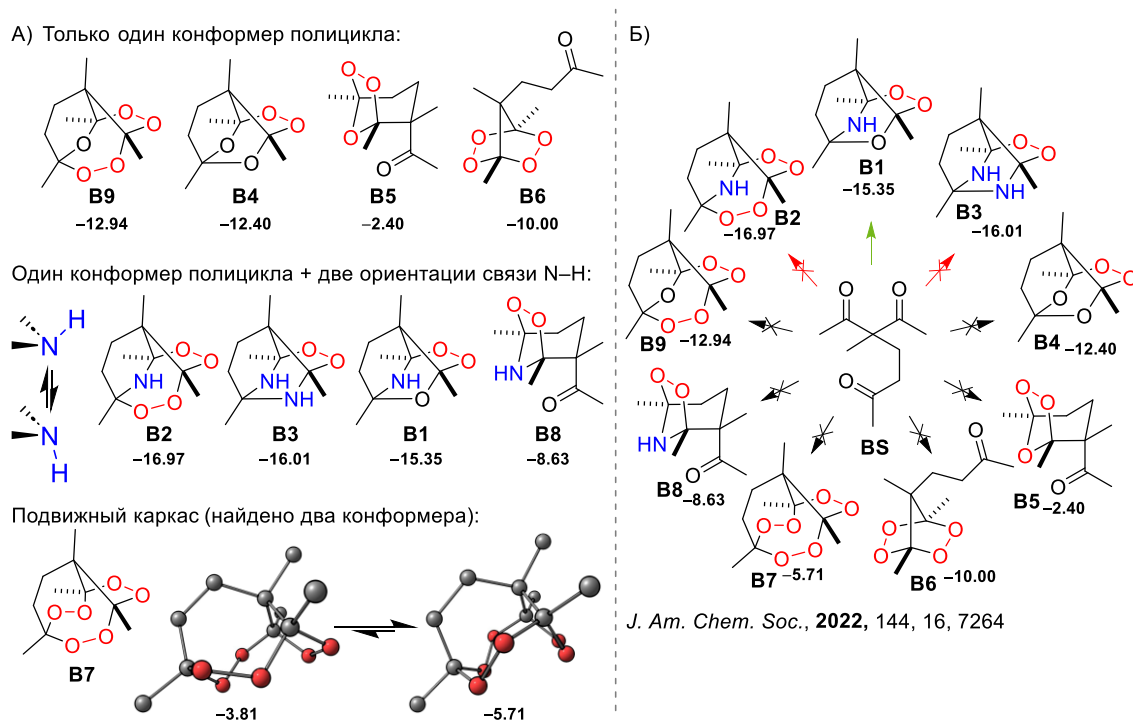
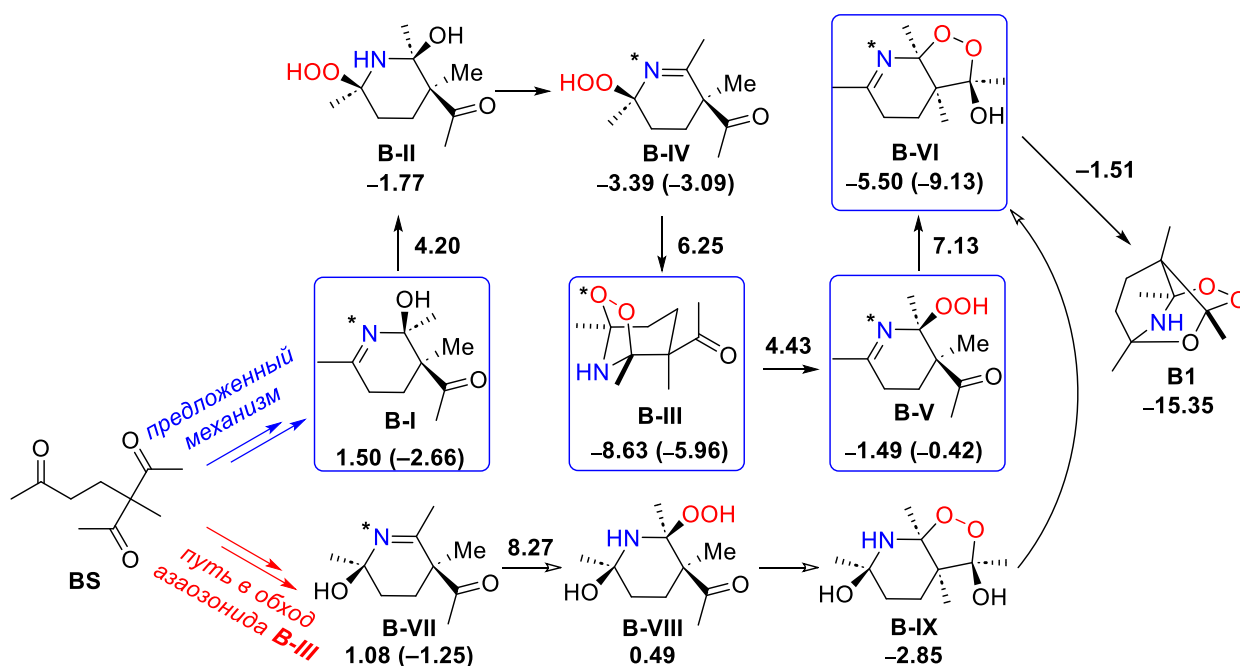


Рисунок 27. (А) – Классификация каркасных пероксидов **V1–V9** по их конформационной подвижности. (Б) – Термодинамический анализ экспериментального (**V1**) и гипотетических (**V2–V9**) продуктов реакции. Для каждой структуры приведены ΔG_f образования из реагентов: β,δ' -трикетона **BS**, H_2O_2 и NH_3 . **V1** – единственный экспериментально наблюдаемый продукт, **V2** и **V3** – не зафиксированные в эксперименте продукты, оказывающиеся более термодинамически выгодными, по сравнению с **V1**, согласно проведённому квантово-химическому моделированию.

Для полученных низкоэнергетических конформеров **V1–V9** была рассчитана ΔG_f сборки из исходных молекул (Рисунок 27Б). Мы обнаружили, что два аминокпероксида **V2** и **V3** (красные стрелки на Рисунке 27Б) более термодинамически выгодны, чем экспериментальный продукт **V1**. Остальные гипотетические продукты **V4–V9** менее термодинамически выгодны, чем **V1**. Поэтому целью дальнейшего моделирования стало объяснение отсутствия термодинамически предпочтительных продуктов **V2** и **V3**.

Для установления кинетических причин, по которым **B2** и **B3** не наблюдаются в эксперименте, было проведено моделирование механизма сборки **B1–B3**. На основании экспериментальных данных (Рисунок 26Б) был предложен механизм сборки **B1** через промежуточное образование азаозонида **B-III**. В рамках этого механизма были рассчитаны энергии интермедиатов и ПС, а также для сравнения проведены расчёты прямой сборки **B1** без образования **B-III** (Рисунок 28).

По результатам моделирования, все стадии от исходного **BS** до продукта **B1** через экспериментально зафиксированные интермедиаты **B-I**, **B-II** и **B-III** являются кинетически разрешенными. Наиболее высокое ПС на данном пути имеет относительную энергию 7.13 ккал/моль, что ниже барьера для альтернативного маршрута (8.27 ккал/моль). Таким образом, результаты квантово-химического моделирования механизма сборки (Рисунок 28) согласуются с экспериментальными наблюдениями.



J. Am. Chem. Soc., 2022, 144, 16, 7264

Рисунок 28. Предложенный механизм сборки аминпероксида **B1** через промежуточный азаозонид **B-III** (путь вверх) и альтернативный механизм без образования **B-III** (путь вниз). Рамкой выделены интермедиаты, из которых возможно ветвление реакционного пути в сторону термодинамически более выгодных продуктов **B2** и **B3**.

Аналогично моделированию сборки пероксидов из β,γ' -трикетонов, квантово-химический анализ был расширен для учёта интермедиатов и ПС, ведущих к экспериментальному (**V1**) и термодинамически более выгодным (**V2** и **V3**) продуктам. Были рассмотрены пять возможных вариантов ветвления процесса сборки **V1** в направлении **V2** и **V3** (Рисунок 30). На всех развилках энергетический барьер в сторону **V1** оказывается ниже. В частности, барьеры присоединения дополнительной молекулы аммиака, необходимого для сборки **V3**, повышены вследствие равновесного протонирования аммиака в реакционной смеси. Стадии присоединения второй молекулы пероксида водорода, требующиеся для образования **V2**, конкурируют с внутримолекулярными перегруппировками на пути к **V1** и также им уступают по энергии активации.

Наконец, перегруппировки уже собранного каркаса **V1** в **V2** и **V3** имеют повышенные энергетические барьеры (28 и 33 ккал/моль, соответственно), так как их механизмы включают образование невыгодных ПКК (Рисунок 29), дестабилизированных ОАЭ.

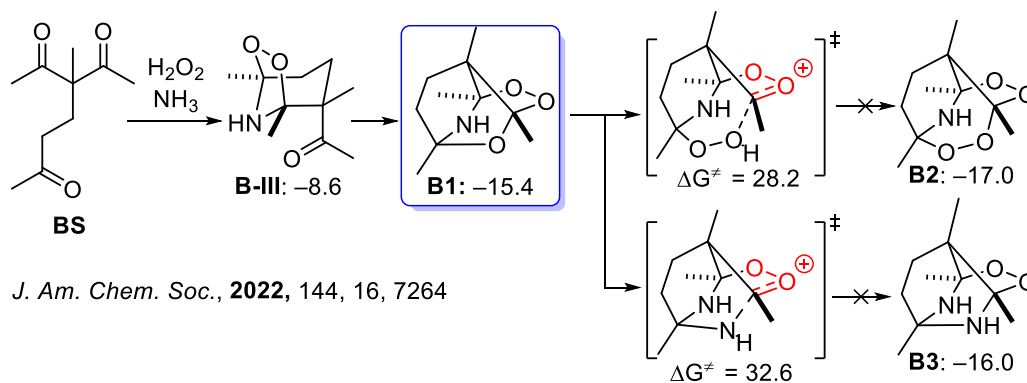


Рисунок 29. Переходные состояния замыкания ПКК, которые блокируют перегруппировку продукта **V1** в более термодинамически выгодные каркасы **V2** и **V3**.

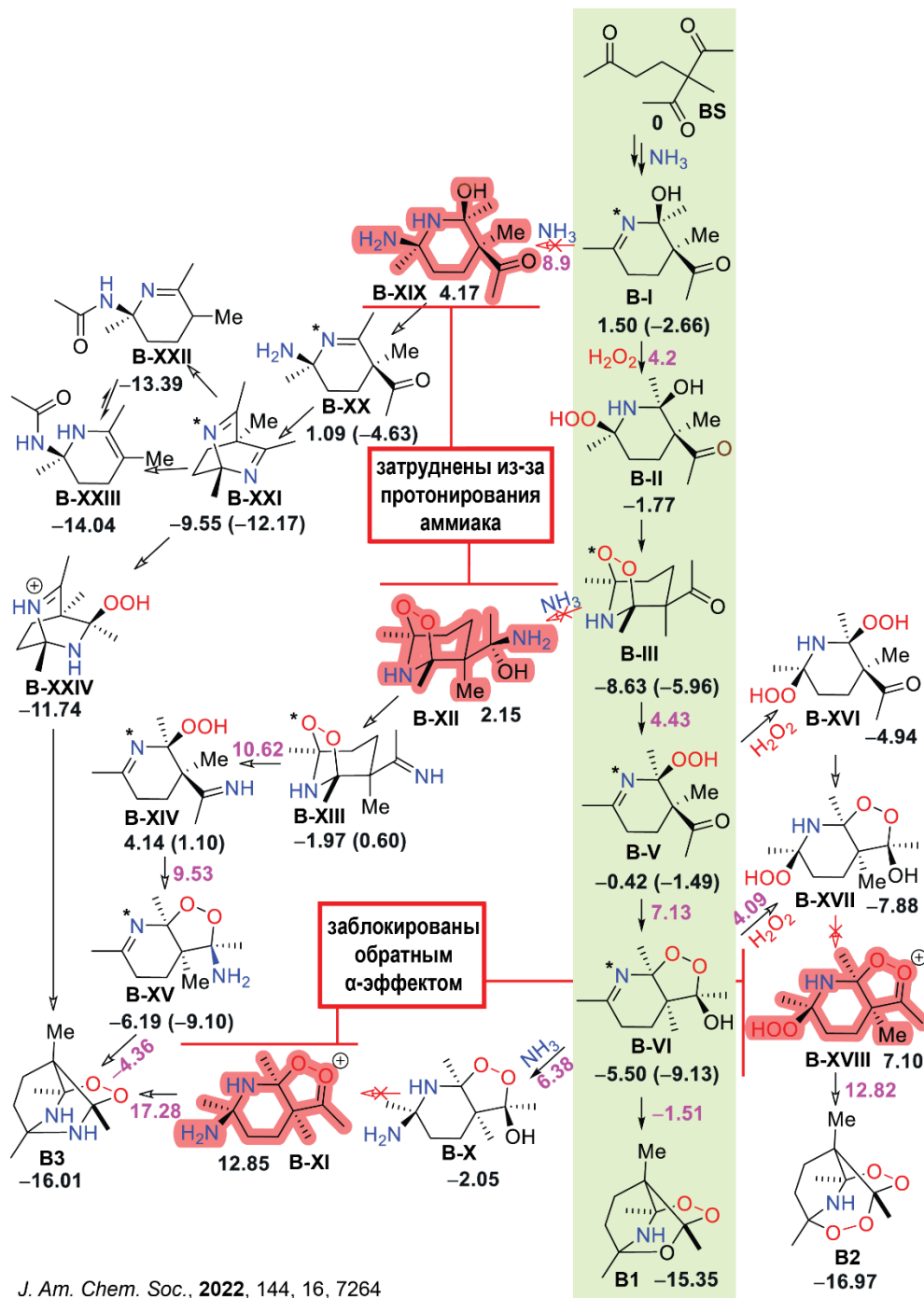


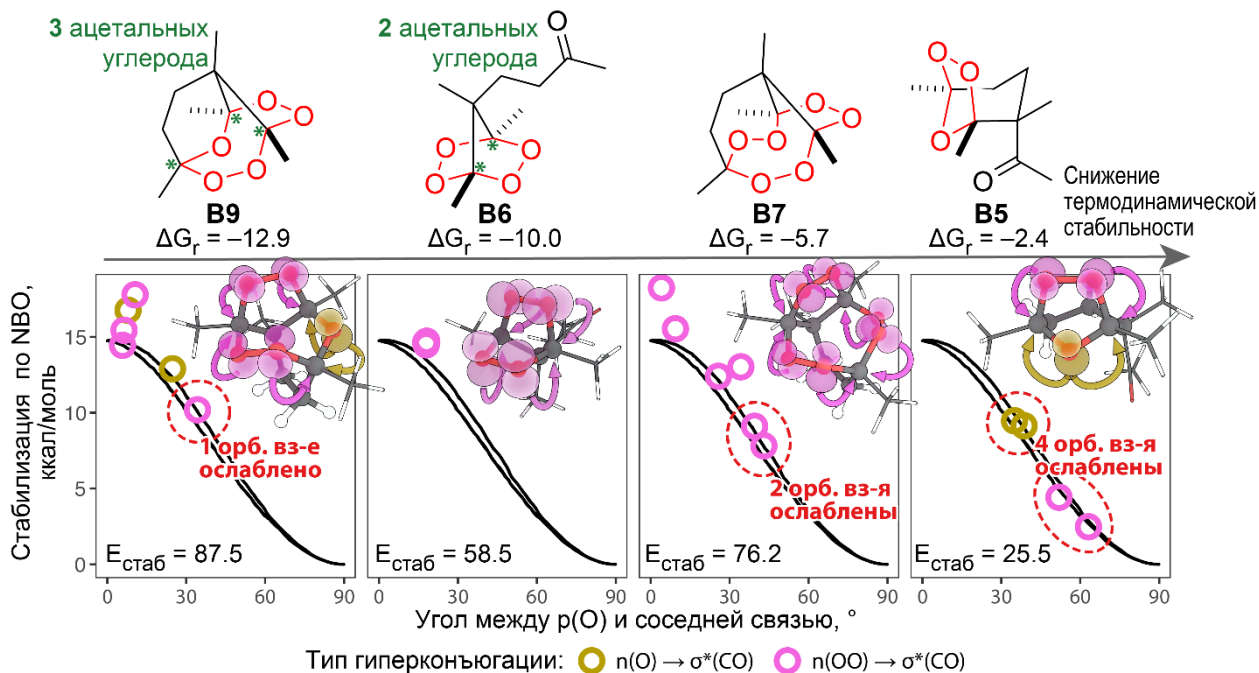
Рисунок 30. Относительные энергии интермедиатов (чёрный) и переходных состояний (фиолетовый, рядом со стрелками), полученные в результате расчётного моделирования механизмов сборки **B1–B3**.

2.2.3. Анализ орбитальных взаимодействий

Полученные выше ΔG_r сборки ряда каркасных пероксидов (Рисунок 27) позволяют продолжить анализ, начатый в подразделе 2.1.4, и проверить на расширенном наборе систем возможность использования энергий гиперконъюгаций $n_X \rightarrow \sigma_{C-Y}^*$ (где X и Y — атомы кислорода или азота) для описания и предсказания термодинамической стабильности каркасных пероксидов. С этой целью был вновь применён анализ NBO.

Во-первых, было обнаружено, что энергия образования каркасов и их стабилизация гиперконъюгацией зависит от количества взаимодействий $n_O \rightarrow \sigma_{C-O}^*$, то есть от числа ацетальных атомов углерода (соединенных с двумя кислородами). Так, трициклический бис-пероксид **B9** с тремя ацетальными углеродами более стабилизирован, чем бициклический бис-пероксид **B6**, у которого только два ацетальных углерода, благодаря большему количеству орбитальных взаимодействий (6 против 4, Рисунок 31).

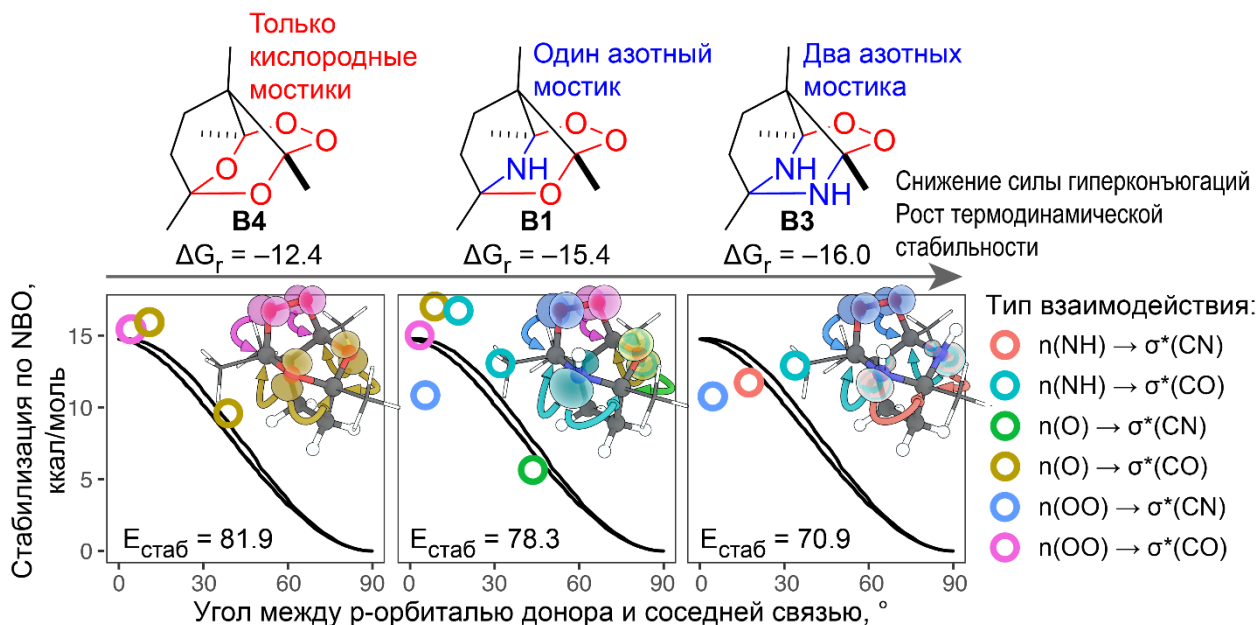
Во-вторых, существенную роль играет конформация каркаса, влияние которой уже обсуждалось в подразделе 2.1.4. Например, в озониде **B5** все четыре взаимодействия $n_O \rightarrow \sigma_{C-O}^*$ ослаблены из-за неоптимальной ориентации орбиталей и отсутствия конформационной подвижности, необходимой для её исправления. Этим объясняется относительно низкая свободная энергия сборки озонида **B5** из реагентов ($\Delta G_r = -2.4$ ккал/моль). Аналогично, в трициклическом трис-пероксиде **B7** две гиперконъюгации ослаблены, что коррелирует с его пониженной термодинамической предпочтительностью ($\Delta G_r = -5.6$ ккал/моль). Таким образом, прослеживается зависимость между термодинамической предпочтительностью продуктов реакции и полнотой активации гиперконъюгаций в них.



J. Am. Chem. Soc., 2022, 144, 16, 7264

Рисунок 31. Ключевые стереоэлектронные взаимодействия в молекулах потенциально возможных продуктов. Внутри графиков также приведены значения сумм энергий стабилизации по NBO от шести изображенных гиперконъюгаций.

Однако ситуация усложняется при появлении в рассматриваемых молекулах азотных мостиков $-N(H)-$. В ряду каркасов **B4**→**B1**→**B3** с увеличением числа азотных мостиков и уменьшением числа кислородных обнаружена тенденция роста термодинамической стабильности (Рисунок 32); при этом, парадоксально, суммарная стабилизация гиперконъюгациями в данном ряду существенно понижается. Ослабление гиперконъюгации связано, в частности, с меньшей электроотрицательностью атома азота, вследствие чего орбитали σ_{C-N}^* являются более слабыми акцепторами, чем σ_{C-O}^* . Следовательно, рост термодинамической стабильности требует наличия иного более сильного эффекта, компенсирующего потери в гиперконъюгации.



J. Am. Chem. Soc., **2022**, 144, 16, 7264

Рисунок 32. Сравнение энергии ключевых стереоэлектронных взаимодействий в молекулах потенциально возможных продуктов с разным количеством групп $-O-$ и $-N(H)-$. Внутри графиков также приведены значения сумм энергий стабилизации по NBO от шести изображенных гиперконъюгаций.

Для выяснения причин роста стабильности азотсодержащих пероксидов на фоне снижения стереоэлектронной стабилизации достаточно сравнить энергии связей. Превращение пары связей $C-O / N-H$ в пару $C-N / O-H$ соответствует замене мостикового кислорода в полициклическом скелете на мостик NH. Расчёт модельной реакции (Рисунок 33А) показывает, что пара связей $C-N / O-H$ действительно выгоднее на 3.6 ккал/моль по электронной энергии. Таким образом, при приблизительно одинаковой стереоэлектронной стабилизации присутствие атомов азота в продукте является термодинамически предпочтительным (Рисунок 33Б).

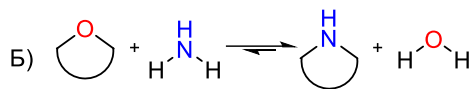
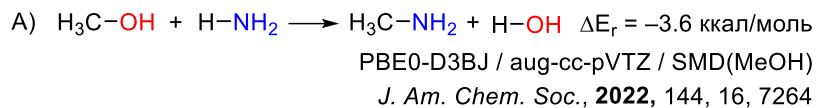


Рисунок 33. (А) Реакция превращения пары связей С–О / N–Н в пару С–N / О–Н; (Б) Термодинамическая предпочтительность продуктов с азотными мостиками, как следствие отрицательной электронной энергии реакции, ΔE_r .

Примечательно, что аддитивная модель, учитывающая только энергии гиперконъюгации из NBO и стабилизацию молекулы за счёт замены связей С–О на связи С–N ($3.6 \cdot 2 = 7.2$ ккал/моль) удовлетворительно описывает относительную стабилизацию молекул. Так, например, при переходе от монопероксида **В4** к аминпероксиду **В1** стабилизация по данным NBO снижается на $81.9 - 78.3 = 3.6$ ккал/моль (Рисунок 32), и с учётом стабилизации 7.2 ккал/моль за счёт ковалентных связей разницу энергий между **В1** и **В4** можно оценить в 3.6 ккал/моль, тогда как по данным из квантово-химического расчёта она составляет 3.0 ккал/моль. Аналогично, разницу энергий между **В3** и **В4** можно оценить в 3.8 ккал/моль, что хорошо согласуется с квантово-химическим расчётом (3.6 ккал/моль). Это наблюдение подтверждает, что обсуждаемые нами эффекты, гиперконъюгация и энергии связей, действительно вносят ключевой вклад в относительные энергии каркасных пероксидов.

Описанные в подразделе 2.2 результаты моделирования опубликованы в *Journal of the American Chemical Society* [175].

Обнаруженные высокие барьеры перегруппировок **В1**→**В2** и **В1**→**В3** (см. Рисунок 29) служат дополнительным подтверждением того, что стадии, протекающие через пероксикарбениевые катионы ($\text{O}-\text{O}-\text{C}^+$), являются кинетически невыгодными. В связи с этим представляет интерес изучение аналогичного эффекта с атомами азота, то есть в азкарбениевых катионах $\text{N}-\text{N}(\text{H})-\text{C}^+$.

2.3. Аналог обратного α -эффекта с атомами азота. Моделирование реакции β,δ' -трикетонов с пероксидом водорода и гидразидом.

В 2023 году нашими коллегами было расширено исследование реакций сборки трициклических аминпероксидов **C1** из β,δ' -трикетонов **BS**, N-нуклеофилов и пероксида водорода в среде метанола на рассмотрение гидразидов в качестве N-нуклеофилов. Введение второго атома азота в нуклеофил меняет природу некоторых интермедиатов, характерных для механизма с аммиаком (Рисунок 34Б). Образование промежуточных иминов **V-I** становится невозможным из-за заместителя при атоме азота; вместо них ожидается образование азакарбениевых катионов **C-I** и **C-II**. Эти катионы содержат фрагмент, являющийся аналогом ПКК, которые дестабилизированы обратным α -эффектом (см. выше), но атомы кислорода пероксида в данном случае заменены на атомы азота (Рисунок 34А).

Удивительно, что, как показали эксперименты, изменение N-нуклеофила не препятствует протеканию реакции и приводит к продукту **C1**, аналогичному продукту **V1**, получаемому при реакции с аммиаком (см. Рисунок 26).

В рамках диссертационного исследования проведено квантово-химическое моделирование данной реакции с целью объяснения экспериментально наблюдаемой селективности реакции, установления наличия или отсутствия обратного α -эффекта в интермедиатах и количественного сравнения его дестабилизирующего эффекта с пероксикарбениевыми катионами, аналогичными рассмотренным в подразделах 2.1 и 2.2.

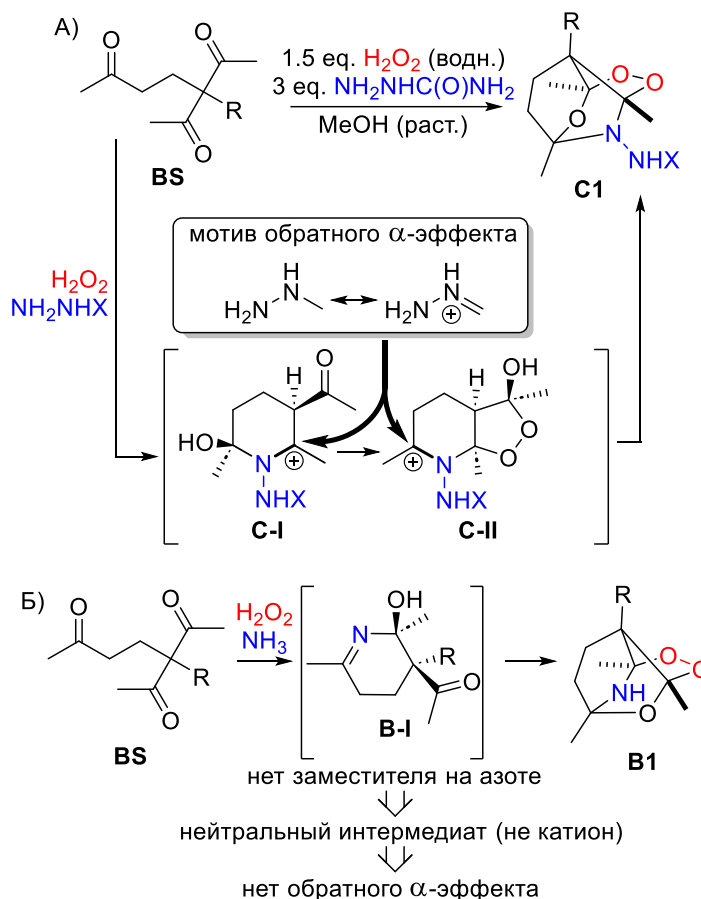


Рисунок 34. (А) – Схема исследуемой реакции β,δ,γ -трикетонов **BS** с пероксидом водорода и гидразидом; иллюстрация мотива обратного α -эффекта с участием атомов азота в катионных интермедиатах. (Б) – Сравнение с аналогичным интермедиатом в реакции с аммиаком из подраздела 2.2 – в данном случае нет возможности для обратного α -эффекта.

2.3.1. Используемые методы

Моделирование выполнено на уровне теории PBE0-D3BJ [19,20] / 6-311++G(d,p) [21,22] / CPCM(MeOH) [24] в Gaussian16 A.03 [18]. Конформационный поиск всех рассмотренных интермедиатов и ПС был проведён с помощью пайплайна iMTD-GC, имплементированного в пакете CREST [27] (см. обзор алгоритма в подразделе 1.1.5).

Моделирование реакции было решено проводить для наименьшего из экспериментально рассмотренных гидразидов – семикарбазида, $NH_2NHC(O)NH_2$ –

что упрощает квантово-химические расчёты и конформационный поиск за счёт уменьшения количества атомов и степеней свободы системы.

2.3.2. Результаты моделирования и анализ орбитальных взаимодействий

Сравнение ΔG_r сборки экспериментального продукта **C1** с энергиями гипотетических продуктов **C2** и **C3** выявило картину (Рисунок 35), аналогичную реакции с аммиаком: продукты замещения третьего атома кислорода являются более термодинамически выгодными, однако не наблюдаются в эксперименте, что указывает на существование кинетических барьеров в их образовании.

Как и в ранее рассмотренных реакциях, необходимость формирования ПКК создаёт кинетически непреодолимые барьеры для образования данных продуктов (Рисунок 35). Даже для **C-III**, являющегося более низкоэнергетическим из двух ПКК, увеличение энергии при переходе из **C1** достигает 26 ккал/моль (моделирование непосредственных ПС для данной реакции не проводилось ввиду схожести с реакцией, описанной в подразделе 2.2). В свою очередь, азакарбениевые катионы **C-I** и **C-II** проявляют себя как очень выгодные интермедиаты: в частности, значения ΔG_r их образования из исходных молекул отрицательны, более того, ΔG_r непосредственных стадий дегидратации с образованием этих катионов (показаны курсивом на Рисунке 35) также отрицательны. Таким образом, результаты расчётов энергии **C1–C3** и ключевых интермедиатов их сборки согласуются с экспериментальными данными и расчётами, выполненными выше для аммиака в качестве нуклеофила.

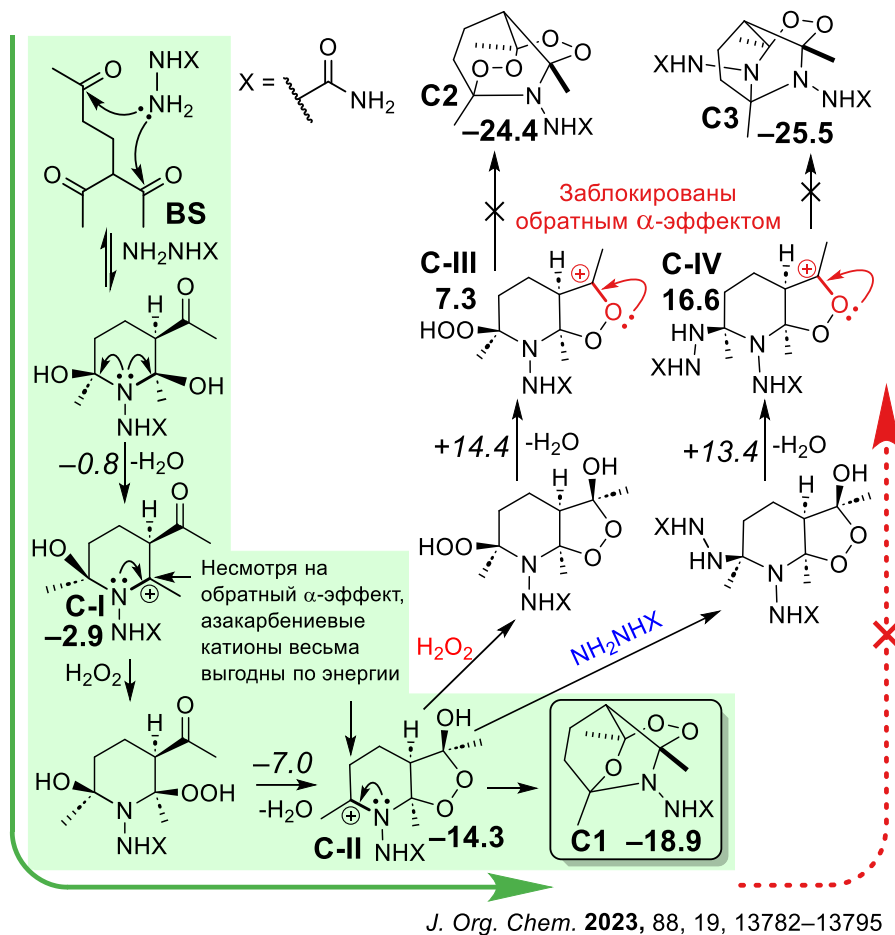


Рисунок 35. Энергии ключевых интермедиатов, рассмотренных в рамках моделирования сборки C1–C3. ΔG_r образования продуктов и интермедиатов из исходных молекул приведены жирным шрифтом рядом с их структурами; ΔG_r стадий дегидратации приведены курсивом рядом со стрелками.

Чтобы более детально выявить закономерности, определяющие различную стабилизирующую способность функциональных групп, был проведён анализ ряда изодесмических реакций замены заместителей при карбокатионном центре (Рисунок 36). Прежде всего отчётливо проявляется различие в π -донорной способности между атомами азота и кислорода. Разница между наименее стабилизирующей из азотсодержащих групп и наиболее стабилизирующей из кислородсодержащих превышает 16 ккал/моль. Кроме того, можно явно наблюдать ОАЭ при переходе от $-OH$ к $-OOH$ (дестабилизация ~ 7 ккал/моль), что согласуется с повышенной ΔG_r образования ПКК из реагентов, которая наблюдалась в подразделах 2.1 и 2.2.

Зависимость стабилизации от природы азотного заместителя позволяет проследить различные варианты баланса между электроотрицательностью, обуславливающей дестабилизирующий ОАЭ, и смешением р-орбиталей, ответственным за стабилизацию (прямым) α -эффектом. В случае семикарбазидного заместителя катион наименее стабилизирован, что указывает на преобладание ОАЭ, вызванного снижением энергии неподелённой электронной пары азота при карбокатионе из-за электроотрицательности соседнего с ним атома азота; при этом прямой α -эффект (повышение энергии данной неподелённой пары за счёт смешения с парой соседнего атома азота) не реализуется, вероятно, из-за вовлечения неподелённой пары соседнего атома азота во взаимодействие $n_N \rightarrow \pi_{C=O}^*$ с карбонильной группой. Несмотря на значительную величину ОАЭ (~8 ккал/моль относительно NH_2 -группы), схожей с эффектом в ПКК (6–15 ккал/моль), семикарбазид всё же обеспечивает хорошую стабилизацию катиона – лучшую, чем гидроксильная группа. В случае гидразиновой группы $-N(H)NH_2$, которая не содержит дополнительного π -акцептора, наблюдается незначительное преобладание (прямого) α -эффекта: стабилизация примерно на 1 ккал/моль выше по сравнению с аминогруппой $-NH_2$.

Таким образом, моделирование данной реакции позволило углубить понимание возможных проявлений ОАЭ в процессах сборки каркасных пероксидов. Реакция с семикарбазидом является ещё одним примером кинетической блокировки сборки каркасов с замещением третьего атома кислорода кетона на нуклеофил за счёт дестабилизации ПКК. Однако, также обнаружено, что в случае азакарбениевых ионов ОАЭ хоть и достигает порядка 8 ккал/моль (схоже с магнитудой данного эффекта в ПКК), но не создаёт кинетических барьеров благодаря сильной стабилизации карбокатиона за счёт π -донорных свойств атома азота.

Описанные в данном подразделе результаты моделирования опубликованы в *The Journal of Organic Chemistry* [176].

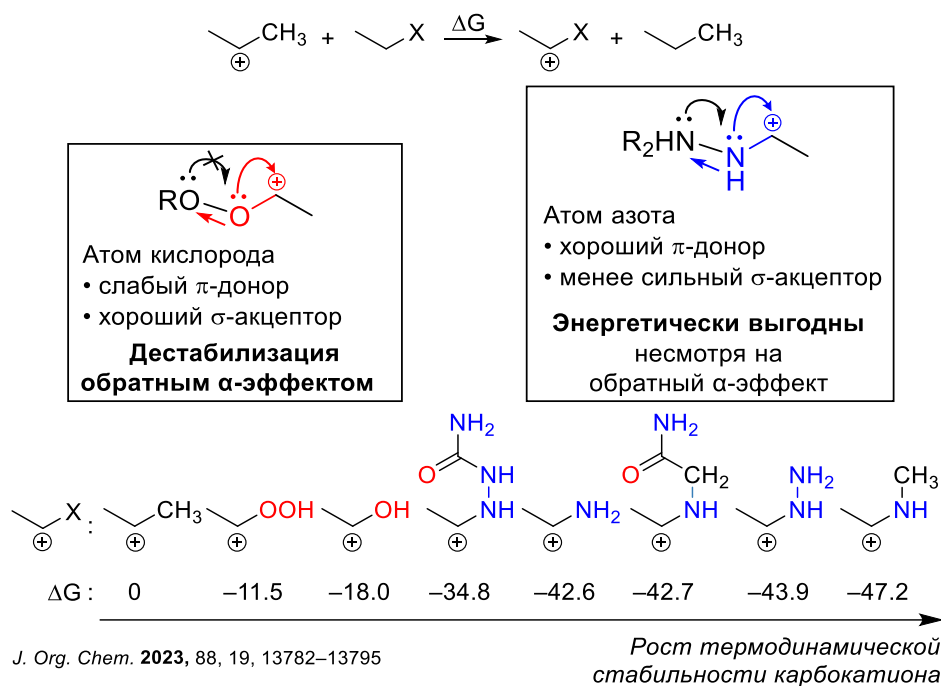


Рисунок 36. Ряд изодесмических реакций для исследования способности различных функциональных групп к стабилизации карбокатионов.

2.4. Другие проявления стереоэлектронных эффектов. Моделирование реакции δ-дикетонов, пероксида водорода и аммиака.

В подразделах 2.1 и 2.2 было показано, что гиперконъюгации $n_X \rightarrow \sigma_{C-Y}^*$ (где X и Y – гетероатомы) являются ключевым орбитальным взаимодействием, определяющим предпочтительность сборки различных каркасов. Однако сильная делокализация неподелённых пар гетероатомов, как следует ожидать, будет проявляться не только в относительных энергиях каркасных пероксидов, но и в других характеристиках – в их реакционной способности и физических свойствах.

В 2020 году нашими коллегами был разработан метод синтеза азаозонидов **D1** и **D2** из δ-дикетонов **DS**, H_2O_2 и NH_3 в среде метанола (Рисунок 37). Ими был экспериментально обнаружен ряд интересных свойств атома азота в **D1** и **D2**. Во-первых, зафиксирована исключительно низкая реакционная способность NH-группы как нуклеофила в реакциях нуклеофильного замещения. Во-вторых, значения химических сдвигов в спектрах ^{15}N ЯМР для атомов азота в **D1** и **D2** (109–113 м.д.) заметно отличаются от типичных для аминов (49.6 и 39.1 м.д. для Et_2NH и

пиперидина соответственно) и, напротив, оказываются близки к диапазону, характерному для амидов (106.4–108.4 и 102.2–104.2 м.д. для N-метилформамида и N-метилацетамида соответственно). Таким образом, парадоксально, атом азота в азаозонидах имеет характер, близкий к амидному.

Кроме этого, согласно данным ЯМР (для случая R = *p*-ClBn, E = CO₂Et), соотношение **D1:D2** меняется в ходе реакции: сигналы обоих соединений регистрируются уже через 5 минут, причём сначала преобладает диастереомер **D2** (соотношение ~30:70). В дальнейшем это соотношение постепенно сдвигается в пользу **D1** и спустя 36 часов достигает значения **D1:D2** = 92:8.

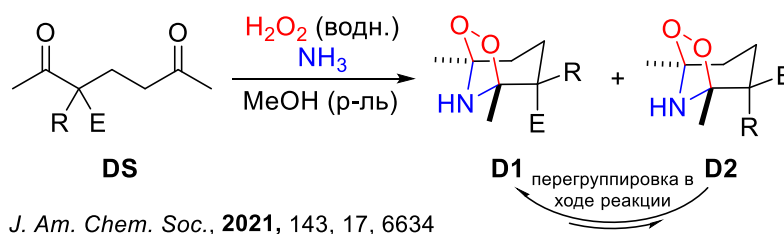


Рисунок 37. Схема реакции получения азаозонидов **D1** и **D2** из δ -дикетонов **DS**.

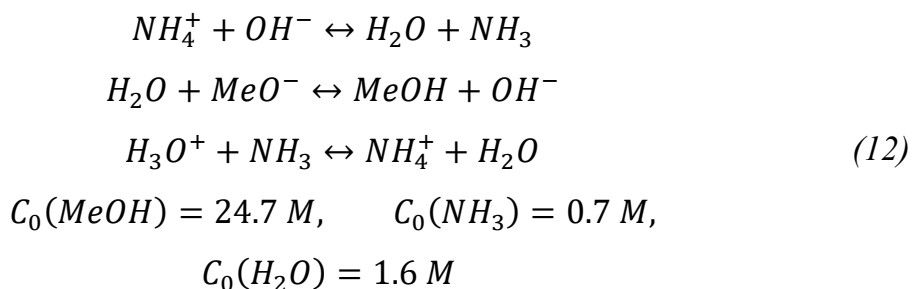
В рамках диссертационного исследования нами было проведено квантово-химическое моделирование этой реакции с целью объяснения амидного характера атома азота, а также определения причин наблюдаемой перегруппировки.

2.4.1. Используемые методы

Моделирование выполнено на уровне теории PBE0-D3BJ [19,20] / 6-311++G(d,p) [21,22,193,194] / SMD(MeOH) [23] в Gaussian16 A.03 [18]. Для всех расчётов свободной энергии применялась квазигармоническая поправка, рассчитываемая в библиотеке GoodVibes [25]. Все термодинамические поправки рассчитаны для температуры 0 °С. Для конформационного поиска молекул и ПС использовался подход, описанный в подразделе 2.1. Для предсказания химических сдвигов ЯМР ¹⁵N

использовался метод GIAO (Gauge-Including Atomic Orbitals, калибровочно-инвариантных атомных орбиталей) [195].

Аналогично методике, описанной в подразделе 2.2, трансляционная энтропия NH_4^+ , NH_3 , H_2O и MeOH рассчитана с учётом их равновесных концентраций на основе следующего равновесия:



Начальные концентрации оценены из экспериментальной методики проведения реакции [174]. Равновесие достигается при следующих концентрациях:

$$\begin{aligned}
 [\text{MeOH}] = 24.69 \text{ M}, \quad [\text{MeO}^-] = 0.01 \text{ M} \\
 [\text{H}_2\text{O}] = 1.6 \text{ M}, \quad [\text{OH}^-] = 2.6 \cdot 10^{-4} \text{ M} \\
 [\text{H}_3\text{O}^+] = 7.5 \cdot 10^{-13} \text{ M} \\
 [\text{NH}_3] = 0.69 \text{ M}, \quad [\text{NH}_4^+] = 0.01 \text{ M}
 \end{aligned}
 \tag{13}$$

Эти концентрации были использованы для расчёта термодинамических поправок к энергиям NH_3 , H_2O и MeOH , а также для оценки энергий протонирования интермедиатов. Предполагалось, что NH_4^+ является наилучшим источником протонов в системе, поэтому разница энергий $G(\text{NH}_4^+) - G(\text{NH}_3)$ использовалась для вычисления относительных энергий положительно заряженных интермедиатов.

2.4.2. Результаты квантово-химического моделирования

Сначала для оценки достижения реакцией термодинамического равновесия были рассчитаны ΔG_r перегруппировки **D2**→**D1**, на основе которых с помощью уравнения Больцмана были определены равновесные соотношения **D1:D2** (Рисунок 38). Расчётные соотношения отлично согласуются с экспериментальными данными для обоих исходных дикетонов. Из этого был сделан вывод, что движущей силой

наблюдавшейся в эксперименте перегруппировки **D2**→**D1** является установление термодинамического равновесия между диастереомерами азаозонидов.

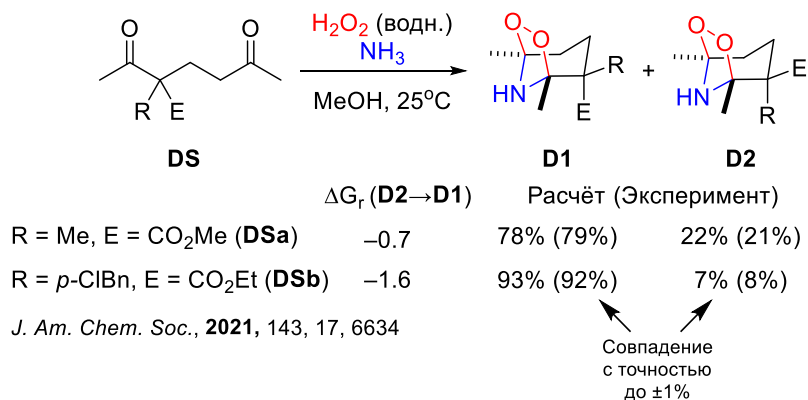
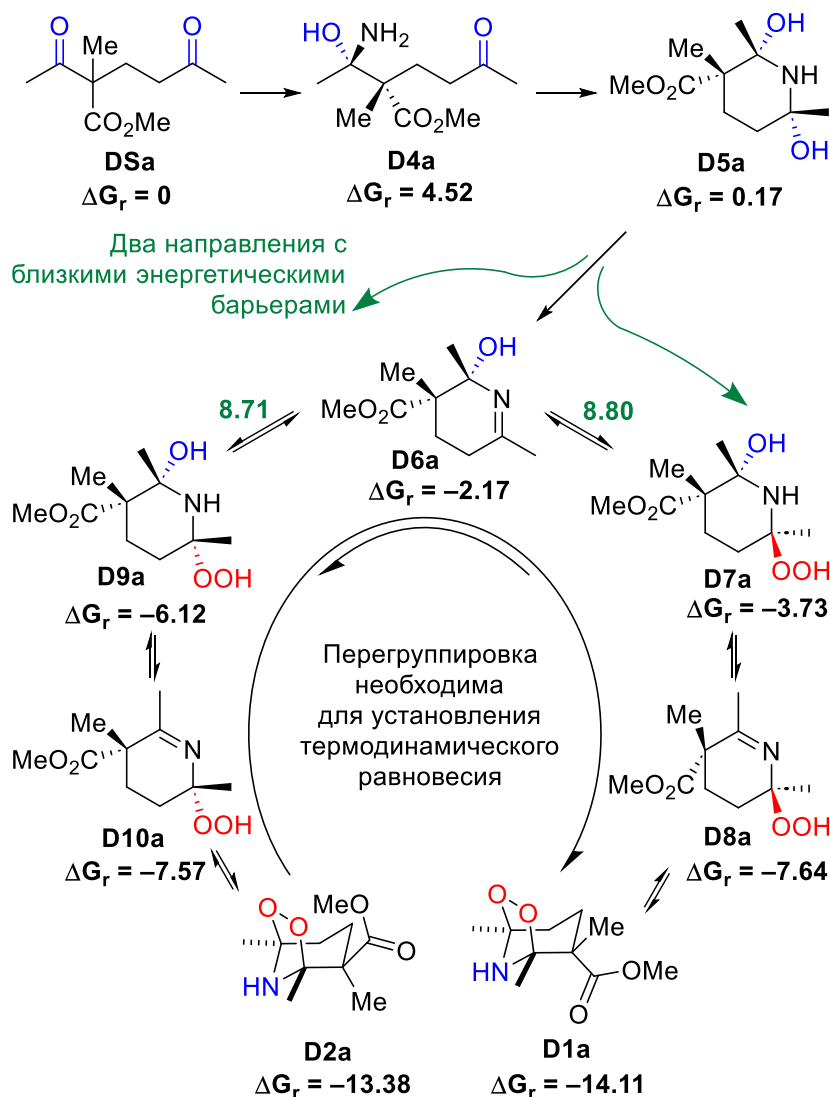


Рисунок 38. Схема реакции получения диастереомерных азаозонидов **D1** и **D2** из δ -дикетонов **DS**. Таблица расчётных и экспериментальных соотношений диастереомеров **D1:D2**. Экспериментальные соотношения получены через 3.5 ч (R = Me) и 36 ч (R = *p*-ClBn) после начала реакции.

Для объяснения того, почему на начальном этапе реакции образуется менее термодинамически выгодный продукт **D2**, предложен механизм реакции для субстрата **DSa** с заместителями R = Me и E = CO₂Me на основе моделирования восьми возможных путей (см. Приложение 2). Согласно полученным результатам (Рисунок 39), энергетические барьеры превращений **DSa** → **D1a** и **DSa** → **D2a** имеют примерно одинаковую высоту, тогда как барьер перегруппировки **D2a** → **D1a** выше на 11.2 ккал/моль. Следовательно, на начальном этапе менее выгодный **D2a** выступает в роли кинетической ловушки: рассчитанный барьер его сборки (8.7 ккал/моль) существенно ниже барьера его превращения в термодинамически более выгодный изомер **D1a** (22 ккал/моль). Таким образом, установление термодинамического равновесия происходит значительно медленнее, чем сборка из **DSa**, что и объясняет наблюдаемую в эксперименте перегруппировку.



J. Am. Chem. Soc., 2021, 143, 17, 6634

Рисунок 39. Результаты моделирования механизма перегруппировки **D2a** → **D1a** в ходе реакции δ -дикетона **DSa** с пероксидом водорода и аммиаком.

2.4.3. Анализ орбитальных взаимодействий

Другой особенностью реакции является селективность образования азаозонидов и отсутствие озонидов. Для установления природы этой селективности мы рассчитали термодинамику перегруппировки озонида **D11b** в азаозонид **D1b** (Рисунок 40А). Отрицательное значение $\Delta G_r = -11$ ккал/моль указывает на то, что азаозониды термодинамически более предпочтительны, чем озониды, что как минимум отчасти обусловлено предпочтительностью замены пар связей C–N / O–N на C–O / N–H (подраздел 2.2.3, Рисунок 33). Для отделения вклада энергии связей от

энергии взаимодействия функциональных групп нами был проведён расчёт изодесмической реакции (Рисунок 40Б). Полученное отрицательное значение $\Delta G_r = -5$ ккал/моль говорит о значительной предпочтительности мостика $-N(H)-$ в качестве соседа пероксидной группы даже при исключении вклада энергий связей.

Аналогично рассуждению в подразделе 2.2.3, энергию замещения (-11.4 ккал/моль) можно приближенно воспроизвести из оценки разницы орбитальных взаимодействий и вклада энергий связей ($-5.0 - 3.6 \cdot 2 = -12.2$ ккал/моль). Это показывает, что термодинамическое предпочтение азаозонидов обусловлено главным образом энергетической выгодой образования связей C–N и стабилизирующим эффектом от соседства мостика $-N(H)-$ с пероксидной группой.

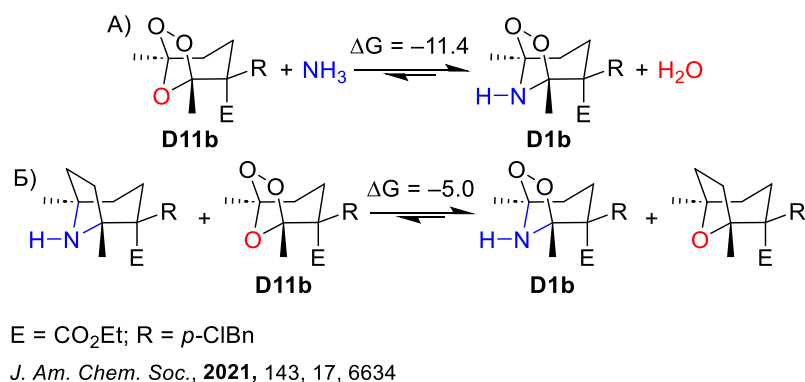


Рисунок 40. Иллюстрация повышенной устойчивости азаозонидов, по сравнению с озонидами. (А) – реакция гипотетической перегруппировки, (Б) – изодесмическая реакция замены кислородного мостика на мостик $-N(H)-$.

Анализ орбитальных взаимодействий в азаозонидах методом NBO показывает наличие двух сильных гиперконъюгаций $n_N \rightarrow \sigma_{C-O}^*$ с энергией ~ 12 ккал/моль каждая (Рисунок 41А,Б). Примечательно, что эти стереоэлектронные эффекты отчётливо проявляются в конформационных предпочтениях: из двух возможных ориентаций связи N–H реализуется та, при которой неподелённая пара азота находится в одной плоскости с двумя орбиталями σ_{C-O}^* для максимальной гиперконъюгации. Этот конформер, **D1a**, оказывается на 4.7 ккал/моль выгоднее конформера **D1a'**, несмотря на наличие внутримолекулярной водородной связи

между NH и сложноэфирной группой в последней (Рисунок 41А). Этот результат также объясняет данные рентгеноструктурного анализа⁶ азаозонида **D1b**, согласно которым связь N–H направлена аксиально относительно пятичленного цикла (Рисунок 41Б).

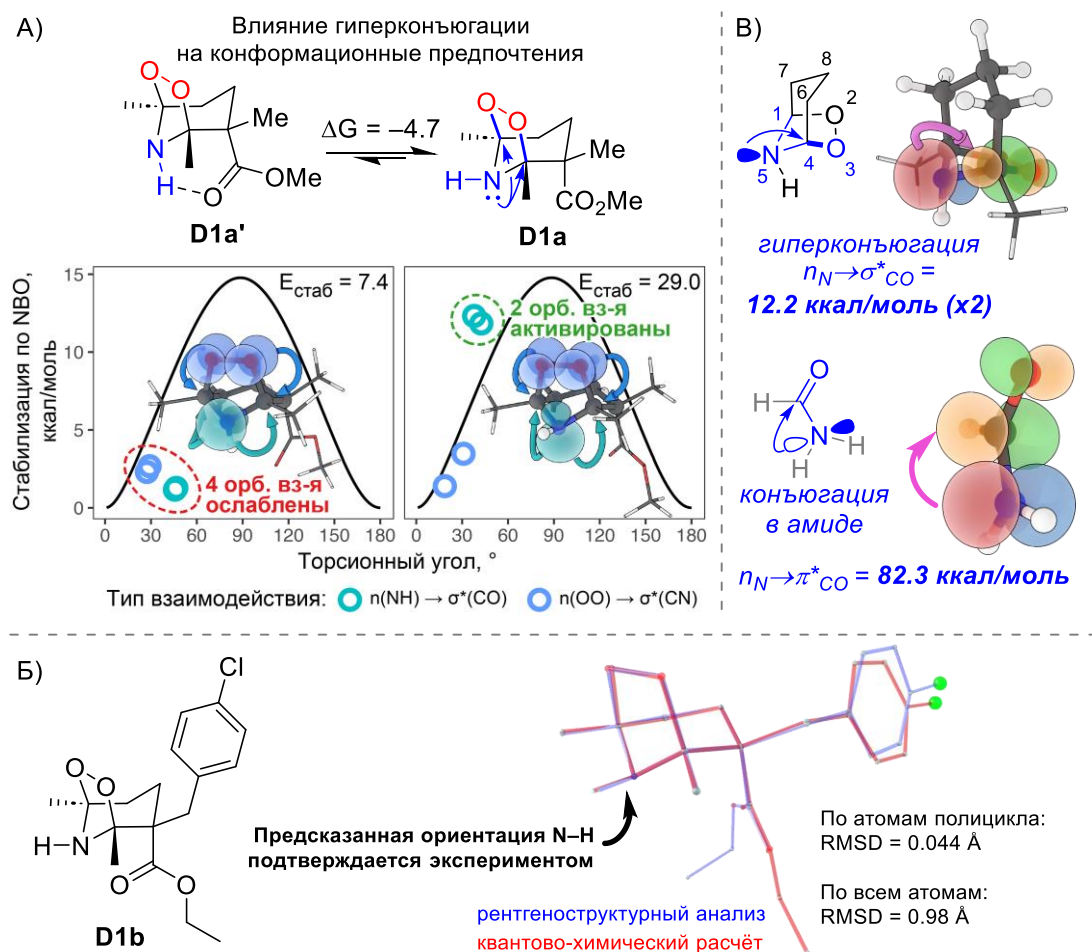


Рисунок 41. (А) – Влияние гиперконъюгаций $n_N \rightarrow \sigma^*_{C-O}$ на конформационные предпочтения азаозонидов на примере **D1a**. (Б) – Сопоставление низкоэнергетического конформера **D1b**, построенного методом RMMC с последующей квантово-химической оптимизацией, с экспериментальной конформацией из рентгеноструктурного анализа (РСА) монокристалла. (В) – Природа амидного характера атома азота в азаозонидах.

⁶ Положение этого атома водорода определено по разностной карте электронной плотности; уточнение проводилось с индивидуальным изотропным параметром смещения. Рентгеновские исследования проведены в Институте элементоорганических соединений им. А. Н. Несмеянова Российской академии наук с помощью дифрактометра Bruker APEX II.

Интересно отметить, что вовлечённость неподелённой пары азота в орбитальные взаимодействия является общей чертой как азаозонидов, так и амидов (Рисунок 41В). Различие заключается лишь в типе акцепторной орбитали: в азаозонидах это две σ_{C-O}^* , в амидах – $\pi_{C=O}^*$. π -акцепторы характеризуются очень высокой энергией взаимодействия с n_N (82 ккал/моль) благодаря их пониженной энергии, большей диффузности и более выгодному пространственному расположению (Рисунок 41В, внизу). Орбитали σ_{C-O}^* , несмотря на менее удачную ориентацию и более высокую энергию, выигрывают в своём числе (два взаимодействия по 12 ккал/моль, Рисунок 41В, вверху). В итоге стереоэлектронная стабилизация неподелённой пары азота в азаозонидах составляет ~30% от стабилизации в амидах.

Это наблюдение навело нас на мысль, что экспериментально наблюдаемый амидный характер азота в спектрах ЯМР и реакциях нуклеофильного замещения также может быть вызван взаимодействиями $n_N \rightarrow \sigma_{C-O}^*$. Для проверки данной гипотезы нами было проведено сравнение нуклеофильности азота в азаозониде **D1b**, его конформере **D1b'** и аналоге **D1b-C** с заменой пероксидной группы на фрагмент CH_2-CH_2 . Согласно расчётным данным (Рисунок 42А), нуклеофильность азота падает в ряду **D1b-C**→**D1b'**→**D1b**: введение электроотрицательных атомов кислорода повышает энергию активации ΔG^\ddagger модельной реакции S_N2 на 4.8 ккал/моль, а дальнейшая активация гиперконъюгаций – ещё на 1.5 ккал/моль. Таким образом, наличие пероксидной группы в азаозонидах снижает нуклеофильность атома азота как за счёт электроотрицательности атомов кислорода, так и вследствие делокализации пары азота.

Дополнительно мы проследили изменение основности атома азота в ряду **D1b-C**→**D1b'**→**D1b** (Рисунок 42А). Полученная зависимость согласуется с приведённым выше рядом нуклеофильности: замена фрагментов CH_2-CH_2 на $O-O$ повышает энергию протонирования азота на 13 ккал/моль, а последующая смена направления связи $N-H$ – ещё на 6 ккал/моль. Полученные зависимости качественно согласуются с химической интуицией, тем не менее, количественно основность изменяется кардинально: соседство азота с пероксидной группой приводит к суммарному росту энергии протонирования примерно на ~18 ккал/моль, что соответствует снижению основности на ~15 единиц pK_b !

Наконец, амидный характер азота в присутствии пероксидной группы удалось проследить в расчёте химических сдвигов ^{15}N (Рисунок 42Б). Активация орбитальных взаимодействий при переходе **D1b-C** \rightarrow **D1b**, согласно расчётам, смещает сигнал ^{15}N в слабое поле на 14 м.д. Таким образом, нестандартный химический сдвиг азота в азаозонидах является спектроскопическим проявлением сильных гиперконъюгаций в их электронной структуре.

Стоит отметить, что полученные выводы о пониженной нуклеофильности и аномальном химическом сдвиге азота в азаозонидах полностью подтверждаются экспериментальными данными. Интересно также отметить, что в литературе описана обратная ситуация – появления «аминного» характера у амида, стерически напряжённый каркас которого запрещает конъюгацию $n_{\text{N}} \rightarrow \pi_{\text{C}=\text{O}}^*$ [196].

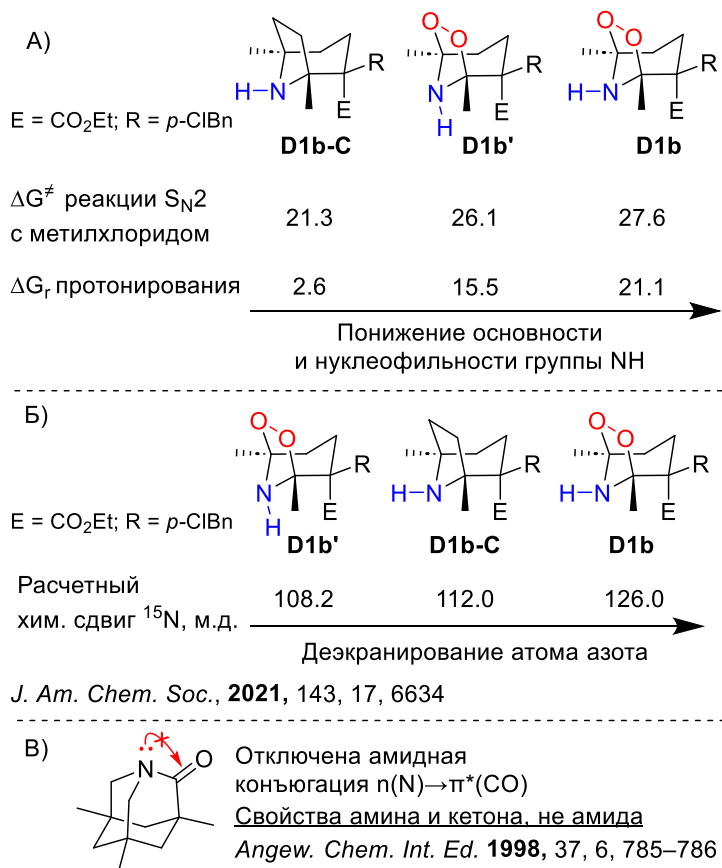


Рисунок 42. (А) – Стереoeлектронные эффекты с участием пероксидной группы понижают нуклеофильность и основность атома азота в азаозонидах.

(Б) – Деэранирование атома азота при активации стереoeлектронных эффектов $n_{\text{N}} \rightarrow \sigma_{\text{C}-\text{O}}^*$. (В) – Аналогичный эффект появления «аминного» характера у амида с отключенной конъюгацией.

Таким образом, стереоэлектронные эффекты $n_N \rightarrow \sigma_{C-O}^*$ в азазонидах изменяют широкий круг свойств молекулы:

- 1) Значительно стабилизируют конформацию с оптимальным перекрыванием взаимодействующих орбиталей (даже несмотря на потерю внутримолекулярной водородной связи);
- 2) Атом азота теряет свои нуклеофильные свойства;
- 3) Основность NH-группы очень сильно понижается;
- 4) Пик в ^{15}N ЯМР смещается в слабое поле.

Описанные в данном подразделе результаты моделирования опубликованы в *Journal of the American Chemical Society* [174].

Таким образом, в данном разделе диссертации представлены квантово-химические модели, описывающие сборку каркасных пероксидов из β,γ' - и β,δ' -трикетонов, а также δ -дикетонов под действием пероксида водорода в присутствии N-нуклеофилов. Моделирование включало анализ термодинамики и энергетических барьеров исследуемых реакций с учётом кислотно-основных равновесий в системе и в случае реакции с β,γ' -трикетоном – с проведением кинетического интегрирования. Построенные квантово-химические модели воспроизводят преимущественные продукты реакций, а результаты конформационного анализа конечных продуктов согласуются с данными рентгеноструктурного анализа.

Термодинамическая стабильность каркасных пероксидов во многом определяется активацией гиперконъюгации с участием неподелённых электронных пар гетероатомов, что напрямую зависит от возможности каркаса принять конформацию с торсионными углами благоприятными для оптимального орбитального перекрывания. Гиперконъюгация также сильно сказывается на конформации (направлении связи N–H) аминопероксидов и на других свойствах их атома азота.

Помимо термодинамической стабильности, селективность сборки каркасных пероксидов определяется кинетическим контролем, связанным с обратным α -эффектом – пониженной стабилизацией карбокатионов пероксидной группой, что блокирует стадию замещения последнего атома кислорода трикетона при сборке

трициклических каркасов. В то же время активация орбитальных взаимодействий с участием неподелённых электронов становится главным фактором термодинамического контроля. Выявленные закономерности можно использовать для прогнозирования продуктов реакций поликетонов с пероксидами водорода и для целенаправленного синтеза каркасных пероксидов.

Раздел 3. Конформационная подвижность циклических молекул

Проведённые в Разделе 2 квантово-химические исследования реакций сборки каркасных органических пероксидов показали, что даже системы малых конденсированных циклов иногда могут обладать конформационной подвижностью и низкоэнергетические конформации определяются компромиссом между стереоэлектронными эффектами и стерическим отталкиванием. Поэтому конформационный поиск являлся необходимым этапом моделирования этих молекул, и на его точность полагаются все дальнейшие выводы из моделирования. Изучавшиеся нами каркасные пероксиды состояли из 5-, 6- и 7-членных циклов, поэтому для них достаточно надёжно работали стандартные алгоритмы конформационного поиска малых циклов: ETKDG, ringmaker Monte-Carlo и iMTD-GC. Однако при проведении подобного исследования для макроциклических молекул способность данных методов к глобальному конформационному поиску может оказаться ограниченной (см. Раздел 1). Поэтому во второй части диссертационного исследования мы проанализировали и нашли решение коренной проблемы применимости многих методов конформационного поиска к макроциклам.

Как показано в Разделе 1, широкий круг методов конформационного поиска молекул основан на приближении жёсткого ротатора, в котором варьирование торсионных углов является первоочередным способом исследования конформационного пространства. К таким методам относится случайный и систематический перебор, алгоритмы байесовской оптимизации и ряд методов машинного обучения, поиск Монте-Карло Метрополиса с варьированием торсионных углов, а также эволюционные алгоритмы на основе скрещивания векторов торсионных углов. Все эти группы алгоритмов наследуют ключевой недостаток приближения жёсткого ротатора – низкую эффективность или невозможность семплирования циклических структур из-за проблемы корректного учёта циклических ограничений при варьировании торсионных углов. Это оставляет

задачу глобального конформационного поиска циклических молекул в значительной степени нерешённой.

В данном разделе представлено решение проблемы параметризации конформационного пространства циклических и полициклических систем для генерации их конформаций. Для этого вводится алгебраическая формулировка задачи построения конформации молекулы по внутренним координатам, – прямой кинематики (ПК, подраздел 3.1), на основе которой условие замкнутости цикла формулируется как обратная задача, – обратная кинематика (ОК, подразделы 3.2 и 3.3). Основные результаты Раздела 3 опубликованы в двух работах и включают в себя метод определения степеней свободы и построения конформаций полициклических молекул с помощью ОК (подраздел 3.4) [197], а также демонстрация этого метода через реализацию алгоритма конформационного поиска Монте-Карло в кинематических степенях свободы полициклической молекулы (подраздел 3.5) [198].

3.1. Механическая модель подвижности молекул. Прямая кинематика.

Для алгебраического описания условия существования молекулярного цикла с заданными параметрами продуктивным оказывается обращение к теории кинематики шарнирных механизмов, изначально разработанной в робототехнике. В этой области алгоритмы кинематики играют важнейшую роль, так как они позволяют рассчитывать требуемые состояния шарниров манипулятора в процессе движения, делая возможным перемещение к требуемой цели по желаемой траектории.

Аналогия между циклом и рукой робота проявляется при сравнении наборов их геометрических параметров. Структура цикла полностью задаётся длинами связей и значениями валентных и торсионных углов внутри цикла, а состояние руки робота может быть однозначно представлено длинами сочленений и состояниями соединяющих их шарниров [199]. В кинематике шарнирных механизмов выделяется два класса задач: расчёт глобального положения частей манипулятора при заданных параметрах сочленений/шарниров (прямая кинематика, ПК), и расчёт параметров сочленений и шарниров, при которых части манипулятора принимают заданную

ориентацию в пространстве (обратная кинематика, ОК). В случае геометрии молекул ПК – это вычисление координат атомов при известных длинах связей, валентных и торсионных углах, а ОК – это, в частности, поиск таких значений торсионных углов, которые обеспечивают замкнутость цикла при заданных длинах связей и валентных углах. С точки зрения ОК и робототехники, задача генерации конформаций цикла эквивалентна задаче «дотронуться концом руки робота до его плеча».

С точки зрения кинематики, и цепь из атомов, и манипулятор робота можно рассматривать как ломаные линии в \mathbb{R}^3 . Для их алгебраического описания применяется метод подвижного репера, где репер – это точка с закреплённым на ней ортонормированным базисом, который перемещается вдоль ломаной от первого атома к последнему под действием группы движений трёхмерного пространства $SE(3)$. Как реперы в \mathbb{R}^3 , так и элементы группы $SE(3)$ представляются блочными матрицами 4×4 с фиксированной последней строкой $(0 \ 0 \ 0 \ 1)$.

$$\begin{pmatrix} \mathbf{R} & \vec{v} \\ 0 & 1 \end{pmatrix}, \text{ где } \mathbf{R}\mathbf{R}^T = \mathbf{I}_{3 \times 3}, \det \mathbf{R} = 1 \quad (14)$$

где $\mathbf{I}_{3 \times 3}$ – это единичная матрица.

В такой матрице правый столбец \vec{v} задаёт начало координат репера (трансляционную компоненту движения), а левый верхний блок 3×3 , \mathbf{R} – ориентацию базиса репера (вращательную компоненту движения). Преобразование репера в его локальных координатах выполняется умножением его матрицы справа на матрицу движения; соответственно, композиция движений выражается умножением их матриц в порядке слева направо.

Чтобы строить геометрии произвольных атомных цепей, вводится три типа движений, соответствующих трём типам параметров: длин связей, валентных углов и торсионных углов (Рисунок 43).

1. Длине связи σ соответствует матрица $D(\sigma)$ трансляции на расстояние σ .
2. Валентному углу θ соответствует матрица $V(\theta)$ поворота на угол $\pi - \theta$ в плоскости предыдущей и текущей связей.
3. Торсионному углу τ соответствует матрица $T(\tau)$ поворота на угол τ в плоскости, перпендикулярной предыдущей связи.

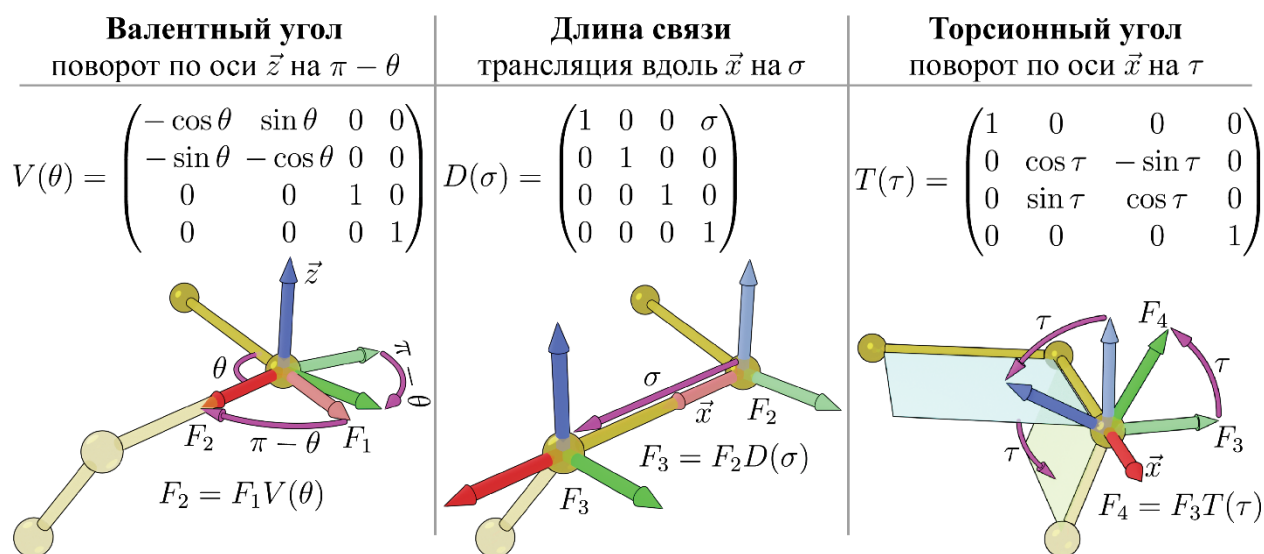


Рисунок 43. Движения, соответствующие матрицам $V(\theta_i)$, $D(\sigma_i)$ и $T(\tau_i)$, используемые для построения атомных цепей по внутренним координатам.

Показано движение репера вдоль атомной цепи, где F_i – это начальное, промежуточные и конечное положения репера.

Преобразование репера при переходе от атома i к атому $i + 1$ в цепи описывается композицией движений $V(\theta_i)D(\sigma_i)T(\tau_i)$. Каждый множитель имеет геометрическую интерпретацию: $V(\theta_i)$ устанавливает направление следующей связи в цепи, $D(\sigma_i)$ переносит центр репера к положению $(i + 1)$ -го атома, $T(\tau_i)$ задаёт ориентацию плоскости, в которой будет лежать $(i + 2)$ -ой атом. Проводя аналогию с геометрией гладких кривых [200], длины связей, валентные углы и торсионные углы являются дискретными аналогами метрики, кривизны и кручения для ломаных (дискретных кривых).

Обход цепочки из $N + 1$ атомов и N связей в общем случае выражается следующим матричным произведением, определяющим движение F_{N+1} от первого атома к $(N + 1)$ -ому:

$$F_{N+1} = D(\sigma_1) \left(\prod_{i=2}^{N-1} V(\theta_i) D(\sigma_i) T(\tau_i) \right) V(\theta_N) D(\sigma_N) \quad (15)$$

Эта формула описывает ПК цепочки атомов, так как положение каждого атома в глобальной системе координат может быть получено из трансляционных компонент частичных произведений при вычислении слева направо. Продолжая аналогию с гладкими кривыми, уравнение выше является дискретным аналогом формул Френе–Серре [200].

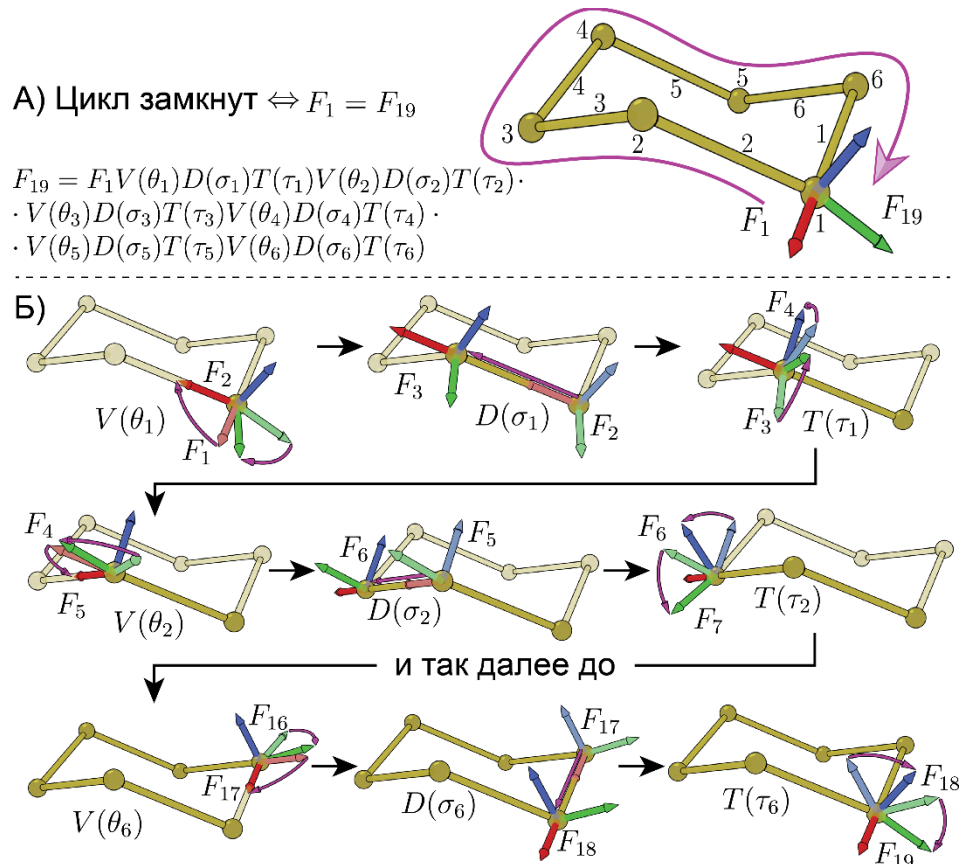


Рисунок 44. (А) – Критерий существования цикла с заданными координатами на примере цикла из 6 атомов. (Б) – Геометрическая интерпретация этого критерия; для каждого движения показаны положения системы координат до и после операции, по итогу обхода цикла система координат возвращается в исходную ориентацию.

Переходя от атомных цепей к циклам, появляется дополнительное требование: композиция движений должна быть равна тривиальному, то есть обходу цикла с сохранением исходного положения и ориентации (Рисунок 44). Это позволяет сформулировать условие существования цикла из N атомов и N связей с заданными параметрами:

$$\prod_{i=1}^N V(\theta_i) D(\sigma_i) T(\tau_i) = I_{4 \times 4}, \quad (16)$$

где $I_{4 \times 4}$ – это единичная матрица.

Таким образом, ПК позволяет восстанавливать геометрии циклов и атомных цепей по их параметрам (длинам связей, валентным и торсионным углам), а также проверять, соответствует ли набор параметров геометрически возможному циклу.

3.2. Степени свободы циклов. Постановка задачи обратной кинематики.

Для генерации конформаций циклов необходимо подобрать значения молекулярных параметров, удовлетворяющих уравнению (16). Для полноценной постановки задачи обратной кинематики (ОК) необходимо также определить, какие параметры можно считать фиксированными, а какие – неизвестными. Следуя химической интуиции (см. подраздел 1.1.1.2), разумным решением является фиксация значений длин связей и валентных углов и рассмотрение торсионных углов как неизвестных переменных. Таким образом, N -членный цикл с заданными длинами связей и валентными углами может быть сгенерирован путём решения матричного уравнения (16) размера 4×4 относительно N неизвестных торсионных углов.

Но сколько решений можно ожидать от уравнения обратной кинематики? Ответ зависит от количества переменных и ограничений. В общем случае любое математическое уравнение или система уравнений накладывает определённое количество ограничений на свои переменные. Это позволяет исследовать пространство решений уравнений, фиксируя значения некоторых независимых переменных, а затем разрешая ограничения для получения значений всех остальных

зависимых переменных. Однако это возможно только в том случае, если система недоопределена, то есть количество ограничений меньше количества переменных. В случае численных уравнений (без использования матриц) одно уравнение обычно добавляет ровно одну зависимую переменную, а все остальные переменные можно рассматривать как степени свободы решений. Таким образом, количество степеней свободы системы уравнений можно получить как разность между количеством переменных и количеством ограничений. В случаях, когда количество степеней свободы равно 0, система имеет только конечное число решений (именно поэтому циклогексан имеет чётко определённые конформации, см. ниже). И, наконец, в случаях, когда количество степеней свободы отрицательно (как в циклопентане или циклах меньшего размера), система называется переопределённой и почти никогда не имеет решений (для его нахождения необходимо точно угадать значения всех валентных углов).

Можно легко понять, что уравнение (16), которое сформулировано через матрицы 4×4 , накладывает шесть ограничений. Это следует из того, что $SE(3)$, как группа Ли, имеет шесть вещественных измерений. С другой стороны, это можно получить, развернув матричное уравнение (16) в систему численных уравнений через вычисление матричного произведения в левой части и рассмотрев каждый элемент матрицы как отдельное уравнение. Это приводит к эквивалентной системе из $4 \cdot 4 = 16$ численных уравнений. Однако в используемом нами представлении группы $SE(3)$ четвертая строка всегда равна $(0 \ 0 \ 0 \ 1)$, поэтому включение этих четырёх элементов является избыточным, поскольку соответствующие уравнения будут тривиальными, $0 = 0$ и $1 = 1$. Система из оставшихся 12 уравнений содержит другие избыточные ограничения в поворотном блоке. Известно, что любой трёхмерный поворот однозначно определяется тремя углами Эйлера или эквивалентной матрицей поворота 3×3 ; следовательно, можно ограничить только три элемента (например, диагональные), чтобы зафиксировать весь поворотный блок на единичной матрице.

Таким образом, уравнение (16) накладывает шесть ограничений на N неизвестных торсионных углов. То есть представление цикла как кинематического

механизма с фиксированными длинами связей и валентными углами приводит к тому, что количество свободно и независимо вращающихся торсионных углов составляет $N_{dof} = N_{ta} - 6$, где N_{dof} – количество свободно вращающихся торсионных углов (степеней свободы), N_{ta} – количество подвижных торсионных углов (равно размеру цикла, если в нём нет кратных связей или общих с другими циклами связей (см. далее)).

Это предполагает следующие шаги для построения конформаций цикла. Во-первых, выбрать N_{dof} торсионных углов цикла, которые будут считаться его степенями свободы (могут быть выбраны любые одинарные связи). Затем сгенерировать значения этих N_{dof} торсионных углов. И, наконец, решить уравнение (16), чтобы найти оставшиеся шесть торсионных углов, удовлетворяющих условию замыкания цикла. Примечательно, что это уравнение может иметь несколько решений, которые порождают все геометрически возможные конформеры цикла (например, конформеры кресла и ванны) при заданных значениях свободно вращающихся торсионных углов. Важно отметить, что $N_{dof} = N_{ta} - 6$ означает, что система уравнений ОК в общем случае является переопределённой для циклов размером менее 6 атомов, что не является проблемой, поскольку конформеры таких малых циклов могут быть перечислены с использованием дистанционной геометрии или баз данных фрагментов.

В целом, описанный подход делает возможным генерацию конформаций циклических молекул через изменение их торсионных углов, то есть обобщает используемую для ациклических молекул модель жёсткого ротатора. Отличие заключается лишь в том, что на вход требуются значения не всех торсионных углов цикла и что в результате может быть получено несколько геометрически возможных конформаций. Для практической реализации этого подхода требуются алгоритмы решения уравнения ОК (вычисления шести недостающих торсионных углов), обсуждению которых посвящен следующий подраздел.

3.3. Решение обратной кинематики одиночных циклов

Применение алгоритмов решения ОК для генерации конформаций циклических молекул были предложены E. Coutsias и соавторами. Ими были реализованы два алгоритма: TLC (Triaxial loop closure) [15] и R6B6 (6 rotors/6 bar) [16,201]. В своей оригинальной формулировке алгоритм TLC решает ОК только для девятичленных циклов при условии задания значений трёх свободных торсионных углов. TLC использует важную особенность девятичленных циклов, заключающуюся в том, что при любом выборе положений свободных торсионных углов оставшиеся шесть связей образуют три пары геминальных связей (связи каждой пары имеют один общий атом, выделены фиолетовым на Рисунке 45А), и торсионные углы этих шести связей должны быть вычислены таким образом, чтобы выполнялось условие замыкания цикла. Поскольку остальные три торсионных угла заданы, цикл можно рассматривать как систему из трёх жёстких фрагментов, каждый из которых состоит из четырёх атомов и трёх связей. Эти фрагменты соединены друг с другом общими шарнирными атомами (P1, P2 и P3 на Рисунке 45Б). Геометрия каждого жёсткого фрагмента определяется шестью фиксированными параметрами: длинами трёх связей, двумя валентными углами и одним наперёд заданным торсионным углом. Это означает, что треугольники, образованные шарнирными атомами P1, P2 и P3, должны быть одинаковыми во всех конформациях с одинаковыми заданными значениями свободных торсионных углов. Следовательно, для генерации цикла необходимо найти подходящий поворот каждого жёсткого фрагмента вокруг соответствующей стороны треугольника, чтобы валентные углы на атомах P1, P2 и P3 имели правильные значения.

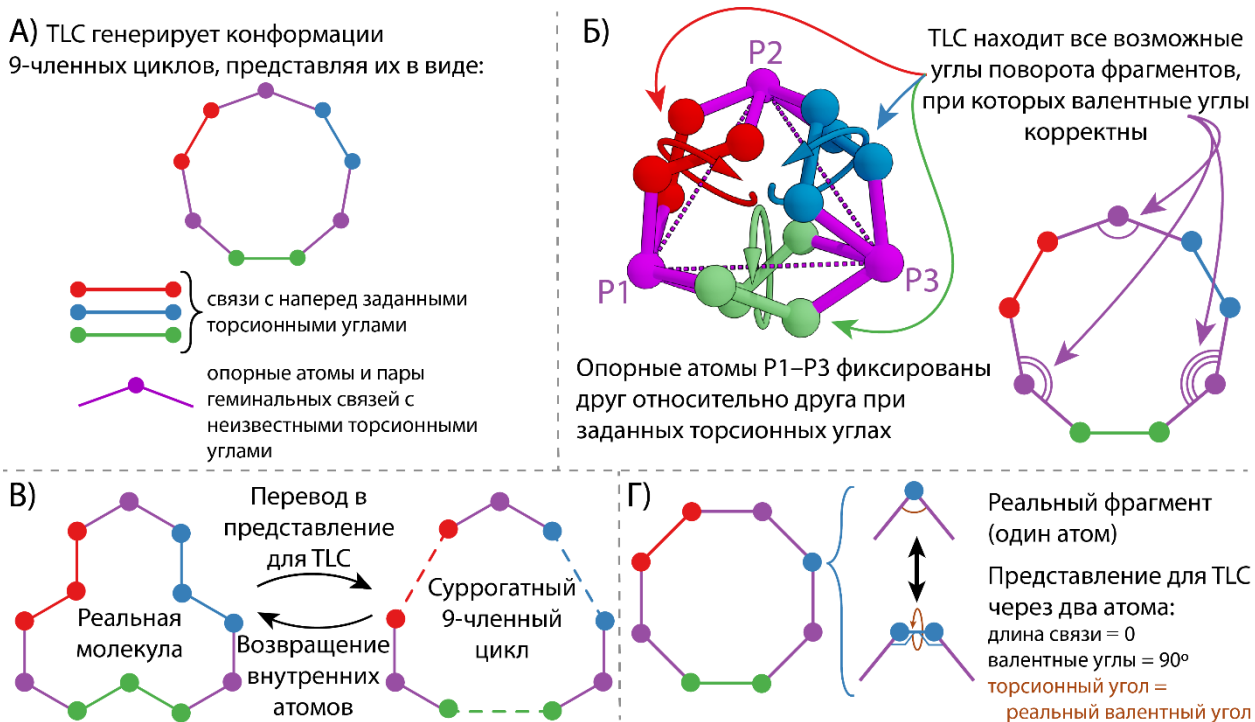


Рисунок 45. Алгоритм triaxial loop closure (TLC) для решения ОК 9-членных циклов. (А, Б) – формулировка задачи замыкания цикла, показывающая необходимость рассмотрения именно 9-членных циклов с тремя парами геминальных связей (фиолетовые). (В, Г) реализованный нами способ применения TLC к циклам из более 9 атомов (В) и к 6-, 7- и 8-членным циклам (Г).

Именно с такой формулировкой задачи ОК работает алгоритм TLC. Это позволяет ему уменьшить число неизвестных до трёх и выразить ограничения через три тригонометрических уравнения, которые затем сводятся к квадратичным, а затем – к одному полиномиальному уравнению 16-й степени. Интересно, что порядок конечного полиномиального уравнения означает, что для заданного конкретного набора независимых торсионных углов может существовать до 16 возможных конформаций цикла (это широко известный факт в робототехнике [202]). Что особенно примечательно, вычисление алгоритма TLC на одном наборе торсионных углов требует всего долей миллисекунды!

На первый взгляд, область применимости TLC кажется довольно ограниченной, так как включает только девятичленные циклы. Чтобы устранить это ограничение,

нами было реализовано расширение TLC на циклы произвольного размера, начиная с шестичленных [197]. Это достигается через построение эквивалентной задачи ОК для суррогатного девятичленного цикла: когда атомов более девяти, каждая последовательность связей с фиксированными торсионными углами группируется в одну псевдосвязь, а когда атомов менее девяти, производится слияние пар атомов в один атом путём установления нулевой длины связи. Таким образом, реализованное нами расширение метода TLC позволяет находить все геометрически возможные конформации произвольного N -членного цикла с наперед заданными длинами связей и валентными углами и торсионными углами, разбитыми на две группы: $(N - 6)$ свободных углов и 6 зависимых – которые рассчитываются с помощью TLC. Причём зависимые торсионные углы должны быть выбраны таким образом, чтобы они были организованы в три пары геминальных связей, то есть каждая пара торсионных углов имела бы один общий атом (фиолетовые связи на Рисунке 45).

Но ограничивает ли это требование область применимости алгоритма TLC? Фактически, да, но только в редких случаях, когда невозможно выбрать три соседние пары геминальных связей с подвижными торсионными углами, например, если связи с подвижными торсионными углами чередуются с неподвижными кратными связями. Чтобы охватить эти особые случаи, Coutsiias и соавторы разработали более универсальный решатель – алгоритм R6B6 [201], который позволяет генерировать циклы без требования, чтобы связи с зависимыми торсионными углами образовывали сочлененные пары; однако данный метод несколько уступает TLC в скорости.

3.4. Обобщение обратной кинематики на полициклические системы

К сожалению, ни один из существующих аналитических решателей ОК не способен работать с несколькими циклическими ограничениями одновременно, поэтому генерация конформаций конденсированных циклов напрямую ими невозможна. Исторически первым алгоритмом конформационного поиска полициклов на основе ОК стал алгоритм BRIKARD [16], который строит остов полициклической молекулы, имеющий хотя бы одну связь в каждом цикле.

Построение конформаций реализуется путём обхода данного остова с последовательной генерацией геометрий встречающихся циклов через решатель ОК для одиночных циклов. Несмотря на свою уникальность, BRIKARD имеет ряд недостатков, которые приведены ниже после обсуждения предложенного нами эффективного и более универсального способа обобщения ОК на полициклические системы [197].

Если в одиночном цикле подвижность одного из торсионных углов ограничена, например, если это двойная связь в *цис*-конфигурации, то для построения корректных конформаций с помощью ОК необходимо отнести данный торсионный угол к набору свободных (независимых) углов и задавать ему значение 0° , чтобы зафиксировать *цис*-конфигурацию во всех получаемых решениях (Рисунок 46А). Мы применили этот принцип для обеспечения корректных торсионных углов в местах конденсации циклов. В простейшем случае конденсации гибкого макроцикла с конформационно жёстким бензольным кольцом входные данные для алгоритма ОК будут идентичны случаю с двойной связью, так как сочленение с бензольным циклом жёстко фиксирует один из торсионных углов макроцикла на значении 0° . Полная полициклическая конформация собирается путём присоединения бензольного фрагмента к геометрии макроцикла, полученной решателем ОК.

Этот подход обобщается на случай сочленения двух гибких макроциклов. Отличие лишь в подвижности торсионных углов в месте конденсации. Для учёта взаимозависимости решатель ОК применяется к циклам последовательно: конформация первого (независимого) цикла генерируется без учёта конденсации, после чего конформация второго (зависимого) цикла строится с фиксацией торсионного угла на общей связи, значение которого определяется конформацией первого цикла (Рисунок 46Б). Такой подход не накладывает дополнительных требований на независимый цикл, однако для зависимого цикла возникает необходимое условие: он должен обладать, как минимум, таким же количеством степеней свободы, сколько связей находится на месте конденсации, поскольку все эти углы определяются конформацией независимого цикла и должны быть предопределены.

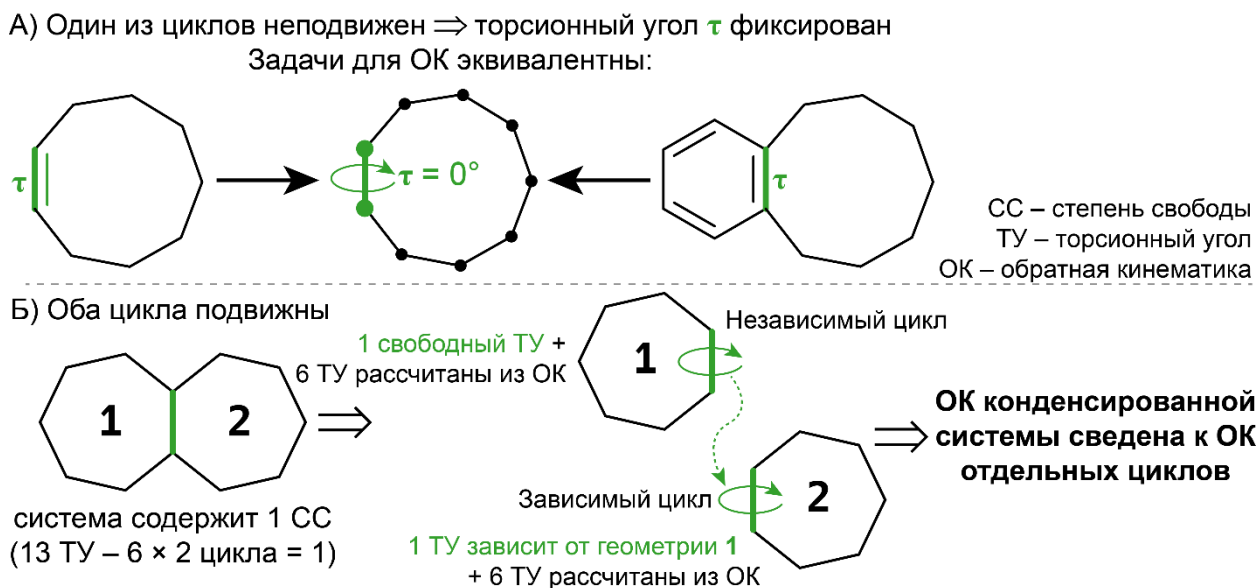


Рисунок 46. Предложенный нами подход к генерации конформаций конденсированных циклов с помощью ОК: (А) – один из циклов неподвижен, (Б) – оба цикла подвижны. Данный подход устраняет фундаментальное ограничение TLC и других решателей ОК, применимых только к одиночным неконденсированным циклам.

Для дальнейшего обобщения этого подхода на произвольные полициклические системы нами был автоматизирован выбор подходящего порядка зависимостей между циклами, чтобы обеспечить выполнение требований к минимальному числу степеней свободы для всех зависимых циклов. Для систем с малым числом циклов (6 и менее) используется прямой перебор всех вариантов; в противном случае применяется жадный алгоритм, который обрабатывает циклы с меньшим числом подвижных торсионных углов в первую очередь.

Для автоматизации всего процесса генерации конформаций с помощью ОК, от анализа полициклической топологии до вызова решателя TLC в требуемом порядке, нами был предложен и реализован алгоритм Ringo. В нашей общедоступной реализации [17], Ringo принимает на вход граф связности молекулы и геометрические параметры, которые могут быть заданы через интерфейс на Python

или извлечены из файла MOL. Последующий анализ топологии молекулы включает следующие ключевые этапы (Рисунок 47):

1. Поиск циклических фрагментов как компонент связности молекулярного графа после удаления из него всех ациклических связей.
2. Разложение циклических фрагментов на образующие их циклы через вычисление базиса циклов минимального веса.
3. Определение последовательности запуска алгоритма TLC для циклов каждого фрагмента путём прямого перебора. Полученные последовательности должны удовлетворять условиям алгоритма TLC и требованиям к минимальному количеству степеней свободы для всех зависимых циклов.
4. Определение последовательности сборки конформации: для циклических фрагментов – порядок сборки циклов в единой системе координат; для всей молекулы – порядок соединения конформаций циклических фрагментов через ациклические связи.

Стоит отметить, что реализованная нами возможность полного перебора последовательностей решения ОК делает Ringo более универсальным алгоритмом по сравнению с вышеупомянутым BRIKARD. Алгоритм BRIKARD, основанный на выделении непрерывного остова, накладывает жёсткие ограничения на допустимые порядки обхода циклов при решении ОК, поскольку они должны последовательно лежать на остове и в этом же порядке впоследствии выполняется сборка конформации полициклической системы. В Ringo же последовательность решения ОК не привязана к порядку сборки молекулы. Это позволяет, во-первых, решать ОК в произвольной последовательности, что даёт максимальную гибкость для обеспечения минимального числа степеней свободы у всех зависимых циклов, а, во-вторых, выполнять сборку полицикла, следуя любой древовидной иерархии, независимо от того, как решались уравнения замыкания. Наконец, важным практическим преимуществом является то, что, в отличие от BRIKARD, который не имеет общедоступной реализации, Ringo распространяется в виде открытой библиотеки для Python.

Пример для реальной молекулы (CSD ID = FINWEE, Такролимус, иммуносупрессивный препарат)
 CC – степень свободы; ТУ – торсионный угол

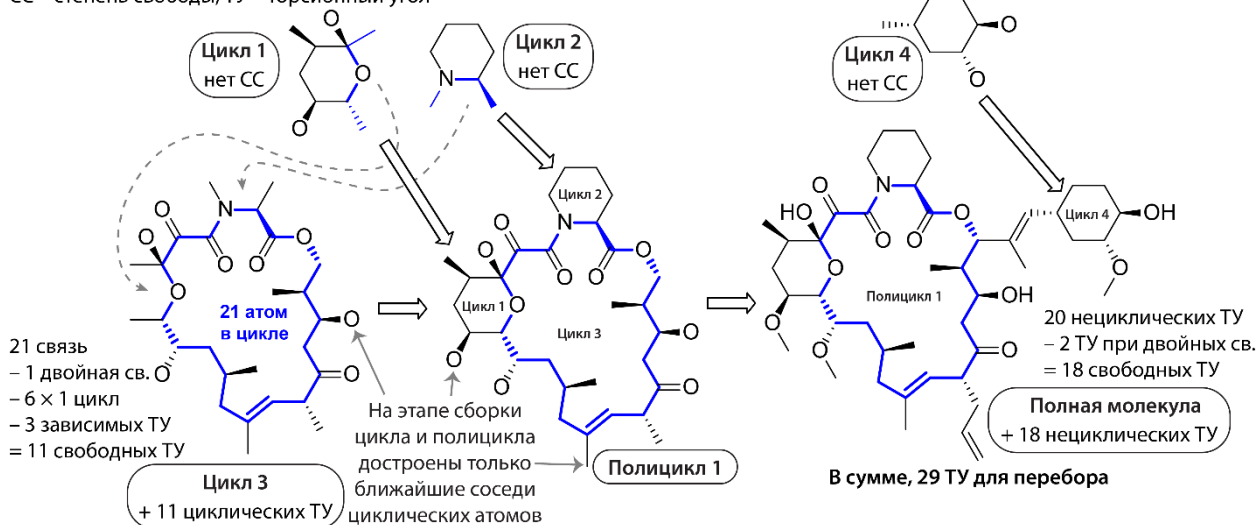


Рисунок 47. Разработанный алгоритм Ringo [17,197] для анализа топологии полициклических молекул для последующей генерации их конформаций с помощью ОК.

Одновременно с анализом топологии Ringo составляет список степеней свободы (CC) молекулы, включающий все ациклические торсионные углы и независимые циклические торсионные углы. При этом торсионные углы в кратных связях и жёстких циклах автоматически фиксируются. Реализовано построение конформаций для заданных извне значений CC через последовательность: (1) решение ОК алгоритмом TLC для каждого цикла с учётом зависимости торсионных углов при конденсированных циклах, (2) сборка геометрий циклических фрагментов из полученных конформаций циклов, и (3) сборка единой конформации молекулы с учётом заданных ациклических торсионных углов. Весь процесс сборки конформации из одного набора значений CC выполняется за доли миллисекунды. Кроме того, Ringo имеет функционал для построения всех возможных конформаций, соответствующих рассматриваемому набору CC (как обсуждалось выше, уравнения ОК могут иметь до 16 решений).

Таким образом, разработанный нами алгоритм Ringo решает проблему параметризации конформационного пространства циклических и полициклических систем для генерации их конформаций путём варьирования торсионных углов.

3.5. Метод конформационного поиска MCR и его тестирование

Чтобы оценить эффективность конформационного поиска с использованием ОК, мы реализовали алгоритм случайного перебора значений СС, получаемых из Ringo [198]. Это потребовало внедрения трёх дополнительных элементов: (1) отсеивание структур с нежелательными стерическими наложениями несвязанных друг с другом атомов, (2) удаление дубликатов конформаций на основе RMSD и (3) оптимизация получаемых конформаций с использованием силового поля MMFF94 [34]. Масштабируемость поиска удалось дополнительно улучшить за счёт корректировки значений СС после неудачной попытки генерации конформации (Рисунок 48). Например, если в некотором фрагменте молекулы обнаружено нежелательное наложение атомов или в некотором цикле отсутствуют решения задачи ОК при заданных значениях СС, то будет сделано несколько попыток уточнения (генерации заново) только тех СС, которые принадлежат проблемным фрагментам. Мы назвали этот алгоритм MCR (Monte Carlo with Refinement, Монте Карло с уточнением) и сделали его доступным в библиотеке Ringo [17,197].

Мы оценили эффективность MCR в сравнении с передовыми алгоритмами для конформационного поиска: RDKit (алгоритм ETKDGV3), CREST (пайплайн iMTD-GC), алгоритм MTD в пакете XTB и два алгоритма в пакете MacroModel (базовый MCMM и макроциклический MMBS). Принципы работы всех этих методов обсуждались в подразделе 1.1. Стоит отметить, что методы, реализованные в MacroModel и CREST, включают в себя некоторые критерии остановки, тогда как остальные методы являются генераторами на основе случайного перебора (MCR и RDKit) или динамики (MTD). Выбрав достаточно короткое время – 10 минут – для методов MCR, ETKDG и MTD нам удалось протестировать их на 150 макроциклах из тестового набора лекарственно-подобных макроциклов [35]. Поскольку CREST и оба метода, реализованных в MacroModel, могут обрабатывать каждую молекулу в течение нескольких часов, их тестирование было ограничено десятью случайно отобранными молекулами из того же тестового набора. Так как все методы сравнения, за исключением MTD, оптимизируют получаемые ими геометрии на

некотором уровне теории, в MCR также была реализована оптимизация получаемых конформаций с использованием силового поля MMFF94 [34].

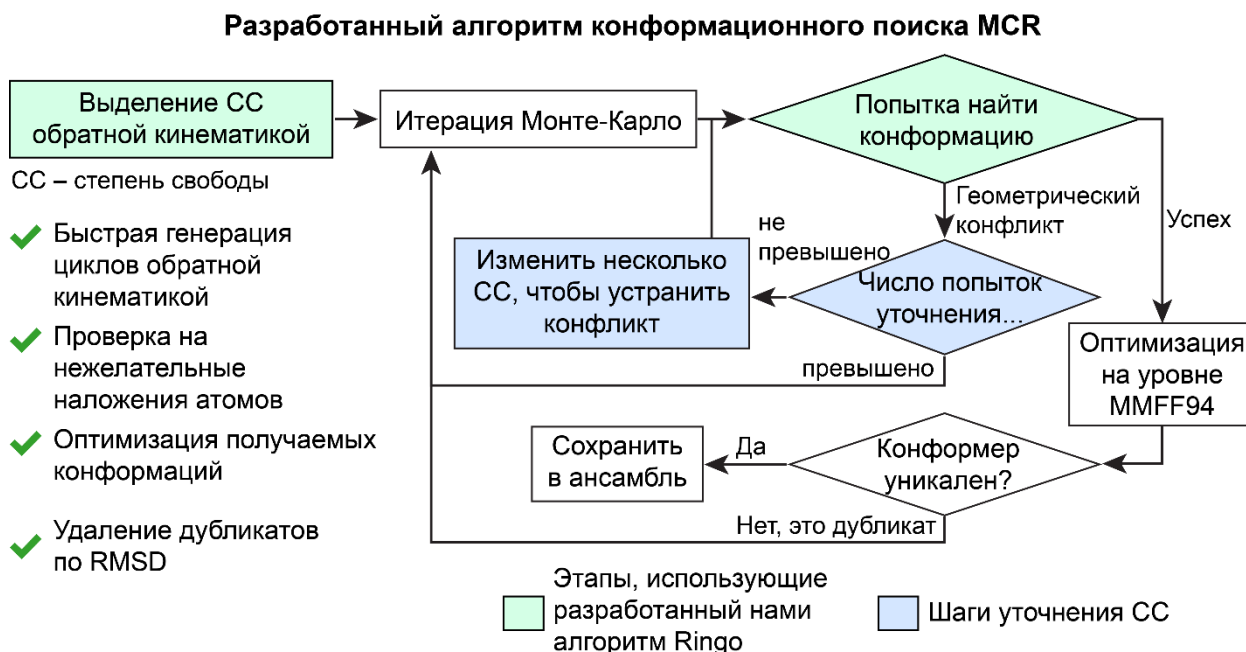


Рисунок 48. Схема разработанного алгоритма MCR. Этапы уточнения СС выделены синим цветом, а этапы применения обратной кинематики на основе разработанного нами алгоритма Ringo выделены зелёным.

В соответствии с обзором, проведённым в подразделе 1.2.2, тестирование проводилось по двум критериям: (1) разнообразие ансамблей согласно двумерным картам конформационных пространств макроциклов, и (2) способность методов генерировать экспериментальные конформации макроциклов, доступные в структурных базах CSD и PDB. Причём в качестве метрики для сравнения конформаций в обоих тестах была выбрана CRMSD (Рисунок 49) – величина RMSD между сравниваемыми структурами по атомам, находящимся внутри (поли)циклов (берётся максимальное значение RMSD в случаях, если циклических фрагментов несколько).

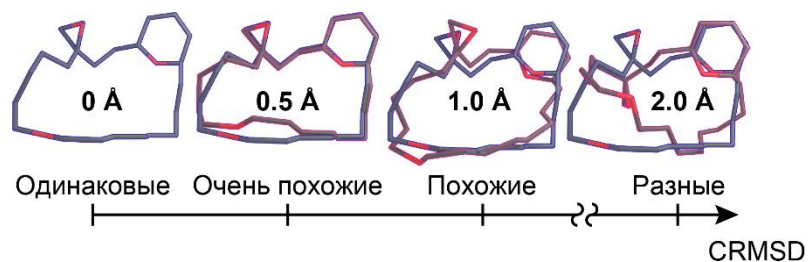


Рисунок 49. Иллюстрация метрики CRMSD, используемой в данной работе для сравнения конформаций полициклов.

Полученные нами результаты показывают, что конформационный поиск MCR превзошел методы MTD и RDKit, успешно воспроизведя экспериментальную конформацию 98 макроциклов с $CRMSD < 0.5 \text{ \AA}$ (Таблица 2). При увеличении порога CRMSD до 1.0 \AA , то есть при сравнении воспроизведения конформаций, похожих на экспериментальные, MCR по-прежнему превосходит MTD, однако немного уступает RDKit для двух макроциклов из 150.

При сравнении с лучшими на сегодняшний день методами, CREST и MMBS, MCR немного превосходит MMBS, найдя на один похожий конформер больше; однако уступает CREST, так как воспроизводит на две экспериментальные конформации меньше при обоих пороговых значениях CRMSD. При этом по сравнению с этими методами MCR имеет преимущество по скорости. Оба метода сравнения намного затратнее: MMBS обрабатывает одну молекулу до 30 раз дольше, чем десятиминутный запуск MCR, а CREST – до 207 раз дольше.

Таблица 2. Оценка способности метода MCR находить экспериментальные конформации макроциклов. Показано количество тестовых молекул, для которых была найдена хотя бы одна конформация близкая к экспериментальной (в заданном диапазоне CRMSD). Жирным выделен новый метод MCR. Зелёный цвет означает результат хуже, чем у MCR, красный – лучше.

Время работы	Метод	Порог CRMSD, Å	Число найденных экспериментальных конформаций	
			из 10 молекул	из 150 молекул
10 минут	MCR	< 0.5	6	98
	новый метод	< 1.0	9	130
	ETKDG (RDKit)	< 0.5	5	90
		< 1.0	9	132
MTD	< 0.5	5	74	
	< 1.0	7	118	
>10 минут (расчёт в X раз продолжительнее, чем MCR)	MacroModel базовый (MCMM) $0.5 \leq X \leq 2.5$	< 0.5 < 1.0	6 8	—
	MacroModel макроциклический (MMBS) $7 \leq X \leq 30$	< 0.5 < 1.0	6 8	
	iMTD-GC (CREST) $3 \leq X \leq 207$	< 0.5 < 1.0	8 10	

С точки зрения разнообразия конформационных ансамблей, мы показали, что MCR глобально покрывает конформационное пространство макроциклов и во многих случаях обнаруживает такие кластеры, которые не были найдены методами сравнения. Можно видеть, что для малоподвижного лиганда (верхние графики на Рисунке 50) MCR посетил все кластеры конформационного пространства макроцикла, которые были найдены методами сравнения, и, более того, обнаружил некоторые конформации, которые были упущены другими методами (зелёные точки). Для более конформационно подвижной молекулы Цетромицина (нижние графики на Рисунке 50) MCR также делает равномерное глобальное семплирование (серые и зеленые точки распределены по всей карте конформационного пространства), однако методам сравнения здесь удаётся найти конформации, отличные от тех, которые сгенерировал MCR. Несмотря на это, MCR тоже находит

довольно много кластеров конформаций, упущенных методами MacroModel, RDKit и MTD.

При сравнении с ансамблями метода CREST, MCR явно упускает множество конформаций, найденных CREST, и находит сильно меньше конформаций, недостающих в ансамблях CREST. Тем не менее MCR имеет большое преимущество по скорости построения ансамбля: расчёт CREST занимает более 12 часов, а MCR воспроизвёл существенную часть этого ансамбля всего за 10 минут. Упущенные конформации мы объясняем большим объемом конфигурационного пространства данной молекулы и лучшей приспособленностью комплексного пайплайна iMTD-GC в CREST к избирательному исследованию их низкоэнергетических областей, по сравнению с геометрическим перебором в MCR.

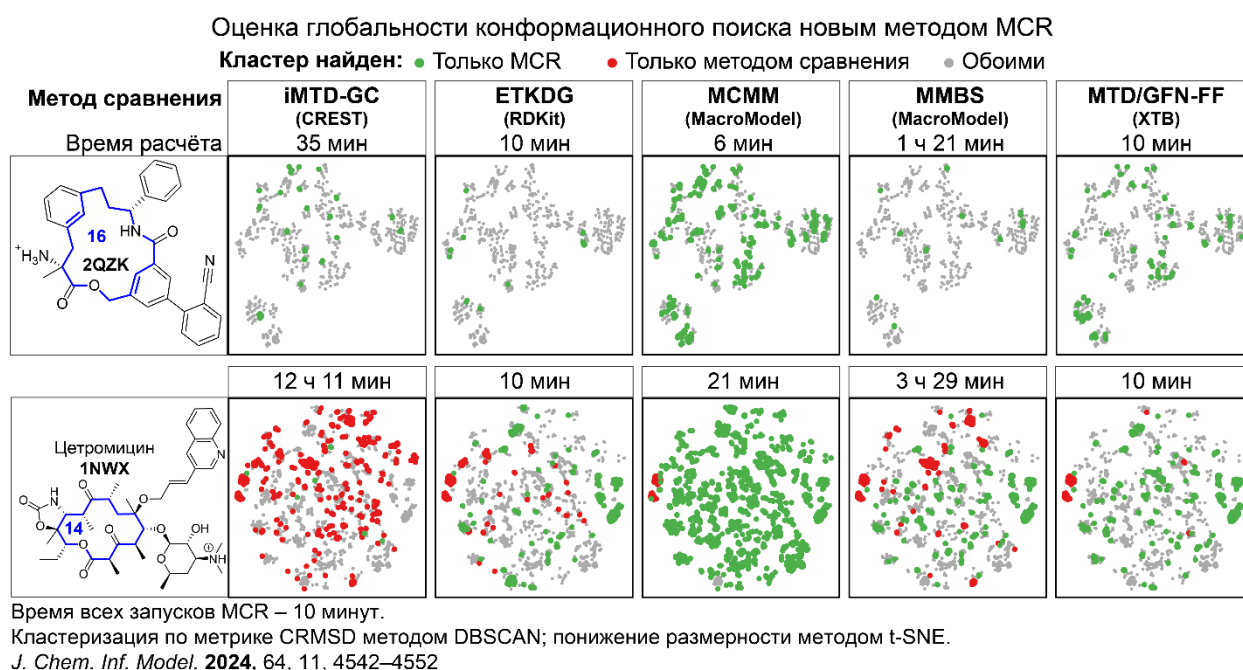


Рисунок 50. Результаты тестирования глобальности и полноты исследования конформационного пространства двух лигандов (2QZK и 1NWX из базы PDB) новым методом MCR и пятью методами сравнения. Метод построения графиков приведен в подразделе «Методология и методы исследования».

Сравнение времени, требуемого для генерации одного конформера в энергетическом окне в 10 ккал/моль от минимума, показало, что методы CREST, MMBS и MMCM генерируют конформации существенно медленнее MCR, в 3.1, 1.7 и 2.7 раз, соответственно; при этом во всех этих методах, включая MCR, используются методы молекулярной механики для оптимизации получаемых геометрий. Интересно отметить, что MCR и RDKit имеют практически одинаковую скорость ($\pm 1\%$) при использовании энергетического окна в 10 ккал/моль, однако сравнение скоростей генерации без отсеивания по энергии показало, что в среднем MCR генерирует конформеры в 1.7 раз быстрее, чем RDKit. То есть при сопоставимом количестве перспективных конформаций метод MCR обладает преимуществом в скорости генерации.

Несмотря на принципиальные отличия генерации стартовых геометрий для оптимизации в методах MCR и RDKit, схожесть их результатов в наших тестах говорит о компенсации преимуществ строгого геометрического подхода MCR и оптимизированного эмпирического подхода RDKit. В отличие от MCR, RDKit использует большой объём эмпирических данных (характерные торсионные углы и эксцентриситеты макроциклов), но его стартовые геометрии всё равно сильно искажены вследствие использования дистанционной геометрии, что повышает продолжительность последующей оптимизации. MCR, напротив, использует обратную кинематику (ОК) и получающиеся геометрии сразу удовлетворяют всем жёстким геометрическим ограничениям на длины связей и валентные углы, соответственно, повышенное качество стартовых конформаций уменьшает число шагов последующей оптимизации. В итоге разработанный нами прототип конформационного поиска на основе ОК через случайный перебор независимых торсионных углов даёт результат, не уступающий методу дистанционной геометрии, который совершенствовался уже более десяти лет [14].

Таким образом, нами был достигнут существенный прогресс в методологии моделирования конформационно подвижных (поли)макроциклических молекул. В основе разработанного алгоритма Ringo лежит предложенный нами универсальный способ параметризации конформационного пространства по значениям

независимых торсионных углов. С использованием этого алгоритма подход конформационного перебора Монте-Карло обобщён на (поли)макроциклические молекулы – полученный алгоритм конформационного поиска назван MCR. Тестирование показало, что MCR обеспечивает высокую полноту ансамблей, точность воспроизведения экспериментальных структур и скорость генерации конформаций, превосходя по одному или нескольким этим показателям популярные на сегодняшний день методы (CREST, ETKDG, MacroModel).

Учитывая, что подход ОК полностью совместим с многочисленными методами конформационного поиска, использующими приближение жёсткого ротатора (Раздел 1), и обобщает эти методы на поиск конформаций (поли)макроциклов, дальнейшее развитие и применение ОК представляет собой не просто возможный, а очевидный путь повышения эффективности конформационного анализа макроциклов – включая макроциклы со стереоэлектронным контролем конформационной подвижности (Раздел 2).

Заключение

В заключение кратко сформулируем основные результаты, полученные в диссертации. Все нижеприведённые результаты получены впервые.

1. Построены квантово-химические модели механизмов сборки полициклических пероксидов из β,γ' -трикетонов, β,δ' -трикетонов и δ -дикетонов под действием пероксида водорода и иных нуклеофилов (аминов, гидразидов), которые воспроизводят экспериментально наблюдаемые продукты реакции. Результаты конформационного анализа конечных продуктов согласуются с данными рентгеноструктурного анализа.

2. Установлена роль стереоэлектронных эффектов (отрицательной гиперконъюгации и обратного α -эффекта) в определении основных продуктов и величин кинетических барьеров сборки трициклических и азотсодержащих пероксидов и азапероксидов из поликетонов. Орбитальные взаимодействия с участием σ^* -орбиталей связей углерод-гетероатом определяют термодинамически наиболее выгодный продукт реакции, в то время как обратный α -эффект создаёт высокие кинетические барьеры на пути механизмов, включающих бимолекулярные стадии с участием пероксикарбениевых ионов.

3. Предложен и реализован алгоритм обратной кинематики для определения независимых торсионных углов полициклических молекул и построения их конформаций по заданному набору значений этих углов при фиксированных длинах связей и валентных углах. Весь процесс сборки конформации для одного набора значений торсионных углов выполняется за доли миллисекунды. Алгоритм является универсальным за исключением редких случаев, когда число связей в мостике между конденсированными циклами превосходит количество независимых торсионных углов в обоих циклах, между которыми он находится.

4. Предложен, реализован и протестирован универсальный алгоритм глобального конформационного поиска полициклических молекул путём случайного перебора независимых торсионных углов с помощью обратной кинематики. Этот метод обеспечивает высокую скорость генерации низкоэнергетических конформаций, полноту ансамблей и точность воспроизведения экспериментальных

структур, превосходя по одному или нескольким этим показателям существующие методы (CREST, ETKDG, MacroModel).

Автор выражает благодарность своему научному руководителю – кандидату физико-математических наук Медведеву Михаилу Геннадьевичу. Также автор благодарит Радулова Петра Сергеевича и Ярёмченко Ивана Андреевича за сотрудничество и их экспериментальную работу по синтезу каркасных пероксидов, Корлюкова Александра Александровича – за проведённые им рентгеноструктурные исследования и Алабугина Игоря Владимировича – за полученные знания в области стереоэлектронных эффектов.

Публикации автора по теме диссертации

Основные результаты диссертационной работы изложены в работах, опубликованных в рецензируемых научных изданиях, рекомендованных для защиты в диссертационном совете ИБХФ РАН по специальности и отрасли наук.

1. **Krivoshchapov N.V.**, Medvedev M.G. Ring kinematics-informed conformation space exploration // WIREs Computational Molecular Science. – 2024. – Vol. 14, № 1. – P. e1690. – DOI: 10.1002/wcms.1690.
2. **Krivoshchapov N.V.**, Medvedev M.G. Accurate and efficient conformer sampling of cyclic drug-like molecules with inverse kinematics // Journal of Chemical Information and Modeling. – 2024. – Vol. 64, № 11. – P. 4542–4552. – DOI: 10.1021/acs.jcim.3c02040.
3. Alabugin I.V., Kuhn L., **Krivoshchapov N.V.**, Mehaffy P., Medvedev M.G. Anomeric effect, hyperconjugation and electrostatics: lessons from complexity in a classic stereoelectronic phenomenon // Chemical Society Reviews. – 2021. – Vol. 50, № 18. – P. 10212–10252. – DOI: 10.1039/D1CS00564B.
4. Alabugin I.V., Kuhn L., Medvedev M.G., **Krivoshchapov N.V.**, Vil' V.A., Yaremenko I.A., Mehaffy P., Yarie M., Terent'ev A.O., Zolfigol M.A. Stereoelectronic power of oxygen in control of chemical reactivity: the anomeric effect is not alone // Chemical Society Reviews. – 2021. – Vol. 50, № 18. – P. 10253–10345. – DOI: 10.1039/D1CS00386K.
5. Yaremenko I.A., Belyakova Y.Yu., Radulov P.S., Novikov R.A., Medvedev M.G., **Krivoshchapov N.V.**, Korlyukov A.A., Alabugin I.V., Terent'ev A.O. Inverse α -effect as the Ariadne's thread on the way to tricyclic aminoperoxides: avoiding thermodynamic traps in the labyrinth of possibilities // Journal of the American Chemical Society. – 2022. – Vol. 144, № 16. – P. 7264–7282. – DOI: 10.1021/jacs.2c00406.
6. Yaremenko I.A., Belyakova Y.Yu., Radulov P.S., Novikov R.A., Medvedev M.G., **Krivoshchapov N.V.**, Korlyukov A.A., Alabugin I.V., Terent'ev A.O. Marriage of peroxides and nitrogen heterocycles: selective three-component assembly, peroxide-preserving rearrangement, and stereoelectronic source of unusual stability of bridged azaozonides // Journal of the American Chemical Society. – 2021. – Vol. 143, № 17. – P. 6634–6648. – DOI: 10.1021/jacs.1c02249.
7. Yaremenko I.A., Radulov P.S., Medvedev M.G., **Krivoshchapov N.V.**, Belyakova Y.Yu., Korlyukov A.A., Ilovaisky A.I., Terent'ev A.O., Alabugin I.V. How to build rigid oxygen-rich

- tricyclic heterocycles from triketones and hydrogen peroxide: control of dynamic covalent chemistry with inverse α -effect // *Journal of the American Chemical Society*. – 2020. – Vol. 142, № 34. – P. 14588–14607. – DOI: 10.1021/jacs.0c06294.
8. Yaremenko I.A., Belyakova Y.Yu., Radulov P.S., Medvedev M.G., **Krivoshchapov N.V.**, Alabugin I.V., Terent'ev A.O. Two-component versus three-component condensations in the race between hydrazide, triketone, and hydrogen peroxide-How do all six reactive centers cooperate to incorporate the most diverse set of heteroatomic bridges in a tricyclic frame? // *The Journal of Organic Chemistry*. – 2023. – Vol. 88, № 19. – P. 13782–13795. – DOI: 10.1021/acs.joc.3c01415.

Список используемых сокращений

BH	– Basin hopping, алгоритм конформационного поиска
CREST	– Conformer-rotamer sampling tool, программа для конформационного поиска
CRMSD	– RMSD по циклическим атомам
CSD	– Кембриджская структурная база данных
EI	– Expected improvement
EVM	– Explorational variance minimizer
FDA	– Federal drug administration
LCB	– Lower confidence bound
LE	– МД с локальным подъёмом потенциала
MCR	– Монте-Карло с уточнением, алгоритм конформационного поиска
ML	– машинное обучение
MMBS	– MacroModel's macrocycle baseline search
MTD	– метод LE, используемый в программах ХТВ и CREST
NBO	– анализ натуральных связевых орбиталей
PDB	– база данных белков
RMMC	– ringmaker Monte Carlo, алгоритм конформационного поиска
RMSD	– среднеквадратичное отклонение координат атомов
SMARTS	– SMILES arbitrary target specification
SMILES	– Simplified Molecular Input Line Entry System
SOS	– Self-Organizing Superimposition, алгоритм конформационного поиска
SPE	– Stochastic Proximity Embedding, алгоритм конформационного поиска
TFD	– Torsion Fingerprint Deviation
TLC	– Triaxial Loop Closure, алгоритм решения обратной кинематики
TPE	– древесно-структурированный оценщик Парзена
T-REMC	– Монте-Карло с марковскими цепями с температурным обменом репликами
T-REMD	– МД с температурным обменом репликами
H-REMD	– МД с гамильтоновым обменом репликами
БО	– Байесовская оптимизация
ДГ	– дистанционная геометрия
МД	– молекулярная динамика

МКМ	– метод Монте-Карло Метрополиса
ОАЭ	– обратный α -эффект
ОК	– обратная кинематика
ПК	– прямая кинематика
ПКК	– пероксикарбениевый катион
ППЭ	– поверхность потенциальной энергии
ПС	– переходное состояние
РСА	– рентгеноструктурный анализ
СС	– степень свободы
СЭ	– стереоэлектронный эффект
ЯМР	– ядерный магнитный резонанс

Список литературы

1. Crees Z.D., Rettig M.P., Jayasinghe R.G., Stockerl-Goldstein K., Larson S.M., Arpad I., Milone G.A., Martino M., Stiff P., Sborov D., Pereira D., Micallef I., Moreno-Jiménez G., Mikala G., Coronel M.L.P., Holtick U., Hiemenz J., Qazilbash M.H., Hardy N., Latif T., García-Cadenas I., Vainstein-Haras A., Sorani E., Gliko-Kabir I., Goldstein I., Ickowicz D., Shemesh-Darvish L., Kadosh S., Gao F., Schroeder M.A., Vij R., DiPersio J.F. Motixafortide and G-CSF to mobilize hematopoietic stem cells for autologous transplantation in multiple myeloma: a randomized phase 3 trial // *Nature Medicine*. – 2023. – Vol. 29, № 4. – P. 869–879. – DOI: 10.1038/s41591-023-02273-z.
2. Drilon A., Ou S.-H.I., Cho B.C., Kim D.-W., Lee J., Lin J.J., Zhu V.W., Ahn M.-J., Camidge D.R., Nguyen J., Zhai D., Deng W., Huang Z., Rogers E., Liu J., Whitten J., Lim J.K., Stopatschinskaja S., Hyman D.M., Doebele R.C., Cui J.J., Shaw A.T. Repotrectinib (TPX-0005) is a next-generation ROS1/TRK/ALK inhibitor that potently inhibits ROS1/TRK/ALK solvent-front mutations // *Cancer Discovery*. – 2018. – Vol. 8, № 10. – P. 1227–1236. – DOI: 10.1158/2159-8290.CD-18-0484.
3. Syed Y.Y. Rezapungin: first approval // *Drugs*. – 2023. – Vol. 83, № 9. – P. 833–840. – DOI: 10.1007/s40265-023-01891-8.
4. Giannangelo C., Fowkes F.J.I., Simpson J.A., Charman S.A., Creek D.J. Ozonide antimalarial activity in the context of artemisinin-resistant malaria // *Trends in Parasitology*. – 2019. – Vol. 35, № 7. – DOI: 10.1016/j.pt.2019.05.002.
5. Vil' V.A., Yaremenko I.A., Ilovaisky A.I., Terent'ev A.O. Peroxides with anthelmintic, antiprotozoal, fungicidal and antiviral bioactivity: properties, synthesis and reactions // *Molecules*. – 2017. – Vol. 22, № 11. – P. 1881. – DOI: 10.3390/molecules22111881.
6. Vil' V.A., Yaremenko I.A., Fomenkov D.I., Levitsky D.O., Fleury F., Terent'ev A.O. Ion exchange resin-catalyzed synthesis of bridged tetraoxanes possessing in vitro cytotoxicity against HeLa cancer cells // *Chemistry of Heterocyclic Compounds*. – 2020. – Vol. 56, № 6. – P. 722–726. – DOI: 10.1007/s10593-020-02722-4.
7. Gao F., Sun Z., Kong F., Xiao J. Artemisinin-derived hybrids and their anticancer activity // *European Journal of Medicinal Chemistry*. – 2020. – Vol. 188. – P. 112044. – DOI: 10.1016/j.ejmech.2020.112044.
8. Yaremenko I.A., Syromyatnikov M.Y., Radulov P.S., Belyakova Y.Yu., Fomenkov D.I., Popov V.N., Terent'ev A.O. Cyclic synthetic peroxides inhibit growth of entomopathogenic fungus *Ascospaera apis* without toxic effect on bumblebees // *Molecules*. – 2020. – Vol. 25, № 8. – P. 1954. – DOI: 10.3390/molecules25081954.
9. Efferth T. Beyond malaria: the inhibition of viruses by artemisinin-type compounds // *Biotechnology Advances*. – 2018. – Vol. 36, № 6. – P. 1730–1737. – DOI: 10.1016/j.biotechadv.2018.01.001.
10. Medvedev M.G., Bushmarinov I.S., Lyssenko K.A. Z-effect reversal in carboxylic acid associates // *Chemical Communications*. – 2016. – Vol. 52, № 39. – P. 6593–6596. – DOI: 10.1039/C6CC01305H.
11. dos Passos Gomes G., Yaremenko I.A., Radulov P.S., Novikov R.A., Chernyshev V.V., Korlyukov A.A., Nikishin G.I., Alabugin I.V., Terent'ev A.O. Stereoelectronic control in the ozone-free synthesis of ozonides // *Angewandte Chemie*. – 2017. – Vol. 129, № 18. – P. 5037–5041. – DOI: 10.1002/ange.201610699.

12. Juaristi E., dos Passos Gomes G., Terent'ev A.O., Notario R., Alabugin I.V. Stereoelectronic interactions as a probe for the existence of the intramolecular α -effect // *Journal of the American Chemical Society*. – 2017. – Vol. 139, № 31. – P. 10799–10813. – DOI: 10.1021/jacs.7b05367.
13. Vil' V.A., Barsegyan Y.A., Kuhn L., Ekimova M.V., Semenov E.A., Korlyukov A.A., Terent'ev A.O., Alabugin I.V. Synthesis of unstrained Criegee intermediates: inverse α -effect and other protective stereoelectronic forces can stop Baeyer–Villiger rearrangement of γ -hydroperoxy- γ -peroxylactones // *Chemical Science*. – 2020. – Vol. 11, № 20. – P. 5313–5322. – DOI: 10.1039/D0SC01025A.
14. Riniker S., Landrum G.A. Better informed distance geometry: using what we know to improve conformation generation // *Journal of Chemical Information and Modeling*. – 2015. – Vol. 55, № 12. – P. 2562–2574. – DOI: 10.1021/acs.jcim.5b00654.
15. Coutsias E.A., Seok C., Jacobson M.P., Dill K.A. A kinematic view of loop closure // *Journal of Computational Chemistry*. – 2004. – Vol. 25, № 4. – P. 510–528. – DOI: 10.1002/jcc.10416.
16. Coutsias E.A., Lexa K.W., Wester M.J., Pollock S.N., Jacobson M.P. Exhaustive Conformational Sampling of Complex Fused Ring Macrocycles Using Inverse Kinematics // *J. Chem. Theory Comput.* – 2016. – Vol. 12, № 9. – P. 4674–4687. – DOI: 10.1021/acs.jctc.6b00250.
17. Кривошапов Н. Ringo – Библиотека Python для построения конформаций на основе обратной кинематики. URL: <https://pypi.org/project/ringo-ik>.
18. Frisch M.J., Trucks G.W., Schlegel H.B., Scuseria G.E., Robb M.A., Cheeseman J.R., Scalmani G., Barone V., Petersson G.A., Nakatsuji H., Li X., Caricato M., Marenich A.V., Bloino J., Janesko B.G., Gomperts R., Mennucci B., Hratchian H.P., Ortiz J.V., Izmaylov A.F., Sonnenberg J.L., Williams-Young D., Ding F., Lipparini F., Egidi F., Goings J., Peng B., Petrone A., Henderson T., Ranasinghe D., Zakrzewski V.G., Gao J., Rega N., Zheng G., Liang W., Hada M., Ehara M., Toyota K., Fukuda R., Hasegawa J., Ishida M., Nakajima T., Honda Y., Kitao O., Nakai H., Vreven T., Throssell K., Montgomery J.A. Jr., Peralta J.E., Ogliaro F., Bearpark M.J., Heyd J.J., Brothers E.N., Kudin K.N., Staroverov V.N., Keith T.A., Kobayashi R., Normand J., Raghavachari K., Rendell A.P., Burant J.C., Iyengar S.S., Tomasi J., Cossi M., Millam J.M., Klene M., Adamo C., Cammi R., Ochterski J.W., Martin R.L., Morokuma K., Farkas O., Foresman J.B., Fox D.J. Gaussian 16 Revision A.03. – 2016.
19. 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 // *The Journal of Chemical Physics*. – 2010. – Vol. 132, № 15. – P. 154104. – DOI: 10.1063/1.3382344.
20. Adamo C., Barone V. Toward reliable density functional methods without adjustable parameters: The PBE0 model // *The Journal of Chemical Physics*. – 1999. – Vol. 110, № 13. – P. 6158–6170. – DOI: 10.1063/1.478522.
21. Krishnan R., Binkley J.S., Seeger R., Pople J.A. Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions // *The Journal of Chemical Physics*. – 1980. – Vol. 72, № 1. – P. 650–654. – DOI: 10.1063/1.438955.
22. Clark T., Chandrasekhar J., Spitznagel G.W., Schleyer P.V.R. Efficient diffuse function-augmented basis sets for anion calculations. III. The 3-21+G basis set for first-row elements, Li–F // *Journal of Computational Chemistry*. – 1983. – Vol. 4, № 3. – P. 294–301. – DOI: 10.1002/jcc.540040303.

23. 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. – Vol. 113, № 18. – P. 6378–6396. – DOI: 10.1021/jp810292n.
24. Mennucci B. Polarizable continuum model // *WIREs Computational Molecular Science*. – 2012. – Vol. 2, № 3. – P. 386–404. – DOI: 10.1002/wcms.1086.
25. Grimme S. Supramolecular Binding Thermodynamics by Dispersion-Corrected Density Functional Theory // *Chemistry – A European Journal*. – 2012. – Vol. 18, № 32. – P. 9955–9964. – DOI: 10.1002/chem.201200497.
26. RDKit: Open-source cheminformatics and Machine Learning. URL: <https://www.rdkit.org/docs/index.html>.
27. Pracht P., Bohle F., Grimme S. Automated exploration of the low-energy chemical space with fast quantum chemical methods // *Physical Chemistry Chemical Physics*. – 2020. – Vol. 22, № 14. – P. 7169–7192. – DOI: 10.1039/C9CP06869D.
28. 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. – Vol. 27, № 3. – P. 224–227. – DOI: 10.1016/j.mencom.2017.05.002.
29. Kozuch S., Shaik S. How to Conceptualize Catalytic Cycles? The Energetic Span Model // *Accounts of Chemical Research*. – 2011. – Vol. 44, № 2. – P. 101–110. – DOI: 10.1021/ar1000956.
30. Stegelmann C., Andreasen A., Campbell C.T. Degree of Rate Control: How Much the Energies of Intermediates and Transition States Control Rates // *Journal of the American Chemical Society*. – 2009. – Vol. 131, № 23. – P. 8077–8082. – DOI: 10.1021/ja9000097.
31. Kavitha T., Mehlhorn K., Michail D., Paluch K.E. An $\tilde{O}(m^2n)$ Algorithm for Minimum Cycle Basis of Graphs // *Algorithmica*. – 2008. – Vol. 52, № 3. – P. 333–349. – DOI: 10.1007/s00453-007-9064-z.
32. Spicher S., Grimme S. Robust Atomistic Modeling of Materials, Organometallic, and Biochemical Systems // *Angewandte Chemie International Edition*. – 2020. – Vol. 59, № 36. – P. 15665–15673. – DOI: 10.1002/anie.202004239.
33. Ehlert S., Stahn M., Spicher S., Grimme S. Robust and Efficient Implicit Solvation Model for Fast Semiempirical Methods // *Journal of Chemical Theory and Computation*. – 2021. – Vol. 17, № 7. – P. 4250–4261. – DOI: 10.1021/acs.jctc.1c00471.
34. Halgren T.A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94 // *Journal of Computational Chemistry*. – 1996. – Vol. 17, № 5–6. – P. 490–519. – DOI: 10.1002/(SICI)1096-987X(199604)17:5/6%3C490::AID-JCC1%3E3.0.CO;2-P.
35. 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. – Vol. 54, № 10. – P. 2680–2696. – DOI: 10.1021/ci5001696.
36. Ester M., Kriegel H.-P., Sander J., Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise // *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*. – 1996. – P. 226–231. – Portland, Oregon: AAAI Press.

37. Cai T.T., Ma R. Theoretical Foundations of t-SNE for Visualizing High-Dimensional Clustered Data // *Journal of Machine Learning Research*. – 2022. – Vol. 23, № 301. – P. 1–54.
38. Xu H., Izrailev S., Agrafiotis D.K. Conformational Sampling by Self-Organization // *Journal of Chemical Information and Computer Sciences*. – 2003. – Vol. 43, № 4. – P. 1186–1191. – DOI: 10.1021/ci0340557.
39. Izrailev S., Zhu F., Agrafiotis D.K. A distance geometry heuristic for expanding the range of geometries sampled during conformational search // *Journal of Computational Chemistry*. – 2006. – Vol. 27, № 16. – P. 1962–1969. – DOI: 10.1002/jcc.20506.
40. Zhu F., Agrafiotis D.K. Self-organizing superimposition algorithm for conformational sampling // *Journal of Computational Chemistry*. – 2007. – Vol. 28, № 7. – P. 1234–1239. – DOI: 10.1002/jcc.20622.
41. Crippen G.M., Havel T.F. Stable calculation of coordinates from distance information // *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography*. – 1978. – Vol. 34, № 2. – P. 282–284. – DOI: 10.1107/S0567739478000522.
42. Havel T.F. Metric matrix embedding in protein structure calculations, NMR spectra analysis, and relaxation theory // *Magnetic Resonance in Chemistry*. – 2003. – Vol. 41, № S1. – P. S37–S50. – DOI: 10.1002/mrc.1242.
43. Spellmeyer D.C., Wong A.K., Bower M.J., Blaney J.M. Conformational analysis using distance geometry methods // *Journal of Molecular Graphics and Modelling*. – 1997. – Vol. 15, № 1. – P. 18–36. – DOI: 10.1016/S1093-3263(97)00014-4.
44. Poongavanam V., Danelius E., Peintner S., Alcaraz L., Caron G., Cummings M.D., Wlodek S., Erdelyi M., Hawkins P.C.D., Ermondi G., Kihlberg J. Conformational Sampling of Macrocyclic Drugs in Different Environments: Can We Find the Relevant Conformations? // *ACS Omega*. – 2018. – Vol. 3, № 9. – P. 11742–11757. – DOI: 10.1021/acsomega.8b01379.
45. Wang S., Witek J., Landrum G.A., Riniker S. Improving Conformer Generation for Small Rings and Macrocycles Based on Distance Geometry and Experimental Torsional-Angle Preferences // *Journal of Chemical Information and Modeling*. – 2020. – Vol. 60, № 4. – P. 2044–2058. – DOI: 10.1021/acs.jcim.0c00025.
46. Lagorce D., Pencheva T., Villoutreix B.O., Miteva M.A. DG-AMMOS: A new tool to generate 3D conformation of small molecules using distance geometry and automated molecular mechanics optimization for in silico screening // *BMC Chemical Biology*. – 2009. – Vol. 9, № 1. – P. 6. – DOI: 10.1186/1472-6769-9-6.
47. Simm G.N.C., Hernández-Lobato J.M. A Generative Model for Molecular Distance Geometry // *arXiv preprint*. – 2020. – № arXiv:1909.11459. – DOI: 10.48550/arXiv.1909.11459.
48. Friedrich N.-O., Meyder A., de Bruyn Kops C., Sommer K., Flachsenberg F., Rarey M., Kirchmair J. High-Quality Dataset of Protein-Bound Ligand Conformations and Its Application to Benchmarking Conformer Ensemble Generators // *Journal of Chemical Information and Modeling*. – 2017. – Vol. 57, № 3. – P. 529–539. – DOI: 10.1021/acs.jcim.6b00613.
49. Leach A.R. *Molecular Modelling: Principles and Applications*. – 2001. – Prentice Hall.
50. Chang G., Guida W.C., Still W.C. An internal-coordinate Monte Carlo method for searching conformational space // *Journal of the American Chemical Society*. – 1989. – Vol. 111, № 12. – P. 4379–4386. – DOI: 10.1021/ja00194a035.

51. Mohamadi F., Richards N.G.J., Guida W.C., Liskamp R., Lipton M., Caufield C., Chang G., Hendrickson T., Still W.C. Macromodel—an integrated software system for modeling organic and bioorganic molecules using molecular mechanics // *Journal of Computational Chemistry*. – 1990. – Vol. 11, № 4. – P. 440–467. – DOI: 10.1002/jcc.540110405.
52. 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. – Vol. 3, № 1. – P. 33. – DOI: 10.1186/1758-2946-3-33.
53. Goodman J.M., Still W.C. An unbounded systematic search of conformational space // *Journal of Computational Chemistry*. – 1991. – Vol. 12, № 9. – P. 1110–1117. – DOI: 10.1002/jcc.540120908.
54. Leite T.B., Gomes D., Miteva M.A., Chomilier J., Villoutreix B.O., Tufféry P. Frog: a FRee Online druG 3D conformation generator // *Nucleic Acids Research*. – 2007. – Vol. 35, № suppl_2. – P. W568–W572. – DOI: 10.1093/nar/gkm289.
55. Li J., Ehlers T., Sutter J., Varma-O’Brien S., Kirchmair J. CAESAR: A New Conformer Generation Algorithm Based on Recursive Buildup and Local Rotational Symmetry Consideration // *Journal of Chemical Information and Modeling*. – 2007. – Vol. 47, № 5. – P. 1923–1932. – DOI: 10.1021/ci700136x.
56. Lipton M., Still W.C. The multiple minimum problem in molecular modeling. Tree searching internal coordinate conformational space // *Journal of Computational Chemistry*. – 1988. – Vol. 9, № 4. – P. 343–355. – DOI: 10.1002/jcc.540090409.
57. O’Boyle N.M., Vandermeersch T., Flynn C.J., Maguire A.R., Hutchison G.R. Confab - Systematic generation of diverse low-energy conformers // *Journal of Cheminformatics*. – 2011. – Vol. 3, № 1. – P. 8. – DOI: 10.1186/1758-2946-3-8.
58. Still W.C., Galynker I. Chemical consequences of conformation in macrocyclic compounds: An effective approach to remote asymmetric induction // *Tetrahedron*. – 1981. – Vol. 37, № 23. – P. 3981–3996. – DOI: 10.1016/S0040-4020(01)93273-9.
59. Barbeau X., Vincent A.T., Lagüe P. ConfBuster: Open-Source Tools for Macrocyclic Conformational Search and Analysis // *Journal of Open Research Software*. – 2018. – Vol. 6, № 1. – DOI: 10.5334/jors.189.
60. Miteva M.A., Guyon F., Tufféry P. Frog2: Efficient 3D conformation ensemble generator for small compounds // *Nucleic Acids Research*. – 2010. – Vol. 38, № suppl_2. – P. W622–W627. – DOI: 10.1093/nar/gkq325.
61. Gasteiger J., Rudolph C., Sadowski J. Automatic generation of 3D-atomic coordinates for organic molecules // *Tetrahedron Computer Methodology*. – 1990. – Vol. 3, № 6, Part C. – P. 537–547. – DOI: 10.1016/0898-5529(90)90156-3.
62. Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N., Bourne P.E. The Protein Data Bank // *Nucleic Acids Research*. – 2000. – Vol. 28, № 1. – P. 235–242. – DOI: 10.1093/nar/28.1.235.
63. Groom C.R., Bruno I.J., Lightfoot M.P., Ward S.C. The Cambridge Structural Database // *Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials*. – 2016. – Vol. 72, № 2. – P. 171–179. – DOI: 10.1107/S2052520616003954.
64. Kothiwale S., Mendenhall J.L., Meiler J. BCL::Conf: small molecule conformational sampling using a knowledge based rotamer library // *Journal of Cheminformatics*. – 2015. – Vol. 7, № 1. – P. 47. – DOI: 10.1186/s13321-015-0095-1.
65. Guba W., Meyder A., Rarey M., Hert J. Torsion Library Reloaded: A New Version of Expert-Derived SMARTS Rules for Assessing Conformations of Small Molecules //

- Journal of Chemical Information and Modeling. – 2016. – Vol. 56, № 1. – P. 1–5. – DOI: 10.1021/acs.jcim.5b00522.
66. Friedrich N.-O., Flachsenberg F., Meyder A., Sommer K., Kirchmair J., Rarey M. Conformer: A Novel Method for the Generation of Conformer Ensembles // Journal of Chemical Information and Modeling. – 2019. – Vol. 59, № 2. – P. 731–742. – DOI: 10.1021/acs.jcim.8b00704.
67. Seidel T., Permann C., Wieder O., Kohlbacher S.M., Langer T. High-Quality Conformer Generation with CONFORGE: Algorithm and Performance Assessment // Journal of Chemical Information and Modeling. – 2023. – Vol. 63, № 17. – P. 5549–5570. – DOI: 10.1021/acs.jcim.3c00563.
68. Bhowmick A., Head-Gordon T. A Monte Carlo Method for Generating Side Chain Structural Ensembles // Structure. – 2015. – Vol. 23, № 1. – P. 44–55. – DOI: 10.1016/j.str.2014.10.011.
69. Rohl C.A., Strauss C.E.M., Misura K.M.S., Baker D. Protein Structure Prediction Using Rosetta // Numerical Computer Methods, Part D. – 2004. – Vol. 383. – P. 66–93. – DOI: 10.1016/S0076-6879(04)83004-0.
70. Shapovalov M.V., Dunbrack R.L. A Smoothed Backbone-Dependent Rotamer Library for Proteins Derived from Adaptive Kernel Density Estimates and Regressions // Structure. – 2011. – Vol. 19, № 6. – P. 844–858. – DOI: 10.1016/j.str.2011.03.019.
71. Watts K.S., Dalal P., Murphy R.B., Sherman W., Friesner R.A., Shelley J.C. ConfGen: A Conformational Search Method for Efficient Generation of Bioactive Conformers // Journal of Chemical Information and Modeling. – 2010. – Vol. 50, № 4. – P. 534–546. – DOI: 10.1021/ci100015j.
72. Rusinko A.I., Sheridan R.P., Nilakantan R., Haraki K.S., Bauman N., Venkataraghavan R. Using CONCORD to construct a large database of three-dimensional coordinates from connection tables // Journal of Chemical Information and Computer Sciences. – 1989. – Vol. 29, № 4. – P. 251–255. – DOI: 10.1021/ci00064a004.
73. Smellie A., Stanton R., Henne R., Teig S. Conformational analysis by intersection: CONAN // Journal of Computational Chemistry. – 2003. – Vol. 24, № 1. – P. 10–20. – DOI: 10.1002/jcc.10175.
74. Hawkins P.C.D., Skillman A.G., Warren G.L., Ellingson B.A., Stahl M.T. Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database // Journal of Chemical Information and Modeling. – 2010. – Vol. 50, № 4. – P. 572–584. – DOI: 10.1021/ci100031x.
75. Ebejer J.-P., Morris G.M., Deane C.M. Freely Available Conformer Generation Methods: How Good Are They? // Journal of Chemical Information and Modeling. – 2012. – Vol. 52, № 5. – P. 1146–1158. – DOI: 10.1021/ci2004658.
76. Friedrich N.-O., de Bruyn Kops C., Flachsenberg F., Sommer K., Rarey M., Kirchmair J. Benchmarking Commercial Conformer Ensemble Generators // Journal of Chemical Information and Modeling. – 2017. – Vol. 57, № 11. – P. 2719–2728. – DOI: 10.1021/acs.jcim.7b00505.
77. Karplus M., Petsko G.A. Molecular dynamics simulations in biology // Nature. – 1990. – Vol. 347, № 6294. – P. 631–639. – DOI: 10.1038/347631a0.
78. Bruccoleri R.E., Karplus M. Conformational sampling using high-temperature molecular dynamics // Biopolymers. – 1990. – Vol. 29, № 14. – P. 1847–1862. – DOI: 10.1002/bip.360291415.

79. Stillinger F.H., Weber T.A. Computer simulation of local order in condensed phases of silicon // *Physical Review B*. – 1985. – Vol. 31, № 8. – P. 5262–5271. – DOI: 10.1103/PhysRevB.31.5262.
80. Ryckaert J.-P., Ciccotti G., Berendsen H.J.C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes // *Journal of Computational Physics*. – 1977. – Vol. 23, № 3. – P. 327–341. – DOI: 10.1016/0021-9991(77)90098-5.
81. Wu X., Wang S. Self-Guided Molecular Dynamics Simulation for Efficient Conformational Search // *The Journal of Physical Chemistry B*. – 1998. – Vol. 102, № 37. – P. 7238–7250. – DOI: 10.1021/jp9817372.
82. Schönborn S.E., Goedecker S., Roy S., Oganov A.R. The performance of minima hopping and evolutionary algorithms for cluster structure prediction // *The Journal of Chemical Physics*. – 2009. – Vol. 130, № 14. – P. 144108. – DOI: 10.1063/1.3097197.
83. Labute P. LowModeMD—Implicit Low-Mode Velocity Filtering Applied to Conformational Search of Macrocycles and Protein Loops // *Journal of Chemical Information and Modeling*. – 2010. – Vol. 50, № 5. – P. 792–800. – DOI: 10.1021/ci900508k.
84. Chen I.-J., Foloppe N. Tackling the conformational sampling of larger flexible compounds and macrocycles in pharmacology and drug discovery // *Bioorganic & Medicinal Chemistry*. – 2013. – Vol. 21, № 24. – P. 7898–7920. – DOI: 10.1016/j.bmc.2013.10.003.
85. Henkelman G., Jónsson H. A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives // *The Journal of Chemical Physics*. – 1999. – Vol. 111, № 15. – P. 7010–7022. – DOI: 10.1063/1.480097.
86. Fröhlich M.G., Sewell T.D. Pivot Algorithm and Push-off Method for Efficient System Generation of All-Atom Polymer Melts: Application to Hydroxyl-Terminated Polybutadiene // *Macromolecular Theory and Simulations*. – 2013. – Vol. 22, № 6. – P. 344–353. – DOI: 10.1002/mats.201300103.
87. Kunitsyn A.Y., Nekrasova N.A., Krivoshchapov N.V., Alexandrov E.V., Pavlov A.A., Medvedev M.G. Hamiltonian replica-exchange method α -REMD for ring spearing elimination in polymers // *The Journal of Chemical Physics*. – 2025. – Vol. 162, № 10. – P. 104114. – DOI: 10.1063/5.0241538.
88. Huber T., Torda A.E., van Gunsteren W.F. Local elevation: A method for improving the searching properties of molecular dynamics simulation // *Journal of Computer-Aided Molecular Design*. – 1994. – Vol. 8, № 6. – P. 695–708. – DOI: 10.1007/BF00124016.
89. Grimme S., Bannwarth C., Shushkov P. A Robust and Accurate Tight-Binding Quantum Chemical Method for Structures, Vibrational Frequencies, and Noncovalent Interactions of Large Molecular Systems Parametrized for All spd-Block Elements ($Z = 1–86$) // *Journal of Chemical Theory and Computation*. – 2017. – Vol. 13, № 5. – P. 1989–2009. – DOI: 10.1021/acs.jctc.7b00118.
90. Wales D.J., Doye J.P.K. Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms // *The Journal of Physical Chemistry A*. – 1997. – Vol. 101, № 28. – P. 5111–5116. – DOI: 10.1021/jp970984n.
91. Li Z., Scheraga H.A. Monte Carlo-minimization approach to the multiple-minima problem in protein folding. // *Proceedings of the National Academy of Sciences of the*

- United States of America. – 1987. – Vol. 84, № 19. – P. 6611–6615. – DOI: 10.1073/pnas.84.19.6611.
92. Jorgensen W.L., Tirado-Rives J. Monte Carlo vs Molecular Dynamics for Conformational Sampling // *The Journal of Physical Chemistry*. – 1996. – Vol. 100, № 34. – P. 14508–14513. – DOI: 10.1021/jp960880x.
93. Lu M., Dousis A.D., Ma J. OPUS-Rota: A fast and accurate method for side-chain modeling // *Protein Science*. – 2008. – Vol. 17, № 9. – P. 1576–1585. – DOI: 10.1110/ps.035022.108.
94. Kolossváry I., Guida W.C. Low Mode Search. An Efficient, Automated Computational Method for Conformational Analysis: Application to Cyclic and Acyclic Alkanes and Cyclic Peptides // *Journal of the American Chemical Society*. – 1996. – Vol. 118, № 21. – P. 5011–5019. – DOI: 10.1021/ja952478m.
95. Parish C., Lombardi R., Sinclair K., Smith E., Goldberg A., Rappleye M., Dure M. A comparison of the Low Mode and Monte Carlo conformational search methods // *Journal of Molecular Graphics and Modelling*. – 2002. – Vol. 21, № 2. – P. 129–150. – DOI: 10.1016/S1093-3263(02)00144-4.
96. Miloshevsky G.V., Jordan P.C. Open-State Conformation of the KcsA K⁺ Channel: Monte Carlo Normal Mode Following Simulations // *Structure*. – 2007. – Vol. 15, № 12. – P. 1654–1662. – DOI: 10.1016/j.str.2007.09.022.
97. Kantarci-Carsibasi N., Haliloglu T., Doruker P. Conformational Transition Pathways Explored by Monte Carlo Simulation Integrated with Collective Modes // *Biophysical Journal*. – 2008. – Vol. 95, № 12. – P. 5862–5873. – DOI: 10.1529/biophysj.107.128447.
98. Wilson S.R., Cui W., Moskowitz J.W., Schmidt K.E. Applications of simulated annealing to the conformational analysis of flexible molecules // *Journal of Computational Chemistry*. – 1991. – Vol. 12, № 3. – P. 342–349. – DOI: 10.1002/jcc.540120307.
99. Weiser J., Holthausen M.C., Fitjer L. HUNTER: A conformational search program for acyclic to polycyclic molecules with special emphasis on stereochemistry // *Journal of Computational Chemistry*. – 1997. – Vol. 18, № 10. – P. 1264–1281. – DOI: 10.1002/(SICI)1096-987X(19970730)18:10%3C1264::AID-JCC2%3E3.0.CO;2-L.
100. Morales L.B., Garduño-Juárez R., Romero D. Applications of Simulated Annealing to the Multiple-Minima Problem in Small Peptides // *Journal of Biomolecular Structure and Dynamics*. – 1991. – Vol. 8, № 4. – P. 721–735. – DOI: 10.1080/07391102.1991.10507841.
101. Hao G.-F., Xu W.-F., Yang S.-G., Yang G.-F. Multiple Simulated Annealing-Molecular Dynamics (MSA-MD) for Conformational Space Search of Peptide and Miniprotein // *Scientific Reports*. – 2015. – Vol. 5, № 1. – P. 15568. – DOI: 10.1038/srep15568.
102. Lee J., Lee I.-H., Lee J. Unbiased Global Optimization of Lennard-Jones Clusters for \mathcal{N} Using the Conformational Space Annealing Method // *Physical Review Letters*. – 2003. – Vol. 91, № 8. – P. 080201. – DOI: 10.1103/PhysRevLett.91.080201.
103. Grebner C., Becker J., Stepanenko S., Engels B. Efficiency of tabu-search-based conformational search algorithms // *Journal of Computational Chemistry*. – 2011. – Vol. 32, № 10. – P. 2245–2253. – DOI: 10.1002/jcc.21807.

104. Ozkan S.B., Meirovitch H. Efficient Conformational Search Method for Peptides and Proteins: Monte Carlo Minimization with an Adaptive Bias // *The Journal of Physical Chemistry B*. – 2003. – Vol. 107, № 34. – P. 9128–9131. – DOI: 10.1021/jp0346615.
105. Ozkan S.B., Meirovitch H. Conformational search of peptides and proteins: Monte Carlo minimization with an adaptive bias method applied to the heptapeptide deltorphin // *Journal of Computational Chemistry*. – 2004. – Vol. 25, № 4. – P. 565–572. – DOI: 10.1002/jcc.10399.
106. de Souza B. GOAT: A Global Optimization Algorithm for Molecules and Atomic Clusters // *Angewandte Chemie International Edition*. – 2025. – Vol. 64, № 18. – P. e202500393. – DOI: 10.1002/anie.202500393.
107. Unsleber J.P., Grimm S.A., Reiher M. Chemoton 2.0: Autonomous Exploration of Chemical Reaction Networks // *Journal of Chemical Theory and Computation*. – 2022. – Vol. 18, № 9. – P. 5393–5409. – DOI: 10.1021/acs.jctc.2c00193.
108. E W., Zhou X. The gentlest ascent dynamics // *Nonlinearity*. – 2011. – Vol. 24, № 6. – P. 1831. – DOI: 10.1088/0951-7715/24/6/008.
109. Sugita Y., Okamoto Y. Replica-exchange molecular dynamics method for protein folding // *Chemical Physics Letters*. – 1999. – Vol. 314, № 1. – P. 141–151. – DOI: 10.1016/S0009-2614(99)01123-9.
110. Hukushima K., Nemoto K. Exchange Monte Carlo Method and Application to Spin Glass Simulations // *Journal of the Physical Society of Japan*. – 1996. – Vol. 65, № 6. – P. 1604–1608. – DOI: 10.1143/JPSJ.65.1604.
111. Fukunishi H., Watanabe O., Takada S. On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction // *The Journal of Chemical Physics*. – 2002. – Vol. 116, № 20. – P. 9058–9067. – DOI: 10.1063/1.1472510.
112. Affentranger R., Tavernelli I., Di Iorio E.E. A Novel Hamiltonian Replica Exchange MD Protocol to Enhance Protein Conformational Space Sampling // *Journal of Chemical Theory and Computation*. – 2006. – Vol. 2, № 2. – P. 217–228. – DOI: 10.1021/ct050250b.
113. Kannan S., Zacharias M. Enhanced sampling of peptide and protein conformations using replica exchange simulations with a peptide backbone biasing-potential // *Proteins: Structure, Function, and Bioinformatics*. – 2007. – Vol. 66, № 3. – P. 697–706. – DOI: 10.1002/prot.21258.
114. Hritz J., Oostenbrink C. Hamiltonian replica exchange molecular dynamics using soft-core interactions // *The Journal of Chemical Physics*. – 2008. – Vol. 128, № 14. – P. 144121. – DOI: 10.1063/1.2888998.
115. Luitz M.P., Zacharias M. Protein–Ligand Docking Using Hamiltonian Replica Exchange Simulations with Soft Core Potentials // *Journal of Chemical Information and Modeling*. – 2014. – Vol. 54, № 6. – P. 1669–1675. – DOI: 10.1021/ci500296f.
116. Bespalov I.A., Krivoshchapov N.V., Lisov A.A., Chaliy V.A., Medvedev M.G. Physics-Informed Bayesian Optimization for Conformational Ensemble Augmentation // *Journal of Chemical Information and Modeling*. – 2025. – Vol. 65, № 12. – P. 6048–6056. – DOI: 10.1021/acs.jcim.5c00522.
117. Kolossváry I., Keserü G.M. Hessian-free low-mode conformational search for large-scale protein loop optimization: application to c-jun N-terminal kinase JNK3 //

- Journal of Computational Chemistry. – 2001. – Vol. 22, № 1. – P. 21–30. – DOI: 10.1002/1096-987X(20010115)22:1%3C21::AID-JCC3%3E3.0.CO;2-I.
118. Aslan M., Davis J.B.A., Johnston R.L. Global optimization of small bimetallic Pd–Co binary nanoalloy clusters: a genetic algorithm approach at the DFT level // *Physical Chemistry Chemical Physics*. – 2016. – Vol. 18, № 9. – P. 6676–6682. – DOI: 10.1039/C6CP00342G.
 119. Davis J.B.A., Shayeghi A., Horswell S.L., Johnston R.L. The Birmingham parallel genetic algorithm and its application to the direct DFT global optimisation of IrN (N = 10–20) clusters // *Nanoscale*. – 2015. – Vol. 7, № 33. – P. 14032–14038. – DOI: 10.1039/C5NR03774C.
 120. Deaven D.M., Ho K.M. Molecular Geometry Optimization with a Genetic Algorithm // *Physical Review Letters*. – 1995. – Vol. 75, № 2. – P. 288–291. – DOI: 10.1103/PhysRevLett.75.288.
 121. Supady A., Blum V., Baldauf C. First-Principles Molecular Structure Search with a Genetic Algorithm // *Journal of Chemical Information and Modeling*. – 2015. – Vol. 55, № 11. – P. 2338–2348. – DOI: 10.1021/acs.jcim.5b00243.
 122. Sun S. A genetic algorithm that seeks native states of peptides and proteins // *Biophysical Journal*. – 1995. – Vol. 69, № 2. – P. 340–355. – DOI: 10.1016/S0006-3495(95)79906-4.
 123. Yang H., Wong M.W. Automatic Conformational Search of Transition States for Catalytic Reactions Using Genetic Algorithm // *The Journal of Physical Chemistry A*. – 2019. – Vol. 123, № 47. – P. 10303–10314. – DOI: 10.1021/acs.jpca.9b09543.
 124. Shankland K., David W.I.F., Csoka T. Crystal structure determination from powder diffraction data by the application of a genetic algorithm // *Zeitschrift für Kristallographie-Crystalline Materials*. – 1997. – Vol. 212, № 8. – P. 550–552. – De Gruyter Oldenbourg.
 125. Payne A.W.R., Glen R.C. Molecular recognition using a binary genetic search algorithm // *Journal of Molecular Graphics*. – 1993. – Vol. 11, № 2. – P. 74–91. – DOI: 10.1016/0263-7855(93)87001-L.
 126. Goto H., Osawa E. Corner flapping: a simple and fast algorithm for exhaustive generation of ring conformations // *Journal of the American Chemical Society*. – 1989. – Vol. 111, № 24. – P. 8950–8951. – DOI: 10.1021/ja00206a046.
 127. Gotō H., Ōsawa E. An efficient algorithm for searching low-energy conformers of cyclic and acyclic molecules // *Journal of the Chemical Society, Perkin Transactions 2*. – 1993. – № 2. – P. 187–198. – DOI: 10.1039/P29930000187.
 128. Glass C.W., Oganov A.R., Hansen N. USPEX—Evolutionary crystal structure prediction // *Computer Physics Communications*. – 2006. – Vol. 175, № 11. – P. 713–720. – DOI: 10.1016/j.cpc.2006.07.020.
 129. Lyakhov A.O., Oganov A.R., Stokes H.T., Zhu Q. New developments in evolutionary structure prediction algorithm USPEX // *Computer Physics Communications*. – 2013. – Vol. 184, № 4. – P. 1172–1182. – DOI: 10.1016/j.cpc.2012.12.009.
 130. Falls Z., Avery P., Wang X., Hilleke K.P., Zurek E. The XtalOpt Evolutionary Algorithm for Crystal Structure Prediction // *The Journal of Physical Chemistry C*. – 2021. – Vol. 125, № 3. – P. 1601–1620. – DOI: 10.1021/acs.jpcc.0c09531.

131. Curtis F., Rose T., Marom N. Evolutionary niching in the GAtor genetic algorithm for molecular crystal structure prediction // *Faraday Discussions*. – 2018. – Vol. 211. – P. 61–77. – DOI: 10.1039/C8FD00067K.
132. Chan L., Hutchison G.R., Morris G.M. Bayesian optimization for conformer generation // *Journal of Cheminformatics*. – 2019. – Vol. 11, № 1. – P. 32. – DOI: 10.1186/s13321-019-0354-7.
133. Chan L., Hutchison G.R., Morris G.M. BOKEI: Bayesian optimization using knowledge of correlated torsions and expected improvement for conformer generation // *Physical Chemistry Chemical Physics*. – 2020. – Vol. 22, № 9. – P. 5211–5219. – DOI: 10.1039/C9CP06688H.
134. Fang L., Makkonen E., Todorović M., Rinke P., Chen X. Efficient Amino Acid Conformer Search with Bayesian Optimization // *Journal of Chemical Theory and Computation*. – 2021. – Vol. 17, № 3. – P. 1955–1966. – DOI: 10.1021/acs.jctc.0c00648.
135. Karniadakis G.E., Kevrekidis I.G., Lu L., Perdikaris P., Wang S., Yang L. Physics-informed machine learning // *Nature Reviews Physics*. – 2021. – Vol. 3, № 6. – P. 422–440. – DOI: 10.1038/s42254-021-00314-5.
136. Gilboa E., Saatçi Y., Cunningham J., Gilboa E. Scaling Multidimensional Gaussian Processes using Projected Additive Approximations // *Proceedings of the 30th International Conference on Machine Learning*. – 2013. – P. 454–461.
137. Andreadi N., Zankov D., Karpov K., Mitrofanov A. Tree Parzen estimator for global geometry optimization: A benchmark and database of experimental gas-phase structures of organic molecules // *Journal of Computational Chemistry*. – 2022. – Vol. 43, № 21. – P. 1434–1441. – DOI: 10.1002/jcc.26947.
138. Axelrod S., Gómez-Bombarelli R. GEOM, energy-annotated molecular conformations for property prediction and molecular generation // *Scientific Data*. – 2022. – Vol. 9, № 1. – P. 185. – DOI: 10.1038/s41597-022-01288-4.
139. Guliyev N.J., Ismailov V.E. On the approximation by single hidden layer feedforward neural networks with fixed weights // *Neural Networks*. – 2018. – Vol. 98. – P. 296–304. – DOI: 10.1016/j.neunet.2017.12.007.
140. Ganea O.-E., Pattanaik L., Coley C.W., Barzilay R., Jensen K.F., Green W.H., Jaakkola T.S. GEOMOL: torsional geometric generation of molecular 3D conformer ensembles // *Proceedings of the 35th International Conference on Neural Information Processing Systems*. – 2021. – Red Hook, NY, USA: Curran Associates Inc.
141. Kuznetsov M., Ryabov F., Schutski R., Shayakhmetov R., Lin Y.-C., Aliper A., Polykovskiy D. COSMIC: Molecular Conformation Space Modeling in Internal Coordinates with an Adversarial Framework // *Journal of Chemical Information and Modeling*. – 2024. – Vol. 64, № 9. – P. 3610–3620. – DOI: 10.1021/acs.jcim.3c00989.
142. Xu M., Wang W., Luo S., Shi C., Bengio Y., Gomez-Bombarelli R., Tang J. An End-to-End Framework for Molecular Conformation Generation via Bilevel Programming // *arXiv preprint*. – 2021. – № arXiv:2105.07246. – DOI: 10.48550/arXiv.2105.07246.
143. Mansimov E., Mahmood O., Kang S., Cho K. Molecular Geometry Prediction using a Deep Generative Graph Neural Network // *Scientific Reports*. – 2019. – Vol. 9, № 1. – P. 20381. – DOI: 10.1038/s41598-019-56773-5.
144. Zhu J., Xia Y., Liu C., Wu L., Xie S., Wang Y., Wang T., Qin T., Zhou W., Li H., Liu H., Liu T.-Y. Direct Molecular Conformation Generation // *arXiv preprint*. – 2022. – № arXiv:2202.01356. – DOI: 10.48550/arXiv.2202.01356.

145. Xu M., Yu L., Song Y., Shi C., Ermon S., Tang J. GeoDiff: a geometric diffusion model for molecular conformation generation // arXiv preprint. – 2022. – № arXiv:2203.02923. – DOI: 10.48550/arXiv.2203.02923.
146. Wang Y., Elhag A.A., Jaitly N., Susskind J.M., Bautista M.A. Swallowing the bitter pill: Simplified scalable conformer generation // arXiv preprint. – 2023. – № arXiv:2311.17932. – DOI: 10.48550/arXiv.2311.17932.
147. Jing B., Corso G., Chang J., Barzilay R., Jaakkola T. Torsional Diffusion for Molecular Conformer Generation // arXiv preprint. – 2023. – № arXiv:2206.01729. – DOI: 10.48550/arXiv.2206.01729.
148. Grambow C.A., Weir H., Diamant N.L., Scalia G., Biancalani T., Chuang K.V. Accurate and Efficient Structural Ensemble Generation of Macrocyclic Peptides using Internal Coordinate Diffusion // arXiv preprint. – 2024. – № arXiv:2305.19800. – DOI: 10.48550/arXiv.2305.19800.
149. Shah S.A., Koltun V. Conformation Generation using Transformer Flows // arXiv preprint. – 2025. – № arXiv:2411.10817. – DOI: 10.48550/arXiv.2411.10817.
150. Xu M., Luo S., Bengio Y., Peng J., Tang J. Learning Neural Generative Dynamics for Molecular Conformation Generation // arXiv preprint. – 2021. – № arXiv:2102.10240. – DOI: 10.48550/arXiv.2102.10240.
151. Shi C., Luo S., Xu M., Tang J. Learning Gradient Fields for Molecular Conformation Generation // arXiv preprint. – 2021. – № arXiv:2105.03902. – DOI: 10.48550/arXiv.2105.03902.
152. Zhang H., Li S., Zhang J., Wang Z., Wang J., Jiang D., Bian Z., Zhang Y., Deng Y., Song J., Kang Y., Hou T. SDEGen: learning to evolve molecular conformations from thermodynamic noise for conformation generation // *Chemical Science*. – 2023. – Vol. 14, № 6. – P. 1557–1568. – DOI: 10.1039/D2SC04429C.
153. Luo S., Shi C., Xu M., Tang J. Predicting Molecular Conformation via Dynamic Graph Score Matching // *Advances in Neural Information Processing Systems*. – 2021. – Vol. 34. – P. 19784–19795. – Curran Associates, Inc.
154. Huang H., Sun L., Du B., Lv W. Learning Joint 2D & 3D Diffusion Models for Complete Molecule Generation // arXiv preprint. – 2023. – № arXiv:2305.12347. – DOI: 10.48550/arXiv.2305.12347.
155. Keng M., Merz K.M.Jr. PEAS: An Application for Autonomous Precision Conformation Sampling // *Journal of Chemical Information and Modeling*. – 2025. – Vol. 65, № 22. – P. 12161–12167. – DOI: 10.1021/acs.jcim.5c02104.
156. Hashim H.M., Harvey J.N. Conformational Analysis of Macrocyclic Compounds Using a Machine-Learned Interatomic Potential // *Journal of Chemical Theory and Computation*. – 2025. – Vol. 21, № 20. – P. 10604–10616. – DOI: 10.1021/acs.jctc.5c01072.
157. Podryabinkin E., Garifullin K., Shapeev A., Novikov I. MLIP-3: Active learning on atomic environments with moment tensor potentials // *The Journal of Chemical Physics*. – 2023. – Vol. 159, № 8. – P. 084112. – DOI: 10.1063/5.0155887.
158. Pracht P., Grimme S., Bannwarth C., Bohle F., Ehlert S., Feldmann G., Gorges J., Müller M., Neudecker T., Plett C., Spicher S., Steinbach P., Wesolowski P.A., Zeller F. CREST—A program for the exploration of low-energy molecular chemical space // *The Journal of Chemical Physics*. – 2024. – Vol. 160, № 11. – P. 114110. – DOI: 10.1063/5.0197592.

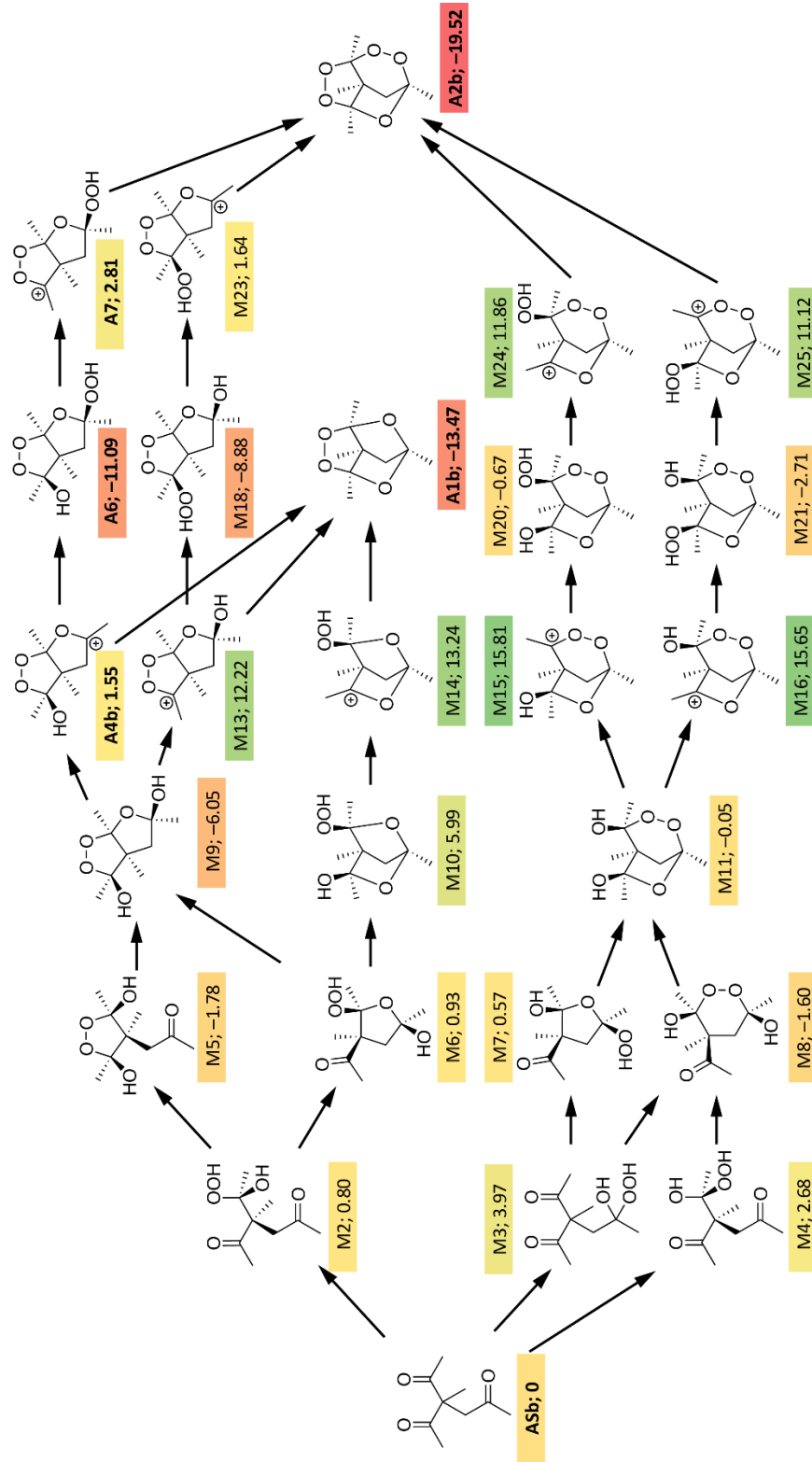
159. Hawkins P.C.D. Conformation Generation: The State of the Art // *Journal of Chemical Information and Modeling*. – 2017. – Vol. 57, № 8. – P. 1747–1756. – DOI: 10.1021/acs.jcim.7b00221.
160. Kabsch W. A discussion of the solution for the best rotation to relate two sets of vectors // *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography*. – 1978. – Vol. 34, № 5. – P. 827–828. – DOI: 10.1107/S0567739478001680.
161. Theobald D.L. Rapid calculation of RMSDs using a quaternion-based characteristic polynomial // *Acta Crystallographica Section A: Foundations of Crystallography*. – 2005. – Vol. 61, № 4. – P. 478–480. – DOI: 10.1107/S0108767305015266.
162. Liu P., Agrafiotis D.K., Theobald D.L. Fast determination of the optimal rotational matrix for macromolecular superpositions // *Journal of Computational Chemistry*. – 2010. – Vol. 31, № 7. – P. 1561–1563. – DOI: 10.1002/jcc.21439.
163. Horn B.K.P. Closed-form solution of absolute orientation using unit quaternions // *JOSA A*. – 1987. – Vol. 4, № 4. – P. 629–642. – DOI: 10.1364/JOSAA.4.000629.
164. Diamond R. A note on the rotational superposition problem // *Acta Crystallographica Section A: Foundations of Crystallography*. – 1988. – Vol. 44, № 2. – P. 211–216. – DOI: 10.1107/S0108767387010535.
165. Walker M.W., Shao L., Volz R.A. Estimating 3-D location parameters using dual number quaternions // *CVGIP: Image Understanding*. – 1991. – Vol. 54, № 3. – P. 358–367. – DOI: 10.1016/1049-9660(91)90036-O.
166. Coutsias E.A., Wester M.J. RMSD and Symmetry // *Journal of Computational Chemistry*. – 2019. – Vol. 40, № 15. – P. 1496–1508. – DOI: 10.1002/jcc.25802.
167. Meli R., Biggin P.C. spyrmsd: symmetry-corrected RMSD calculations in Python // *Journal of Cheminformatics*. – 2020. – Vol. 12, № 1. – P. 49. – DOI: 10.1186/s13321-020-00455-2.
168. Fukutani T., Miyazawa K., Iwata S., Satoh H. G-RMSD: Root Mean Square Deviation Based Method for Three-Dimensional Molecular Similarity Determination // *Bulletin of the Chemical Society of Japan*. – 2021. – Vol. 94, № 2. – P. 655–665. – DOI: 10.1246/bcsj.20200258.
169. Allen W.J., Rizzo R.C. Implementation of the Hungarian Algorithm to Account for Ligand Symmetry and Similarity in Structure-Based Design // *Journal of Chemical Information and Modeling*. – 2014. – Vol. 54, № 2. – P. 518–529. – DOI: 10.1021/ci400534h.
170. Bell E.W., Zhang Y. DockRMSD: an open-source tool for atom mapping and RMSD calculation of symmetric molecules through graph isomorphism // *Journal of Cheminformatics*. – 2019. – Vol. 11, № 1. – P. 40. – DOI: 10.1186/s13321-019-0362-7.
171. Kufareva I., Abagyan R. Methods of Protein Structure Comparison // *Homology Modeling: Methods and Protocols* / ed. Orry A.J.W., Abagyan R. – 2012. – P. 231–257. – DOI: 10.1007/978-1-61779-588-6_10.
172. Schulz-Gasch T., Schärfer C., Guba W., Rarey M. TFD: Torsion Fingerprints As a New Measure To Compare Small Molecule Conformations // *Journal of Chemical Information and Modeling*. – 2012. – Vol. 52, № 6. – P. 1499–1512. – DOI: 10.1021/ci2002318.
173. Yaremenko I.A., Radulov P.S., Medvedev M.G., Krivoshchapov N.V., Belyakova Y.Yu., Korlyukov A.A., Ilovaisky A.I., Terent'ev A.O., Alabugin I.V. How to Build

- Rigid Oxygen-Rich Tricyclic Heterocycles from Triketones and Hydrogen Peroxide: Control of Dynamic Covalent Chemistry with Inverse α -Effect // *Journal of the American Chemical Society*. – 2020. – Vol. 142, № 34. – P. 14588–14607. – DOI: 10.1021/jacs.0c06294.
174. Yaremenko I.A., Belyakova Y.Yu., Radulov P.S., Novikov R.A., Medvedev M.G., Krivoshchapov N.V., Korlyukov A.A., Alabugin I.V., Terent'ev A.O. Marriage of Peroxides and Nitrogen Heterocycles: Selective Three-Component Assembly, Peroxide-Preserving Rearrangement, and Stereoelectronic Source of Unusual Stability of Bridged Azaazonides // *Journal of the American Chemical Society*. – 2021. – Vol. 143, № 17. – P. 6634–6648. – DOI: 10.1021/jacs.1c02249.
175. Yaremenko I.A., Belyakova Y.Yu., Radulov P.S., Novikov R.A., Medvedev M.G., Krivoshchapov N.V., Korlyukov A.A., Alabugin I.V., Terent'ev A.O. Inverse α -Effect as the Ariadne's Thread on the Way to Tricyclic Aminoperoxides: Avoiding Thermodynamic Traps in the Labyrinth of Possibilities // *Journal of the American Chemical Society*. – 2022. – Vol. 144, № 16. – P. 7264–7282. – DOI: 10.1021/jacs.2c00406.
176. Yaremenko I.A., Belyakova Y.Yu., Radulov P.S., Medvedev M.G., Krivoshchapov N.V., Alabugin I.V., Terent'ev A.O. Two-Component versus Three-Component Condensations in the Race between Hydrazide, Triketone, and Hydrogen Peroxide—How do All Six Reactive Centers Cooperate to Incorporate the Most Diverse Set of Heteroatomic Bridges in a Tricyclic Frame? // *The Journal of Organic Chemistry*. – 2023. – Vol. 88, № 19. – P. 13782–13795. – DOI: 10.1021/acs.joc.3c01415.
177. Alabugin I.V., Kuhn L., Krivoshchapov N.V., Mehaffy P., Medvedev M.G. Anomeric effect, hyperconjugation and electrostatics: lessons from complexity in a classic stereoelectronic phenomenon // *Chemical Society Reviews*. – 2021. – Vol. 50, № 18. – P. 10212–10252. – DOI: 10.1039/D1CS00564B.
178. Alabugin I.V., Kuhn L., Medvedev M.G., Krivoshchapov N.V., Vil' V.A., Yaremenko I.A., Mehaffy P., Yarie M., Terent'ev A.O., Zolfigol M.A. Stereoelectronic power of oxygen in control of chemical reactivity: the anomeric effect is not alone // *Chemical Society Reviews*. – 2021. – Vol. 50, № 18. – P. 10253–10345. – DOI: 10.1039/D1CS00386K.
179. 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. – Vol. 355, № 6320. – P. 49–52. – DOI: 10.1126/science.aah5975.
180. Stewart J.J.P. Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements // *Journal of Molecular Modeling*. – 2007. – Vol. 13, № 12. – P. 1173–1213. – DOI: 10.1007/s00894-007-0233-4.
181. Kütt A., Leito I., Kaljurand I., Sooväli L., Vlasov V.M., Yagupolskii L.M., Koppel I.A. A Comprehensive Self-Consistent Spectrophotometric Acidity Scale of Neutral Brønsted Acids in Acetonitrile // *The Journal of Organic Chemistry*. – 2006. – Vol. 71, № 7. – P. 2829–2838. – DOI: 10.1021/jo060031y.
182. Eckert F., Leito I., Kaljurand I., Kütt A., Klamt A., Diedenhofen M. Prediction of acidity in acetonitrile solution with COSMO-RS // *Journal of Computational Chemistry*. – 2009. – Vol. 30, № 5. – P. 799–810. – DOI: 10.1002/jcc.21103.

183. Petzold L. Automatic Selection of Methods for Solving Stiff and Nonstiff Systems of Ordinary Differential Equations // *SIAM Journal on Scientific and Statistical Computing*. – 1983. – Vol. 4, № 1. – P. 136–148. – DOI: 10.1137/0904010.
184. Virtanen P., Gommers R., Oliphant T.E., Haberland M., Reddy T., Cournapeau D., Burovski E., Peterson P., Weckesser W., Bright J., van der Walt S.J., Brett M., Wilson J., Millman K.J., Mayorov N., Nelson A.R.J., Jones E., Kern R., Larson E., Carey C.J., Polat İ., Feng Y., Moore E.W., VanderPlas J., Laxalde D., Perktold J., Cimrman R., Henriksen I., Quintero E.A., Harris C.R., Archibald A.M., Ribeiro A.H., Pedregosa F., van Mulbregt P. SciPy 1.0: fundamental algorithms for scientific computing in Python // *Nature Methods*. – 2020. – Vol. 17, № 3. – P. 261–272. – DOI: 10.1038/s41592-019-0686-2.
185. Clayden J., Greeves N., Warren S. *Organic chemistry*. – 2012. – Oxford university press.
186. International Union of Pure and Applied Chemistry (IUPAC). hyperconjugation. – DOI: 10.1351/goldbook.H02924.
187. International Union of Pure and Applied Chemistry (IUPAC). n- σ^* delocalization. – DOI: 10.1351/goldbook.N04080.
188. Wiberg K.B., Bailey W.F., Lambert K.M., Stempel Z.D. The Anomeric Effect: It's Complicated // *The Journal of Organic Chemistry*. – 2018. – Vol. 83, № 9. – P. 5242–5255. – DOI: 10.1021/acs.joc.8b00707.
189. Vil' V.A., dos Passos Gomes G., Bityukov O.V., Lyssenko K.A., Nikishin G.I., Alabugin I.V., Terent'ev A.O. Interrupted Baeyer–Villiger Rearrangement: Building A Stereoelectronic Trap for the Criegee Intermediate // *Angewandte Chemie International Edition*. – 2018. – Vol. 57, № 13. – P. 3372–3376. – DOI: 10.1002/anie.201712651.
190. Weinhold F., Landis C.R., Glendening E.D. What is NBO analysis and how is it useful? // *International Reviews in Physical Chemistry*. – 2016. – Vol. 35, № 3. – P. 399–440. – DOI: 10.1080/0144235X.2016.1192262.
191. Ponomarev D.A., Takhistov V.V. What are Isodesmic Reactions? // *Journal of Chemical Education*. – 1997. – Vol. 74, № 2. – P. 201. – DOI: 10.1021/ed074p201.
192. Fan X.-W., Ju X.-H., Xia Q.-Y., Xiao H.-M. Strain energies of cubane derivatives with different substituent groups // *Journal of Hazardous Materials*. – 2008. – Vol. 151, № 1. – P. 255–260. – DOI: 10.1016/j.jhazmat.2007.05.075.
193. McLean A.D., Chandler G.S. Contracted Gaussian basis sets for molecular calculations. I. Second row atoms, Z=11–18 // *The Journal of Chemical Physics*. – 1980. – Vol. 72, № 10. – P. 5639–5648. – DOI: 10.1063/1.438980.
194. Spitznagel G.W., Clark T., von Ragué Schleyer P., Hehre W.J. An evaluation of the performance of diffuse function-augmented basis sets for second row elements, Na-Cl // *Journal of Computational Chemistry*. – 1987. – Vol. 8, № 8. – P. 1109–1116. – DOI: 10.1002/jcc.540080807.
195. Wolinski K., Hinton J.F., Pulay P. Efficient implementation of the gauge-independent atomic orbital method for NMR chemical shift calculations // *Journal of the American Chemical Society*. – 1990. – Vol. 112, № 23. – P. 8251–8260. – DOI: 10.1021/ja00179a005.
196. Kirby A.J., Komarov I.V., Wothers P.D., Feeder N. The Most Twisted Amide: Structure and Reactions // *Angewandte Chemie International Edition*. – 1998. – Vol. 37, № 6. – P. 785–786. – DOI: 10.1002/(SICI)1521-3773(19980403)37:6%3C785::AID-ANIE785%3E3.0.CO;2-J.

197. Krivoshchapov N.V., Medvedev M.G. Ring kinematics-informed conformation space exploration // WIREs Computational Molecular Science. – 2024. – Vol. 14, № 1. – P. e1690. – DOI: 10.1002/wcms.1690.
198. Krivoshchapov N.V., Medvedev M.G. Accurate and Efficient Conformer Sampling of Cyclic Drug-Like Molecules with Inverse Kinematics // Journal of Chemical Information and Modeling. – 2024. – Vol. 64, № 11. – P. 4542–4552. – DOI: 10.1021/acs.jcim.3c02040.
199. Denavit J., Hartenberg R.S. A Kinematic Notation for Lower-Pair Mechanisms Based on Matrices // Journal of Applied Mechanics. – 1955. – Vol. 22, № 2. – P. 215–221. – DOI: 10.1115/1.4011045.
200. Kühnel W. Differential Geometry: Curves - Surfaces - Manifolds. – 2006. – 394 p. – American Mathematical Soc.
201. Cao X., Coutsias E.A., Pollock S. Numerically stable solution to the 6R problem of Inverse Kinematics // Advances in Computational Science and Engineering. – 2023. – Vol. 1, № 2. – P. 123–161. – DOI: 10.3934/acse.2023006.
202. Kucuk S., Bingul Z. Inverse kinematics solutions for industrial robot manipulators with offset wrists // Applied Mathematical Modelling. – 2014. – Vol. 38, № 7. – P. 1983–1999. – DOI: 10.1016/j.apm.2013.10.014.

Приложение 1. Результаты квантово-химических расчётов относительных энергий предполагаемых интермедиатов вдоль нескольких предложенных путей сборки пероксидов **A1b** и **A2b** (см. подраздел 2.1).



Приложение 2. Результаты квантово-химических расчётов относительных энергий предполагаемых интермедиатов и ПС вдоль нескольких предложенных путей сборки азаозонидов **D1a** и **D2a** (см. подраздел 2.4).

