

Моделирование процессов в биомолекулярных системах методами квантовой и молекулярной механики (КМ/ММ)*

А.В. Немухин^{1,2}, Б.Л. Григоренко¹, Д.И. Морозов¹

Химический факультет МГУ им. М.В. Ломоносова¹, Институт биохимической физики им. Н.М. Эмануэля РАН²

Обсуждаются алгоритмы и программные реализации комбинированного метода квантовой и молекулярной механики (КМ/ММ), активно используемого для моделирования процессов в биомолекулярных системах. Основное внимание уделено расчетам строения и оптических спектров перспективных флуоресцирующих белков, а также расчетам энергетических профилей реакций ферментативного катализа. Анализируется опыт расчетов КМ/ММ на суперкомпьютерах СКИФ-Полигон.

1. Введение

Метод моделирования механизмов реакций ферментативного катализа на основе комбинированного подхода квантовой и молекулярной механики (КМ/ММ), предложенного в 1976 г. [1], быстро завоевал популярность. Впоследствии подход КМ/ММ стал применяться для расчетов структуры и свойств других биомолекулярных процессов, включая прогнозирование спектров фоторецепторных и флуоресцирующих белков. Согласно подходу КМ/ММ та часть белковой системы, в которой происходят перераспределения химических связей или рассматриваются электронные возбуждения, включается в квантовую подсистему, и энергии и силы в ней рассчитываются различными приближениями квантовой химии. Большинство атомов, окружающую выделенную центральную часть, относится к молекулярно-механической подсистеме. Таким образом, в методе КМ/ММ учитывают конформационные изменения в белковой матрице, сопровождающие химическую реакцию или переходы между электронными состояниями. В приближениях КМ/ММ энергия каждой точки на поверхности потенциальной энергии (ППЭ) складывается из энергии квантовой части в поле ММ-подсистемы и молекулярно-механической энергии [2,3].

Анализ стационарных точек на ППЭ, т.е. геометрических конфигураций локальных минимумов, отвечающих реагентам, продуктам, возможным интермедиатам, а также седловых точек, через которые происходят переходы между минимумами составляет предмет исследования механизма реакции ферментативного катализа. Особое внимание уделяется расчетам энергетических барьеров на пути от реагентов к продуктам. Переход к экспериментально значимым величинам, таким как константы скорости химических реакций, осуществляется, как правило, в рамках теории переходных состояний.

При расчетах структуры и спектров фоторецепторных белков основную ценность представляет прогнозирование новых линеек биомаркеров в живых системах с улучшенными свойствами (высоким квантовым выходом флуоресценции, повышенной стабильностью, нетоксичностью и проч.) за счет рационального введения точечных мутаций, т.е. замен отдельных аминокислотных остатков в белковой матрице. Перебор огромного числа вариантов чисто экспериментальными средствами потребует слишком много времени и больших материальных затрат. Компьютерное моделирование с использованием современных квантовых методов расчетов структуры и спектров биологических хромофоров в белковых матрицах оказывает существенную поддержку подобным исследованиям. Вычислительные задачи связаны с решением уравнений КМ/ММ для большого фрагмента белковой частицы, содержащей хромофорную молекулу и ближайшие аминокислотные остатки с целью определить координаты всех атомов (общим числом до нескольких тысяч), соответствующие минимумам на поверхностях потенциальной энергии основного и электронно-возбужденных состояний, а также как можно точнее рассчитать разности энергий между различными электронными состояниями модельной моле-

* Работа выполнена при поддержке РФФИ (проекты 10-03-00085 и 10-03-00139)

кулярной системы. Даже для одной композиции подобные расчеты крайне ресурсозатратны, но их необходимо выполнить для каждого потенциально перспективного варианта цветного белка [4].

2. Протокол вычислений для моделирования цветных белков

Опишем подробнее стратегию и протокол вычислений методом КМ/ММ на примере прогнозирования оптимального состава флуоресцирующих белков.

(1) Для выбранной композиции белковой макромолекулы найти равновесную геометрическую конфигурацию на поверхности потенциальной энергии основного состояния в рамках метода КМ/ММ с использованием алгоритмов теории функционала электронной плотности в квантовой подсистеме.

(2) Методами квантовой механики на основе многоконfigurационного метода самосоглавленного поля (МК ССП) рассчитать вертикальные энергии возбуждений из найденной равновесной точки и оценить интенсивности соответствующих полос в оптическом спектре.

(3) Для выбранного интенсивного электронного перехода сканировать сечения поверхностей потенциальной энергии возбужденного электронного состояния, найти геометрические конфигурации точек минимумов и конических пересечений и соответствующие разности энергий, используя варианты метода МК ССП.

(4) Сформулировать заключение о перспективности данной композиции.

Шаги (1-4) повторяются для каждого варианта белковой макромолекулы с точечными мутациями.

Самыми ресурсозатратными процедурами являются расчеты методом МК ССП. Алгоритм МК ССП включает в себя следующие циклически повторяющиеся шаги:

(1) приведение интегралов из атомного базиса в текущий молекулярный;

(2) генерацию матрицы гамильтониана и оптимизацию коэффициентов многоконfigurационного разложения при частичной давидсоновской диагонализации;

(3) генерацию матрицы плотности первого и второго порядков;

(4) внесение улучшений в молекулярные орбитали, проверка критериев сходимости. При использовании небольших активных пространств молекулярных орбиталей, либо при использовании одного из алгоритмов МК ССП (ORMAS) лимитирующей расчёт стадией является стадия преобразования интегралов (1).

Данные по масштабируемости подобных расчетов на суперкомпьютере СКИФ МГУ «Чебышёв» приведены в Таблице 1.

Таблица 1. Масштабируемость преобразования интегралов при расчетах МК ССП

Количество процессоров	Время, мин
8	32.8
16	25.5
24	12.4
32	7.2
64	3.2

Лучшая масштабируемость достигается при больших размерах квантовой системы в силу большего количества молекулярных орбиталей. Кроме того, для больших систем в пакете квантовохимических программ GAMESS(US), активно используемом для расчетов, существует специальный алгоритм преобразования, использующий преимущества модели распределённой памяти DDI, таким образом улучшая масштабируемость.

Для использования в процессе МКССП в программном пакете GAMESS(US) присутствуют три вида прямого детерминантного варианта конфигурационных разложений: (1) полное конфигурационное взаимодействие (КВ) в активном пространстве (ALDET, Ames Laboratory Determinant Full CI); (2) полное КВ во множественных активных пространствах, с ограничениями по заселенности Occupation Restricted Multiple Active Spaces (ORMAS), используемое обычно для ограничения возбуждения и, тем самым, упрощения задачи полного КВ; (3) метод КВ с выбранным списком детерминантов (GENCI) [5,6].

ALDET: Полное КВ в выбранном активном пространстве. Основным ограничением для данного метода является количество памяти, доступной программе. В реальных условиях на суперкомпьютере СКИФ-«Чебышёв» при использовании 4 узлов (32 процессорных ядра) доступно порядка 128 Гб оперативной памяти, что ограничивает размеры доступных активных пространств до 18 электронов на 16 орбиталях. Данный код является наиболее масштабируемым кодом метода конфигурационного взаимодействия (КВ), поскольку активно использует модель распределённой памяти DDI для хранения векторов КВ [5].

ORMAS: Метод, используемый для ограничения размера задачи полного КВ. Используется, когда задача полного КВ становится слишком большой (>1000000 детерминантов), а также если необходимо ограничить или выделить только нужные переходы электронов между орбиталями. Разбиение большого активного пространства на несколько частей, с заданием возможных границ заполнения этих подпространств электронами, приводит к значительному уменьшению задачи полного КВ. Тем не менее, остаётся возможность сохранить многие эффекты полного КВ, поскольку в каждом из маленьких подпространств решается задача полного КВ. Являясь очень гибким вариантом метода КВ, ORMAS позволяет использовать достаточно большие пространства (например 20 на 16 орбиталях и более) при относительно небольших затратах памяти и практически без затрат дискового пространства (полностью прямой метод), при его использовании, обычно, лимитирующей стадией расчёта становится первый шаг процедуры МКССП – преобразование интегралов. ORMAS, благодаря использованию распределённой памяти, также как и ALDET обладает хорошей масштабируемостью [6].

При применении метода ORMAS для моделирования реакций с переносом электрона в зеленом флуоресцентном белке модельная система состояла из 81 атома (соответственно включала 805 молекулярных орбиталей). Полное пространство, необходимое для моделирования реакции переноса электрона с аниона глутаминовой кислоты Glu222 на хромофор зеленого флуоресцирующего белка составляет 18 электронов на 17 орбиталях, при этом количество детерминантов составляет порядка 25 миллионов. Задача становится технически невозможной на текущих системах при использовании метода ALDET. Поэтому использовался метод ORMAS со следующими параметрами: большое пространство разбивалось на 3 подпространства (занятые орбитали аниона глутаминовой кислоты, занятые орбитали хромофора, вакантные орбитали), переходы между подпространствами ограничивались до 1 электрона с занятых орбиталей глутамата на вакантные и полной заморозкой занятых орбиталей хромофора. Ограничение вводилось для того чтобы самое низшее возбуждённое соответствовало переходу с переносом заряда с аниона на хромофор.

3. Расчеты реакций ферментативного катализа

В настоящее время метод КМ/ММ на основе конформационно-подвижных эффективных фрагментов может быть применен для расчетов энергетических профилей реакционного пути (сечений поверхностей потенциальной энергии) основного электронного состояния с использованием наиболее востребованных квантово-химических подходов – приближений теории функционала электронной плотности, многоконфигурационных приближений самосогласованного поля, теории возмущений Мёллера-Плессе, в КМ-подсистеме. Число атомов, относимых к КМ-части, может превышать 100; более важным параметром является число базисных функций для представления молекулярных орбиталей – не более 1500. ММ-подсистема может содержать до 3000 атомов, сгруппированных не более чем в 700 эффективных фрагментов. Для расчетов вкладов в энергию и градиенты энергии по ядерным координатам в ММ-части могут быть использованы параметры популярных силовых полей AMBER, CHARMM, OPLSAA и других. Приведенные ограничения на параметры вычислительной схемы не являются принципиальными и могут быть изменены в следующих версиях программы [7].

Нами рассматривалась молекулярная система, моделирующая фермент-субстратный комплекс пенициллин-ацилазы (ПА) с пенициллином G. Исходная информация о координатах тяжелых атомов была получена из базы данных белковых структур (PDB) для кристаллической структуры фермента ПА с ингибирующим субстратом SOX (пенициллин G сульфоксид) – код PDB:1GM9. Далее методами компьютерного моделирования ингибирующий субстрат был заменен на пенициллин G, добавлены атомы водорода и проведены расчеты равновесных геомет-

рических конфигураций фермент-субстратного комплекса методом КМ/ММ с полной оптимизацией координат в субстрат-связывающей области. ММ-подсистема включала 2400 атомов, сгруппированных в 670 эффективных фрагментов. Расчеты энергий и сил в КМ-части проводились в приближении В3LYP/6-31G*, в ММ-части – с использованием параметров силового поля AMBER. Тщательное сопоставление равновесных геометрических параметров (расстояний между тяжелыми атомами, валентных углов), рассчитанных методом КМ/ММ, и измеренных в соответствующей кристаллической структуре показывает хорошее соответствие теоретических и экспериментальных величин.

Как для всех сериновых гидролаз, первая стадия реакции гидролиза пенициллина ферментом ПА (стадия ацилирования) проходит через образование первого тетраэдрического интермедиата. Особенности реакционного механизма для данного фермента и данного субстрата с приведением структур интермедиатов будут описаны в отдельной работе. Здесь мы обращаем внимание на сопоставление результатов двух методик КМ/ММ: приближении внедренного кластера и приближении конформационно-подвижных эффективных фрагментов. Следует отметить, рассчитанные в приближении В3LYP/6-31G*/AMBER геометрические параметры и для фермент-субстратного комплекса, и для тетраэдрического интермедиата в обоих подходах достаточно близки. Однако, разность энергий, вычисленная в подходе внедренного кластера, +5 ккал/моль, явно не выглядит правдоподобной. Пересчет разности энергий в варианте КМ/ММ с конформационно-подвижными фрагментами (-17 ккал/моль) позволяет получить ожидаемый выигрыш энергии на первом сегменте реакции.

4. Резюме

Таким образом, известны алгоритмы, позволяющие проводить масштабные расчеты комбинированным методом КМ/ММ как для реакций ферментативного катализа, так и для моделирования структуры и спектров фоторецепторных белков.

Литература

1. A. Warshel, M. Levitt. Theoretical studies of enzymatic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. // *J. Mol. Biol.* 1976. V. 103. P. 227-249.
2. F. Bernardi, M. Olivucci, M. A. Robb. Simulation of MC-SCF results on covalent organic multi-bond reactions: molecular mechanics with valence-bond (MM-VB). // *J. Am. Chem. Soc.* 1992. V. 114. P. 1606-1616.
3. J. Aqvist, A. Warshel. Simulation of enzyme reactions using valence bond force fields and other hybrid quantum-classical approaches. // *Chem. Rev.* 1993. V. 93. P. 2523-2544.
4. M. Matz, K. Lukyanov and S. Lukyanov. Family of the green fluorescent protein: journey to the end of the rainbow. *Bioessays*, 2002, V. 24(10). P. 953-962
5. J.Ivanic, K.Ruedenberg. Full CI (ALDET) and general CI (GENCI) // *Theoret.Chem.Acc.*, 2001, V. 106, P. 339-351
6. J.Ivanic. Occupation restricted multiple active space (ORMAS). // *J.Chem.Phys.*, 2003, V. 119, P. 9364-9385
7. A. Nemukhin, B. Grigorenko, A. Bochenkova. A QM/MM approach with effective fragment potentials applied to the dipeptide-water structures. // *J.Molec.Struct.(Theochem)*, 2002, V. 581, p.167