

Лекция шестая «Методы молекулярной динамики конформационного анализа биополимеров»

Лектор д.ф.-м.н., проф. Ю.Н. Воробьев.

Метод молекулярной динамики и конформационный анализ биополимеров будут рассмотрены с точки зрения физика. Макромолекула, как физическая система, это система взаимодействующих частиц или статистическая система. Экспериментально доказано, что молекулы белков обратимо переходят из уникального, для каждого белка, нативного стабильного структурного состояния в другое или другие неупорядоченные состояния с неупорядоченной структурой. Проявляется прямая аналогия с плавлением молекулярного кристалла. Статистические равновесные свойства системы частиц определяются поверхностью потенциальной энергии этой системы в фазовом пространстве координат частиц. Минимумы на ППЭ отвечают метастабильным конфигурациям или конформациям – конформации при условии сохранения ковалентной химической структуры молекулы. Глобальный (самый глубокий минимум) соответствует стабильной нативной структуре. Пути переходов между метастабильными конформациями, это пути конформационных перестроек. Таким образом конформационный анализ это изучение свойств ППЭ.

На этом рисунке схематически изображена ППЭ белка лизоцима. Ордината F это свободная энергия белка в зависимости от двух обобщенных координат реакции самосборки – числа нативных контактов между остатками, принадлежащими двум доменам  и . Видно, что самосборка (фолдинг) белка может происходить по разным путям, быстрым и более медленным, включая попадание интермедиатов в метастабильные ловушки. В настоящее время доказано, что существует специальный клеточный механизм, с участием огромных белковых тел, молекул шаперонов, которые связывают метастабильные интермедиаты, вынуждают обратный фолдинг, и предоставляют молекуле новую попытку самосборки и новых начальных условий.

Если старая физика 19-го и первой половины 20-го века развивалась как аналитическая теория, т.е. явления описывались набором законов и уравнений, которые можно было решить и получить интерпретацию явлений, то начиная со второй половины 20-го века этого перестало хватать для потребностей науки и технологии. Уравнения математической физики для сложных граничных условий стало возможным решать только численно на компьютере. Более того, для ряда физических явлений нет возможности составить достаточно точные системы уравнений, т.е. Аналитические теории. Например, теория конденсированных жидкостей. Для простых жидкостей типа инертных газов, удалось получить удовлетворительные аналитические уравнения для корреляционных функций, однако, для практически важных систем, молекулярных жидкостей и особенно воды, не существует удовлетворительных аналитических статистико-механических теорий. На помощь теоретикам аналитикам, приходит компьютерный эксперимент, который моделирует свойства вещества путем симбиоза моделей взаимодействия между частицами и прямым расчетом статистических ансамблей наиболее вероятных состояний системы. Роль КЭ в современной науке огромна, она срединна между аналитической теорией и экспериментом и позволяет получать данные о системе недоступные для экспериментальных измерений. Особенно плодотворна роль КЭ в исследовании свойств биополимеров в растворе.

Итак основная задача КЭ состоит в генерации наиболее вероятных конфигураций частиц. Возможны различные подходы: тотальное сканирование фазового пространства для малых

молекул; Для макромолекул число конформаций астрономически огромно, необходимы более рациональные методы. Метод Монте-Карло генерирует марковскую цепочку конфигураций системы, переход из одного состояния в другое происходит согласно принципу детального равновесия по критерию Метрополиса, так что изображающие точки распределены в фазовом пространстве пропорционально Больцмановскому фактору от потенциальной энергии системы в этом состоянии. Известны методы псевдо-глобальной оптимизации, которые в состоянии найти один из глубоких минимумов, однако не гарантируют его глобальности. Наиболее эффективным оказывается метод молекулярной динамики, который дает прямое моделирование термического движения частиц системы в собственном поле сил. Информация получаемая из метода МД содержит серию микросостояний системы в фазовом пространстве координат и скоростей частиц, т.е. Полную статистическую информацию, обработка которой позволяет получить термодинамические и динамические характеристики системы.

Принципы МД достаточно просты.

Атомы биополимера и молекул растворителя рассматриваются как система взаимодействующих частиц, движущихся под влиянием сил, действующих со стороны всех частиц системы. Рассматриваются классические уравнения движения для каждого атома в поле сил остальных частиц системы, которые решаются численно для всех атомов и получаются \Rightarrow траектории атомов $\{r_i(t)\}$ от времени, это основная первичная информация метода МД. Статистико-механическая обработка этой информации от МД траектории позволяет рассчитать различные свойства системы. Например:

- Конформационная динамика –
 - основные МОДЫ существенных конформационных движений
 - равновесные величины конформационных параметров
 - термические флуктуации конформационных параметров
- Термодинамические свойства –
 - энергии конформаций
 - энергии связывания лиганда с макромолекулой – drug design

История метода МД могла начаться только с появлением достаточных компьютерных средств. Возможности и области применения метода МД постоянно расширяются от простейшей системы твердых шаров, кристаллических систем с малым числом независимых частиц, до молекул биополимеров в водном растворе в присутствии ионов соли. Эти системы содержат порядка сотни тысяч независимых частиц - атомов. Применение метода МД для исследования конформационной динамики биополимеров и конструирования лекарственных препаратов, в том числе и дизайн новых молекул белков является наиболее перспективным.

Метод МД имеет ограничения связанные с областью применимости физических моделей, лежащих в его основе.

Применение классических уравнений движения корректно только в условиях когда длина волны Де Бройля много меньше размеров атомом и характерных межатомных расстояний. Это условие соблюдается для всех атомов органических веществ при комнатной температуре. Более существенными являются размерные эффекты, поскольку невозможно по техническим причинам симулировать слишком большую систему частиц. С ростом размерности системы растет множественность числа состояний, времена релаксации неравновесных параметров. Все это требует больших времен МД траекторий. Реалистичность моделей взаимодействия и расчетов траекторий. Преодоление этих ограничений потребовало значительных алгоритмических разработок. Так что реалистичные пакеты МД алгоритмов первого поколения были разработаны к середине 90-х годов.

Поэтому массированное применение метода МД для решения задач молекулярной биологии практически только начинается.

Потенциалы взаимодействия между атомами биополимера и растворителя определяют силовое поле. Современные молекулярно механические поля основаны на минималистской философии. Представлены только главные члены различных типов взаимодействий. Потенциальная энергия биополимера состоит из двух слагаемых: деформации валентного остова и взаимодействий между валентно не связанными атомами. В свою очередь энергия деформации валентного остова состоит из деформации длин валентных связей, валентных углов и вращений вокруг валентных связей – это первые три члена в формуле для энергии. Эти деформации аппроксимированы в гармоническом приближении и в виде главных членов Фурье ряда. Взаимодействия валентно не связанных атомов, т.е. удаленных по цепи аппроксимировано ван дер Ваальсовыми взаимодействиями, аналогично взаимодействию атомов инертных газов. Электростатические взаимодействия аппроксимированы монополярным приближением, как взаимодействие точечных зарядов, центрированных на атомах. Совокупность констант деформации валентных связей и углов, величины барьеров торсионных потенциалов, параметров ван дер Ваальсовых взаимодействий и зарядов на атомах составляют параметры силового поля. Эти параметры определяются на основе совокупности термодинамических и спектроскопических экспериментальных данных для молекулярных кристаллов и жидкостей. А также и квантово-химических расчетов на большом наборе малых молекул, включающих всевозможные химические группы, характерные для биополимеров. Применение силового поля, полученного из свойств малых молекул, для биополимеров, основано на гипотезе трансферабельности. Т.е. Подобные химические группы в различных молекулах имеют одинаковые взаимодействия. Современные силовые поля второго поколения:

- AMBER, OPLS, GROMOS, CHARMM

живой клетке макромолекулы находятся в растворе воды в присутствии ионов солей. Поэтому симуляция конформационно-динамических характеристик также должна выполняться в среде. Моделирование раствора представляем определенные трудности. Если взять большую каплю, то она не может содержать слишком большое число молекул воды, малая капля не в состоянии моделировать условия в растворе по причине больших пограничных эффектов. Аппроксимация неупорядоченного раствора как периодической системы, с повторяющейся единицей – основным блоком, позволяет рассматривать относительно небольшое число молекул раствора и минимизировать пограничные эффекты. Однако и в методе периодических граничных условий необходим учет множества молекул воды – слой в 8-10 ангстрем. В среднем оказывается, что молекулы воды составляют до 70% атомов системы биополимер-раствор.

Реалистичное описание структуры раствора и энергетики взаимодействий раствор – биополимер чрезвычайно важно.

Для этих целей создано несколько моделей силового поля для молекул воды, которые удовлетворительно воспроизводят экспериментальные структурные и термодинамические характеристики жидкой воды: тетраэдрический ближний порядок, энергию испарения, коэффициент самолиффузии.

Типичный протокол молекулярно-динамического численного эксперимента приведен на рис.11, и заключается в серии последовательных этапов. После построения исходной модели структуры и ее предварительной релаксации методами локальной оптимизации следует медленный разогрев системы, т.е. Назначение начальных скоростей атомам системы в

соответствии с Максвелловским распределением при заданной температуре с последующими короткими МД симуляциями. При достижении требуемой температуры, проводится более длинная МД симуляция до достижения равновесия при заданных температуре и давлении. После этого выполняется длинная МД симуляция для результативного расчета конформационно-динамических характеристик.

Симуляция биополимера в растворе возможна в различных термодинамических ансамблях.

NEV – это симуляция с постоянным числом частиц, при постоянном объеме и полной энергии, т.е. изолированный микроканонический ансамбль.

Контроль температуры системы – средней кинетической энергии атомов, позволяет вести симуляцию при постоянной температуре, а контроль температуры и давления соответствующим масштабированием скоростей частиц и объема главного бокса, воспроизводит фазовые траектории биополимера в растворе при постоянных температуре и давлении, например комнатных, то что имеется в реальном молекулярно-биологическом эксперименте.

Моделирование конформационной динамики не возможно без расчета реалистичных траекторий движения атомов системы. В микроканоническом ансамбле полная энергия симулируемой системы должна сохраняться, т.е. быть постоянной во времени. Сохранение полной энергии накладывает ограничение на максимальный шаг интегрирования уравнений движения ~ 1 фемтосек. Такой малый шаг интегрирования делает расчет содержательной МД траектории вычислительно трудоемкой задачей. Для ускорения интегрирования уравнений движения применяются многшаговые схемы интегрирования, которые основаны на разделении сил действующих на частицу на быстро и медленно меняющиеся. Например деформация ковалентного остова биополимера под быстро изменяющиеся силы, тогда как электростатические взаимодействия с удаленными атомами изменяются более медленно. Быстрые и медленные силы можно пересчитывать с разными частотами и таким образом уменьшить операционную сложность вычислений.

Вторым моментом определяющим содержательные возможности метода МД является мощность вычислителей, скорость процессоров, емкость оперативной и дисковой памяти, возможность вести параллельный расчет конфигураций. Минимальная содержательная МД траектория должна продолжаться несколько времен релаксации системы. ~ 100 псек. Наносекундные времена МД симуляции биополимеров в растворе, систем из нескольких десятков тысяч атомов, являются рутинными в настоящее время в том числе на персональном компьютере с процессорами 1 ГГц и более. Однако, наносекундные траектории являются недостаточными на три порядка для моделирования конформационных переключений и на шесть порядков для моделирования самоорганизации молекулы белка в 100-150 остатков. Активное применение многпроцессорных систем и создание эффективных параллельных алгоритмов метода МД позволят, согласно прогнозам, в ближайшие 20 лет выйти на времена симуляции в 1 сек. Это будет представлять качественный скачок в информационной полезности методов МД в приложении к биологическим системам.

Из стат. Механики известно, что усреднение по ансамблю состояний системы эквивалентно усреднению вдоль фазовой траектории, т.е. по состояниям, которые вычисляются методом МД. Поэтому расчитав эволюцию системы как серию мгновенных конфигураций системы, легко получить средние значения некоторого свойства, зависящего от конфигурации атомов. Средние квадратичные флуктуации свойства A также могут быть легко вычислены. Например величина термических флуктуаций конформационных параметров определяют жесткость соответствующих деформаций. Флуктуации потенциальной энергии определяют теплоемкость системы при постоянной температуре. Метод МД позволяет также расчитать

плотность состояний (конформационных) системы и соответственно относительные свободные энергии конформаций системы. Поскольку метод МД генерирует вероятные состояния системы, упорядоченные во времени, то это дает возможность рассчитать временные корреляционные функции, временные Фурье трансформации, которых дают спектральные частотные характеристики колебаний системы.

Алгоритм расчета относительной конформационной свободной энергии состояний приведен на этом рисунке. Основная рассчитываемая величина, в этом случае, это плотность состояний (конформационных). Однако статистически достоверный расчет распределения плотности состояний требует большого числа МД визитов в каждую ячейку фазового пространства, т.е. Больших времен МД симуляции и потому применяется только для малых молекул.

Например, на этом рисунке приведена плотность МД визитов на карте Рамачанрана для дипептида в течение 2.5 нсек. Видно, что распределение изображающих точек сосредоточено в основном в области минимумов энергии. Однако для выполнения кинетических условий равновесия системы, система должна многократно, за время наблюдения переходить из одной конформации в другую, т.е. Заполнять также и область перешейка между областями низкой энергии.

Этот рисунок демонстрирует, что альфа-бэта переход случается в среднем с частотой в 50 пикосек для аланин-дипептида. Поэтому за время наблюдения в 5 нано сек, можно ожидать порядка 100 конформационных переходов, что обеспечивает равновесность распределения МД визитов разных конформационных состояний.

Более мощным и универсальным методом расчета разности свободной энергии для двух состояний системы, является метод термодинамической теории возмущений или термодинамического интегрирования. Метод ТТВ состоит в том, что разность свободных энергий системы в двух состояниях есть работа обобщенной силы вдоль пути перехода из состояния А в состояние В. Универсальным методом конструирования пути перехода из состояния А в В, является метод параметра связи. Так что предельные значения параметра связи λ описывают исходно и конечное состояния. Тогда расчет обобщенной силы для химерной системы при промежуточных значениях параметра связи и интегрирование этой силы по интервалу изменения параметра связи, позволяет рассчитать разность свободных энергий.

Метод МД дает возможность симуляции химерной системы и расчета средних по ансамблю $\langle \lambda \rangle$ состояний химерной системы и таким образом дает возможность расчета величины ΔG_{A-B} .

Метод ТТВ имеет разнообразные применения, - оценка влияния аминокислотных замен на **стабильность** молекулы белка, - зависимость **энергии связывания** лиганда от строения лиганда.

При оценке влияния аминокислотных замен на стабильность молекулы белка, рассматривается термодинамический цикл, приведенный на рис. 20. Поскольку полное изменение свободной энергии при полном обходе цикла равно нулю, система возвращается в исходное состояние, для свободных энергий перехода между состояниями системы, имеем уравнение (1).

Для изменения стабильности молекулы белка в результате мутации имеем уравнение (2). Прямая симуляция процесса самоорганизации пока не возможна, ввиду огромной операционной емкости. Однако симуляция вертикальных процессов мутации аминокислотного остатка в сложенном или развернутом состоянии намного проще. Именно

эти трансформации и симулируются методом МД, вычисляются соответствующие свободные энергии, разность которых и определяет изменение стабильности мутантного белка.

Другой пример области применения ТТВ – это конструирование лекарственных препаратов и оптимизация строения лиганда (малой молекулы лекарственного препарата) для максимально селективного связывания с данным белком. Оптимизация селективности важна как для проявления лекарственного эффекта, так и для уменьшения побочных эффектов. Аналогично, рассматривается термодинамический цикл, приведенный на рис.21. Далее методом ТТВ и МД вычисляются свободные энергии вертикальных процессов, мутация лиганда из химической структуры А в структуру В в свободном состоянии в растворе и в связанном состоянии на белковой молекуле. Искомая разность свободных энергий вычисляется из уравнения (2). Таким образом имея серию лигандов можно выбрать лиганд с максимальной энергией связи.

Основным недостатком метода ТТВ является его большая операционная емкость, поскольку необходима медленная трансформация одной структуры в другую, с многократным расчетом обобщенной средней силы по параметру трансформации.

Другая возможность расчета свободной энергии биополимера в растворе состоит в использовании физической модели раствора и взаимодействия раствор-биополимер, вместо явной молекулярной модели растворителя.

Свободная энергия биополимера в растворе выражается уравнением приведенным на рис.22. Где первый член это средняя потенциальная внутренняя энергия биополимера, второй член – это конформационная энтропия, а третий член представляет среднюю свободную энергию сольватации. Последний метод расчета свободной энергии использует метод МД в явном растворителе, чтобы получить конформационные микросостояния биополимера совместимые с молекулярной структурой растворителя.

Свободная энергия сольватации статистически выражается интегралом ТТВ по параметру включения взаимодействия биополимер-растворитель. Энергия сольватации может быть разложена на компоненты – элементарные физические акты описывающие процесс сольватации как трех-этапный термодинамический процесс – 1) образование полости в растворе. С энергией пропорциональной площади поверхности полости, 2) включение короткодействующих ван дер Ваальсовых взаимодействий биополимер-раствор, которые легко вычисляются явно, 3) электростатическая поляризация полярного раствора зарядами биополимера.

Поляризационная энергия дает наибольший вклад в энергию сольватации молекулы биополимера в водном растворе. Эта энергия вычисляется на основе решения уравнения Пуассона для модели двух непрерывных диэлектрических сред с поверхностью раздела эквивалентной поверхности молекулярной полости. Полость вытесненного растворителя имеет низкую диэлектрическую константу, а окружающий полярный растворитель высокую, 80 для воды.

Метод ESIS оказывается на несколько порядков быстрее метода ТТВ и не значительно уступает последнему в точности.

Метод ESIS позволил показать, что нативная конформация молекулы белка соответствует глобальному минимуму свободной энергии.

На рис. 24 приведены расчеты свободной энергии порядка 1000 компактных структур нескольких глобулярных белков, в зависимости от отклонения структуры от нативной. Видно, что нативные структуры всех рассмотренных белков соответствуют минимальной

свободной энергии. Отклонение конформации от нативной ведет к росту свободной энергии, т.е. не стабильным структурам.

Одно из приложений метода МД это анализ конформационной динамики и существенных обобщенных движений. Такой анализ делается с помощью расчета матрицы позиционных флуктуаций C_{ij} . Диагонализация матрицы C позволяет разложить конформационные движения на набор ортогональных обобщенных мод – квазигармонических осцилляторов. Собственные значения матрицы C дают амплитуды флуктуаций обобщенных координат, а собственные векторы определяют формы обобщенных мод.

Расчеты для биополимеров показывают, что до 80% всех движений сосредоточены в не большом количестве обобщенных мод.

На этом рисунке показаны конформационные движения ДНК дуплексов в водном растворе. Видно, что термические флуктуации ДНК дуплекса в основном состоят в контексто-зависимых изгибных колебаниях за наносекундные времена.

Другой пример существенных конформационных движений в 14-звенных ДНК дуплексах, полученный из анализа 1.5 нсек МД траекторий, показан на этом рисунке. Видно, что поли – А тракты имеют наименьшую амплитуду флуктуаций. Направления изгиба также контексто-зависимы. А-тракты гнутся спонтанно в сторону малой бороздки ДНК, тогда как ТАТА последовательность гнется с торону большой бороздки.

Метод МД помогает получить механистическую картину процессов связывания лиганда молекулой белка. Например МД симуляция комплекса белок-лиганд при высокой температуре позволяет инициировать ускоренный процесс отрыва лиганда от сайта связывания и получить траекторию отрыва, которая близка к траектории связывания лиганда. Такая информация полезна для конструирования оптимальных лекарственных препаратов.

Другой пример возможностей метода МД это симуляция биомембран, т.е. Двойного липидного слоя в водном растворе. Такое моделирование позволяет моделировать процессы проникновения малых молекул через мембрану или взаимодействие мембран с мембранными белками и формирование трансмембранных каналов, которые обеспечивают обмен внутриклеточного вещества с окружающей средой.

В заключение можно сказать, что современные методы моделирования биополимеров достигли к 2000 году уровня способности получать новое знание о биологической системе недоступное экспериментальными методами, однако их широкое применение еще только начинается. Фактически предстоит еще огромная работа по повышению уровня достоверности методов, разработке более аккуратных силовых полей и более эффективных параллельных алгоритмов для многопроцессорных систем. Это позволит в будущем предсказывать оптимальную структуру белков по их аминокислотной последовательности, решить проблему предсказания трехмерных структур РНК и субмолекулярных белково-нуклеиновых комплексов.