

Лекция 6. Модификации молекулярной динамики. Методы Монте-Карло.

Курс: Структурная Биоинформатика и моделирование лекарств (ВШЭ)

Головин А.В.¹

¹МГУ им М.В. Ломоносова, Факультет Биоинженерии и Биоинформатики

Москва, 2017

Уравнение Шредингера

$$\left(-\frac{\hbar^2}{m} \left[\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right] + V \right) \Psi(r, t) = i\hbar \frac{\partial \Psi(r, t)}{\partial t}$$

Или:

$$H\Psi = E\Psi; \quad H = \frac{-\hbar^2}{m} \nabla^2 - \frac{Ze^2}{4\pi\epsilon_0 r}$$

В молекулярной механике где аппроксимируем электронную плотность уравнениями классической физики.

$$F = m \frac{\partial^2 r}{\partial t^2}$$

Осталось придумать как следить за эволюцией системы во времени.

Простое уравнение силового поля (СП)

$$\begin{aligned}
 U = & \sum_{bonds} \frac{k_i}{2} (l_i - l_0)^2 + \sum_{angles} \frac{k_i}{2} (\phi_i - \phi_0)^2 + \sum_{torsions} \frac{V_n}{2} (1 + \cos(n\omega - \gamma)) + \\
 & + \sum_{i=1}^N \sum_{j=i+1}^N \left(4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right)
 \end{aligned}$$

Молекулярная динамика

Монте-Карло

$$v(t + \frac{\Delta t}{2}) = v(t - \frac{\Delta t}{2}) + \frac{F(t)}{m} \Delta t$$

$$acc(o \rightarrow \eta) = \min \left(1, \exp \left\{ -\beta \left[U(r^N) - U(r^N) \right] \right\} \right)$$

Молекулярная динамика

Расчёт силы действующей на атом

$$F_i = \sum_j F_{ij}$$

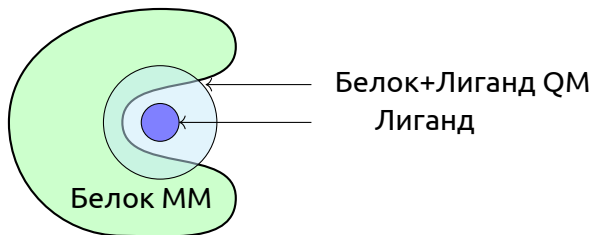
Δt

Расчёт новых координат

$$\frac{\partial^2 r_i}{\partial t^2} = \frac{F_i}{m_i}$$
$$\frac{\partial r_i}{\partial t} = v_i, \quad \frac{\partial v_i}{\partial t} = \frac{F_i}{m_i}$$

Гибридное QM/MM моделирование

- Основная идея: разделить большую систему на квантовую и молекулярную части.
- Электростатическое окружение из MM части чувствуется QM частью.
- MM часть принимает силы из QM части и соответственно адаптируется.



Гибридное QM/MM моделирование

Простейший Гамильтониан для QM/MM системы:

$$H = -\frac{1}{2} \sum_i^{elect} \nabla^2 + \sum_i^{nucl} \sum_j^{elect} \frac{1}{r_{ij}} + \sum_i^{nucl} \sum_j^{nucl} \frac{Z_i Z_j}{R_{ij}} - \sum_i^{elect} \sum_j^{MM_q} \frac{Q_j}{R_{ij}} + \sum_i^{nucl} \sum_j^{MM_q} \frac{Z_i Q_j}{R_{ij}}$$

К QM/MM части можно добавить и VdW составляющую:

$$H_{QM/MM} = - \sum_i^{elect} \sum_j^{MM_q} \frac{Q_j}{R_{ij}} + \sum_i^{nucl} \sum_j^{MM_q} \frac{Z_i Q_j}{R_{ij}} +$$

$$\sum_i^{nucl} \sum_j^{MM_{atoms}} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]$$

Реализации описания QM/MM

Встраивание электронов

$$H^{QM/MM} = H_e^{QM} - \sum_i^n \sum_J^M \frac{e^2 Q_J}{4\pi\epsilon_0 r_{iJ}} + \sum_A^N \sum_J^M \frac{e^2 Z_A Q_J}{e\pi\epsilon_0 R_{AJ}}$$

- Ковалентные взаимодействия между QM и MM системами описываются соответствующими параметрами из MM
- Для QM системы в месте разрыва связи добавляют протон для восстановления системы до полного состояния.
- Сила действующая на этот протон "заглушку" в QM системе распределяется между атомами между которыми происходит раздел QM и MM системами

Реализации описания QM/MM

ONIOM

- Расчитываем энергию и градиент для QM системы с желаемым уровнем теории.
- Расчитываем энергию и градиент для MM системы с учетом ранее рассчитанных данных для QM системы.
- Рассчитываем MM энергию и градиенты для QM системы и вычитаем.

$$E_{tot} = E_I^{QM} + E_{I+II}^{MM} - E_I^{MM}$$

Этот подход можно использовать не только для двух уровней теории, но и больше.

Реализация QM/MM в Gromacs

- Добавляем атомы "заглушки":

```
[ virtual_sites2 ]  
LA QAtom MAtom 1 0.65
```

а связь описывается:

```
[ constraints ]  
QAtom MAtom 2 0.153
```

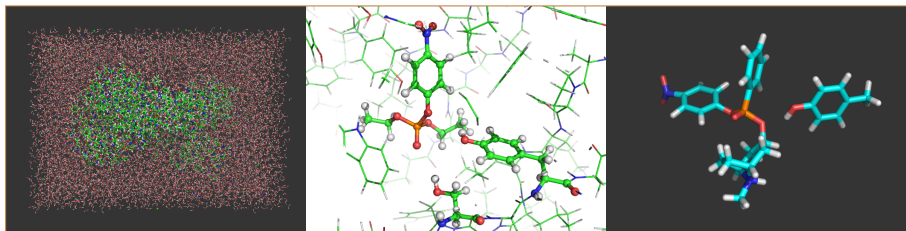
- Для атомов в QM системе надо поправить описание ковалентных связей:

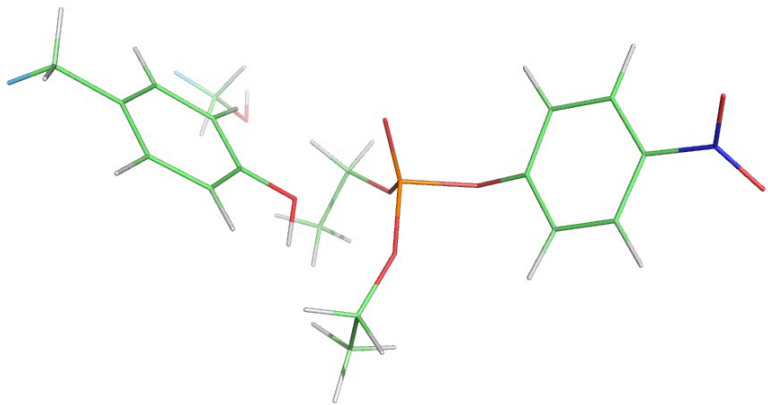
```
[ bonds ]  
QAtom1 QAtom2 5  
QAtom2 QAtom3 5
```

- В mdp файле описываем параметры для QM системы.

Гибридное QM/MM моделирование

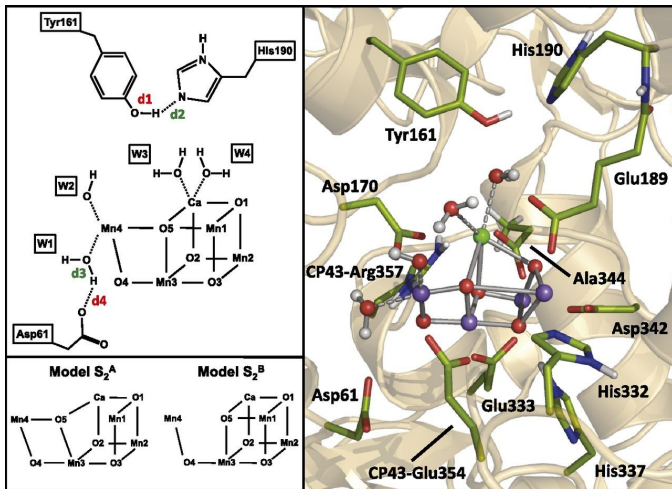
- Атомы связки
- Специальные орбитали







Ab initio QM/MM model of photosystem II. (Right) The QM region, consisting of 224 atoms, is shown in balls and sticks representation.

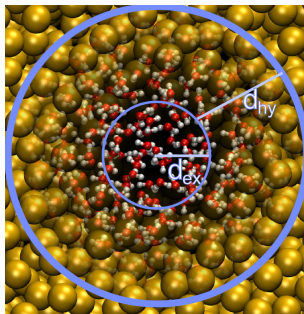


Daniele Narzi et al. PNAS 2014;111:8723-8728

Adaptive resolution scheme

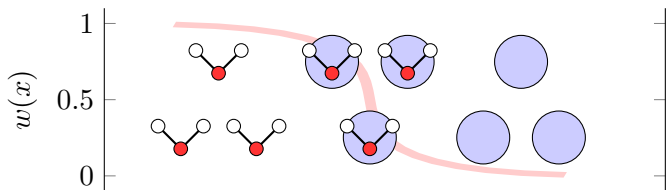
Описание взаимодействия между полноатомным и крупнозернистым описаниями системы

- В отличие от QM/MM уровень описания молекул системы может меняться "на лету"



Реализация AdResS

$$\vec{F}_{\alpha\beta} = w_{\alpha}w_{\beta}\vec{F}_{\alpha\beta}^{\text{ex,mol}} + [1 - w_{\alpha}w_{\beta}]\vec{F}_{\alpha\beta}^{\text{cg,mol}},$$



$$w(x) = \begin{cases} 0 & : & x > d_{\text{ex}} + d_{\text{hy}} \\ \cos^2\left(\frac{\pi}{2d_{\text{hy}}}(x - d_{\text{ex}})\right) & : & d_{\text{ex}} + d_{\text{hy}} > x > d_{\text{ex}} \\ 1 & : & d_{\text{ex}} > x \end{cases},$$

Метод обмена репликами (REMD)

- Основная идея: запустить параллельно несколько счётов с разными температурами.
- Мы можем выбрать правило когда производить обмен конформациями.
- Если мы проводим обмен когда потенциальная энергия одной из реплик ниже чем других, то это похоже на моделирование отжига.
- Такой подход часто используется для моделирования самосборки.

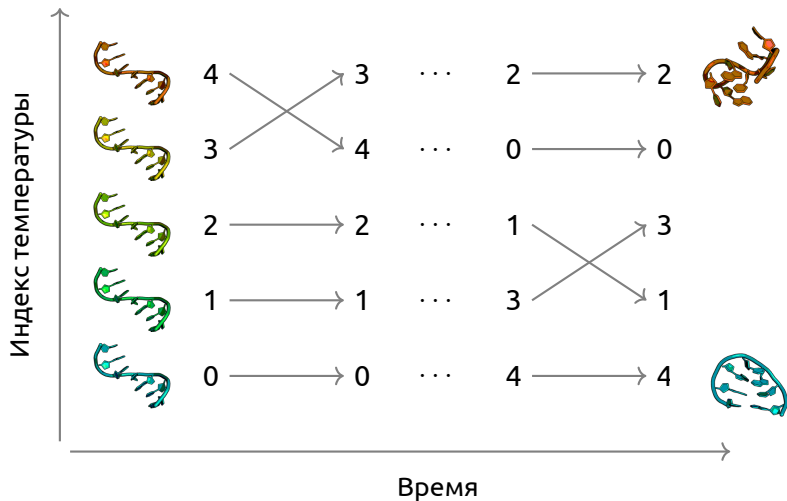
REMD

- Цель метода это ускорить сканирование (sampling) конформационного пространства.
- Применимо к переходам через значимые энергетические барьеры.
- В Gromacs обмен между репликами происходит случайно по условию:

$$P(1 \leftrightarrow 2) = \min \left(1, \exp \left[\left(\frac{1}{k_B T_1} - \frac{1}{k_B T_2} \right) (U_1 - U_2) \right] \right)$$

А скорости масштабируются: $(T_1/T_2)^{\pm 0.5}$

REMD



REMD: количество реплик

Разница между температурами (репликами):

$$U_1 - U_2 = N_{df} \frac{c}{2} k_B (T_1 - T_2)$$

где N_{df} это количество степеней свободы и "с" это величина от 1 до 2 для системы белок вода.

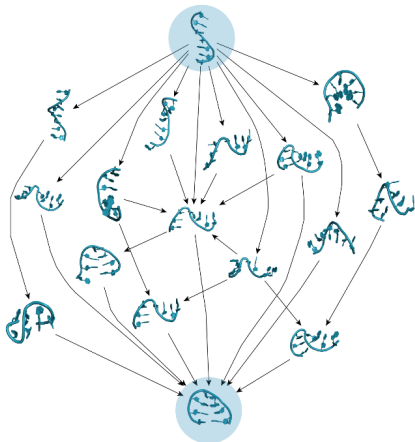
Если $T_2 = (1 + \epsilon)T_1$ тогда вероятность обмена:

$$P(1 \leftrightarrow 2) = \exp\left(-\frac{\epsilon^2 c N_{df}}{2(1 + \epsilon)}\right) \approx \exp\left(-\epsilon^2 \frac{c}{2} N_{df}\right)$$

Таким образом для вероятности обмена $e^{-2} \approx 0.135$ получаем $\epsilon \approx 2/\sqrt{c N_{df}}$.

И если мы контролируем длину связей, то: $N_{df} \approx 2 N_{atoms}$ и при $c = 2$ надо использовать: $\epsilon = 1/\sqrt{N_{atoms}}$.

Результат REMD



Коллективные переменные (CV)

- Положение всех атомов в пространстве явно избыточная информация для описания некоторых простых процессов
- Коллективные переменные это некоторые значения которые наиболее значительно изменяются в ходе процесса, который нас интересует
- Трудно предположить заранее все важные коллективные переменные
- Недостаток описания приводит к гистерезису

Примеры коллективных переменных

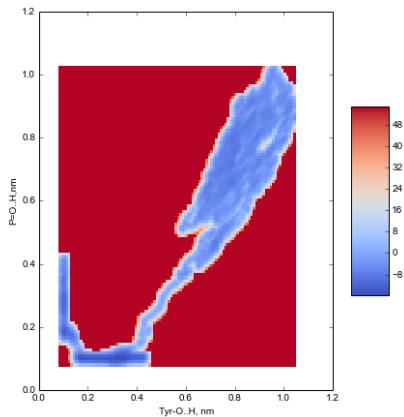
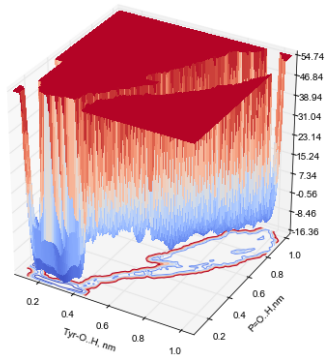
ALPHABETA
ANTIBETARMSD
CONSTANT
DHENERGY
DISTANCE
NOE
PATH
RDC

ALPHARMSD
CELL
CONTACTMAP
DIHCOR
ENERGY
PARABETARMSD
POSITION
TORSION

ANGLE
CH3SHIFTS
COORDINATION
DIPOLE
GYRATION
PATHMSD
PROPERTYMAP
VOLUME

Пример анализа на основе CV

$$E \approx -\frac{1}{\beta} \log(P)$$



Использование CV для влияния на МД, ABMD

adiabatic biased MD

$$V(\rho(t)) = \begin{cases} \frac{K}{2} (\rho(t) - \rho_m(t))^2, & \rho(t) > \rho_m(t) \\ 0, & \rho(t) \leq \rho_m(t), \end{cases}$$

$$\rho(t) = (CV(t) - TO)^2$$

Суть метода состоит в движении системы к заданным значениям CV используя гармонический потенциал, который изменяет термические флуктуации если система не движется к заданным значениям CV.

Использование CV для влияния на МД, steered MD

Позволяет добавлять временно зависимый гармонический потенциал на одну или несколько переменных.

$$V(\vec{s}, t) = \frac{1}{2} \kappa(t) (\vec{s} - \vec{s}_0(t))^2$$

или

$$\begin{aligned} H_\lambda(X, t) &= H(X) + U_\lambda(X, t) \\ &= H(X) + \frac{k(t)}{2} (s(X) - \lambda(t))^2 \\ &= H(X) + \frac{k(t)}{2} (s(X) - s_0 - vt)^2. \end{aligned}$$

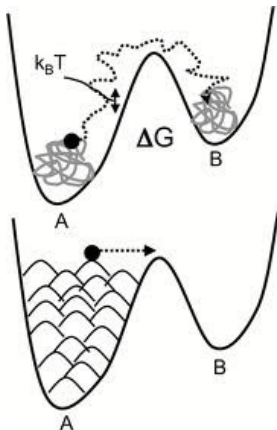
Использование CV для влияния на МД, WALLS

Суть очень проста выставление "стен" в виде штрафа за пределами значения CV

$$\sum_i k_i ((x_i - a_i + o_i) / s_i)_i^e$$

Метадинамика

$$V(\vec{s}, t) = \sum_{k\tau < t} W(k\tau) \exp\left(-\sum_{i=1}^d \frac{(s_i - s_i^{(0)}(k\tau))^2}{2\sigma_i^2}\right).$$



Well tempered metadynamics

$$V(s, t) = \sum_{t'=0, \tau_G, 2\tau_G, \dots}^{t' < t} W e^{-V(s(q(t')), t') / \Delta T} \exp \left(- \sum_{i=1}^d \frac{(s_i(q) - s_i(q(t')))^2}{2\sigma_i^2} \right)$$

Этот вариант метадинамики позволяет добавлять энергию более аккуратно, что приводит к тому, что добавленное количество энергии начинает стремиться к энергии перехода между состояниями.

Обсуждение

Преимущества

- Ускорение событий путём выталкивания системы из известной области
- Знание результата не нужно, возможно все, что может быть в этой системе
- Возможность восстановить профиль поверхности потенциальной энергии

Недостатки

- Динамика и температура изменены человеком.
- Данные о кинетике процессов не доступны (а может и нет : arXiv:1309.5323, Tiwari & Parrinello)

Выбор CV

Правильный выбор CV это критический момент, вот некоторые требования:

- CV должны описывать процесс и интереса
- Включать все медленно изменяющиеся степени свободы
- Количество CV должно быть не большим
- Используйте химическую/физическую интуицию
- Подход проб и ошибок

Метадинамика с обменом между CV

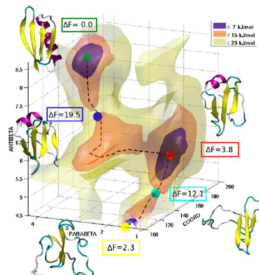
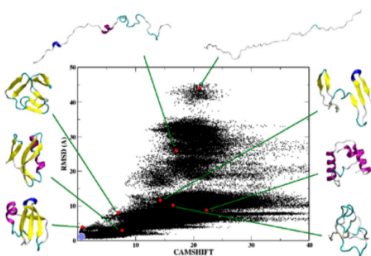
Используя правило Метрополиса:

$$\min \left(1, \exp \left[\beta (V_G^a(x^a, t) + V_G^b(x^b, t) - V_G^a(x^b, t) - V_G^b(x^a, t)) \right] \right)$$

Запускаются несколько параллельных траекторий с разными CV, которые обмениваются конформациями как в REMD

Примеры с использованием метадинмики

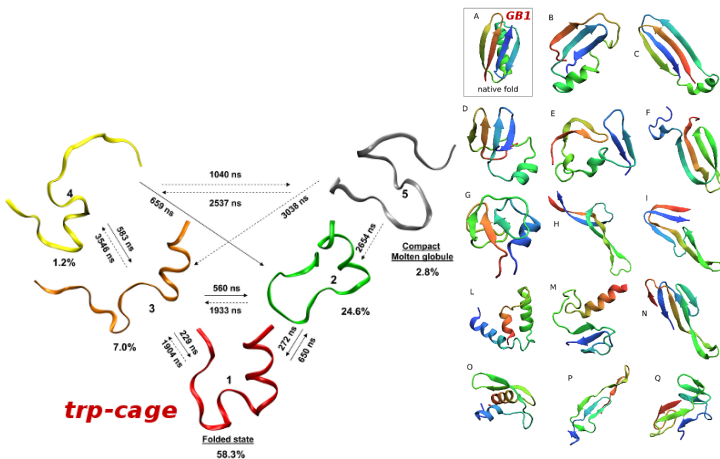
Образование структуры белка на основе знания химических сдвигов и реконструкция FES белка :



doi: 10.1073/pnas.1218350110.

Примеры с использованием метадинамики

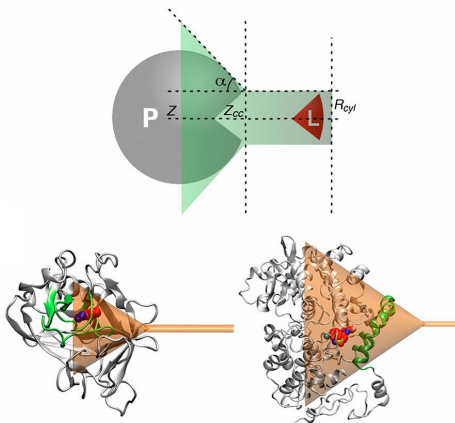
Изучение процесса фолдинга белка:



JOURNAL OF CHEMICAL THEORY AND COMPUTATION, 5, 2197 (2009)

Примеры с использованием метадинамики

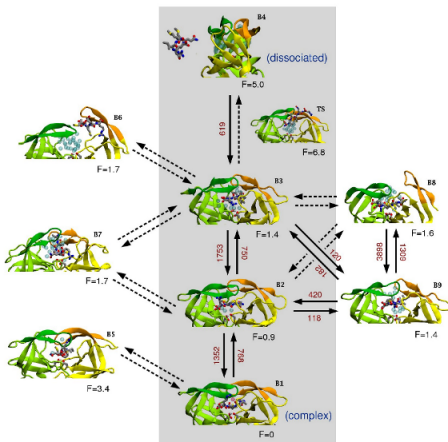
Определение пути лиганда к сайту связывания:



doi: 10.1073/pnas.1303186110

Примеры с использованием метадинмики

Аккуратное определение механизма связывания лиганда:



JOURNAL OF THE AMERICAN CHEMICAL SOCIETY, 131, 11811 (2009)

Простое уравнение силового поля (СП)

$$\begin{aligned}
 U = & \sum_{bonds} \frac{k_i}{2} (l_i - l_0)^2 + \sum_{angles} \frac{k_i}{2} (\phi_i - \phi_0)^2 + \sum_{torsions} \frac{V_n}{2} (1 + \cos(n\omega - \gamma)) + \\
 & + \sum_{i=1}^N \sum_{j=i+1}^N \left(4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right)
 \end{aligned}$$

Молекулярная динамика

Монте-Карло

$$v(t + \frac{\Delta t}{2}) = v(t - \frac{\Delta t}{2}) + \frac{F(t)}{m} \Delta t$$

$$acc(o \rightarrow \eta) = \min \left(1, \exp \left\{ -\beta \left[U(r^N) - U(r^N) \right] \right\} \right)$$

Метод Монте-Карло

- Основная идея предполагает поиск конформаций с низкой энергией на основе случайного изменения координат.
- Скорости у атомов не рассчитываются

Представим, что канонический ансамбль можно разбить :

$$Q_{NVT} = Q_{NVT}^{ideal} Q_{NVT}^{excess}; \quad Q_{NVT}^{excess} = \frac{1}{V^N} \int dr^N \exp \left[\frac{-U r^N}{k_b T} \right]$$

Энергия Гемгольца:

$$A = -k_b T \ln Q_{NVT} \rightarrow A = A^{ideal} + A^{excess}$$

Расчёт термодинамических свойств

Формально нам надо интегрировать по всему пространству конформаций, это невыполнимо.

Приближение:

- 1 Генерируем конформацию
- 2 Считаем потенциальную энергию: $\exp\left(\frac{U(r^N)}{k_b T}\right)$
- 3 Считаем фактор Больцмана.
- 4 Суммируем известные факторы и находим среднее и переходим к пункту 1.

В результате средняя энергия определяется как:

$$\langle U(r^N) \rangle = \frac{\sum_1^{N_{trial}} U_i(r^N) \exp\left[\frac{U(r^N)}{k_B T}\right]}{\sum_1^{N_{trial}} \exp\left[\frac{U(r^N)}{k_B T}\right]}$$

Расчёт термодинамических свойств

- Неудобно, что основной счёт приходит на ненужные состояния.
- Метрополис и соавторы предложили использовать цепи Маркова для генерации конформаций.

Пример:

- 1 генерируем: $x_{new} = x_{old} + (2\xi - 1)\delta r_{max}$
- 2 если новая энергия меньше старой то принимаем конформацию и используем дальше
- 3 если энергия выше то сравниваем фактор Больцмана со случайным числом 0:1 и если фактор Больцмана изменения энергии больше, то принимаем конформацию

Моделирование жестких молекул (rigid body)

- Так как молекула не точка, мы смещаем центр масс.
- Нам важно не только смещение молекулы, но и её ориентация.
- Удобно использовать сферические координаты и вектор отображающий основную ось молекулы.

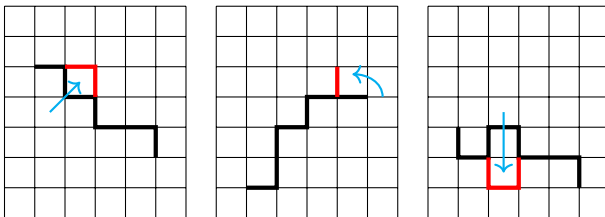
Моделирование гибких молекул

- Монте-Карло моделирование гибких молекул проводить сложнее.
- Реалистично только для малых молекул и молекул с замороженными степенями свободы.
- Часто замораживают длину валентных связей и валентные углы.
- Для больших молекул незначительное изменение торсионного угла может привести к высоким энергиям.

Моделирование полимеров

Решёточные модели полимеров.

Представим что один мономер занимает один кубик (аналог крупно-зернистого моделирования)



Методы выборочного поиска (Biased Monte-Carlo)

Суть идеи: сканировать не все пространство конформаций, а только объект интереса.

Пример: Нас интересует поведение молекулы в воде. Поведение воды в далеке от молекулы не важно.

Существуют:

- Силовое предпочтение — направляет движение молекулы вдоль действия силы.
- Умное (Smart) предпочтение — тоже использует силы для определения движения. В случае силового предпочтения есть ограничения на движение. В большинстве случаев оба метода очень похожи.

Проблема квази-эргодичности

Суть проблемы: При наличии высокоэнергичных барьеров метод Метрополиса оказывается в ловушке локальных минимумов энергии. Происходит накопление данных для весьма ограниченной области фазового пространства.

Существуют два подхода для решения этой проблемы:

- Подход “прыжок-ход” (Jump-walking или J-walking).
- Мультиканонический подход.

J-walking

Суть идеи: При движении в пространстве при заданной температуре, разрешены переходы в пространство с более высокой температурой.

Простейшее решение: проводить 2 и более моделирования параллельно при разных температурах. Моделирование при высокой температуре делают с предпочтением в фазовое пространство при низкой температуре.

Мультиканонический ансамбль

При каноническом ансамбле вероятность (P) оказаться в точке с энергией E , пропорциональна фактору Больцмана умноженному на плотность состояний между точками E_1 и E_2 .

Фактор Больцмана: $w_B = \exp\left[-\frac{E}{k_B T}\right]$

$$P_{\text{canon}}(T, E) \sim n(E)w_B(E)$$

Этот вариант предполагает существенные скачки энергии при переходах.

Мультиканонический ансамбль

В мультиканоническом ансамбле вероятность не зависит от энергии в определённых диапазонах энергии. Это равносильно замене фактора Больцмана на некий мультиканонический фактор w .

$$P_{mu}(E) \sim n(E)w_{mu}(E) = constant$$

Определение этого фактора не тривиальная задача. Сначала делают поиск при высокой температуре. На основе результатов определяют состояния близкие к наименьшей энергии и наоборот.

$$w_{mu}(E) = exp[-\beta(E)E - \alpha(E)]$$

Большой канонический ансамбль

Большой канонический ансамбль предполагает постоянство химического потенциала.

$$\mu = k_B T \ln(\Lambda^3 z); \quad \Lambda = \sqrt{\frac{h^2}{2\pi m k_B T}}$$

Ключевые особенности большого канонического ансамбля:

- Смещение частицы происходит методом Метрополиса
- Частица может быть уничтожена
- Частица может появиться в случайном положении
- Необходимо аккуратно подобрать частоту уничтожений/появлений.

Что выбрать молекулярную динамику или Монте-Карло?

Молекулярная динамика: незаменима когда нужно найти свойство базирующиеся на эволюции системы во времени.

Монте-Карло: гораздо удобнее для работы с заданным ансамблем. Работать с большими молекулами не удобно, но возможность «нефизических” переходов позволяет проводить эффективный поиск в фазовом пространстве.

Методы комплиментарны: не удивительно, что существуют подходы с использованием обоих методов одновременно.

Фреймворк Rosetta

- Rosetta это унифицированный пакет для предсказания структуры белка и функционального дизайна.
- Для решения этих задач надо:
 - Исследовать конформационное пространство белка
 - Для дизайна надо еще исследовать пространство последовательностей
 - Используется эмпирически адаптированный алгоритм Монте-Карло и эмпирические функции для расчёта энергии (PDB)

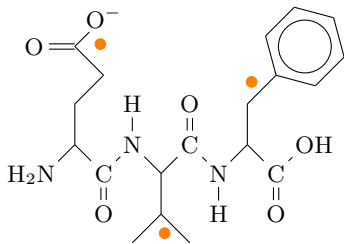
Исследование конформационного пространства остова

- Изменения в конформации остова белка делятся на локальные и глобальные
- Глобальные изменения описываются обменом конформаций между фрагментами определенной длины (3-6-9). Конформации фрагментов получают из PDB
- Локальная оптимизация достигается перебором ϕ, ψ углов, так, что бы не происходило глобального изменения фолда.

Конформации боковых радикалов

- Для значительного числа аминокислот перебрать все конформации боковых цепей невозможно.
- Rosetta предлагает значительно сократить количество конформаций состояниями из PDB - rotamers
- Используется Монте-Карло моделирование отжига для поиска глобального минимума

Функция для расчёта энергии, centroid



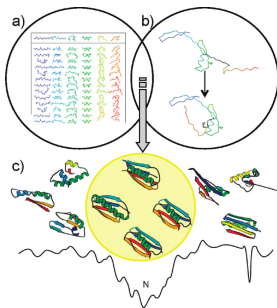
- Сольватация описывается на основе вероятности такой конформации в PDB
- Электростатика описывается вероятностью расстояния между центроидами в PDB
- Водородные связи между β тяжами описываются на основе геометрии расположения тяжей

Полноатомные функции энергии

- VdW : Lennard-Jones
- Гидрофобика: SAS
- Эмпирический потенциал для водородной связи на основе конформации
- Эмпирический потенциал для электростатики (заряды)
- Важная особенность: потенциалы попарно разложимы (оптимизация)

De Novo Folding Simulation

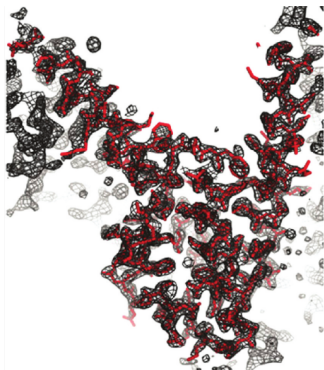
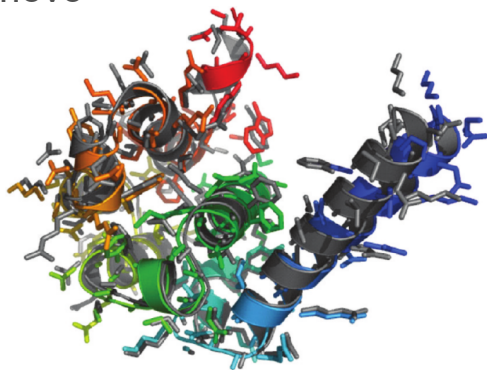
- В линейную конформацию вставляют фрагменты и быстро сканируют пространство укладывая остов белка. 9-мер 30000 раз затем 10000 раз 3-мер. Таким образом собирают 25000-50000 моделей.
- Выбранные модели оптимизируют на полноатомном уровне.



Релаксация и улучшение моделей

- В ходе релаксации незначительно изменяются углы остова.
- Боковые радикалы укладываются с помощью Монте-Карло моделирования отжига.
- В конце применяется градиентная оптимизация по всем торсионным углам.
- Возможно последовательное увеличение коэффициента при потенциале отталкивания
- Есть работы где в ходе релаксации накладывались ограничения из ЯМР или РСА

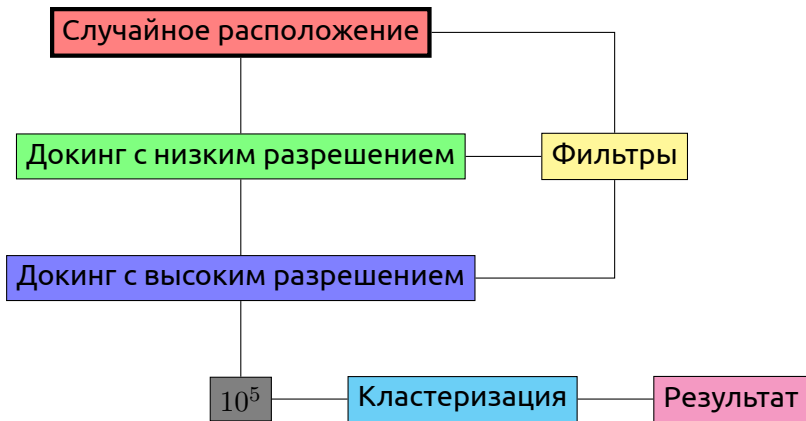
Использование экспериментальных данных в de novo



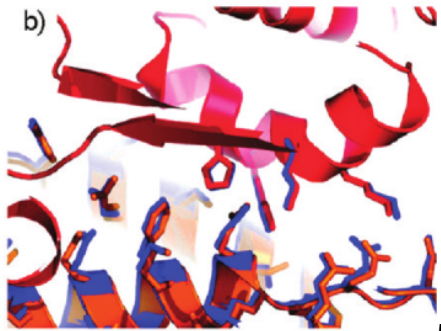
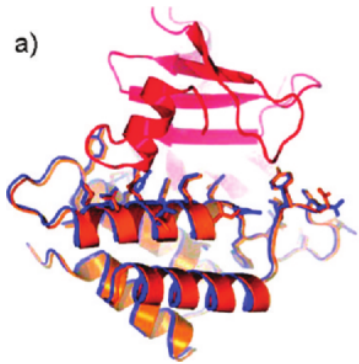
DOI: 10.1021/bi902153g

- Слева ЭПР данные + Rosetta, серым отмечена структура по данным PCA
- Справа: решение проблемы фазы с помощью моделей Rosetta и автоматического алгоритма молекулярного

Белок-Белок докинг: Rosettadock



Rosettadock, результат



дель окрашена
синим

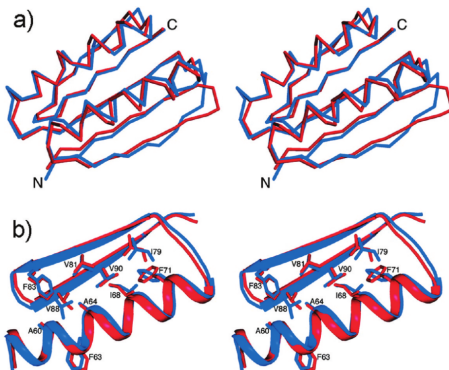
DOI: 10.1021/bi902153g

RosettaLigand

- Основное преимущество это одновременный учет подвижности лиганда и белка
- Надо знать конформеры лиганда
- Можно учитывать не только подвижность боковых цепей, но и остова
- Надо использовать в системах где предполагается индуцированное конформационное изменение

Дизайн de novo

- Надо менять не только конформацию в ходе моделирования но и последовательность.
- Была построена модель белка, укладки которого не было в PDB



Другие способы дизайна

- Редизайн известных белков для увеличения устойчивости структуры
- Изменение места контакта с другими биополимерами полимерами (нуклеаза-ДНК)
- Дизайн ферментов, оптимизация структуры и последовательности белка для улучшения заданного переходного состояния.

Использование Rosetta

AbinitioRelax AnchoredDesign
antibody_graft BuildPeptide
docking_protocol enzyme_design
FlexPepDocking FloppyTail
hbs_design homodimer_design
ligand_dock loopmodel
loops_from_density

membrane_abinitio2 relax
remodel rna_denovo rna_design
rosettaDNA rosetta_scripts
rotamer_recovery SymDock
UnfoldedStateEnergyCalculator
zinc2_homodimer_design

Подробнее:

<https://www.rosettacommons.org/docs/latest/Application-Documentation.html>

Разработка Rosetta, PyRosetta

- Язык программирования в Rosetta C++
- К основным функциям и протоколам есть Python bindings

```
In [1]: import sys
import os
import operator
from itertools import *

sys.path.append('/home/anur/pyrosetta')

from rosetta import *
from toolbox import *
from rosetta.utility import vector1_bool
init()
#from prody import *
#from pylab import *
import numpy as np

from numpy.random import randint

from joblib import Parallel, delayed
```

let use pymol for visual anlysis and debugging

```
In [2]: pymover = PyMOL_Mover()
```