

Молекулярная динамика. Моделирование самосборки белка. Структурная Биоинформатика

Головин А.В.¹

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

Москва, 2013

Содержание

Введение

Силовые поля

Ковалентные взаимодействия

Нековалентные взаимодействия

Варианты ММ

Молекулярная динамика

Температура и давление

Растворитель в МД

REMD

Самосборка белка

Folding@home

FoldIt Game



Структурная биоинформатика

Структурная биоинформатика является разделом биоинформатики, который связан с анализом и предсказанием трехмерной структуры биологических макромолекул.

Основные задачи структурной биоинформатики:

- Обобщение данных о 3D структуре макромолекул.
- Поиск принципов укладки трехмерной структуры.
- Поиск соотношения структура-функция.



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:

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

Современные базисы предполагают примерно 60 функций на атом. Итого: 900 функций на аминокислоту.

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

Можно аппроксимировать электронную плотность уравнениями классической физики.



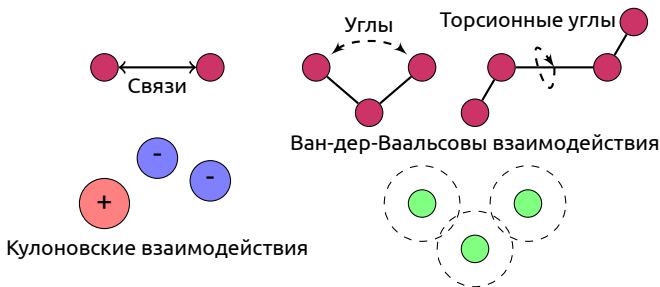
Молекулярная механика (ММ)

- В ММ электронная структура атома замещается на достаточно простые уравнения с параметрами.
- Наборы параметров называются силовыми полями.
- Используется допущение Борна-Оппенгеймера
- Расчёт энергии происходит на основе положения ядер.
- Упрощения позволяют работать с большими системами
- В некоторых случаях ММ подходы могут давать результаты, сравнимые по точности с методами КМ.



Простое уравнение силового поля (СП)

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



Основные особенности силовых полей

- Большинство параметров неотделимо от поля.
- Параметризация ММ сильно зависит от целей исследования.
- Большинство силовых полей параметризованы для воспроизведения структуры.
- Силовые поля — это результат оптимизации параметров.
- Силовые поля — это эмпирически найденные данные.



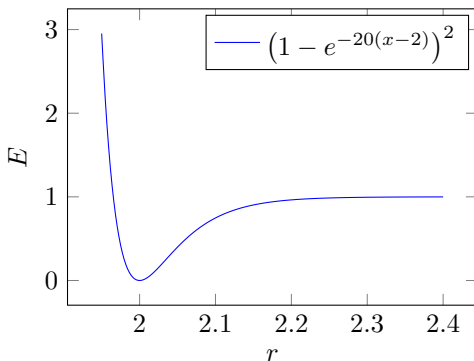
Потенциал для описания связи

Потенциал Морзе

$$U(l) = D_e \{1 - e^{[-a(l-l_0)]}\}^2$$

Более распространённый

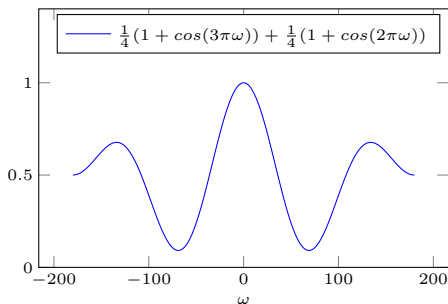
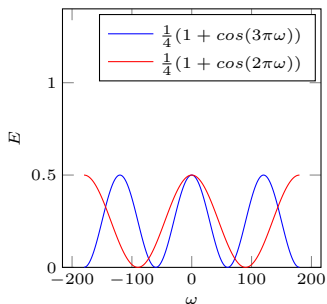
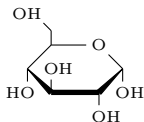
$$U(l) = \frac{k_i}{2} (l_i - l_0)^2$$



Потенциал торсионного угла

$$U(\omega) = \sum_{\text{torsions}} \frac{V_n}{2} (1 + \cos(n\omega - \gamma))$$

Рассмотрим O-C-C-O (сахар в ДНК)



Нековалентные взаимодействия

- Нековалентные взаимодействия являются определяющими в формировании структуры биополимеров.
- Так как эти взаимодействия реализуются через пространство, то часто они описываются как функции, обратно пропорциональные расстоянию между двумя атомами.



Электростатические взаимодействия

Допустим, что поверхность единичного потенциала можно представить зарядами в центрах атомов.

Тогда электростатические взаимодействия будут описываться по закону Кулона:

$$U(q_1, q_2) = \frac{q_1 q_2}{4\pi\epsilon_0\epsilon_r r_{ij}};$$

$$U = \sum_{i=1}^{N_A} \sum_{j=1}^{N_B} \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r_{ij}}$$



Электростатические взаимодействия

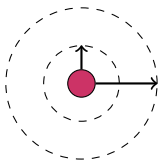
Очевидно, что количество вычислений растет значительно быстрее количества частичных зарядов.

Существуют следующие упрощения:

- Двойное обрезание
- Потенциал реакционного поля
- Суммирование Эвальда



Двойное обрезание



Для сферы А мы считаем все частичные заряды, а для сферы А-В мы будем считать взаимодействие групп зарядов с нашим атомом.

$$U_1 = \sum_{i=1}^{N_A} \frac{q_1 q_i}{4\pi\epsilon_0\epsilon_r r_{1i}} + \sum_{j=1}^{N_{group}} \frac{q_1 q_j}{4\pi\epsilon_0\epsilon_r r_{1j}}$$



Суммирование Эвальда

Основная идея: нам нужно учитывать не только заряды в ближайшем окружении, но и, как в кристалле, заряды, находящиеся в соседних ячейках.

$$U_{ij} = \sum_{x=1}^{N_x} \sum_{y=1}^{N_y} \sum_{z=1}^{N_z} \sum_{i=1}^N \sum_{j=1}^N N \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r_{ij}}$$

Это сходится, но очень медленно.



Суммирование Эвальда

Эвальд предложил перевести этот ряд в сумму 2-ух быстро сходящихся рядов и константы.

$$U = U_{dir} + U_{rec} + U_0$$

$$U_{dir} = f/2 \sum_{i,j}^N \sum_{x=1}^{N_x} \sum_{y=1}^{N_y} \sum_{z=1}^{N_z} q_i q_j \frac{\operatorname{erfc}(\beta r_{ij,n})}{r_{ij,n}}$$

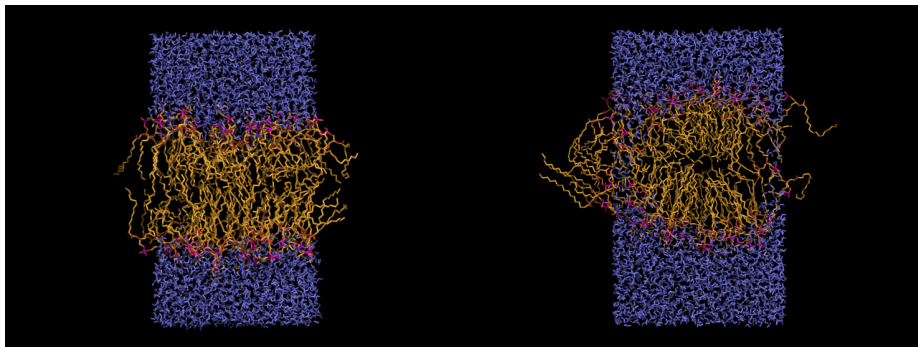
$$U_{rec} = \frac{f}{2} \pi V \sum_{i,j}^N q_i q_j \sum_{m_x} \sum_{m_y} \sum_{m_z} \frac{\exp(-\pi m/\beta)^2 + 2\pi i m (r_i \cdot r_j)}{m}$$

$$U_0 = \frac{f\beta}{\sqrt{\pi}} \sum_i^N q_i^2 \quad (1)$$

Где бета - это параметр, определяющий соотношение прямого и обратного взаимодействий



Суммирование Эвальда vs двойное обрезание



Self-assembly with PME

Self-assembly with Cut-off



Ван-дер-Ваальсовы взаимодействия

- В основе природы Ван-дер-Ваальсовых взаимодействий лежат электронные эффекты: дисперсионные и обменные.
- В принципе, рассчитать такие эффекты можно в QM, но это далеко не тривиальная задача.
- В ММ нам надо считать такие взаимодействия быстро, на сегодняшний день наиболее часто используют потенциал Леонарда-Джонса:

$$U_{VdW} = \sum_{i=1}^N \sum_{j=i+1}^N 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]$$



Водородные связи

- В силовых полях водородная связь часто описывается как комбинация Ван-дер-Ваальсовых и Кулоновских взаимодействий
- Существуют силовые поля, где водородная связь задаётся своим потенциалом на основе потенциала Леонарда-Джонса 10-12:

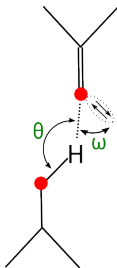
$$U_{HB} = \frac{A^{10}}{r} - \frac{C^{12}}{r}$$



Водородные связи

Для точного описания водородной связи вносят поправки, учитывающие геометрию водородной связи:

$$U_{HB} = \left(\frac{C}{d^6} - \frac{D}{d^4} \right) \cos^m \theta$$



$$U_{HB} = \left(\frac{A}{r_{H..Ac}^{10}} - \frac{C}{r_{H..Ac}^{12}} \right) \cos^2 \theta_{Don-H...Acc} \cos^4 \omega_{LP-Acc...H}$$



Модели воды

- Вода - достаточно сложный объект.
- Важно, что модель воспроизводит как свойства одной молекулы, так и свойства фазы.

Существуют три основных класса моделей:

- Простые модели
- Поляризуемые модели
- Ab initio модели



Простое уравнение силового поля (СП)

$$U = \sum_{bonds} \frac{k_i}{2} (l_i - l_0)^2 + \sum_{angles} \frac{k_i}{2} (\phi_i - \phi_0)^2 + \sum_{torsions} \frac{V_n}{2} (1 + \cos(n\omega - \gamma)) +$$

$$+ \sum_{i=1}^N \sum_{j=i+1}^N \left(4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right)$$

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

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

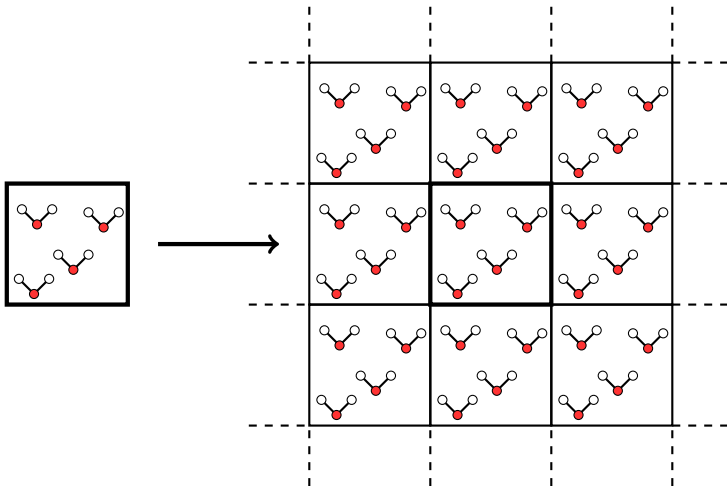
Δt

Расчёт новых координат

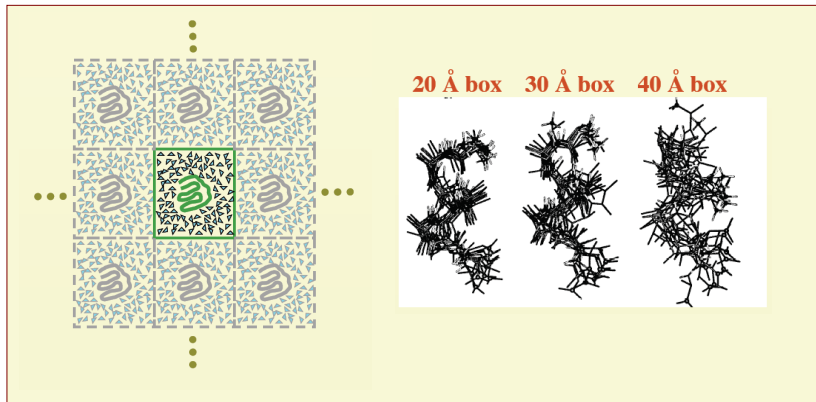
$$\frac{\partial^2 r_i}{\partial t^2} = \frac{F_i}{m_i}$$
$$\frac{\partial r_i}{\partial t} = v_i, \quad \frac{\partial v_i}{\partial t} = \frac{F_i}{m_i}$$



Периодические граничные условия



Периодические граничные условия



МД поли-аланина показала искусственную стабилизацию альфа спирали, при использовании маленькой ячейки.

Производительность

Система из 80000 атомов, компьютер Core2Quad:

Скорость:	24 шага/сек
Время наблюдения за системой:	1 мкс
Число шагов:	$5 \cdot 10^8$
Длина шага:	2 фс.
Время симуляции:	$5 \cdot 10^8 / 24$ сек 24000 часов 1000 суток около 3 лет

Кластер (96 процессоров) примерно 25 дней, можно до 2000 процессоров.



Температура

При МД часто происходит релаксация структуры и появляется излишек кинетической энергии.

$$E_{NVT} = \frac{3}{2} N k_b T$$

Самый простой способ сохранить температуру — это масштабирование скоростей

$$\frac{\partial T}{\partial t} = \frac{T_0 - T}{\tau} \quad \lambda = \left[1 + \frac{n_{TC} \Delta t}{\tau_T} \left\{ \frac{T_0}{T(t - \frac{1}{2} \Delta t)} - 1 \right\} \right]^{1/2}$$

Кроме масштабирующих термостатов, существуют столкновительные термостаты и термостаты с дополнительной степенью свободы.



Контроль давления в системе

Баростат Берендсена

$$\frac{\partial P}{\partial t} = \frac{P_0 - P}{\tau_p},$$

Баростат Паринелло-Рахмана

$$\frac{\partial \mathbf{b}^2}{\partial t^2} = \mathbf{V}\mathbf{W}^{-1}\mathbf{b}'^{-1} (\mathbf{P} - \mathbf{P}_{\text{ref}}).$$

\mathbf{b} - матрица ветров ячейки

\mathbf{V} - объём

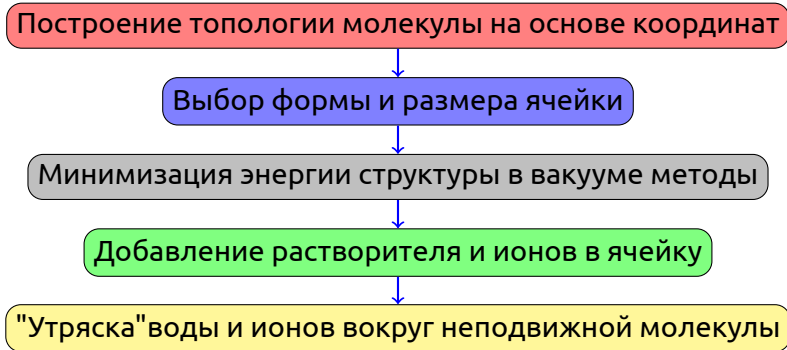
\mathbf{W} - матрица параметров определяющих силу сопряжения



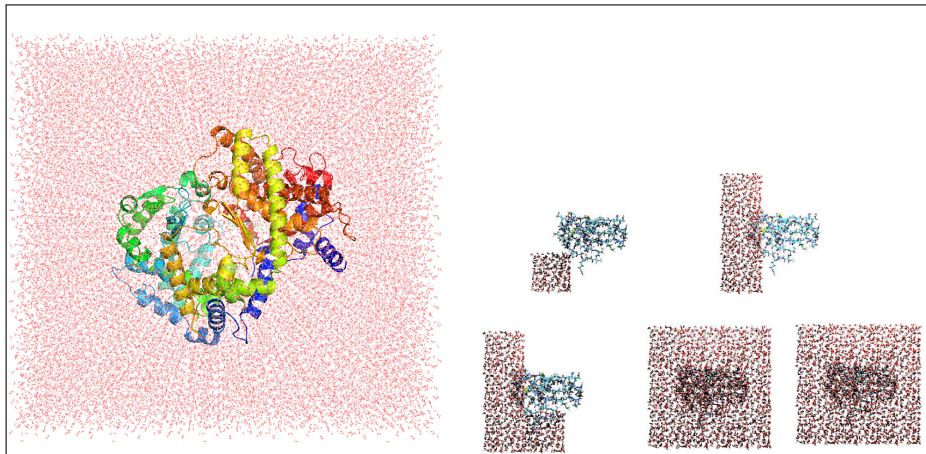
Пример контроля давления



Методология подготовки системы для МД



Добавление воды в ячейку



Молекулярная динамика и неявный растворитель

- Часто явно заданный растворитель, т.е. молекулы воды заменяют потенциалами.
- Методы на основе поверхности доступной растворителю.

$$\Delta G_{solv} = \sum_i \sigma_i ASA_i$$

- Метод Пуассона-Больцмана и его производные



Молекулярная динамика и неявный растворитель

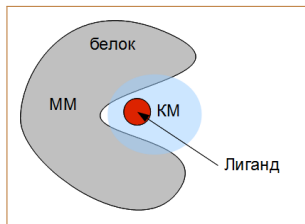
Основные недостатки неявного растворителя:

- РВ и GB учитывают в основном электростатическую составляющую.
- Гидрофобный эффект не учитывается.
- Вязкость как результат столкновений и скоростей не рассчитывается и не учитывается.
- Водородные связи воды с объектом интереса не могут быть учтены.
- Исчезает возможность учёта водных мостиков.



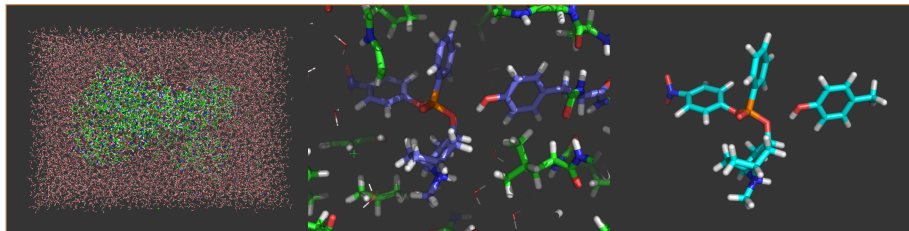
Гибридное QM/MM моделирование

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

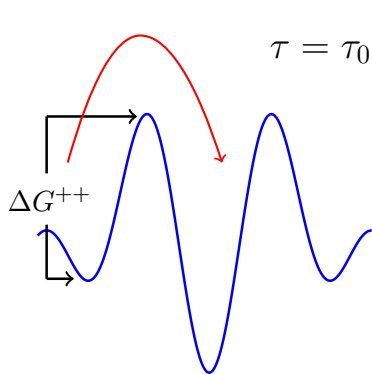


Гибридное QM/MM моделирование

- Атомы связки
- Специальные орбитали



Длина траектории МД



$$\tau = \tau_0 e^{\frac{\Delta G^{++}}{k_b T}}$$

$\Delta G^{++}, \tau \sim 10^{-12} s$	t_{MD}
1 kcal/mol	~ 1.2 ps
5 kcal/mol	~ 1.5 ns
10 kcal/mol	>1 ms



Конформации

- В 1950 году Бартон показал, что химические и физические свойства определяются конформациями молекул.
- Это справедливо и для биополимеров.
- Важны не только оптимальные конформации.



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: 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:

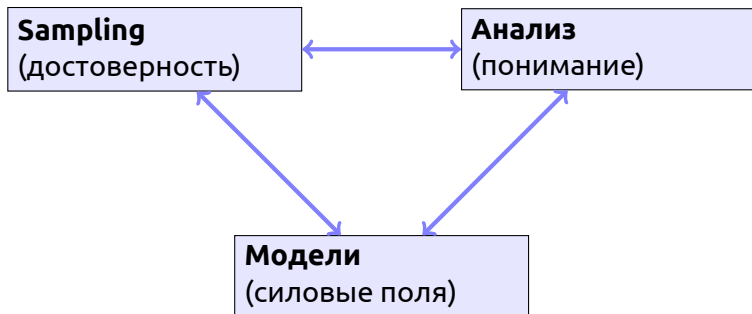
$$P(1 \leftrightarrow 2) = \exp\left(-\frac{\epsilon^2 c N_{df}}{2(1 + \epsilon)}\right) \approx \exp\left(-\epsilon^2 \frac{c}{2} N_{df}\right)$$

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: $\epsilon = 1/\sqrt{N_{atoms}}$.



Вызовы молекулярного моделирования

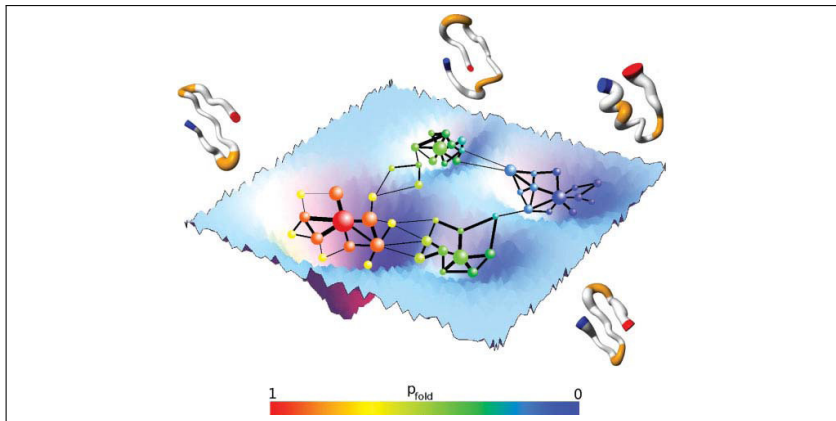


Моделирование самосборки белка

- Как иметь дело с большими временами?
Самые "быстрые" белки собираются за 10-100 μs ,
моделирование должно быть на порядок длинее.
- На сколько хороши наши силовые поля?
Сможем ли мы получать нативное состояние белка без
знания структуры.
Будут ли предсказанные структуры иметь правильные
параметры: ΔG , K_f .
- Сможем ли мы изучать сборку моделированием?
Механизмы, Теория ...



МД в моделировании самосборки белка



PMID: 21387345



МД в моделировании самосборки белка

- Традиционный подход: несколько длинных траекторий
- Альтернатива: методы стохастического кинетического сканирования.

Фолдинг это стохастический процесс с экспоненциальной кинетикой, т.е. количество молекул, которые собрались:

$$f(t) = M[1 - \exp(-kt)]$$

для малых времён:

$$f(t) \sim Mkt \quad M \sim 10,000 \text{procs}, k \sim 1/10,000 \text{ns}, t \sim 20 \text{ns/proc}$$

ожидается, что можно увидеть 20 раз сборку белка.

- Это эффективно.
- Эргодично



Folding@home

English

Folding@home DISTRIBUTED COMPUTING

Home Learn Stats About Us Search

Help unlock the mysteries of disease.

READY.

- What is protein folding?
- Why does it matter?
- How can I help?

SET.

Download
Folding@home

64bit Linux
.deb + .rpm

- Other platforms
- Older versions

FOLD.

- Start folding now!
- Get help.
- Earn points, join a team.

You can help scientists studying Alzheimer's, Huntington's, and many cancers by simply running a piece of software on your computer or **game console**.

Join others around the world to form the world's largest distributed supercomputer.

Make a gift now!

[Click here to make a gift.](#)



Folding@home

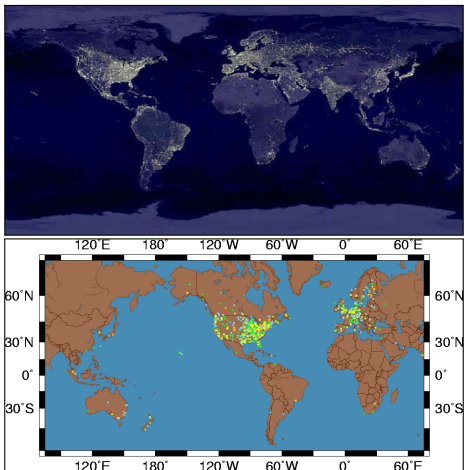


Illustration from Stefan Larson



Результаты

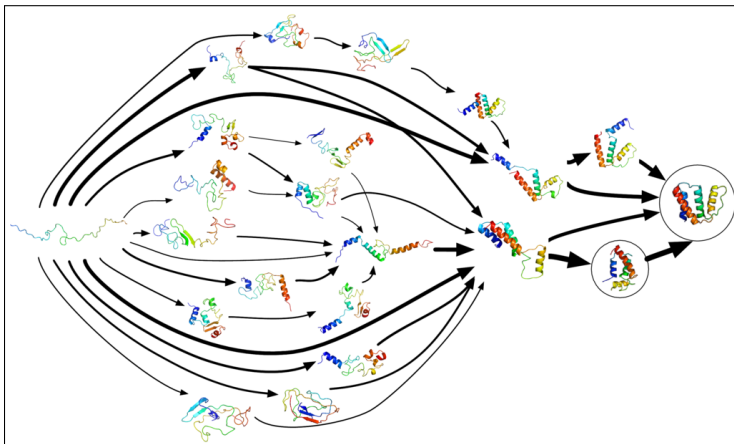


Illustration from Wikipedia

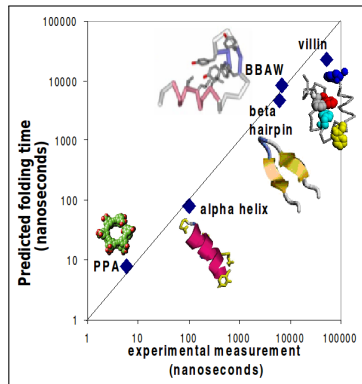


Результаты и ответы

Ответы на наши вопросы:

- Выборка достаточна.
- Силовые поля работают
Сборка происходит правильно
и с нужной скоростью
- Механизм
Вероятно, что самосборка
это не универсальный
механизм и индивидуален
для каждого белка

Illustration from Stefan Larson



FoldIt Game

Основная идея:

- Использование **crowdsourcing** для поиска оптимальной геометрии.
- Пользователь может изменять торсионные углы и добиваться оптимальной энергии.
- В ходе игры для прохождения уровня пользователь должен достичь нужного уровня энергии для структуры
- Создатели программы собираются утилизировать способы придуманные людьми для разработки новых алгоритмов.



FoldIt

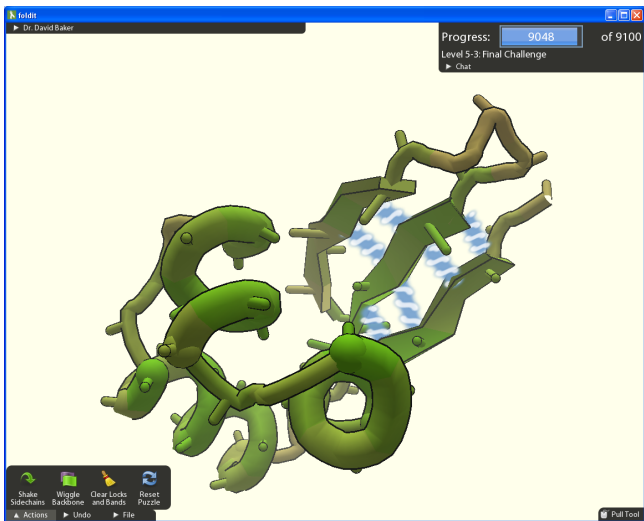


Illustration from Wikipedia

Вопросы?

