

Белковые силовые поля и потенциалы Потенциалы крупноблочных моделей

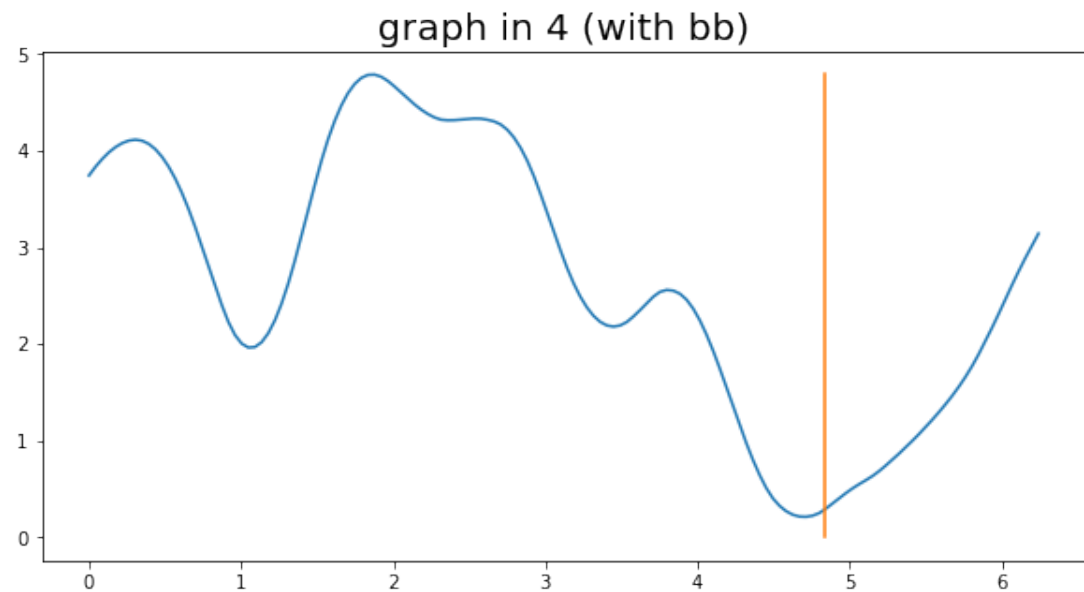
Игнатов Андрей Дмитриевич

aignatov@hse.ru

t.me/a_ignatov

Фолдинг: основная идея

- Когда мы говорим о процессе фолдинга белка, мы предполагаем, что этот процесс представляет собой переход молекулы в наиболее энергетически выгодное состояние (**конформацию**). Такое состояние называется **нативным**.
- Предполагается¹, что нативной конформации соответствует минимум потенциальной энергии белка.



¹ Scott, R. A. Conformational Analysis of Macromolecules. III. Helical Structures of Polyglycine and Poly-L-Alanine [текст] / R. A. Scott, H. A. Scheraga // The Journal of Chemical Physics. — 1966. — т. 45, No 6. — с. 2091—2101.

Силовые поля и потенциалы

- Как вычисляется белковый потенциал?
 - Есть ли теоретически обоснованная и практически подтвержденная формула?
 - Есть ли методология вычисления потенциала?
 - Есть ли какие-то ограничения? Для всех ли белков его возможно вычислить?
 - Насколько легко (вычислительно) вычислить потенциал для произвольного белка?

Силовые поля и потенциалы

- Как вычисляется белковый потенциал?



Единственно верно –

никак.

Вместо этого – есть ряд эмпирически полученных
белковых потенциалов.

Силовые поля и потенциалы

- Главная причина противоречивости потенциалов – неоднозначная взвешенность взаимодействий.
 - Потенциал – взвешенная сумма взаимодействий, которые считаются независимыми друг от друга.
 - Веса для этих взаимодействий могут быть получены только эмпирическим путем, что не имеет под собой строгого теоретического обоснования.
 - Наиболее часто используемые сегодня потенциалы (OPLS, Rosetta) продолжают уточняться.

Силовые поля и потенциалы

- Еще один эмпирический вклад в белковые потенциалы – наличие слагаемых для ускорения оптимизации и создания искусственных ограничений.
- Такие слагаемые не несут явного теоретического значения.
- В результате получается, что любой белковый потенциал уже не столько моделирует реальную природу, сколько является инструментом для построения этой модели.



Силовые поля и потенциалы

- С другой стороны, потенциалы учитывают наибольшее число факторов из всех методов предсказания белковой геометрии.
- Этот факт делает их наиболее точным инструментом предсказания, что сопровождается большой вычислительной сложностью из-за гигантского числа степеней свободы.
- Вывод:
 - ~~Глобальная минимизация~~ 🙅
 - Поиск близкого начального приближения и локальная минимизация 👍

Потенциалы: резюме

+

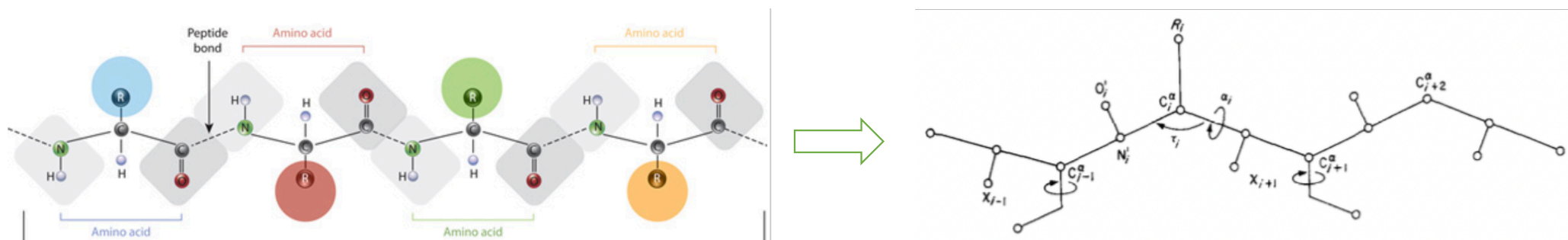
- Высокая точность
- Гибкость метода в приложении к задачам

—

- Высокая вычислительная сложность
- Сильная зависимость от близости начальной точки к минимуму
- Сильная эмпирика подхода

Крупноблочные модели белковой геометрии

- Геометрия белка очень сложна, поскольку подразумевает большое число степеней свободы.
- Для упрощения геометрии предлагаются крупноблочные модели, которые подразумевают обобщение групп атомов:



- Обобщение может происходить на разных уровнях.

Облегчение минимизации: крупноблочные модели

- Крупноблочные модели:
 - Упрощают геометрию
 - Уменьшают число степеней свободы
 - Уменьшают число переменных в задаче
 - Ускоряют минимизацию!
- Однако здесь классические потенциалы мало применимы: разработчики таких моделей предлагают свои потенциалы.
 - Главная причина этого – предельная точность модели: точность 4 – 5 Å обычна для них!
 - В силу разных обобщений потенциалы разных обобщенных моделей также слабо применимы к другим моделям.

Статистические потенциалы

- Статистические потенциалы строятся по логарифмированной статистике значений расстояния между атомами.
 - Важную роль здесь играет нормировка, которая может существенно отличаться в разных подходах.
- Главная проблема: взаимодействие двух атомов рассматривается независимо от всех остальных, что, вообще говоря, не так.
- В то же время, попытка рассмотреть все атомы белка в статистическом потенциале обычно приводит к ухудшению точности предсказания.

Отличие потенциалов

- А сильно ли отличаются полноатомные потенциалы и потенциалы крупноблочных моделей?
 - Логично ожидать, что у крупноблочных моделей потенциал должен быть проще, геометрия же упростилась!

Отличие потенциалов

- А сильно ли отличаются полноатомные потенциалы и потенциалы крупноблочных моделей?
 - Логично ожидать, что у крупноблочных моделей потенциал должен быть проще, геометрия же упростилась!
Увы. Это не всегда так 😊
- Посмотрим на примере, сравним два примера:
 - Полноатомный потенциал OPLS.
 - Потенциал крупноблочной модели Левитта (1976).

Пример: OPLS

- Потенциал OPLS имеет 4 составляющих:

$$E(\vec{r}) = E_{bonds} + E_{angles} + E_{dihedrals} + E_{nonbonded}$$

- $E_{nonbonded}(\vec{r}) = \sum_{i>j} f_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{C_{ij}}{r_{ij}^6} + \frac{q_i q_j e^2}{4\pi\epsilon_0 r_{ij}} \right)$

Можно заметить, что здесь используется запись, похожая на потенциал ван дер Ваальса. Эта часть потенциала моделирует притяжение и отталкивание атомов, которые не имеют химической связи. Сюда же входит кулоновское взаимодействие атомов (электростатическое).

- Два слагаемых моделируют малое изменение длин связей и планарных углов:

- $E_{bonds}(\vec{r}) = \sum_{bonds} K_r (r - r_0)^2$

- $E_{angles}(\vec{r}) = \sum_{angles} k_\theta (\theta - \theta_0)^2$

Пример: OPLS (продолжение)

- Потенциал OPLS имеет 4 составляющих:

$$E(\vec{r}) = E_{bonds} + E_{angles} + E_{dihedrals} + E_{nonbonded}$$

- $E_{torsion}(\vec{r}) = \sum_{torsions} \left(\frac{V_1}{2} [1 + \cos(\phi - \phi_1)] + \frac{V_2}{2} [1 - \cos(2\phi - \phi_2)] + \right.$
 $\left. + \frac{V_3}{2} [1 + \cos(3\phi - \phi_3)] + \frac{V_4}{2} [1 - \cos(4\phi - \phi_4)] \right)$

// Замечание: Эта часть потенциала зависит не от расстояний, а от значений двугранных углов. Это слагаемое играет роль 'ограничителя' возможных значений двугранных углов, что помогает при минимизации.

Пример: OPLS + GB/SA

- В модели GB/SA учитывается свободная энергия сольватации, которая представляется в виде суммы трех слагаемых:

$$G_{sol} = G_{cav} + G_{vdw} + G_{pol}$$

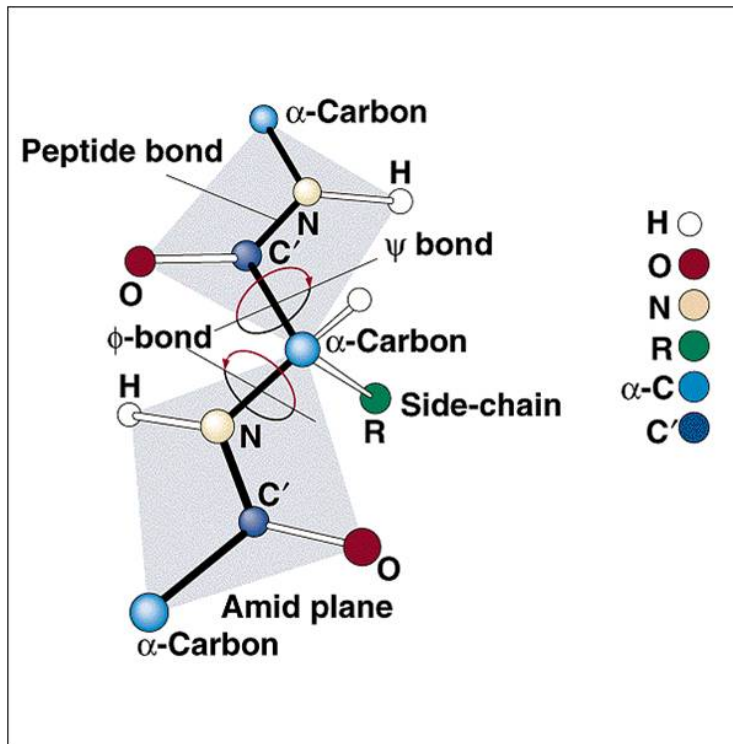
- G_{cav} и G_{vdw} учитывают свободную энергию для формирования полости в растворителе и для любых дисперсионных взаимодействий.

$$G_{cav} + G_{vdw} = \sum \sigma_k SA_k,$$

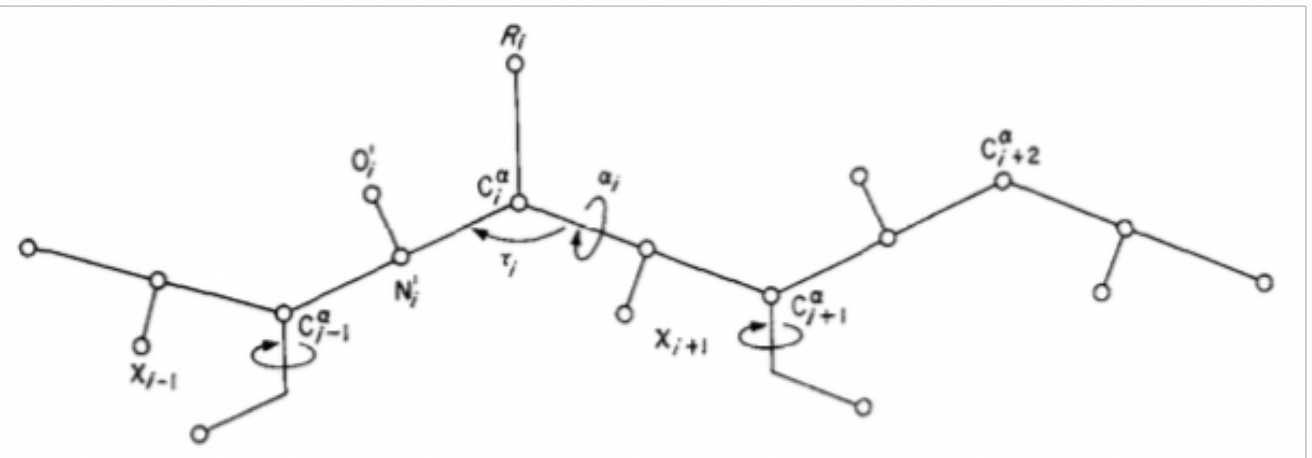
где SA_k - общая площадь поверхности атома k , доступная растворителю. Эта часть энергии называется **полярной** или **гидрофобной**.

- $G_{pol} = 166 \cdot 4.184 \cdot \left(\frac{1}{\epsilon} - 1\right) \sum_{i=1}^n \sum_{j=1}^n \frac{q_i q_j}{\sqrt{r_{ij}^2 + \alpha_{ij}^2} e^{-D_{ij}}}$ - энергостатическая энергия поляризации.

Пример: Крупноблочная модель Левитта²



Copyright © 1997 Wiley-Liss, Inc.



$$\alpha_i = \text{dihedral}(C_{\alpha}^{i-1}, C_{\alpha}^i, C_{\alpha}^{i+1}, C_{\alpha}^{i+2})$$

$$\tau_i = 106^{\circ} + 13^{\circ} \cos(\alpha_i - 45^{\circ})$$

² Michael Levitt. A simplified representation of protein conformations for rapid simulation of protein folding. *Journal of molecular biology*, 104(1):59– 107, 1976.

Пример: Крупноблочная модель Левитта²

$$\begin{aligned} V_{total}(\alpha) = & \sum_{i,j} \epsilon_{i,j} \{3(r_{i,j}^o/r_{i,j})^8 - 4(r_{i,j}^o/r_{i,j})^6\} \\ & + \sum_{i,j:r_{i,j} < 9\text{\AA}} (s_i + s_j)g(r_{i,j}) + \sum_{SSbonds} K_{SS}(r_{i,j}^{SS} - r_o^{SS})^2 \\ & + \sum_{i,j} \{\epsilon_p (r_p^o/r_{NO})^{12} - 2(r_p^o/r_{NO})^6 + (r_p^o/r_{ON})^{12} - 2(r_p^o/r_{ON})^6\} \\ & + 332 \sum_{i,j} q_p^2 \{1/r_{NN} + 1/r_{OO} - 1/r_{NO} + 1/r_{ON}\} \\ & \sum_i \{2 \sum_{k=1}^6 A_k^i \cos[(k-1)\alpha_i] + B_k^i \sin[(k-1)\alpha_i]\} \end{aligned}$$

Пример: Крупноблочная модель Левитта²

Взаимодействие с растворителем

Энергия ван дер Ваальса

Энергия S-S связей

Водородные связи пептидных групп

Энергия non-bonded взаимодействий

$$V_{total}(\alpha) = \sum_{i,j} \epsilon_{i,j} \{3(r_{i,j}^o/r_{i,j})^8 - 4(r_{i,j}^o/r_{i,j})^6\}$$

$$+ \sum_{i,j:r_{i,j} < 9\text{\AA}} (s_i + s_j)g(r_{i,j}) + \sum_{SSbonds} K_{SS}(r_{i,j}^{SS} - r_o^{SS})^2$$

$$+ \sum_{i,j} \{\epsilon_p (r_p^o/r_{NO})^{12} - 2(r_p^o/r_{NO})^6 + (r_p^o/r_{ON})^{12} - 2(r_p^o/r_{ON})^6\}$$

$$+ 332 \sum_{i,j} q_p^2 \{1/r_{NN} + 1/r_{OO} - 1/r_{NO} + 1/r_{ON}\}$$

$$\sum_i \{2 \sum_{k=1}^6 A_k^i \cos[(k-1)\alpha_i] + B_k^i \sin[(k-1)\alpha_i]\}$$

Пример: Крупноблочная модель Левитта²

- Контакт радикалов с растворителем:
 - Так как точное вычисление поверхности контакта с водой затруднено, мы аппроксимируем долю воды, потерянную вокруг атома i из-за близости атома j :

$$g(r_{ij}) = \begin{cases} 1 - \frac{1}{2}(7x^2 - 9x^4 + 5x^6 - x^8), & x < 1, \\ 0, & x \geq 1 \end{cases},$$

где $x = \frac{r_{ij}}{r_{max}}$; $r_{max} = 9\text{\AA}$.