

ПОИСК СОСТОЯНИЙ МАКРОМОЛЕКУЛ С НАИМЕНЬШЕЙ ЭНЕРГИЕЙ

К. Н. Файзуллина
НИУ ВШЭ,

департамент прикладной математики
МИЭМ НИУ ВШЭ

Аннотация

В работе для изучения процесса фолдинга белка используется простейшая гидрофобно-полярная (HP) модель на квадратной решетке. Реализован алгоритм Монте-Карло, основанный на перестройке случайно выбранного фрагмента. Алгоритм применялся для поиска нативных структур для последовательностей вплоть до длины 60.

Введение

Гидрофобно-полярная модель протеина [1] является простейшей решеточной моделью биополимера. Протеин представляет собой цепочку из n аминокислот. Каждая аминокислота относится к одному из двух типов: H (гидрофобная) или P (полярная). Для учета исключенного объема, конформация цепочки моделируется как блуждание на решетке без самопересечений. Топологическими соседями называется пара мономеров, если они являются смежными в пространстве, но не являются последовательными в цепочке. Каждый контакт HH между топологическими соседями имеет энергию взаимодействия ϵ , которую без потери общности примем равной -1. Каждый топологический контакт PP или HP имеет энергию взаимодействия 0. Энергия всей цепочки есть сумма энергий всех топологических контактов в конформации. Нативным состоянием называется состояние, при котором цепочка имеет минимум свободной энергии, то есть число топологических контактов HH максимально.

Поиск минимума функции энергии для HP-модели является NP-полной задачей [2], поэтому для ее решения применяются методы Монте-Карло. В данной работе используется метод последовательной фрагментарной перестройки цепочки на основе изменения энергии (FRESS) [3].

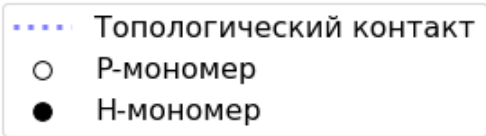
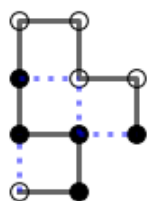


Рис. 1. Пример модели последовательности PHHHHPPPPH

Алгоритм FRESS

Метод FRESS относится к методам Монте-Карло по схеме марковской цепи. В рамках HP-модели набор состояний процесса Марковской цепи — набор всех конформаций фиксированной длины. Метод основан на использовании выборки по значимости, генерируемой при

помощи последовательного изменения случайно выбранного фрагмента.

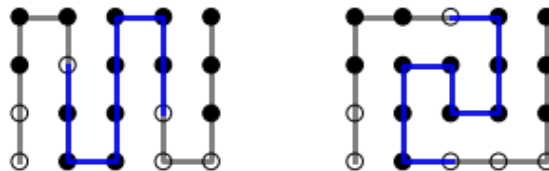


Рис. 2. Пример перестройки фрагмента длины 10.

На каждой итерации алгоритма происходит попытка перестроить случайно выбранный фрагмент цепочки при фиксированной температуре T .

Пусть C - текущая конформация. Одна итерация перестройки фрагмента состоит из нескольких шагов.

1. Случайным образом выбирается $l \in [L_{min}; L_{max}]$ - длина фрагмента, который будет перестраиваться. l выбирается с обратно пропорциональной вероятностью. Наиболее оптимальными параметрами являются $L_{min}=2$ и $L_{max}=12$ [3].

2. Случайно выбирается число s от 1 до $(n-l+1)$, где n - длина цепочки. Это позиция, начиная с которой в цепочке будет удален. Номера выбираются с равной вероятностью.

3. Пусть $C_{t,e}$ - конформация в которой отсутствуют элементы начиная от номера t до e . Начиная с номера s последовательно вставляем остатки, выращаявая удаленную часть.

Для того, чтобы вставить один остаток под номером t в текущую конформацию $C_{t,e}$ с энергией $E_{t,e}$, нужно рассчитать вероятности для возможных позиций. Если позиция уже занята другим мономером цепочки или из нее нельзя за оставшееся число шагов достигнуть точки X_{e+1} (точка, до которой был удален фрагмент), то вероятность принять ее ноль. Из точки X_t можно достичь точку X_{e+1} , если выполняется следующее условие:
 $D(X_t, X_{e+1}) \leq |e-t+1|$ где $D(X_t, X_{e+1})$ - манхэттенское расстояние на решетке.

Каждую из возможных позиций j вставляем в текущую конформацию и вычисляем энергию $E_{t+1,e}$. Позиция выбирается пропорционально $\exp(-(E_{t+1,e}^j - E_{t,e})/T)$. Если подходящих позиций j не оказалось, то начинается новая итерация.

Если выстраивание нового фрагмента прошло успешно, то изменения принимаются согласно правилу Метрополиса:

$$p(C \rightarrow C_{new}) = \min(1, \exp(-(E_{C_{new}} - E_C)/T))$$

Где $E_{C_{new}}$ - энергия новой конформации и E_C - энергия конформации C .

Имитация отжига

Применение метода имитации отжига позволяет постепенно уменьшать вероятность принятия конформации с большей энергией при изучении алгоритмом пространства структур. В данной работе используется простейший протокол: фиксируется $T_{max}=3.5$ и $T_{min}=0.1$. Начиная с максимального значения величина температуры изменяется геометрически (умножая на число меньше единицы, например, на 0.98) через определенное число шагов.

Результаты

Алгоритм применялся для поиска состояний с минимальной энергией для четырех последовательностей. Для упрощения вычислений использовалась квадратная решетка. Найденные минимумы согласуются с

результатами из литературы[4]. На рисунках 3 и 4 показаны примеры найденных конформаций. Данные свернутые конформации имеют сформированное гидрофобное ядро.

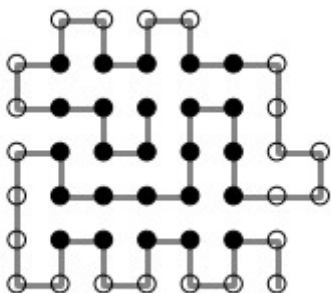


Рис.3. Найденная конформация для последовательности длины 48.

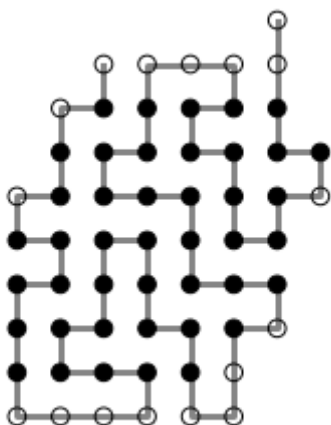


Рис.4. Найденная конформация для последовательности длины 60.

Заключение

В ходе работы был рассмотрен и применен метод фрагментарной перестройки цепочки (FRESS). Была проведена валидация алгоритма с использованием цепочек длиной до 60 мономеров. В дальнейшей работе реализованный программный код применен для последовательностей большей длины. Алгоритм размещен в свободном доступе [5].

Работа выполнена в рамках проекта 18-05-0024 Программы «Научный фонд Национального исследовательского университета „Высшая школа экономики“ (НИУ ВШЭ)» в 2018 — 2019 гг. и в рамках государственной поддержки ведущих университетов Российской Федерации «5-100».

Список литературы:

1. Ken A. Dill, A Lattice Statistical Mechanics Model of the Conformational and Sequence Spaces of Proteins / Kit Fun Lau, Ken A. Dill // *Macromolecules J.* – 1989 – Vol 22 – P. 3986-3997.
2. B. Berger, T. Leighton, Protein Folding in the Hydrophobic-Hydrophilic (HP) Model is NP-Complete / B.

Berger, T. Leighton // *Journal of Computational Biology* – 1998 – Vol 5 – P. 27-40.

3. Z. Jinfeng, S. Kou, J. Liu, Biopolymer Structure Simulation and Optimization via Fragment Regrowth Monte Carlo / Z. Jinfeng, S. Kou, J. Liu // *Journal of Chemical Physics* – 2007 – Vol 126(22) 225101.

4. F. Liang, W.Wong, Evolutionary Monte Carlo for protein folding simulations / F. Liang, W.Wong // *Journal of Chemical Physics* – 2001 – Vol 115 - P.3374-3380.

5. K. Faizullina (2019), <https://github.com/kamilla0503/fress>, DOI: 10.5281/zenodo.3550647