

Zwijanie białek

Zadanie zaliczeniowe z przedmiotu "Programowanie i Metody Numeryczne"

1 Wstęp

Białka to duże cząsteczki syntetyzowane w komórkach wszystkich organizmów żywych. Białka zbudowane są z łańcucha aminokwasów połączonych ze sobą wiązaniami kowalencyjnymi. Kluczowe znaczenie dla funkcji danego białka ma jego struktura przestrzenna. O ile stosunkowo łatwo eksperymentalnie określić kolejność aminokwasów, z których składa się dane białko, o tyle trudne jest eksperymentalne zbadanie kształtu białka. Dlatego do ustalenia kształtu białek ("zwijania białek") wykorzystuje się również modelowanie numeryczne. Celem tego zadania jest napisanie programu, który będzie modelował zwijanie białek, przyjmując prosty model białek (*HP lattice*) i wykorzystując algorytm Metropolis'a do znalezienia najsilniej związanej struktury.

2 Algorytm Metropolis'a

Algorytm Metropolis'a to metoda typu Monte Carlo służąca do próbkowania z rozkładu prawdopodobieństwa za pomocą błędzenia losowego. W przypadku zwijania białek polega ona na losowym generowaniu możliwych struktur przestrzennych (konformacji) w celu znalezienia tej o największym prawdopodobieństwie. W algorytmie Metropolis'a zwijanie białek modeluje się jako łańcuch Markowa, co oznacza, że czas jest dyskretyzowany, a prawdopodobieństwo przejścia do kolejnej konformacji zależy jedynie od tej aktualnej, a nie od poprzednich (brak pamięci). Ogólnie, próbkowanie Monte Carlo za pomocą łańcuchów Markowa stosuje się do rozwiązywania problemów o dużej złożoności (co w przypadku zwijania białek związane jest z dużą przestrzenią możliwych konformacji).

Założmy, że naszym celem jest próbkowanie z rozkładu o gęstości prawdopodobieństwa $P(x)$, a znamy funkcję $f(x)$ proporcjonalną do $P(x)$. Zakładamy też symetryczne prawdopodobieństwo przejścia między stanami x i y : $g(x|y) = g(y|x)$. Algorytm Metropolis'a obejmuje następujące kroki:

1. Inicjalizacja - losowy wybór początkowej wartości x_0 .
2. Iteracja (indeks dolny i numeruje kroki)
 - Wybór proponowanej nowej wartości $x_{i+1}^{(p)}$ próbując z rozkładu $g(x_{i+1}^{(p)}|x_i)$
 - Przyjęcie proponowanej wartości z prawdopodobieństwem $\min(1, f(x_{i+1}^{(p)})/f(x_i))$

W przypadku zwijania białek x oznacza konformację, a rozkład z którego próbujemy to rozkład Boltzmann'a:

$$P(x) \propto \exp(-E(x)/(k_B T)),$$

gdzie $E(x)$ to energia konformacji, k_B to stała Boltzmann'a, a T to temperatura.

W algorytmie Metropolis'a problemem może być utknięcie w lokalnym minimum energetycznym. Jeśli początkowa konformacja jest blisko takiego minimum, algorytm z dużym prawdopodobieństwem będzie generował trajektorię w jego kierunku. Tym samym nie zostanie znalezione globalne minimum energetyczne. Aby temu zapobiec, stosuje się technikę symulowanego wyżarzania (*simulated annealing*). Polega ona na

zmniejszaniu temperatury w trakcie symulacji. Na początku, gdy temperatura jest wysoka, algorytm próbuje dużą część przestrzeni konformacji, ponieważ wysoce prawdopodobne są przejścia do konformacji o wyższej energii. Zwiększa to szanse na znalezienie się w okolicy globalnego minimum. Następnie, z powodu spadku temperatury, algorytm będzie próbował konformacje coraz bliżej tego minimum.

3 Model HP na siatce regularnej

Energia konformacji zależy od oddziaływań pomiędzy aminokwasami oraz między aminokwasami a "rozpuszczalnikiem" (płynem otaczającym białko). Prostym modelem zwijania białek jest model HP, w którym uwzględnia się jedynie oddziaływania hydrofobowe. Aminokwasy dzieli się na dwie grupy: H - aminokwasy hydrofobowe i P - aminokwasy polarne (hydrofilowe). Ten prosty model pozwala uchwycić jedną z podstawowych struktur białek, tj. grupowanie się aminokwasów hydrofobowych (tak, by zminimalizować ich styczność z rozpuszczalnikiem).

Dodatkowym uproszczeniem jest dyskretyzacja przestrzeni na regularnej siatce. Aminokwasy mogą znajdować się jedynie w określonych punktach przestrzeni trójwymiarowej: $x, y, z = \dots, -1, 0, 1, \dots$. Aminokwasy sąsiadujące ze sobą w łańcuchu muszą zajmować najbliższe punkty, tj. odległość między nimi musi wynosić 1. Do energii konformacji wkład mają jedynie pary sąsiednich (odległość 1) aminokwasów typu H niebędących kolejnymi elementami łańcucha (niezwiązanych kowalencyjnie). Energia każdej takiej pary wynosi -1 . Przyjmuje się wartość stałej Boltzmanna równą jeden.

Początkową konformację można wygenerować w dowolny sposób. Propozycje kolejnych konformacji można tworzyć, stosując ruchy takie, jak:

- Przesunięcie końca - zmiana położenia jednego z ostatnich aminokwasów w łańcuchu.
- Obrót narożnika - przykładowo, w dwóch wymiarach, dla trzech kolejnych aminokwasów: $(0, 0), (1, 0), (1, 1) \rightarrow (0, 0), (0, 1), (1, 1)$.
- Obrót - przykładowo, w dwóch wymiarach, dla czterech kolejnych aminokwasów $(0, 1), (0, 0), (1, 0), (1, 1) \rightarrow (0, 1), (0, 2), (1, 2), (1, 1)$.

Rodzaj ruchu wykonywanego w danym kroku czasowym może być wybierany losowo.

4 Zadanie

Celem projektu jest napisanie programu do modelowania zwijania białek przy użyciu modelu HP na trójwymiarowej siatce regularnej. Zaimplementuj trzy wyżej opisane rodzaje ruchów aminokwasów. Program powinien zostać użyty do przewidzenia struktury ubikwityny. Jest to niewielkie, powszechnie występujące białko, które w modelu HP można opisać łańcuchem:

PHPHHHHHHPHPHPHHPPPPHPPPPHHHPPPPHPPPHHPHPHHHHPPPPHHHHPPHPHPHHHHHHHHHHHPHHPP.

Zastosuj symulowane wyżarzanie zaczynając od temperatury T_0 , a kończąc na T_∞ , ze zmianą temperatury opisaną $T_{i+1} = \max(\alpha T_i, T_\infty)$. W celu znalezienia konformacji o najniższej energii, przetestuj czułość wyników na następujące parametry (w nawiasach podane są orientacyjne wartości parametrów, od których dobrze zacząć testy):

- $T_0(10)$
- $T_\infty(1)$
- $\alpha(0.999)$
- liczbę kroków (10^4)
- Sposób inicjalizacji położeń: w jednej linii vs. losowy (dla losowego warto wykonać zbiór symulacji)

Napisz sprawozdanie, zawierające opis programu i wykonanych symulacji. Przygotuj wizualizację końcowej konformacji oraz zmian energii w kolejnych krokach algorytmu. Kończącą konformację zapisz w pliku tekstowym w formacie "x y z", zapisując położenie kolejnych aminokwasów w kolejnych liniach. Podaj liczbę proponowanych i zaakceptowanych ruchów każdego rodzaju. Dodatkowo punktowane będzie przygotowanie animacji, przedstawiającej proces zwiłania białka.

Opracowanie: Piotr Dziekan.