

Model HP

Z BioFizInfo
Model HP

Model HP

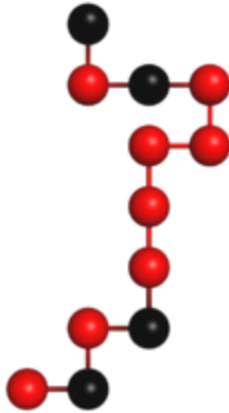
Model HP (*hydrophobic-polar protein folding model*) to model polimeru wykorzystywany w badaniach nad ogólnymi zasadami rządzącymi procesem zwijania białek. Badania tego typu w przypadku modeli pełnoatomowych wiążą się ze znacznymi kosztami obliczeniowymi, podczas gdy w modelu HP, ze względu na uproszczoną charakterystykę układu, możliwe jest przeprowadzenie krótkiej symulacji (trwającej od kilku minut do kilku godzin), w trakcie której układ jest w stanie osiągnąć wszystkie możliwe mikrostanu [\[1\]](#).

Spis treści

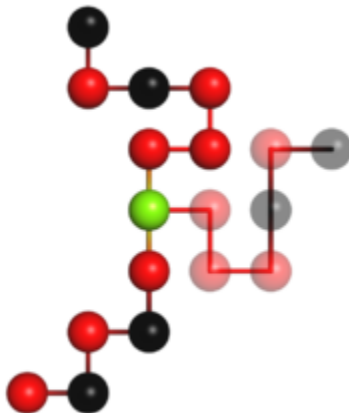
[\[ukryj\]](#)

- [1 Model HP](#)
 - [1.1 Wstęp](#)
 - [1.2 Sieć](#)
 - [1.3 Sekwencja](#)
 - [1.4 Oddziaływania](#)
 - [1.5 Średnia po zespole](#)
 - [1.6 Metoda Monte Carlo - algorytm Metropolis](#)
 - [1.7 Transformacje](#)
 - [1.8 Symulowane wyżarzanie \(Simulated Annealing\)](#)
 - [1.9 Zamiana replik \(Replica Exchange Monte Carlo\)](#)
 - [1.10 Symulacje planowane na ćwiczeniach](#)
 - [1.10.1 Model HP "na piątke"](#)
 - [1.11 Linki zewnętrzne](#)

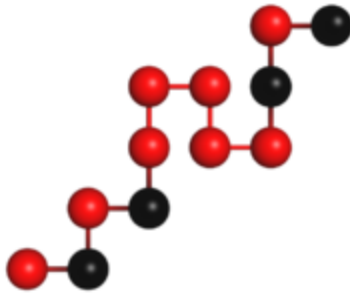
Wstęp



(1) Dwuwymiarowy model HP o sekwencji: HPHPHHHHHHPHP (wizualizacja w PyMOLu). Na czerwono zaznaczono aminokwasy hydrofobowe (H), na czarno aminokwasy polarne (P). Ponieważ w powyższym mikroście nie występują kontakty H-H, energia wynosi 0.



(2) Obrót wokół szóstego aminokwasu (zaznaczono na zielono) skutkuje utworzeniem kontaktu H-H między ósmym i piątym aminokwasem.



(3) Transformacja została zaakceptowana, liczba kontaktów H-H wynosi 1, zatem nowa energia układu wynosi $-\varepsilon$.

Idea modelu HP opiera się na obserwacji, iż kluczową rolę w procesie zwijania białek pełni efekt hydrofobowy (w tym kontekście spotkać się można z terminem: "oddziaływania hydrofobowe"). W podstawowym modelu HP polimer zbudowany jest z monomerów H (hydrofobowych) oraz P (polarnych), przy czym wkład do energii pochodzi jedynie od H. Można więc myśleć o modelu HP jak o modelu białka, w którym alfabet aminokwasów ograniczony został do zbioru {H,P}. Aminokwasy znajdują się w węzłach sieci kwadratowej (*square lattice*) w przypadku modelu dwuwymiarowego (2D), bądź w węzłach sieci sześcienniej (*cubic lattice*) w przypadku modelu trójwymiarowego (3D). Dwa aminokwasy nie mogą znajdować się w tym samym węźle. Natomiast jeśli dwa aminokwasy połączone są wiązaniem (przez analogię do wiązania peptydowego między aminokwasami w białkach), to muszą się one znajdować w sąsiednich węzłach. **Mikrostan** układu można określić poprzez: sekwencję peptydu oraz współrzędne poszczególnych aminokwasów.

Ewolucję układu w modelu HP zadaje zestaw dozwolonych transformacji struktury oraz rozkład prawdopodobieństwa przejść między mikrostanami. Przykładem dozwolonej transformacji może być obrót części białka o pewien kąt wokół wybranego aminokwasu (przykład przedstawiono po prawej). W przypadku modelu 2D istnieją trzy możliwe nietrywialne obroty. Jeżeli po dokonaniu obrotu żadne dwa aminokwasy nie zajmują tego samego punktu w przestrzeni, obrót uznajemy za dozwolony.

Po dokonaniu dozwolonej transformacji prawdopodobieństwo akceptacji nowego mikrostanu zależy jedynie od zmiany wartości energii. Innymi słowy: to, czy zaakceptujemy mikrostan uzyskany w wyniku transformacji zależy

jedynie od mikrostanu przed transformacją; wcześniejsza historia układu nie ma tu znaczenia. Zatem ewolucja peptydu (ciąg mikrostanów wygenerowany w toku symulacji) jest realizacją **procesu stochastycznego**, w którym prawdopodobieństwo zdarzenia (akceptacja nowego mikrostanu) zależy jedynie od wyniku poprzedniego. Proces stochastyczny tego typu w przypadku dyskretnej przestrzeni stanów nazywany jest **łańcuchem Markowa**.

Sieć

Niech:

$$\mathbf{e}_x = (1, 0), \mathbf{e}_y = (0, 1)$$

będą wektorami bazowymi w przypadku dwuwymiarowym, zaś:

$$\mathbf{e}_x = (1, 0, 0), \mathbf{e}_y = (0, 1, 0), \mathbf{e}_z = (0, 0, 1)$$

wektorami bazowymi w przypadku trójwymiarowym. Siecią kwadratową nazywać będziemy zbiór:

$$LATTICE_{2D} = \{x\mathbf{e}_x + y\mathbf{e}_y \mid x, y \in \mathbb{Z}\}$$

zaś zbiór:

$$LATTICE_{3D} = \{x\mathbf{e}_x + y\mathbf{e}_y + z\mathbf{e}_z \mid x, y, z \in \mathbb{Z}\}$$

nazwiemy siecią sześcienną. Element sieci (węzeł) opisujemy przez podanie dwóch, bądź trzech liczb całkowitych (współrzędnych węzła), przykładowo dla sieci sześcienniej: (0,1,-10).

Powiemy, że węzły **a** i **b** sąsiadują ze sobą na siatce (ozn. **a** ~ **b**), jeżeli istnieje wektor bazowy **e** taki, że:

$$\mathbf{a} = \mathbf{b} + \mathbf{e} \quad \vee \quad \mathbf{b} = \mathbf{a} + \mathbf{e}.$$

Niech $CHAIN_n = \{1, \dots, n\}$ będzie zbiorem aminokwasów tworzących peptyd, gdzie n - długość peptydu. Wówczas strukturę przestrzenną wyrażać będziemy przez funkcję:

$$s: CHAIN_n \rightarrow LATTICE$$

spełniającą warunki:

$$\begin{aligned} \mathbf{s}(1) &= (0, 0, 0) \\ \forall_{i < n} \mathbf{s}(i+1) &\sim \mathbf{s}(i) \\ \forall_{i \neq j} \mathbf{s}(i) &\neq \mathbf{s}(j) \end{aligned}$$

Podanie struktury (w postaci funkcji \mathbf{s}) nie wystarcza do określenia mikrostanu układu, potrzebna jest jeszcze sekwencja.

Sekwencja

Sekwencja łańcucha określona jest przez wzorzec hydrofobowy

$Pat: CHAIN_n \rightarrow \{H, P\}$. Rozważany model dzieli aminokwasy ze względu na właściwości oddziaływań dalekozasięgowych na dwie kategorie: hydrofobowe (H) oraz polarne (P). *Dalekozasięgowość* oddziaływań odnosi się do wzajemnych położeń aminokwasów w sekwencji, a nie w przestrzeni. Przykładowo: o obecności oddziaływań dalekozasięgowych możemy mówić w przypadku pary aminokwasów o numerach 1 i 4, bądź: 2 i 9, ale nie w przypadku par: 1 i 3, czy też 4 i 5. Szczegóły w poniższej sekcji *Oddziaływania*.

Oddziaływania

W najprostszym modelu HP rozważa się jedynie oddziaływania dalekozasięgowe pomiędzy aminokwasami hydrofobowymi. Energia danego mikrostanu zależy od liczby kontaktów występujących między aminokwasami H, niesąsiadującymi w peptydzie.

Niech

$$K_{HH}(\mathbf{s}) = \#\{\{i, j\}: |i - j| > 1, \quad \mathbf{s}(i) \sim \mathbf{s}(j), \quad Pat(i) = Pat(j) = H\}$$

będzie liczbą kontaktów H-H w peptydzie o strukturze \mathbf{s} . Energia układu wyraża się przez:

$$E(\mathbf{s}) = -\varepsilon \cdot K_{HH}(\mathbf{s})$$

gdzie $\varepsilon > 0$.

Interakcje pomiędzy aminokwasami hydrofobowymi odzwierciedlają ich tendencję do kierowania się do wewnątrz białka i tym samym unikania kontaktu z wodą. Należy podkreślić, że model HP uwydatnia jeden aspekt procesu zwijania białek (efekt hydrofobowy), ignoruje natomiast oddziaływania lokalne występujące w rzeczywistym białku - "sztywność" łańcucha (objawiająca się niedozwolonymi wartościami kątów φ - ψ na wykresie Ramachandrana) oraz

wiązania wodorowe (istotne w α -helisach i β -karkach). Proste modele, jak model HP, skłaniają do zadawania pytań: *Które z własności białek udaje się odtworzyć pomimo poczynionych przybliżeń?*

Średnia po zespole

Niech A będzie pewną własnością fizyczną badanego układu. Mikrostan układu oznaczmy przez $\mathbf{x} = (x_1, \dots, x_n)$, gdzie n jest liczbą stopni swobody. Przyjmujemy, że własność A objawia się jako średnia po próbce pewnej przestrzeni mikrostanów, tzn.:

$$\langle A \rangle = Z^{-1} \int_{\Omega} A(\mathbf{x}) f(\mathcal{H}(\mathbf{x})) d\mathbf{x}$$

gdzie f jest funkcją rozkładu gęstości prawdopodobieństwa, Ω jest przestrzenią dostępnych stanów układu (nazywana również w szerszym kontekście: przestrzenią fazową), zaś:

$$Z = \int_{\Omega} f(\mathcal{H}(\mathbf{x})) d\mathbf{x}$$

to **sumą statystyczną** nazywana również **funkcją podziału**. Rozkład f określa odpowiedni zespół statystyczny (mikrokanoniczny, kanoniczny,...).

W przypadku modelu HP liczba mikrostanów układu jest skończona (ozn. N), zaś średnią wartość A wyraża się w postaci sumy po dostępnych mikrostanach układu:

$$\langle A \rangle = \sum_{i=1}^N A_i \cdot p_i$$

gdzie p_i jest prawdopodobieństwem uzyskania przez układ i -tego mikrostanu. Prawdopodobieństwo, że układ o określonej, stałej temperaturze T (używa się też określenia: w kontakcie z termostatem o temperaturze T) osiągnie i -ty mikrostan o energii E_i , dane jest rozkładem Boltzmanna:

$$p_i = \frac{e^{-E_i/kT}}{\sum_{j=1}^N e^{-E_j/kT}}$$

gdzie k - stała Boltzmanna. Suma w mianowniku zapewnia normalizację rozkładu p_i :

$$\sum_{j=1}^N p_j = 1$$

Metoda Monte Carlo - algorytm Metropolis

W celu wyznaczenia $\langle A \rangle$ dla układu o temperaturze T wystarczy dysponować metodą do generowania mikrostanów zgodnie z rozkładem Boltzmann. Metodą tego typu jest algorytm Metropolis.

Wprowadźmy następujące oznaczenie:

$$\pi(a) = e^{-E_a/kT}$$

gdzie a jest mikrostanem o energii E_a . Istotą algorytmu Metropolis jest stworzenie ciągu mikrostanów, będący realizacją łańcucha Markowa z prawdopodobieństwem przejść, zależącym od różnicy energii kolejnych mikrostanów. W przypadku modelu HP algorytm Metropolis przebiega następująco:

1. Zainicjuj ciąg mikrostanów, tworząc pierwszy, dowolny mikrostan X .
2. Oblicz energię E_X .
3. Dokonaj dozwolonej transformacji peptydu (transformacje opisano dalej, dla ustalenia uwagi - dokonujemy obrotu części peptydu wokół losowo wybranego aminokwasu).
4. Wyznacz energię E_Y uzyskanego w wyniku transformacji mikrostanu Y .
5. Zaakceptuj nowy mikrostan ($X:=Y$) z prawdopodobieństwem

$p(X, Y) = \min \left\{ 1, \frac{\pi(X)}{\pi(Y)} \right\}$ i wróć do 3. albo zakończ, jeśli wygenerowano ciąg o długości M .

Dysponując ciągiem (\mathbf{x}_n) mikrostanów uzyskanych w algorytmie Metropolis, możemy wyznaczyć średnią wartość A :

$$\langle A \rangle \approx \frac{1}{M} \sum_{i=1}^M A(\mathbf{x}_i)$$

Transformacje

...

Symulowane wyżarzanie (*Simulated Annealing*)

Najprostszym sposobem znajdowania konformacji o minimalnej energii jest systematyczne obniżanie temperatury podczas symulacji. Wadą tego rozwiązania jest to, że układ może łatwo zatrzymać się w lokalnym minimum energii, z którego wyjście przy obniżonej temperaturze okaże się niemożliwe (precyzyjniej: niezwykle mało prawdopodobne). Ponadto, zbieżność algorytmu przy niskich temperaturach jest dosyć wolna. Układ może stracić dużo czasu (kroków symulacji) w niecce reprezentującej lokalne minimum, bądź oscylując między stanami o tej samej energii.

Zamiana replik (*Replica Exchange Monte Carlo*)

W tym podejściu równolegle symuluje się wiele kopii układu, każdy w innej, stałej temperaturze. Załóżmy, że w pewnym momencie symulacji algorytmu Metropolis'a i -ta replika o temperaturze T_i jest w mikroście \mathbf{x}_i o energii $E(\mathbf{x}_i)$, zaś j -ta replika w odpowiednio: temperaturze T_j , mikroście \mathbf{x}_j i energii $E(\mathbf{x}_j)$. Z rozkładu jednostajnego losujemy parę kolejnych replik (i, j) , które z prawdopodobieństwem p_s zostaną zamienione miejscami:

$$p_s = \min\{1, e^{-\Delta}\},$$

gdzie

$$\Delta = \left(\frac{1}{kT_j} - \frac{1}{kT_i} \right) (E(\mathbf{x}_i) - E(\mathbf{x}_j))$$

Po zamianie i -ta replika symulowana jest w temperaturze T_j , a j -ta w temperaturze T_i . Ponieważ prawdopodobieństwo zamiany maleje wykładniczo wraz ze wzrostem różnicy temperatur, rozważamy wyłącznie repliki sąsiednie.

Zamiana temperatur zmienia krajobraz energetyczny. W bardzo wysokich temperaturach bariery energetyczne znikają i można domniemywać, że prawdopodobieństwo odwiedzenia mikrostanu jest zadane rozkładem jednostajnym. Repliki, które utknęły w lokalnych minimach mogą zostać z nich wyzwolone przez przeniesienie do wyższej temperatury.

Wymiany nie powinny być zbyt częste. Po zmianie temperatury układ przez pewien czas się stabilizuje i przemieszcza w najbardziej prawdopodobny region krajobrazu energetycznego.

Po zakończeniu symulacji średnia A w temperaturze T_i może zostać oszacowana wzorem:

$$\langle A \rangle \approx \frac{1}{M} \sum_{i=1}^M A(\mathbf{x}_i)$$

Symulacje planowane na ćwiczeniach

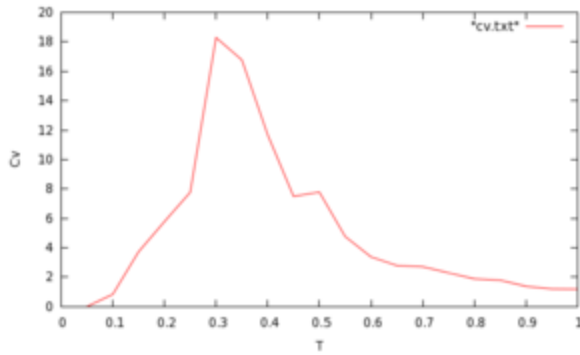
Celem ćwiczeń jest zaimplementowanie modelu HP (domyślnie w języku *Java*) i sprawdzenie wyników przedstawionych w publikacji K. A. Dilla z 1995 roku. W pierwszym etapie przeprowadzimy symulowane wyżarzania modeli dwuwymiarowych trzech polimerów o sekwencjach:

- **PHPPHPPHHPPHHPPHPPHP**
- **HPPHPPHPHPPHPPHPPHH**
- **HPPPHHPPHPPHPPHPPHH**

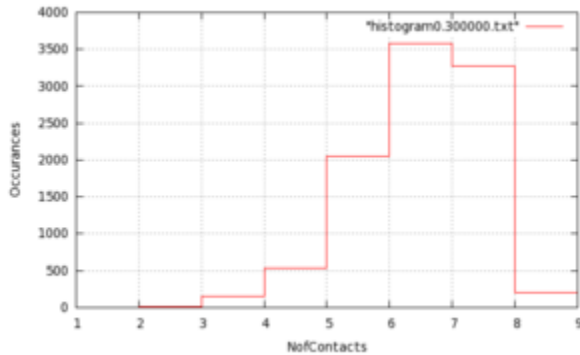
(na ostatnich ćwiczeniach zawęziłem symulacje obowiązujące na 25. marca do jednej sekwencji, pierwszej na powyższej liście). W przypadku każdego peptydu temperaturą początkową będzie $T_{max} = 1$, a temperaturą końcową $T_{min} = 0.1$. Temperatura będzie w trakcie symulacji maleć o czynnik $\delta = 0.05$. Po przeprowadzeniu $x = 10000$ transformacji (dokonując akceptacji/odrzuceń wygenerowanych konformacji) w danej temperaturze T , przeprowadzamy kolejnych x kroków w temperaturze $T - \delta$ i tak dalej, aż do osiągnięcia T_{min} .

W każdej temperaturze wygenerujemy $m(T) < x$ konformacji, które akceptować będziemy na drodze algorytmu Metropolis'a i dla każdej takiej próby możemy wyznaczyć ciepło właściwe układu C_v , średni moment bezwładności I oraz histogram wystąpień stanów w zależności od liczby kontaktów.

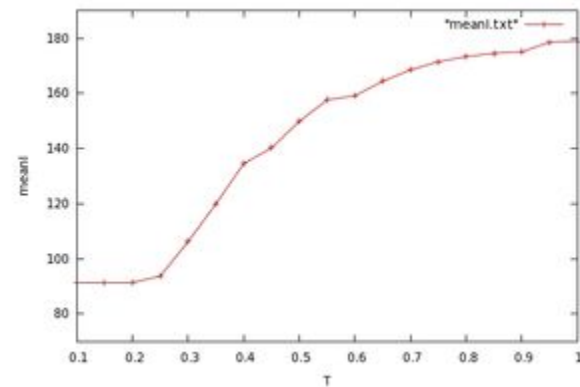
(z ostatnich ćwiczeń: na 25. marca proszę wykonać wykres $C_v(T)$ oraz histogramy dla poszczególnych temperatur. Załączone rysunki przedstawiają czego spodziewać się można po wynikach).



Ciepło właściwe w funkcji temperatury, uzyskane w symulowanym wyżarzaniu.



Histogram wystąpień stanów w zależności od liczby kontaktów.



Średni moment bezwładności w funkcji temperatury.

Ciepło właściwe wyraża się wzorem:

$$C_V(T) = \frac{\langle E(T)^2 \rangle - \langle E(T) \rangle^2}{kT^2}$$

Warto zauważyć, że w C_V jest proporcjonalne do wariancji energii i można ją

interpretować jako "miarę rozrzutu energii" stanów, jakie osiąga układ w danej temperaturze.

Moment bezwładności definiujemy następująco:

$$I = \sum_{i=1}^n (s_i - s_0)^2$$

gdzie s_0 jest środkiem ciężkości, a n liczbą aminokwasów w strukturze.

Natywna struktura większości białek jest globularna. Można więc przyjąć, że moment bezwładności jest dobrym przybliżeniem "stopnia zwinięcia białka".

Model HP "na piątkę"

Na ocenę bardzo dobrą z tej części ćwiczeń należy przeprowadzić dodatkowo (poza symulacją na zaliczenie, na 25. marca) dwie symulacje, dla białek o sekwencjach:

- HPPPHHPPHPHHHHHH
- HPPPHHPPHPHHPHHH

które różnią się sekwencyjnie tylko na jednej pozycji (praca Dilla, str. 20). Specyfikacje symulacji są te same, co poprzednio. Należy wykonać analogiczne histogramy wystąpień kontaktów, wykresy $C_v(T)$ oraz - analogiczny do wykresu C_v - wykres $\langle I(T) \rangle$.

Czy różnica sekwencyjna na jednej pozycji powoduje zmianę minimalnej wartości energii dla tych dwóch białek? Czy obydwa białka "zwijają się" do minimum energii? A jak nie - to do minimum jakiego potencjału?

Linki zewnętrzne

- [Model HP w Wikipedii](#)
- [Mikrostan w Wikipedii](#)
- [Proces stochastyczny w Wikipedii](#)
- [Łańcuch Markowa w Wikipedii](#)
- [Suma statystyczna \(funkcja podziału\) w Wikipedii](#)
- [Algorytm Metropolis'a](#)
- [Praca K. A. Dilla z 1995 roku](#)
- [Strona WWW projektu PyMOL](#)

Źródło „http://essa/wiki/Model_HP”