

Bioinforamtyka 2  
Sprawozdanie 8

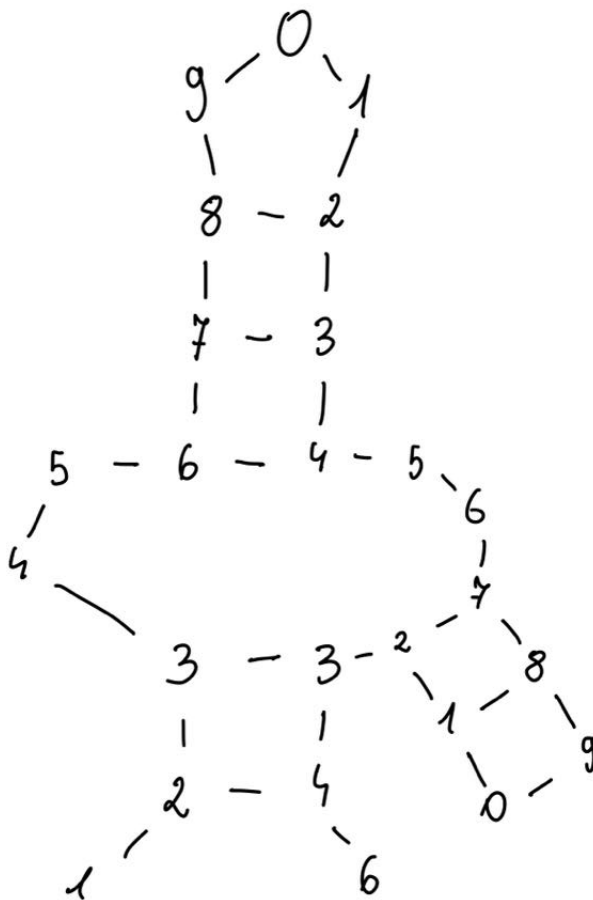
Kaja Wróblewska

21 grudnia 2024

## 1 Zadanie 1

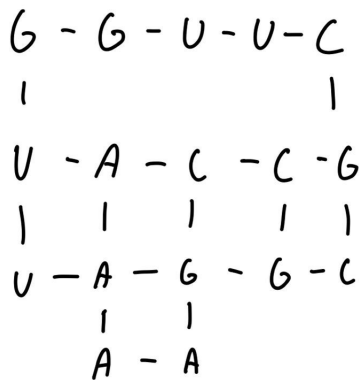
Jaka będzie struktura dla:

1234567890123456789012346

$$.(((..(((..)))..((..))))).$$


## 2 Zadanie 2

Jaka będzie struktura dla:  
 UUGGUUCGCCAAAAGGC  
 ((.....[[[D))..]])



## 3 Zadanie 3

Czy rzeczywista liczba struktur będzie większa czy mniejsza od przewidzianej w zad. 3.  
 Proszę podać ostateczny wzór. Odpowiedź uzasadnij.

Przykład CCGG:

$C_1$  - 2 opcje przyłączenia

$C_2$  - 1 opcja przyłączenia

największe skojarzenie:  $2 * 1 = 2! = 2$

Przykład CCGGG:

$C_1$  - 3 opcje przyłączenia

$C_2$  - 2 opcja przyłączenia

największe skojarzenie:  $3 * 2 = 3! = 6$

Przykład CCGGAAUUU:

CCGG - największe skojarzenie = 2

AAUUU - największe skojarzenie = 6

CCGGAAUUU - największe skojarzenie =  $2 * 6 = 12$

Przykład CCCAAAGGG:

CCCGGG:

$C_1$  - 3 opcje

$C_2$  - 2 opcje

$C_3$  - 1 opcja

największe skojarzenie:  $3 * 2 * 1 = 3! = 6$

AAA:

0 opcji

największe skojarzenie:  $0! = 1$

CCCAAAGGG:

największe skojarzenie:  $6 * 1 = 6$

Wzór ogólny:

1° jeżeli liczba zasad z odpowiadających sobie par jest równa liczymy ze wzoru doskonałego

2° jeżeli liczba jednego rodzaju zasad z pary równa się 0 to największe skojarzenie wynosi  $0!$

3° jeżeli liczba zasad z odpowiadających sobie par nie jest równa, skojarzenie wynosi  $\binom{n}{k} * 2$ , gdzie  $n > k$  i  $n, k$  - ilość zasad z pary

W przypadku CCCAAAGGG największe skojarzenie jest równe skojarzeniu doskonałemu, ponieważ mamy po równo C i G, a A nie mają żadnej pary (brak par AU, wzór doskonały daje  $0! = 1$ )

## 4 Zadanie 4

1. Wykonać predykcje dla sekwencji:

(a) Nussinov:

....(((((((.....)))))))))

(b) RNAfold

.....((((((.....))))))

2. Porównać te struktury i uzasadnić dlaczego się od siebie różnią.

RNAfold uwzględnia dokładne modele energetyczne (np. uwzględniając energie pętli), które mogą preferować stabilniejsze układy, podczas gdy Nussinov optymalizuje jedynie liczbę par zasad.

## 5 Zadanie 5

Zmodyfikuj implementację algorytmu Nussinov (Python) tak, aby:

1. uwzględniał pary wobble G-U

2. zamiast maksymalizacji liczby par realizowana była maksymalizacja punktacji przypisanej różnym parom. Np.  $\text{score}('G', 'C')=3$ ,  $\text{score}('A', 'U')=2$ ,  $\text{score}('G', 'U') = 1$

Odpowiedź uzyskana przy zmodyfikowaniu funkcji pair\_check:

(((((((.....))))))....

## 6 Zadanie 6

### 1. Struktura 1

(a) Wzór ogólny:

$$\Delta G_{\text{Strucutre 1}}^{\circ} = \Delta G_{\text{CCGCGG}}^{\circ} + \Delta G_{\text{Watson-Crick}}^{\circ} + \Delta G_{\text{Dangling end}}^{\circ} + \Delta G_{\text{Penalty AU}}^{\circ}$$

(b) Pętla:

$$\Delta G_{\text{CCGCGG}}^{\circ} = 3.6$$

(c) Watson-Crick helices:

$$\begin{aligned}\Delta G_{\text{Watson-Crick}}^{\circ} &= \Delta G_{37}^{\circ}(\text{intermolecular initiation}) + \Delta G_{37}^{\circ} \left( \frac{5' \text{AU} 3'}{3' \text{UA} 5'} \right) + (\dots) \\ &= 4,09 + 0,45 + [-1,10 - 0,93 - 1,33 - 0,93 - 2,24] = -1,44\end{aligned}$$

(d) Dangling end:

$$\text{C połączzone z AU} = -0.1$$

(e)  $\Delta G_{\text{Penalty AU}}^{\circ} = 0.45$

(f) Oszacowana energia swobona =  $3.6 - 1.99 - 0.1 + 0.45 = 1.41$

### 2. Struktura 2

(a) Wzór ogólny:

$$\Delta G_{\text{Strucutre 2}}^{\circ} = \Delta G_{\text{GUUAAU}}^{\circ} + \Delta G_{\text{Watson-Crick}}^{\circ} + \Delta G_{\text{Dangling end}}^{\circ}$$

(b) Pętla:

$$\Delta G_{\text{GUUAAU}}^{\circ} = 3.9$$

(c) Watson Crick:

$$\Delta G_{\text{Watson-Crick}}^{\circ} = \Delta G_{37}^{\circ}(\text{intermolecular initiation}) + \Delta G_{37}^{\circ} \left( \frac{5' \text{CG} 3'}{3' \text{GC} 5'} \right) + (\dots) = -3.62$$

(d) Dangling end:

$$\Delta G_{\text{Dangling end}}^{\circ} = -0.3$$

(e) Oszacowana energia swobodna =  $3.9 - 3.62 - 0.3 = -0.02$

## 7 Zadanie 7

Wykorzystując doświadczenie zdobyte w zadaniu 6 zaproponuj sekwencje dla podanej struktury w formacie Vienna, która uzyskuje możliwie najniższą energię (weryfikacja poprzez RNA-fold). W odpowiedzi proszę też zawrzeć uzyskaną energię.

Odpowiedź:

AGGGGGGAAACCCCCAGGCCCCGCCA, energia -14.40

## 8 Zadanie 8

Policzyć prawdopodobieństwo każdej struktury. Stworzyć dwa wykresy zależności  $P(E)$ . Jeden dla unikalnych wartości energii. Drugi sumując wartości prawdopodobieństwa dla struktur o takiej samej energii. Kod do wygenerowania wykresu pierwszego poniżej.

1. Prawdopodobieństwa dla wszystkich struktur posortowane po malejącej energii:

```
[0.1558612808735608, 0.11262810794948216, 0.09574168402862908, 0.09574168402862908,  
0.06918462791788149, 0.058811720327603074, 0.036126567889258464, 0.036126567889258464,  
0.036126567889258464, 0.036126567889258464, 0.03071008215323663, 0.026105694533439246,  
0.026105694533439246, 0.026105694533439246, 0.026105694533439246, 0.026105694533439246,  
0.01603607479266155, 0.01603607479266155, 0.01603607479266155, 0.01603607479266155]
```

Suma prawdopodobieństw wynosi  $0.9999999999999997 \approx 1$

