

# Metody anotowania struktury drugorzędowej

Tomasz Żok

## 1 Rozpoznawanie par kanonicznych

1. Poprawna klasyfikacja pary zasad potrzebuje sprawdzenia kilku parametrów: odległości, wartości kątów płaskich i kątów między płaszczyznami
2. Na potrzeby ćwiczenia zaimplementujemy uproszczony model, który sprawdzi wyłącznie pary kanoniczne na podstawie odległości najważniejszych atomów
3. Aby rozpoznać parę G-C, należy sprawdzić, czy odległości między poszczególnymi atomami spełniają poniższe wymagania:

Atom 1	Atom 2	Min	Max
G:N1	C:N3	2.492	3.150
G:N2	C:O2	2.409	3.135
G:O6	C:N4	2.485	3.185

4. Aby rozpoznać parę A-U, należy sprawdzić, czy odległości między poszczególnymi atomami spełniają poniższe wymagania:

Atom 1	Atom 2	Min	Max
A:N1	U:N3	2.390	3.027
A:N6	U:O4	2.503	3.224

5. Aby rozpoznać parę G-U, należy sprawdzić, czy odległości między poszczególnymi atomami spełniają poniższe wymagania:

Atom 1	Atom 2	Min	Max
G:O6	U:N3	2.440	3.220
G:N1	U:O2	2.400	3.180

6. Napisz program, który wczyta plik w formacie PDB i wypisze strukturę drugorzędową w formacie BPSEQ

```
$ wget http://www.pdb.org/pdb/files/2ZY6.pdb
$ ./program 2ZY6.pdb > output.bpseq
```

output.bpseq

```
1 G 35
2 C 34
3 C 33
4 C 32
5 G 31
6 G 30
7 A 29
8 U 0
9 A 0
10 G 25
11 U 0
12 G 24
13 U 0
14 C 0
15 C 0
16 U 0
17 U 0
18 G 0
19 G 0
20 G 0
21 A 0
22 A 0
23 A 0
24 C 12
25 C 10
26 G 0
27 A 0
28 G 0
29 U 7
30 C 6
31 C 5
32 G 4
33 G 3
34 G 2
35 C 1
36 A 0
37 C 0
38 C 0
```

## 2 Wyszukiwanie niekanonicznych motywów 3D (modułów)

1. Pary niekanoniczne w pewnych konfiguracjach tworzą stabilne elementy, które znaleźć można w wielu strukturach przestrzennych RNA z różnych organizmów – są to tzw. motywy 3D lub moduły
2. Przegląd znanych motywów 3D dostępny jest na stronie RNA 3D Motif Atlas
3. Na ćwiczeniach skupimy się na bardzo znanym motywie GNRA – guanina, dowolny, puryna, adenina (np. GAGA, GCAA, GAAA)
4. Motyw tworzy pętlę do stemów i charakteryzuje się niekanoniczną parą typu trans Sugar-Hoogsteen (tSH) pomiędzy G-A oraz niesparowanymi dwoma wewnętrznymi nukleotydami
5. W ćwiczeniu wykorzystamy sygnaturę z RNA 3D Motif Atlas: sekwencja NNNNNN, parowania cWW-tSH-R-R. Jak widać, motyw ten jest nawet ogólniejszy niż sekwencyjnie GNRA, choć z niego się wywodzi i nazwa pozostała
6. A zatem szukamy (1) sześciu dowolnych, ale kolejnych po sobie nukleotydów (2) takich, że między pierwszym a ostatnim jest para typu cWW, a między drugim i przedostatnim jest para typu tSH i (3) nukleotydy trzeci i czwarty pozostają niesparowane
7. Użyjemy narzędzia DSSR, które poprawnie klasyfikuje pary kanoniczne i niekanoniczne
8. Przykład:
  - (a) Wejdź na stronę DSSR
  - (b) Wpisz identyfikator struktury 1RLG, pozostałe pola pozostaw bez zmian
  - (c) Na wynikowej stronie można zauważyć, że (1) mamy parę cWW pomiędzy nukleotydami o indeksach 9 i 14, (2) mamy parę tSH pomiędzy indeksami 10 i 13, (3) nukleotydy o indeksach 11 i 12 są niesparowane (tj. nie ma ich na liście "base pairs")
  - (d) A zatem udało się odnaleźć motyw GNRA o sekwencji CGAAAG (rdzeń to GAAA, czyli "klasyczny" motyw GNRA) na pozycji od 9 do 14

- (e) Wróć do strony głównej DSSR i powtórz te same kroki zaznaczając na formularzu "JSON output". Na stronie wynikowej jest opcja download the output file
9. Napisz program, który wczyta wynik działania DSSR w formacie JSON i wypisze wszystkie motywy typu GNRA w formacie indeks-początkowy sekwencja indeks-końcowy
10. **Uwaga! Należy wypisać indeksy takie jak używa DSSR. W przykładzie 1RLG mamy motyw GNRA w łańcuchu C od pozycji 9-14 i w łańcuchu D od pozycji 9-14. Jednak w unikalnej indeksacji DSSR są to numery 9-14 i 34-39 (bez podawania nazw łańcuchów)**

```
$ wget -O 1RLG.json http://dssr.x3dna.org/pdb/1rlg?json=yes
$ ./program 1RLG.json
9 CGAAAG 14
34 CGAAAG 39
```