

Określanie struktury drugorzędowej na podstawie modelu 3D

Tomasz Żok

1 Format BP

1. Wejdź na stronę <http://rna.bgsu.edu/FR3D/AnalyzedStructures/>
2. Wybierz strukturę np. 1DDY
3. Wybierz link *List of basepair interactions in ...*
4. Poniżej poziomej linii jest tabela z informacjami o parowaniach między zasadami azotowymi w nukleotydach
5. Napisz program, który wczyta plik tekstowy zawierający tabelę w takim formacie i zwróci strukturę drugorzędową w formacie BP
6. Uwagi:
 - Z tabeli należy odczytać tylko pary A-U, G-C lub G-U oznaczone jako cWW
 - W danych testowych będą wyłącznie pary z jednej nici RNA

input.txt

1	G	7(A)	-	C	22(A)	-	cWW	-	4
2	G	8(A)	-	C	21(A)	-	cWW	-	4
3	G	8(A)	-	A	25(A)	-	tsS	-	4
4	U	9(A)	-	A	20(A)	-	cWW	-	4
5	G	10(A)	-	G	28(A)	-	cWH	-	2
6	C	11(A)	-	G	30(A)	-	cWW	-	0
7	G	12(A)	-	C	33(A)	-	cWW	-	1
8	C	13(A)	-	G	32(A)	-	cWW	-	1
9	A	14(A)	-	A	31(A)	-	tWW	-	1
10	U	15(A)	-	C	29(A)	-	cWW	-	0
11	A	17(A)	-	C	29(A)	-	bif	-	0
12	C	18(A)	-	G	28(A)	-	cWW	-	0
13	C	19(A)	-	G	26(A)	-	cWW	-	0
14	A	20(A)	-	U	9(A)	-	cWW	-	4
15	C	21(A)	-	G	8(A)	-	cWW	-	4
16	C	22(A)	-	G	7(A)	-	cWW	-	4
17	A	25(A)	-	G	8(A)	-	tSs	-	4
18	G	26(A)	-	C	19(A)	-	cWW	-	0
19	G	28(A)	-	G	10(A)	-	cHW	-	2
20	G	28(A)	-	C	18(A)	-	cWW	-	0
21	C	29(A)	-	U	15(A)	-	cWW	-	0
22	C	29(A)	-	A	17(A)	-	bif	-	0
23	G	30(A)	-	C	11(A)	-	cWW	-	0
24	A	31(A)	-	A	14(A)	-	tWW	-	1
25	G	32(A)	-	C	13(A)	-	cWW	-	1
26	C	33(A)	-	G	12(A)	-	cWW	-	1

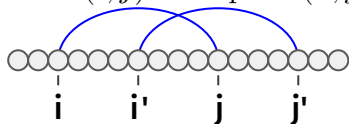
```
$ ./program input.txt > output.txt
```

output.txt

```
7 22
8 21
9 20
11 30
12 33
13 32
18 28
19 26
```

2 Pseudowęzły

1. Para (i, j) oraz para (i', j') tworzą pseudowęzł jeżeli $i < i' < j < j'$



2. Napisz program, który wczyta plik w formacie BP i zwróci listę par tworzących pseudowęzł

```
input.txt
7 22
8 21
9 20
11 30
12 33
13 32
18 28
19 26
```

```
$ ./program input.txt > output.txt
```

```
output.txt
(7, 22) (11, 30)
(7, 22) (12, 33)
(7, 22) (13, 32)
(7, 22) (18, 28)
(7, 22) (19, 26)
(8, 21) (11, 30)
(8, 21) (12, 33)
(8, 21) (13, 32)
(8, 21) (18, 28)
(8, 21) (19, 26)
(9, 20) (11, 30)
(9, 20) (12, 33)
(9, 20) (13, 32)
(9, 20) (18, 28)
(9, 20) (19, 26)
(11, 30) (12, 33)
(11, 30) (13, 32)
```

3 Motyw *stem*

1. Motyw *stem* zdefiniowany jest jako pary (i, j) , $(i + 1, j - 1)$, ... $(i + k, j - k)$ dla $k > 0$ (tzn. *stem* ma co najmniej dwie pary)
2. Napisz program, który wczyta plik w formacie BP i zwróci listę opisów motywów typu *stem*. Opis motywu to trójka liczb: indeks nukleotydu z 5' końca z pierwszej pary, indeks nukleotydu z 3' końca z pierwszej pary, liczba par tworzących motyw

input.txt

```
7 22
8 21
9 20
11 30
12 33
13 32
18 28
19 26
```

```
$ ./program input.txt > output.txt
```

output.txt

```
7 22 3
12 33 2
```