

Organizacja

Zadania w ramce są do wykonania samodzielnie i należy sporządzić z nich krótki raport. Dokumenty elektroniczne (najlepiej PDF) proszę przesłać na mój adres emailowy do 20 marca, do godziny 20.

Formaty zapisu struktury drugorzędowej

- BP: lista par:

```
1 20
2 21
3 22
7 27
```

- BPSEQ: indeks, nukleotyd, para (0 jeśli brak):

```
1 C 0
2 A 0
3 G 103
4 C 102
```

- Dot-bracket: sekwencja oraz zapis kropkowo-nawiasowy (kropka = bez pary):

```
CAGC
. ( ) .
```

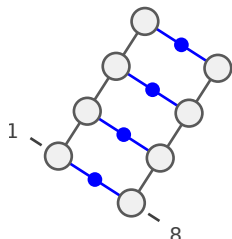
Zadanie 1

Napisz program, który konwertuje zapis w postaci dot-bracket na BPSEQ oraz dodatkowo podaje sekwencję i numery nukleotydów tworzących najdłuższy niesparowany region. Jakiej struktury danych najlepiej jest użyć?

Motywy obecne w strukturze drugorzędowej

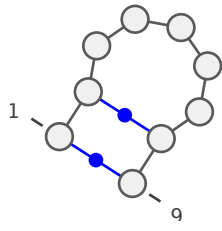
- Stem:

```
(( ( ( ) ) ) )
```



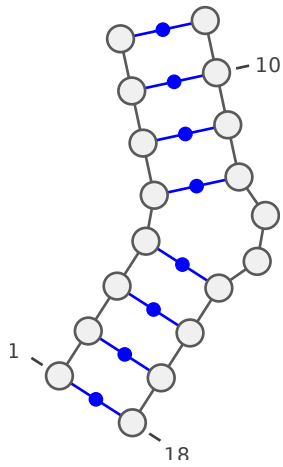
- Hairpin loop:

```
(( . . . . . ) )
```



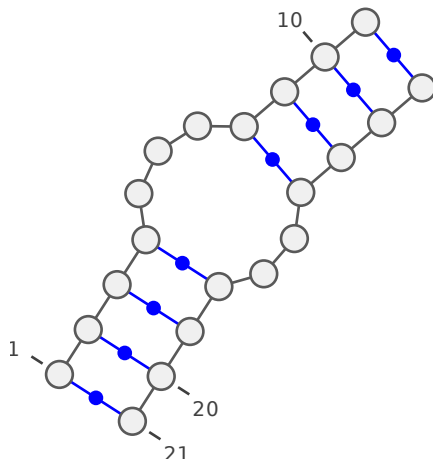
- Bulge:

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



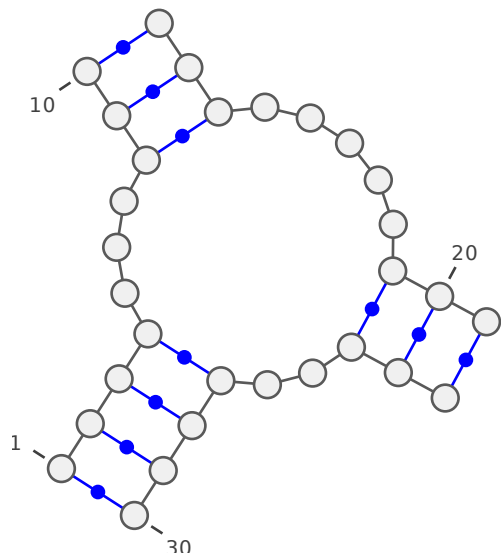
- Interior loop:

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



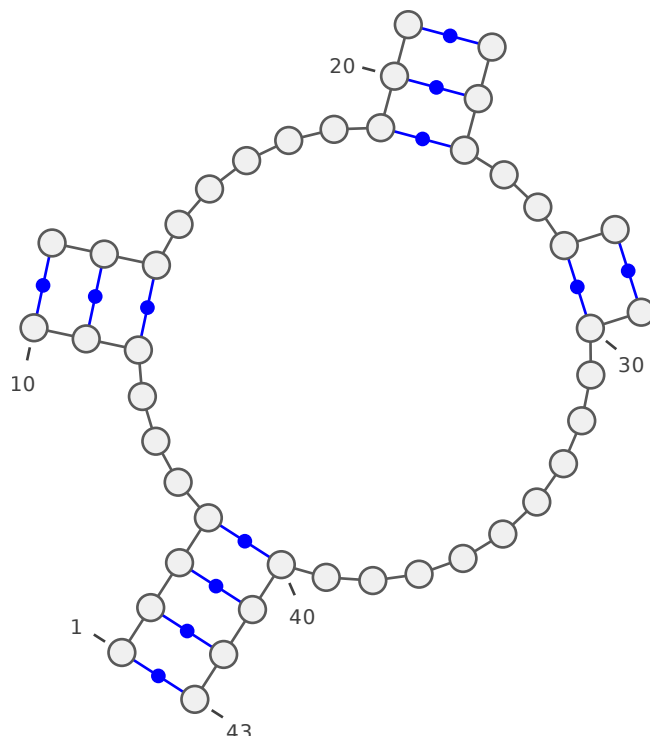
- 3-way junction:

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



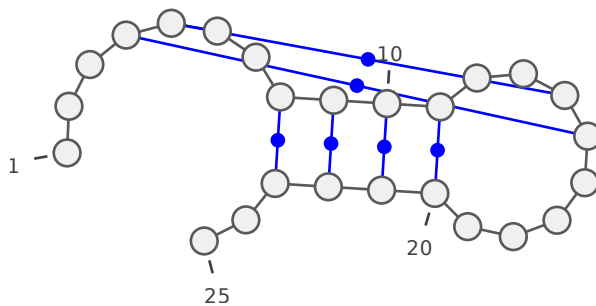
- 4-way junction:

$(((((\dots((()))\dots((()))\dots((()))\dots))))))$



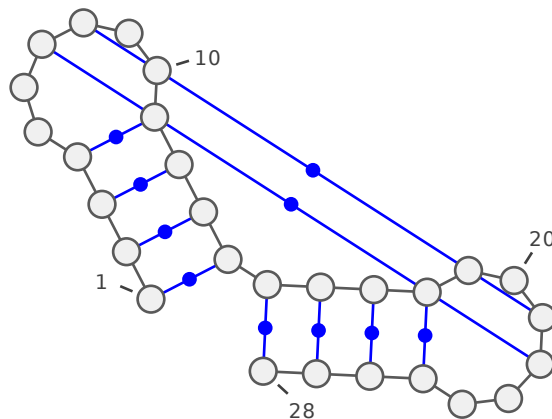
- Pseudoknot:

$\dots[[\dots(((\dots))\dots))\dots]]\dots$



- Kissing loops:

(((((... [[...]]))) (((...]]...))))



Narzędzia do wizualizacji struktury drugorzędowej

- [VARNA](#)

1. Po uruchomieniu aplikacji wczytane są dwie przykładowe struktury, otwórz drugą
2. Zapisz ją we wszystkich dostępnych formatach i sprawdź jak wyglądają wygenerowane pliki
3. Spróbuj “chwycić” i przeciągnąć fragment dwuniciowy
4. Zaznacz dowolny fragment, zobacz jak pokazane jest to w polu z sekwencją i zapisem kropkowo-nawiasowym
5. Kliknij w któryś nukleotydy prawym przyciskiem myszy i poruszając kursorem po dostępnych opcjach w `Edit`→`Selections` zobacz ponownie jak wizualizowane są na bieżąco te ruchy
6. Dodaj wszystkie automatyczne adnotacje

Zadanie 2

Narysuj ponownie dowolną z domyślnie wczytywanych struktur każdym z dostępnych algorytmów: *Linear*, *Circular*, *Radiate* i *NAView*.

Następnie zmodyfikuj ją poprzez zamianę kilku nukleotydów z dowolnej pętli na pseudowęzły (zamiast odpowiadających im kropek wpisz nawiasy kwadratowe). Jaki wpływ ma to na wizualizację dla poszczególnych algorytmów? W jaki sposób można w programie VARNA ukryć te połączenia?

• PseudoViewer

1. Wyślij żądanie narysowania motywów *Pseudoknot* i *Kissing loops* (notacja kropkowo-nawiasowa jak powyżej zostało to opisane, sekwencja nie jest istotna)
2. Po obejrzeniu wygenerowanych grafik, zastąp pseudowęzły niesparowanymi nukleotydami (w miejsce nawiasów klamrowych wpisz kropki)
3. Wygeneruj obrazki ponownie. Czy PseudoViewer lepiej wizualizuje pseudowęzły niż VARNA? A jak wygląda to z większymi strukturami? Przykładowo:

```
CAGCACGACACUAGCAGUCAGUGUCAGACUGCAIACAGCACGACACU
. [ ((((((... ((((((... ((((((..... ((((((... ((((((...
AGCAGUCAGUGUCAGACUGCAIACAGCACGACACUAGCAGUCAGUGU
..))))))...)))))).....)))))).....))))))
CAGACUGCAIA
).....)))))..
```

4. Zasymuluj “braki” w strukturze. Przełącz typ wejścia na *PseudoViewer input file type*. Jest to format kropkowo-nawiasowy, w którym można przeplatać na zmianę pary linii sekwencja-struktura. Jeśli linie poprzedzone są numerem, to na wygenerowanej grafice nukleotydy mają nowe numerowanie. Sprawdź przykład:

```
1 CCUUACCUCGGGUAGAGGCCCA
1 (((.....))..(())....
120 CCUUACCUCGGGUAGAGGCCCA
120 (((.....))..(())....
```

RNA FRABASE

- FRABASE bazując na usłudze PseudoViewer udostępnia mechanizm wizualizacji struktur drugorzędowych wzbogacony o:
 - Trójpoziomowe rozróżnienie pseudowęzłów przez nawiasy kwadratowe, klamrowe i trójkątne
 - Braki/błędy wyrażone bezpośrednio w notacji kropkowo-nawiasowej (znak minus) i rysowane w specjalny sposób
 - Automatyczny podział całej struktury na niezależne nici
- Sprawdź struktury drugorzędowe o następujących identyfikatorach:
 - 1RY1: pseudowęzeł, wiele nici
 - 3E3J: wszystkie poziomy pseudowęzłów, braki w strukturze
 - 3G4S: bardzo duża jednostka

Inne reprezentacje

Dana jest sekwencja RNA S oraz jej struktura drugorzędowa w postaci zbioru par zasad P :

$$\begin{aligned} S &= \text{UCCCGCGUUUUUCGUUGGAAACCAAGGCGAA} \\ P &= \{(1; 29); (2; 28); (3; 26); (4; 25); \\ &\quad (5; 14); (6; 13); (7; 12); (15; 23); \\ &\quad (16; 22); (17; 21)\} \end{aligned}$$

Przedstaw strukturę RNA za pomocą poniższych reprezentacji:

- Notacji kropkowo-nawiasowej (ang. *dot-bracket*)
- Klasycznej reprezentacji graficznej (ang. *standard graphical representation*)