

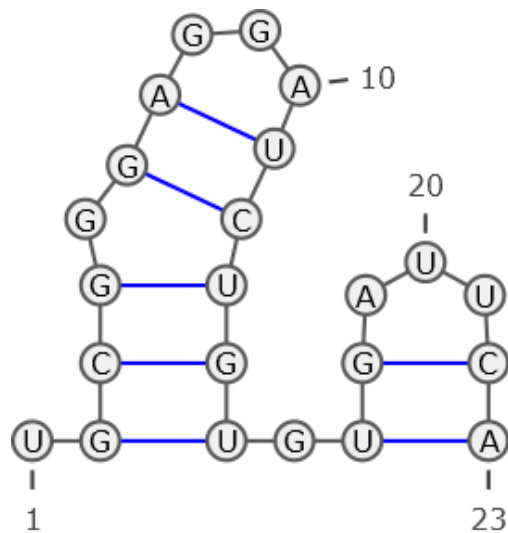
## Exercise sheet 11: RNA Bioinformatics

---

### Exercise 1

RNA secondary structures can be represented using a graph notation, where nodes represent nucleotides and edges encode the molecule's backbone or intramolecular base pairs between nucleotides. Below, an RNA molecule graph is depicted that encodes base pairs in blue.

Decide for the following properties whether they are correct or wrong given the RNA secondary structure graph.



**Question 1A** Position 23 represents the 5'-end

**Solution** wrong; We always encode in the 5' (=1) to 3' (=n) direction.

**Question 1B** the graph is invalid

**Solution** wrong

**Question 1C** contains invalid base pairs

**Solution** wrong; all base pairs depicted in the graph are valid

**Question 1D** contains a pseudoknot

**Solution** wrong; There are no pseudoknot structures in this graph.

**Question 1E** non-crossing

**Solution** correct; There are no pseudoknot structures in this graph.

**Question 1F** nested

**Solution** correct

**Question 1G** contains base pair (5,12)

**Solution** wrong; Position 5 is unpaired.

**Question 1H** contains base pair (4,13)

**Solution** correct

**Question 1I** base pair (1,10) would be crossing

**Solution** correct

**Question 1J** obeys a minimal loop length of 4

**Solution** wrong; The minimum loop length for this graph is 3. (number of unpaired bases in loops)

**Question 1K** encoded by  $((\dots)).((((\dots))).)$ .

**Solution** wrong

**Question 1L** encoded by  $.(((.((\dots))))).((\dots))$

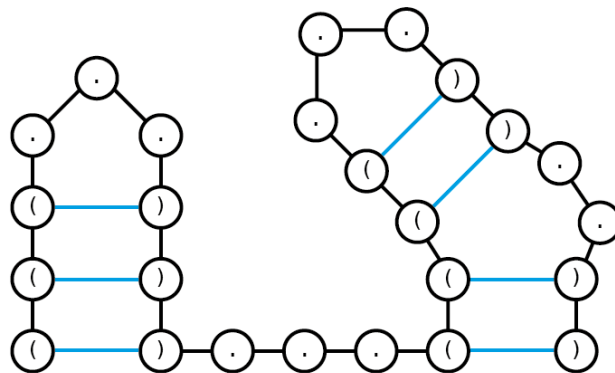
**Solution** correct

## Exercise 2

You are given the following dot-bracket string:  $(((\dots)))\dots((((\dots))).)$

**Question 2A** Draw graph representations of all nested structures that can be encoded by the dot-bracket string. Assume a minimal loop length of 3.

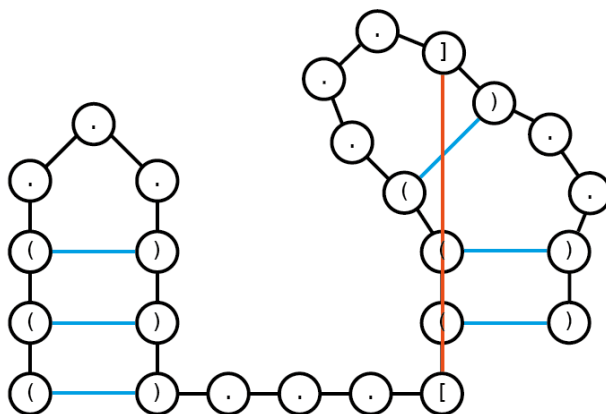
**Solution**



This is the only possible nested structure based on the dot-bracket string given.

**Question 2B** Draw a graph representation of one possible crossing structure that can be encoded by the dot-bracket string. Assume a minimal loop length of 3.

**Solution**



Dot-bracket string: (((...)))...[(((...))]..)

There are multiple other possible crossing structures. (e.g. (((...)))...[(((<...]))..> )

### Exercise 3

Given the following partially filled Nussinov matrix  $N$  using a minimal loop length  $l = 0$ , i.e. neighbored nucleotides are allowed to pair.

		G	C	C	G	G	A	C	G	
1	0	0	1	1	?					G
2		0	0	0	1	2				C
3			0	0	1	1	1			C
4				0	0	0	0	?		G
5					0	0	0	1	1	G
6						0	0	0	1	A
7							0	0	1	C
8								0	0	G
	0	1	2	3	4	5	6	7	8	

**Question 3A** What are the values of the green and red entry?

**Solution** green: 2

red: 1

**Question 3B** How many tracebacks exist for the red entry using the original recursion by Nussinov?

**Solution** There exist two tracebacks for the red entry; pairing with either G.

## Exercise 4

Given any matrix  $N$  filled by Nussinov's algorithm for an RNA sequence  $S$  of length  $n$ . Discuss which of the following statements are correct or wrong.

The entry  $N_{1,n}$  of the Nussinov matrix encodes ...

**Question 4A** ... the optimal structure.

**Solution** wrong; It only encodes the base pair number for the optimal structure.

**Question 4B** ... the minimum free energy (mfe) structure.

**Solution** wrong; No energy minimization is done.

**Question 4C** ... the maximal number of base pairs for sequence  $S_1..S_n$ .

**Solution** correct

**Question 4D** ... the traceback end.

**Solution** wrong; It encodes the traceback start.

**Question 4E** ... the maximal number of base pairs for any structure.

**Solution** correct

**Question 4F** ... information for a unique structure.

**Solution** wrong; Typically there is more than one optimal structure with maximal number of base pairs.

## Exercise 5

$$\begin{aligned}
 N_{i,j} &= \max \left\{ \begin{array}{c} \text{diagram of } B_{i,j} \\ \text{diagram of } D_{i,j} \end{array} \right\} \\
 B_{i,j} &= \begin{cases} N_{i+1,j-1} + 1 & \text{if } S_i \text{ and } S_j \text{ are compl.} \\ 0 & \text{else} \end{cases} \\
 D_{i,j} &= \max_{i \leq k < j} \left\{ \begin{array}{c} \text{diagram of } B_{i,k} \\ \text{diagram of } N_{k+1,j} \end{array} \right\}
 \end{aligned}$$

**Question 5A** Provide all recursion and initialization details for the following recursion depictions. Note, also ensure a minimal loop length  $l$  within your recursions.

**Solution**

$$\begin{aligned}
 N_{i,i} &= N_{i,i-1} = 0 \quad (\text{no init for } B \text{ and } D \text{ needed}) \\
 \forall 1 \leq i < j \leq n : N_{i,j} &= \max \{B_{i,j}, D_{i,j}\} \\
 B_{i,j} &= \begin{cases} N_{i+1,j-1} + 1 & \text{if } i + l < j \wedge S_i, S_j \text{ compl.} \\ 0 & \text{else} \end{cases} \\
 D_{i,j} &= \max_{i \leq k < j} \{B_{i,k} + N_{k+1,j}\} \quad (\text{only valid for } i < j)
 \end{aligned}$$

