

COSC 2P03 Advanced Data Structures: Assignment 1

Instructor: Yifeng Li^{*1}

¹Department of Computer Science, Brock University

September 18, 2021

1 Introduction to RNA Secondary Structure Prediction

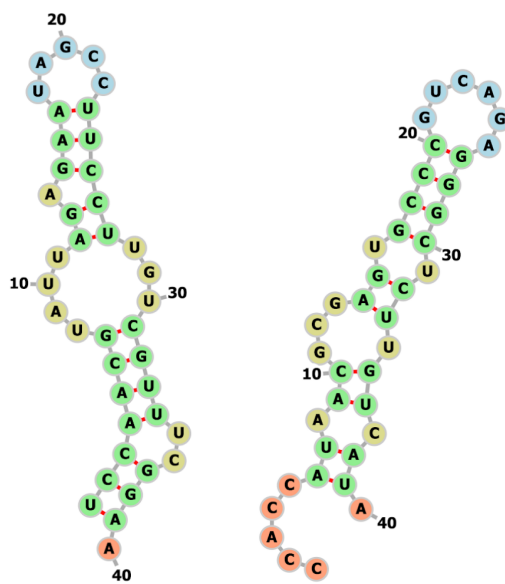
Ribonucleic acid (RNA) is a nucleic acid molecule that plays essential roles in many biological processes in protein coding and gene expression regulation. It can also be designed as therapeutic agents (e.g., drugs or vaccines). The primary sequence of an RNA is a string of nucleobases (G, U, A, and C). Each base is chained to the string by covalent bonds. Each base may form hydrogen bond with another base, which is called base pairing. Given the primary sequence of an RNA, the prediction of its secondary and tertiary structures can help understand the function of this RNA. The secondary structure of an RNA can be represented in dot-bracket notation and visualized in planar plot. A dot-bracket notation is a string composed of repeated use of three characters: “.”, “(”, and “)”, where “.” means unpaired bases, and matched “(” and “)” for paired bases. In this assignment, you are provided with predicted secondary structures of a few RNAs from a tool and are tasked to analyze these results. To help you understand the primary sequence, dot-bracket notation, and planar plot, two examples are provided below.

RNA 1:

UCCAACGUAAUAGAGAAUAGCCUCCUUGUCGUUUCGGAA
(((((((....((.((((.....))))))....)))))).

RNA 2:

CCACCAUAACGCGAGUGCCCGUCAGAGGGCUCUUGUCAUA
.....((.(((....((.((((.....)))))).)).)).



*E-mail address: yli2@brocku.ca yifeng.li.cn@gmail.com

2 Your Tasks

You should define a class named `rnaSecStrcAnalysis` with the following requirements.

1. Define a method named `readData` to load the provided data to your class instance. Each data item, including a sequence and a dot-bracket string, should be stored in a node of a doubly linked list with header and tail. This list instance should be the data attribute of your class. Key information in a node of the list includes: `primarySequence`, `dotBracketString`, and `validity`. (2 marks)
2. Define a method named `checkValidity` to check the validity of each dot-bracket string in your data by using a stack data structure. If all the opening brackets and closing brackets match, this string should be labelled as 1; otherwise 0. Assign this result to the `validity` of your list node mentioned above. (2 marks)
3. To tackle the task in `checkValidity`, you should define a class for the stack named `stackForDotBracket` with all necessary methods (`push`, `pop`, `isEmpty`, `isFull`, `size`, and `top`), and use an instance of this class as a data attribute in your `rnaSecStrcAnalysis` class. (2 marks)
4. Write back your result in a text file named `rna_data_result.txt` in the following format: (2 marks)
primary sequence of data item 1
dot-bracket notation of data item 1
0/1
primary sequence of data item 2
dot-bracket notation of data item 2
0/1
:
:

3 Submission

- Your source code.
- A PDF printout of your source code.
- The text file with result required in Task 4.
- Compress the above files in a zipped folder named `COSC2P03_A1_Firstname_Lastname_StudentNumber.zip` and submit it through Sakai before indicated due time.
- Late submissions will not be accepted.

4 Academic Integrity

This assignment should be tackled individually. Outsourcing or teamwork is not allowed. Violation of this requirements will be seriously processed in accordance with university policies.