

# **Mapping RNA Secondary Structure into an Alignment**

MSC ON BIOINFORMATICS FOR HEALTH SCIENCES -  
PROGRAMMING IN PERL

Gerard Pradas  
Course 2018/2019 Term 1

# Introduction

The aim of the project is to get the contact strings for a set of aligned RNA secondary structures. To do so, we created a program that only needs a valid input file containing a reference sequence and its contact string, and then N aligned sequences. The software removes the gaps of the sequences, so in order to print the result with the gaps, we can use the option -g.

## Data Structures

During the development we have used 3 types of data structures, arrays, hashes and stacks (in perl are the same as arrays).

We used the arrays mainly to store data, starting from the input files, and later, to save the result of the contact strings for the secondary structures.

We have used a hash to store the hairpins, where the key is the position in the array where the hairpin is found, and the value is the open/closing hairpin. That way we only focus on the hairpins of the contact string.

Finally, we have use an array as a stack, to find the complementary position of the open/closing hairpin.

## Algorithms

There are few sections of the code that need a special mention regarding the algorithms involved.

### Input data

For entering the data, we dumped all the file at once, because the input file we have is quite small, so for bigger files, it might be better, to parse line by line.

### Traversing hairpin hash

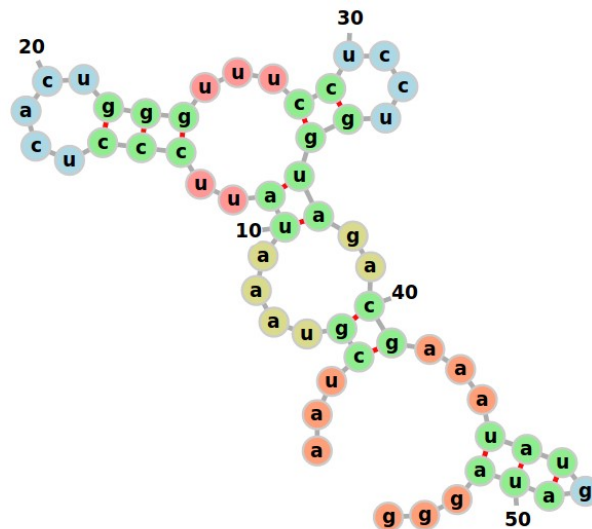
To generate the new contact string for the secondary structure, we copied the initial one, and only look at the points where an hairpin is open or close, so we reduce the number of iterations done.

### Getting complementary position

To get the complementary position for an open/closing hairpin, we used a stack, pushing where we found an opening hairpin, and popping where we found a closing hairpin, thus when we end, we end up with the position of the last closing hairpin, that will be the complementary position.

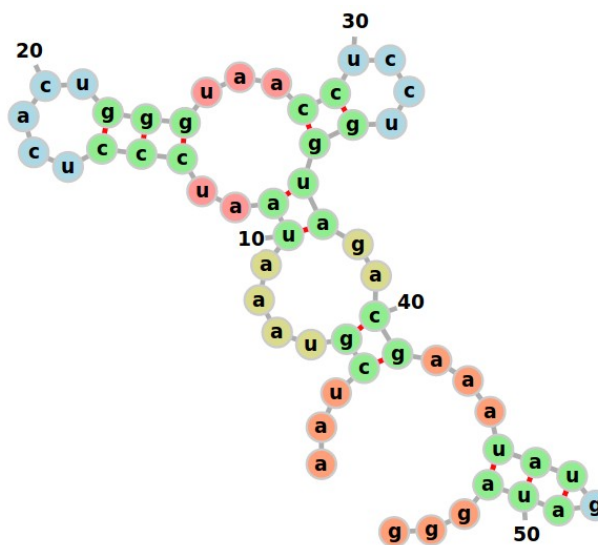
# Appendix

RNA Secondary Structure Visualization (<http://rna.tbi.univie.ac.at/forna/>)



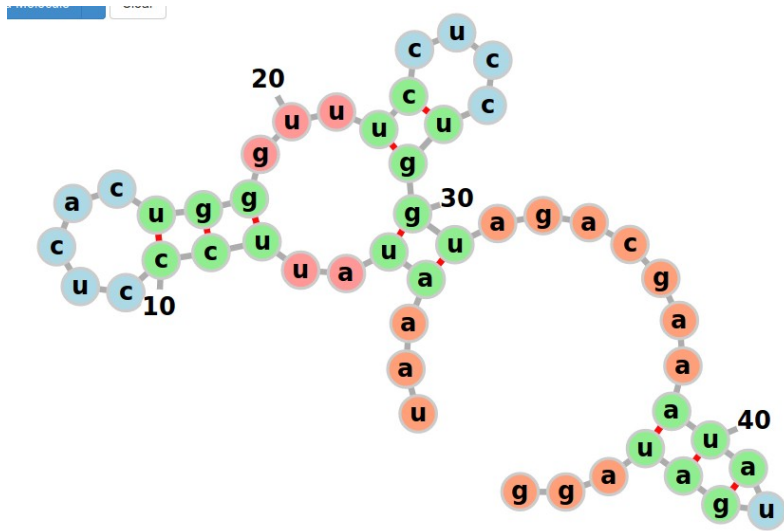
SQref

aaucguaaaauauucccucacuggguuuccuccugguagacgaaauaugauaggg  
 ...(( .....(( ..((( .....))) ...(( .....))) ...)) ...((( .))) ...



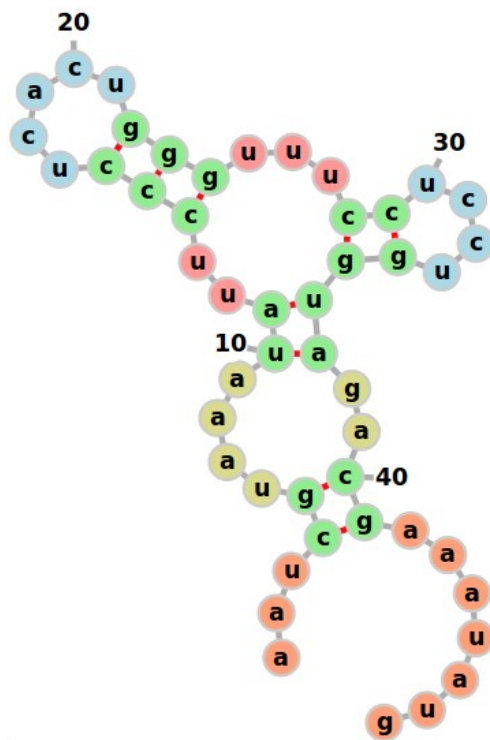
SQv01

aaucguaaaauauucccucacuggguaaccuccugguagacgaaauaugauaggg  
 ...(( .....(( ..((( .....))) ...(( .....))) ...)) ...((( .))) ...



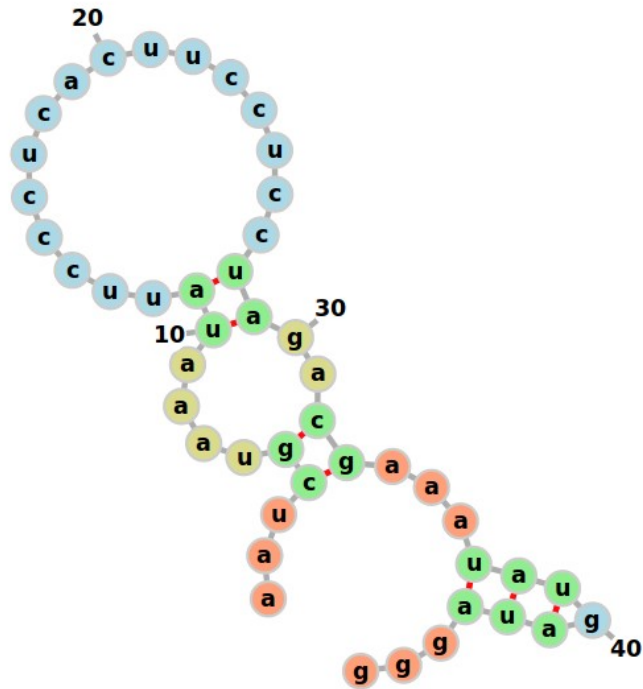
SQv02

uaaauauucccucacuggguuuccuccugguagacgaaaauaugauaggg  
 ....((..(((.....)))...((.....))).....(((.)))...



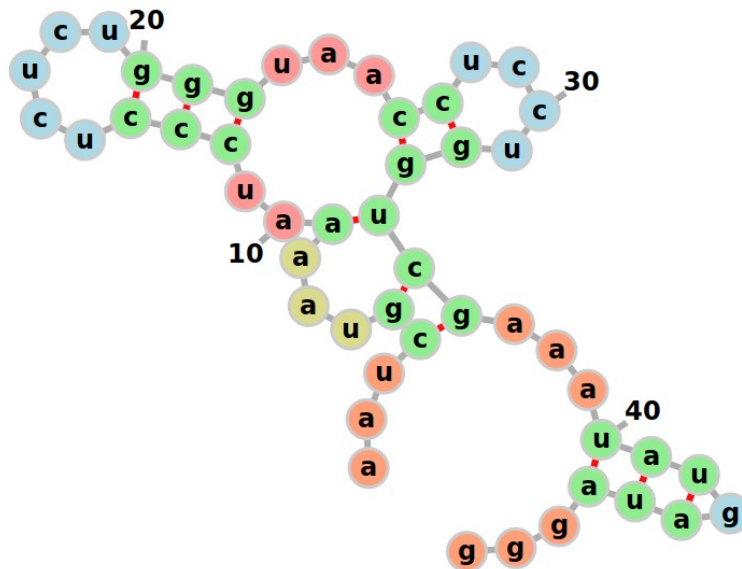
SQv03

aaucguaaaaauucccucacuggguuuccuccugguagacgaaaauaug  
 ...((.....((..(((.....)))...((.....)))...)).....



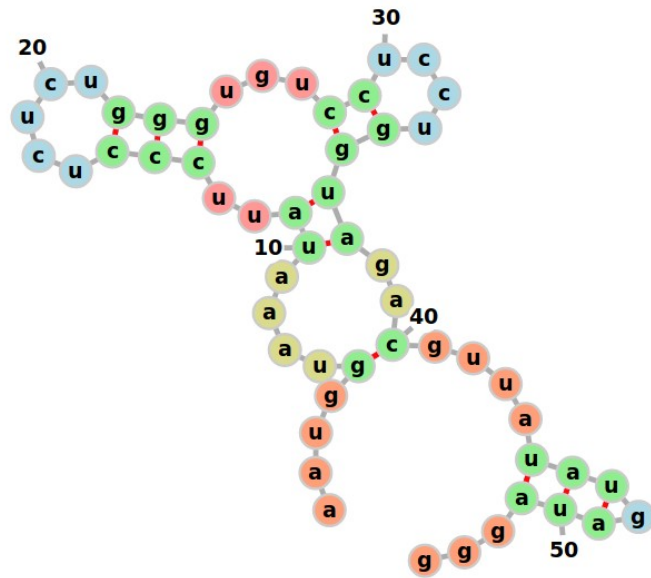
SQv04

aaucguaaaaauucccucacuuuccuccuagacgaaauaugauaggg  
 ...((( ...(( ...)))) ...((( .))) ...



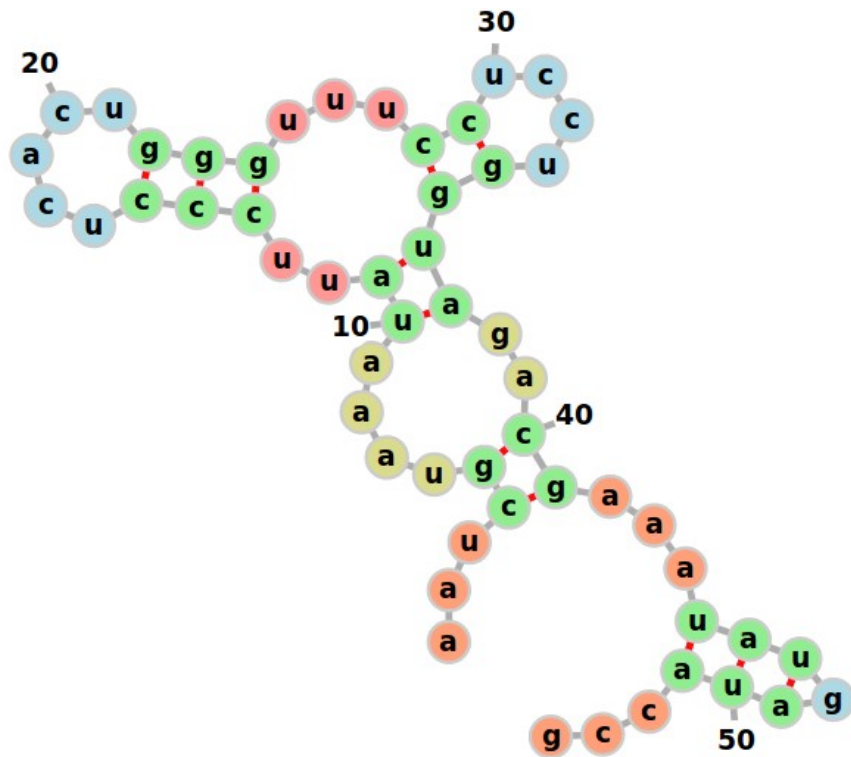
SQv05

aaucguaaaauucccucucuggguaaccuccuggucgaaauaugauaggg  
 ...(( ... ( ..((( .....))) ...((( .....)))) ...((( .))) ...



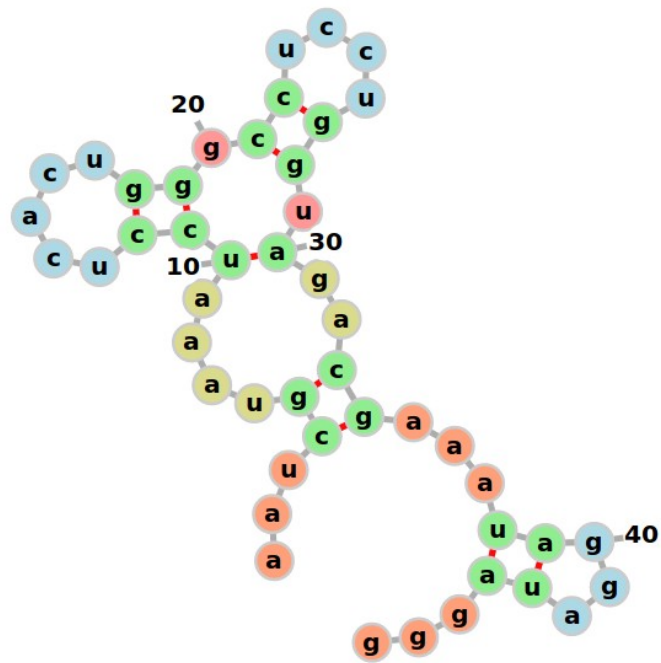
SQv06

aaugguaaaauauucccucucuggguguccuccugguagacguuauaugauaggg  
 ....( ....(( ..((( .....))) ...(( .....))) ..) ....((( .))) ...



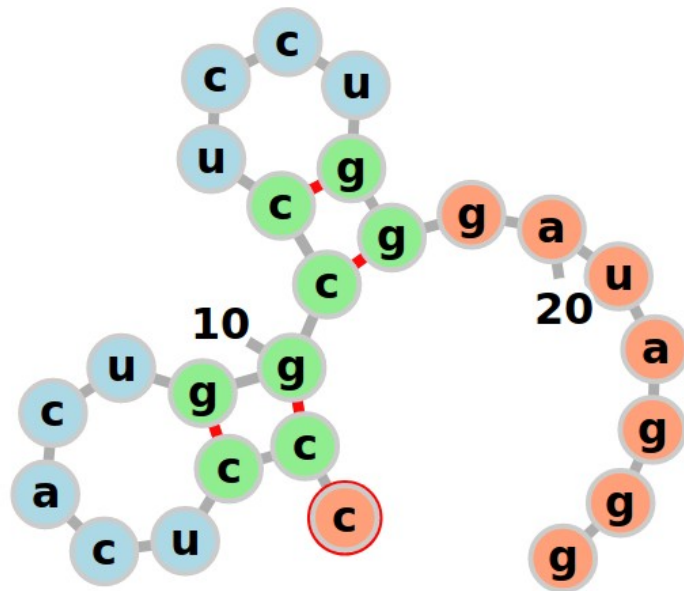
SQv07

aaucguaaaauauucccucacuggguuuccuccugguagacgaaauaugauaccg  
 ...(( ....(( ..((( .....))) ...(( .....))) ..) ....((( .))) ...



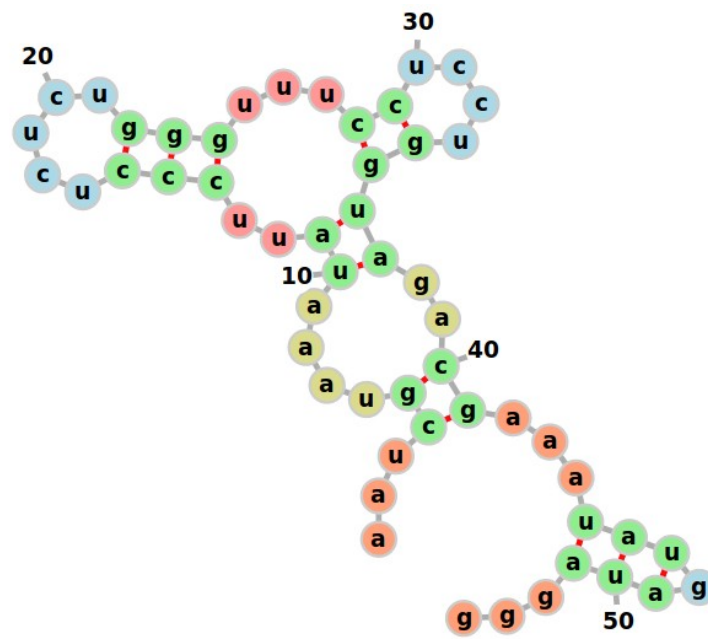
SQv08

aaucguaaauccucacuggggccuccugguagacgaaauaggauaggg  
 ...(( .....((( .....)) .(( .....)) .) ..)) ...(( .....)) ...



SQv09

cccucacugggccuccuggggauggg  
 .(( .....))(( .....)) .....



SQv10

aaucguaaaauauucccucucuggguuuccuccugguagacgaaaauaugauaggg  
 ...(( .....(( ..((( .....))) )...(( .....))) )... )...((( .))) ...