



## Lecture: Grundlagen der Bioinformatik

**SoSe 2022**

### Assignment 11

(20 points)

Hand out:

Thursday, July 14

Hand in due:

Thursday, July 21 18:00

Direct inquiries via the ILIAS forum or to your respective tutor at:

Mathias Witte Paz: [iizwi01@uni-tuebingen.de](mailto:iizwi01@uni-tuebingen.de)[theresa-anisja.harbig@uni-tuebingen.de](mailto:theresa-anisja.harbig@uni-tuebingen.de)[meret.haeusler@student.uni-tuebingen.de](mailto:meret.haeusler@student.uni-tuebingen.de)[jules.kreuer@student.uni-tuebingen.de](mailto:jules.kreuer@student.uni-tuebingen.de)[simon.heumos@qbic.uni-tuebingen.de](mailto:simon.heumos@qbic.uni-tuebingen.de)

## Theoretical Assignments

### 1. RNA Secondary Structure Theory

(3P)

Non-coding RNAs (ncRNAs) fulfill different molecular functions. One of them is the inhibition of the translation of protein coding genes via riboswitches. Explain how the secondary structure of this class of ncRNAs allows the regulation of protein translation (write at most 100-150 words). In addition, please also hand in a sketch of this process that you refer to within your description. Be creative.

### 2. Nussinov algorithm - Traceback

(3P)

Construct an RNA sequence such that it contains at least one bifurcation, incl. one hairpin with a stem containing two basepairs and a loop of at least 3 nucleotides. For this sequence compute the secondary structure using the Nussinov algorithm, incl. the traceback. Hand-in the computed dynamic programming matrix with traceback (you do not need to hand intermediate steps of the computation). Moreover, try to show how the *traceback path* is connected with the RNA secondary structure of the sequence (in particular the bifurcation).

## Practical Assignments

### 3. Characterization of protein secondary structures

(4p)

Imagine your biology cooperation partner has sequenced a protein molecule and deduced its mRNA sequence (see file `proteinMRNA.fasta`). Please help him to find out (using the PDB database, among other resources) whether the (secondary) structure of this molecule has been experimentally resolved. Hand in a short summary of your approach, and if you identified a structure, what type of protein this is incl. a brief description of its function (altogether, at most 1 page).

#### 4. Implementation of Nussinov

(10P)

For this task we ask you to implement the Nussinov algorithm in **Python**. In detail:

- (a) Your program should expect a path to a FASTA file with RNA sequence(s) as command line argument.
- (b) Implement the Nussinov algorithm that maximizes the number of basepairs as discussed in the lecture. This includes the initialization, recursion and traceback step. **Hint:** You can reuse and adapt accordingly the code from Assignment 2 where you implemented the Needleman-Wunsch algorithm.
- (c) Using the traceback, provide **one** possible optimal dot-bracket notation for the given sequence. Furthermore, print out the maximal number of basepairs computed for each RNA sequence.
- (d) Extend your program so that it visualizes the RNA structure as a graph. You might want to take a look at the package **forgi**<sup>1</sup> for this task.
- (e) Apply your program to the given sequences in the file **nussinov.fasta**
- (f) **Bonus (+2P):** Extend your program to only compute loops with a minimal length  $l$ .  $l$  is a parameter that should be read from the command line.

Please read the questions carefully. If there are any questions, you may ask them during the tutorial session or in the forum of ILIAS. You will usually get an answer in time, but late e-mails (e.g. the evening of the hand-in) might not be answered in time. Please upload all your solutions to ILIAS. Don't forget to put your names on every sheet **and** in your source code files. Please pack both your source code as well as the theoretical part into one single archive file and give it a name using this scheme: <name1>\_<name2>\_<Assignment>\_<#>.zip. The program should run without any modification needed.

---

<sup>1</sup><https://viennarna.github.io/forgi/index.html>