

8. ÜBUNG "BIOINFORMATIK", SS 16

Aufgabe 1: (2+3=5 Credits)

Use the RNAfold WebServer* to compute the MFE-structure of the sequences

$s1 = \text{AAAUGCGGUCCAAGUAACC}$

$s2 = \text{CCAAUGAACCUUGGCGUAAA}$

$s3 = \text{ACGUACGUACGUACGUACGU}$

- (a) Give the respective MFE-structures $S1$, $S2$ and $S3$ for the sequences $s1$, $s2$ and $s3$ in *bracket-notation*.
- (b) Prove or disprove: There is a sequence $s \in \{A, C, G, U\}^{20}$ that realizes all three MFE-structures. Give such a sequence if one exists.

* <http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi>

Use the options: *minimum free energy (MFE) only* and *avoid isolated base pairs*

Aufgabe 2: (8+2=5 Credits)

Implement the Nussinov-Traceback-Algorithm in C++ with the modification that $\Theta = 1$. Send the source-code via email to the tutor Nikolai Nøjgaard.

- (a) Use your algorithm to compute the secondary structure for the sequence

$s = \text{AAAUGCGGUCCAAGUAACCCCAAUGAACCUUGGCGUAAAACGUACGUACGUACGUACGU}$

Give the computed matrix and the traceback, as well as the final secondary structure that maximizes the number of basepairs as dot-bracket notation. Send the matrix and the secondary structure in dot-bracket notation in a text-file via email to Nikolai Nøjgaard.

- (b) Discuss shortly the differences and similarities between your computed structure and the MFE-structure for s computed with RNAfold WebServer by means of their loop and stacking regions.

Aufgabe 3: (5 Credits)

Given the undirected graph $G = (V, E)$ with $V = \{1, 2, 3, 4, 5\}$ and $E = \{(1, 2), (1, 4), (2, 3), (2, 4), (3, 4), (3, 5), (4, 5)\}$. Let S_1, \dots, S_4 be secondary structures ($\Theta = 0$) such that if $(i, j) \in S_l$ then $(i, j) \in E$. Show that there is no shape graph $G(S_1, S_2, S_3, S_4)$ s.t. $G \simeq G(S_1, S_2, S_3, S_4)$.

Deadline: Monday - June 20th, 2016 - 2pm