Grundlagen der Bioinformatik

SoSe 2022

Tutor: Theresa/ Mathias

1	2	3	4	$\sum$

Marina Dittschar & Clarissa Auckenthaler

Blatt 11

(Abgabe am 21.07.2022)

Theoretical Assignments

Task 1: RNA Secondary Structure Theory(3)

In finding out more about translation control with riboswitches paper [2] was our main source. The structure that enables riboswitches to affect the translation of RNA with non-coding RNA is the aptamer. The aptamer is located upstream of the coding region of the RNA. It overlaps with the expression platform. The aptamer reacts to ligand-binding by changing its secondary structure. This causes the expression platform to also change secondary structure. The translation is then inhibited through various mechanisms, for example the occlusion of the ribosome binding site. If this process happens, the ribosome binding site becomes occupied by the "Anti-ribosome-binding-site", such that the ribosome cannot attach to the RNA. You can see an illustration of this process in Figure 1. Importantly, this translation control depends on the presence of the ligand. If the ligand is not present, the translation can occur uninhibited. Note there are also other, more indirect mechanisms by which riboswitches affects translation.

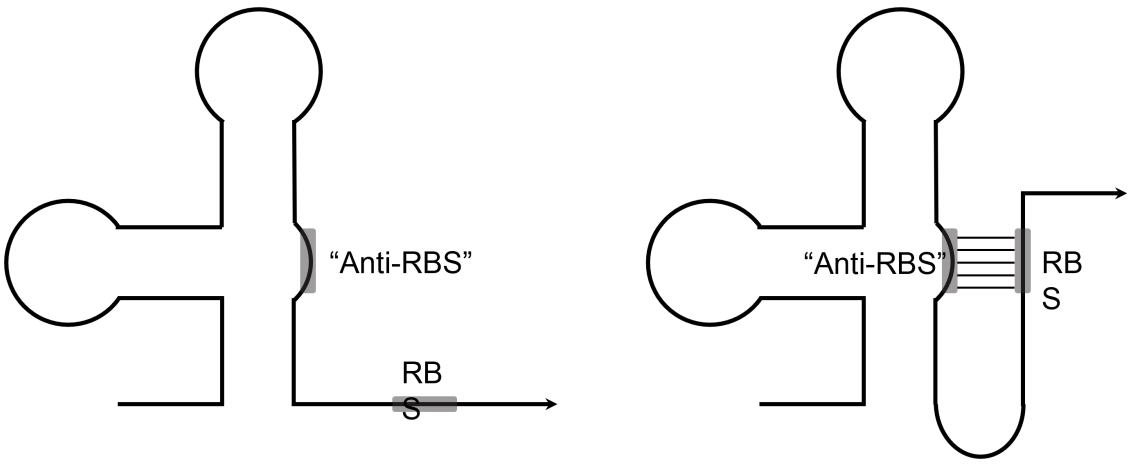


Figure 1: Illustration for riboswitch inhibiting RNA translation by blocking the ribosome binding site

Task 2: Nussinov algorithm - Traceback (3)

We computed the matrix of the nussinov algorim based on the following sequence:

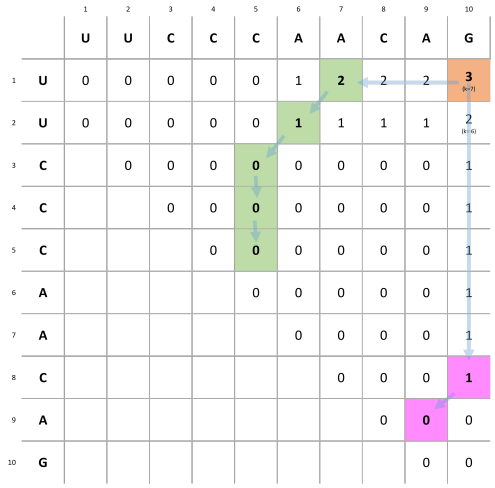
U U C C C A A C A G

The values in our matrix 2a are calculated by the recursion equation:

$$\text{Set}\gamma(i,j) = max \begin{cases} \gamma(i+1,j), \\ \gamma(i,j-1), \\ \gamma(i+1,j-1)+\delta(i,j), \\ max_{i<k<j}[\gamma(i,k)+\gamma(k+1,j)], \end{cases} \tag{1}$$

		1 U	2 U	3 C	4 C	5 C	6 A	7 A	8 C	9 A	10 G
1	U	0	0	0	0	0	1	2	2	2	3 (k=7)
2	U	0	0	0	0	0	1	1	1	1	2 (k=6)
3	C		0	0	0	0	0	0	0	0	1
4	C			0	0	0	0	0	0	0	1
5	C				0	0	0	0	0	0	1
6	A					0	0	0	0	0	1
7	A						0	0	0	0	1
8	C							0	0	0	1
9	A								0	0	0
10	G									0	0

(a) Computed Matrix for our sequence



(b) Traceback for our sequence

Figure 2: Nussinov algorithm by hand

Result and dot notation based on the traceback matrix 2b:

1	2	3	4	5	6	7	8	9	10
U	U	C	C	C	A	A	C	A	G
(	(	.	.	.	)	)	(	.	)

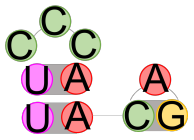


Figure 3: Nussinov algorithm by hand

Practical Assignments

Task 3: Characterization of protein secondary structures (4)

As a fist step we translate the mRNA sequence from the file proteinMRNA.fasta with the tool ExPASy (`https://web.expasy.org/translate/`)[4]. We received 6 frames and selected the one displayed in figure 4. The last symbol is a gap, because of the reading frame. Because we are not able to enter gaps into the PDB database (`https://www.rcsb.org/`) [1], we only inserted the letters without the last gap in the search tool. We used the following settings:

- Target: Protein
- Identity Cutoff: 90%
- E-Value Cutoff: 0.05
- Return: Polymer Entities

With this setting we received overall 9 polymer Entities and all of them have an Sequence Identity of 100%. Explaining the important parameters: The sequence identity is the ratio between the number of identical amino acids between the two aligned sequences. The E-value shows if the observe sequence was a chance match or if it has any evolutionary significance, a high value expressing a chance match. The region indicates the position of amino acids in the database protein that were taken into account for the comparison.[1] We explored three entities with the same parameters for sequence identity (100%), E-value (4,49 $e^{-}88$) an region (1-144). All refer to the tmRNA-binding small protein B of the 70S ribosome of the bacteria species Thermus Thermophilus. The tmRNA-small protein b has the function of releasing stalled ribosomes and "targeting incomplete proteins for degradation" [3].

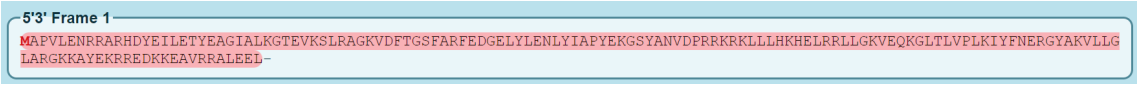


Figure 4: Selected result of translated mRNA sequence to protein.

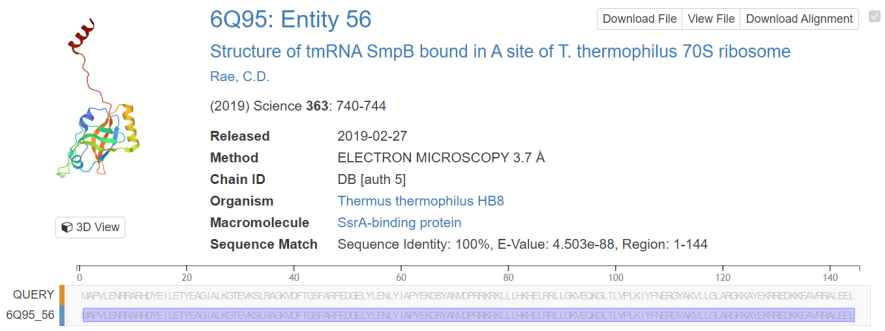


Figure 5: First entity result from PDB database

Task 4: Implementation of Nussinov (10)

We used Python 3.8.8, and the following libraries **io, Bio, sys, getopt, and numpy**. Enter the following code in the command line to run the file:

```
python dittschar_auckentaler_assignment_11.py -f nussinov.fasta
```

We exported the output as a txt-file named:

- auckenthaler_dittschar_sequences.txt

The structures are stored in a txt-file named:

- auckenthaler_dittschar_structures.txt

We could not get viennarna run on our own machines (even with virtual linux machines), which is why we cooperated with Lea Heinen and Marit Bockstedte who helped us in how to write and run the program. Run the following command to run the second file, which takes in structures file from above:

```
python dittschar_auckenthaler_plotting_assignment_11.py
```

The output will be two .ps files. You can open the resulting .ps-files for example with Inkscape (see Figures 6 and 7). The output files are:

- sequence1_ss.ps
- sequence2_ss.ps

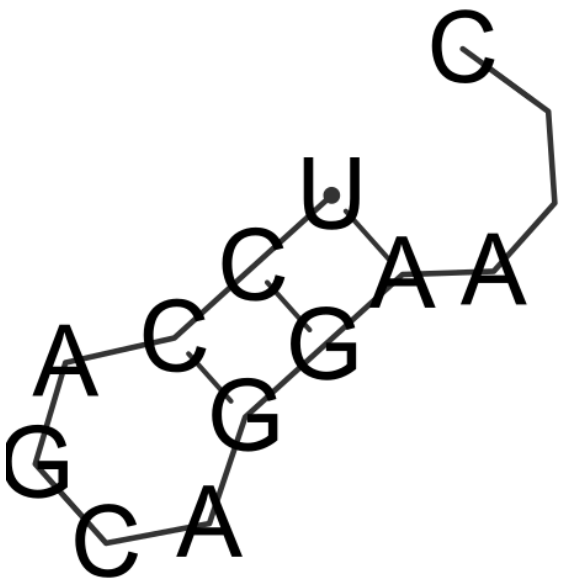


Figure 6: Visual result of the first sequence plot

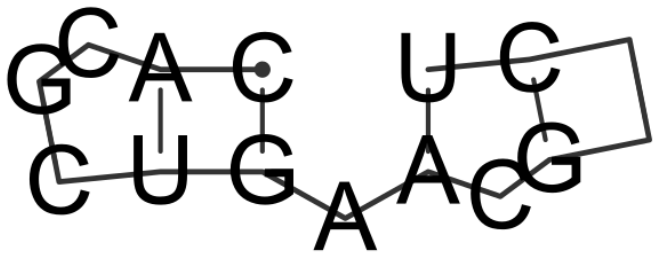


Figure 7: Visual result of the second sequence plot

References

[1] Helen M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov, and Philip E. Bourne. The protein data bank. Nucleic acids research, 28(1):235–242, 2000.

[2] Ronald R Breaker. Riboswitches and translation control. Cold Spring Harbor perspectives in biology, 10(11):a032797, 2018.

[3] Gang Dong, Jacek Nowakowski, and David W Hoffman. Structure of small protein b: the protein component of the tmrna–smpb system for ribosome rescue. The EMBO journal, 21(7):1845–1854, 2002.

[4] Jon Ison, Matús Kalas, Inge Jonassen, Dan Bolser, Mahmut Uludag, Hamish McWilliam, James Malone, Rodrigo Lopez, Steve Pettifer, and Peter Rice. Edam: an ontology of bioinformatics operations, types of data and identifiers, topics and formats. Bioinformatics (Oxford, England), 29(10):1325–1332, 2013.