RNA REPORT

Exercises 1 & 2

Sequence 1:

AUCAGUUCUAGCAGGAGCUGUACUCAGAGACUCGGGAAAUUUUCCCGG AAUUUUACCCGGGUUUUUACGU

As can be seen in Figure 1, its secondary structure is formed by the following sub-structures (from top to the bottom of the image):

hairpin loop – duplex – interval loop – duplex – 2 adjacent internal loops – duplex – hairpin loop

Sequence 2:

AUCGGUUCCAGCAGGAACUGUACUCGGGGGCUCGGGAAACCCUCCCG GGGUUUUACCCGGGUUUUUACGU

Its secondary structure is formed by the following sub-structures (from top to the bottom of the image):

hairpin loop – duplex – interval loop – duplex –internal loop – duplex – hairpin loop

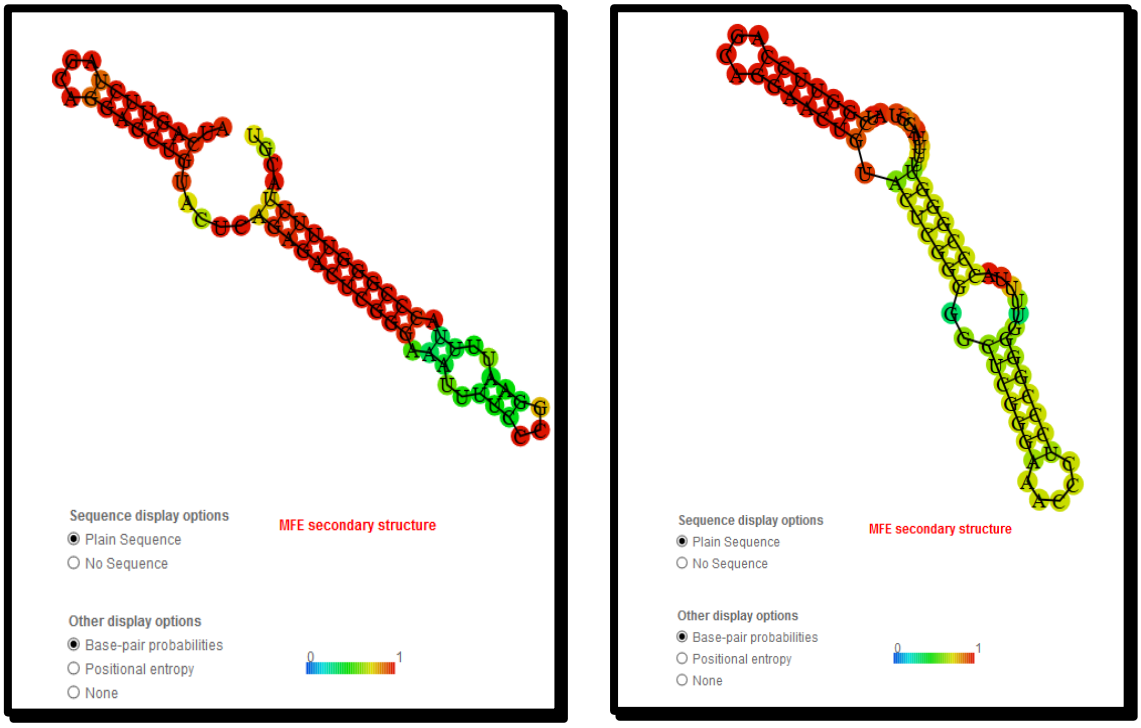

***Figure 1**: The two corresponding MFE structures obtained with RNAfold. The first (left) shows a concentrated zone around the double loop at the bottom with some base-pairs with lower probability (around 50%). The second (right) shows a spread zone around the bottom-half of the structure with several base-pairs with lower probability (around 60%), plus a slight torsion between the first and the second duplexes.*

1st SEQUENCE		2nd SEQUENCE	
BP index	**Probability**	**BP index**	**Probability**
[31, 61]	0.999933062	**[8 15]**	0.999960213
[33, 59]	0.999819173	**[5 18]**	0.999902333
[34, 58]	0.999704943	**[7 16]**	0.999621379
[35, 57]	0.999603073	**[6 17]**	0.999476892
[32, 60]	0.999304321	**[9 14]**	0.998518681
[30, 62]	0.999159668	**[4 19]**	0.994732334
[5, 18]	0.998548557	**[3 20]**	0.963067864
[6, 17]	0.997941879	**[26 58]**	0.844576643
[8, 15]	0.997787830	**[25 59]**	0.844520394
[7, 16]	0.997778302	**[27 57]**	0.844511677
[29, 63]	0.997371946	**[35 45]**	0.843816543
[4, 19]	0.996840463	**[36 44]**	0.843759360
[28, 64]	0.992920702	**[34 46]**	0.843717468
[36, 56]	0.992891738	**[24 60]**	0.843452012
[3, 20]	0.985527242	**[23 61]**	0.843052788
[27, 65]	0.981682106	**[28 56]**	0.841616016
[9, 14]	0.949761569	**[33 47]**	0.831778152
[26, 66]	0.874445058	**[32 48]**	0.818898899
		[37 43]	0.818237004
		[31 49]	0.806949843

***Table 1**: Table with the base-pairs with probability > 80% sorted from high to low in the first and second sequences, with 18 and 20 base pairs respectively. These data have been obtained with the corresponding Python script from the .eps files obtained through RNAfold.*

Exercises 3 & 4

Hamming distance	Base-pair distance
22	**30**

***Table 2**: Table with the Hamming and base-pair distances between the two sequences. These data have been obtained with the corresponding Python script.*

Exercise 5

Both sequences lead to two MFE foldings that share a common part corresponding to the top hairpin loop and duplex (Figure 1), so it is likely that the base composition variation given in the corresponding sequence zones is not so meaningful speaking in terms of the biological structure conservation. In fact, the top duplex could correspond to the 7 base-pairs observed in sequence 2, all of them with P > 95% of appearance in that structure (Table 1). However, from the first internal loop to the very bottom of both structures, a more varied zone between them is presented, as clearly shown with the Hamming and base-pair distances parameters, as well as the color-scale variation. This could be due to a specific base change near the internal loop involved in an important variation between both structures, and even in a change in the torsion of the second molecule, as cited before.

In the first one, a much more stable secondary structure is reached, with 16 base-pairs with P > 98%, most of them located in the upper and middle duplexes. Apart from this, it is important to point out a more variable zone concentrated at the bottom zone, with single bases and base-pairs appearing with a probability of around 50%, as commented before. In the second one, a less stable secondary structure is reached, with only 6 base-pairs with P > 98%, and most likely all of them located in the top part. Below the internal loop the structure shows single bases and base-pairs appearing with a probability of around 60% spread at the bottom-half of the structure. However, this does not necessarily mean that it is an unstable structure, mostly because of its 20 base-pairs with P > 80% of appearance.

In conclusion, it can be deduced that the first secondary structure is more stable that the second one, because despite it only has 18 base pairs with P > 80% of appearance instead of the 20 of the second structure, 16 of them are formed with 98% probability, three times more than those formed with the same probability in the second structure.

Exercise 6

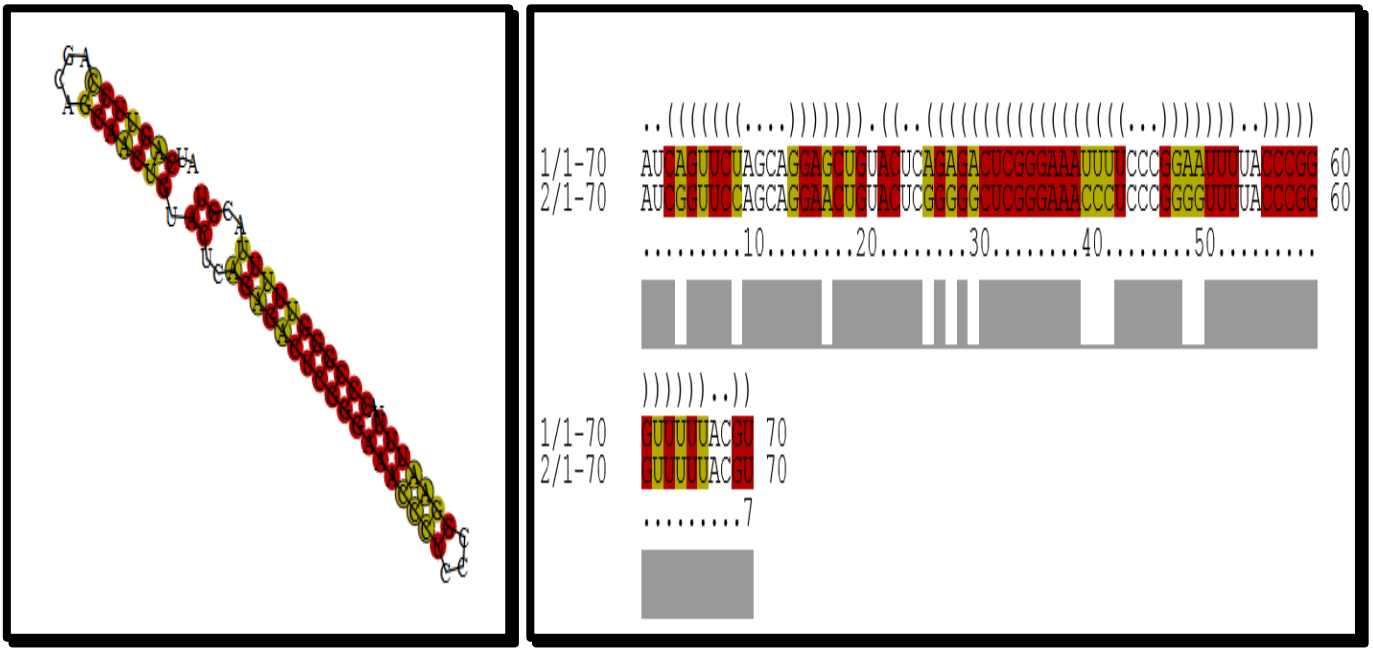

***Figure 2**: The consensus MFE structure obtained with RNAalifold (left) and its corresponding annotated alignment between the two sequences of analysis (right).*

1st SEQUENCE - CONSENSUS		2nd SEQUENCE - CONSENSUS	
Hamming distance	**Base-pair distance**	**Hamming distance**	**Base-pair distance**
12	**14**	**23**	**34**

***Table 3**: Table with the Hamming and base-pair distances between the consensus sequence and each of the sequences. These data have been obtained with the same Python script used to get these parameters in Table 2.*

It can be seen in the alignment at the right part of Figure 2 that there are 9 base pair changes in the consensus structure (colored in yellow). From these, 7 of them are consistent, where only one base changes. The other two are compensatory base pair changes, where both bases of the pair change, and they are located at [40,50] and [41,49], at the very end of the bottom duplex, corresponding with the more variable zone observed in the secondary structure of the first sequence. As this information is used by *RNAalifold* to obtain the consensus MFE structure (observed at the left part of Figure 2), this could mean that the software uses more compensatory mutations to reach a consensus stabilized structure at the more variable secondary structure zones of the single sequences.

REFERENCES

1. Bernhart SH, Hofacker IL, Will S, Gruber AR, Stadler PF. **RNAalifold: improved consensus structure prediction for RNA alignments.** BMC Bioinformatics. 2008 Nov 11;9:474.
2. Gruber AR, Lorenz R, Bernhart SH, Neuböck R, Hofacker IL. **The Vienna RNA Websuite.** Nucleic Acids Res. 2008