Exercises in RNA structure predictions

Submit your answers to a.p.goultiaev@liacs.leidenuniv.nl before 25.03.2019, 23:59.
Email subject: CMB2019 exercises.
It is sufficient to describe what has been done and to give answers to the questions.
Please don't send large files in the attachments.
The answers given as URL's will not be considered.

1. The 5' untranslated regions (5'UTR) of coronavirus RNA genomes contain regulatory RNA structures. Using the program mfold (http://mfold.rna.albany.edu/?q=mfold/RNA-Folding-Form), predict the low-energy secondary structures of the 5'UTR of two coronaviruses: Middle East respiratory syndrome coronavirus (accession NC_019843) and related coronavirus isolated from hedgehogs (NC_039207). Compare the mfold predictions reported as "Structure 1": give the nucleotide positions of the structural RNA elements that are likely to be homologous to each other.

2. Using the RNAalifold program provided by The ViennaRNA Web Services (http://rna.tbi.univie.ac.at), predict the consensus structure for 5'UTR's of several related coronavirus genomes. Make the dataset consisting of four coronavirus 5'UTR's: two sequences of the previous exercise (accessions NC_019843 and NC_039207) and the 5'UTR's of two bat coronaviruses (NC_009019 and NC_009020). Compare predicted consensus structure with the MERS coronavirus 5'UTR structure prediction by mfold from previous exercise: give examples of similarities and differences. In the RNAalifold prediction, try to identify structural elements supported by covariations (explore the structure-annotated alignment of the RNAalifold output). Give examples of covariations, e.g. reporting their nucleotide position numbers in the MERS virus genome sequence and covaried nucleotides in other three viruses.
Hint: In order to use RNAalifold, you have to align the 5'UTR sequences by e.g. Clustal Omega program (https://www.ebi.ac.uk/Tools/msa/clustalo/). The alignment yielded by Clustal Omega can be used as an input for RNAalifold.

3. A domain in the mRNA of plasmid R483 *pnd* gene (accession D00364, positions 184-240), has been suggested to undergo a rather slow conformational transition between two alternative secondary structures. Using the Barriers server of ViennaRNA Web Services (http://rna.tbi.univie.ac.at), generate the lowest free energy structure (global minimum) and a suboptimal structure in a local minimum separated from the global one by a barrier of at least 10 kcal/mol (change the default 0.1 value to consider only minima with a barrier higher than 10.0). Upon obtaining the results (~15 minutes), you will get the visualized energy landscape in the form of a tree showing the minima and connecting saddle points. You can also watch the animated folding path between two structures and save the folding path in a file with bracket views of intermediates.
What is the main difference between two main alternative structures? What are their free energy values? What is the value of predicted energy barrier between the local minimum and the global one?

4. Using the Rfam database collection of RNA families (http://rfam.org), determine the RNA family that could include the following RNA sequence:
`UUCCUGCUUCAACAGUGCUUGGACGGAA`
What is the name of this family? What is the sequence of hairpin loop in the target RNA upon the folding into the found motif structure? Generate the structure of this RNA using free energy minimization by mfold program (http://mfold.rna.albany.edu/?q=mfold/RNA-Folding-Form). What is the difference between this structure and the one suggested by the Rfam motif modeling?

5. Using the microRNA database miRBase (mirbase.org), identify the precursor stem-loop structure of human miRNA hsa-mir-22. What are the sequences of two miRNAs produced from the 5'proximal and the 3'-proximal parts of the precursor (hsa-miR-22-5p and hsa-miR-22-3p, respectively)? The database reports the deep sequencing evidence for these miRNA's: which of them is more likely to be real? Explore the hsa-miR-22-3p target mRNAs, predicted by the TargetScanVert algorithm. Give an example of target 3'UTR: target gene name, representative transcript accession (ENST...), target sequence paired to miRNA.