Exercise: RNA 3D modules discovery

Stefan E. Seemann seemann@rth.dk
(prepared together with Corinna Theis and Adrian Geissler)

The U4 spliceosomal RNA is a small nucleolar RNA (snRNA) involved in the splicing process through binding to other RNAs and proteins. This compactly structured RNA includes a kink-turn module which is involved in protein binding. In this exercise we will predict the RNA secondary structure of the U4 spliceosomal RNA and integrate 3D modules. To do so we will investigate in their structure conservation.

1. RNA secondary structure

   - Find the U4 spliceosomal RNA in the Rfam database `http://rfam.xfam.org/` and get the human sequence `M15957.1` from the alignment!
   - Fold the (ungapped) sequence with the RNAfold webserver `http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi` to get its RNA structure with minimum free energy. Compare the shape (stems and loops) of the MFE structure and the annotated structure in Rfam. Do you have an explanation why they may differ?

2. RNA 3D modules

   We screened the Rfam alignment of the U4 family (`http://rfam.xfam.org/family/RF00015/alignment?acc=RF00015&format=stockholm&download=0`) for RNA 3D modules with RMDetect. We subsequently clustered the predicted RNA 3D modules with RMCluster. The detected clusters are:

Cluster	Model	Count	Occurrences (%)	Score	BPP	MI	Start	End
1	KT	140	83.33	15.626	0.394	0.172	47	24
2	TGA	118	69.64	10.432	0.028	0.234	29	45
3	CL	16	6.55	2.899	0.042	0.518	77	64

   Please answer the following questions:

   - Familiarize yourself with the Rfam alignment. What data format is it? How many sequences are there?
   - What does RMDetect compute, and what do the columns in the output mean? Have a look at Figure 2 of the tools' publication (reading material: Cruz and Westhof (Nature Methods, 2011)) to look up the abbreviations.
   - Argue for the different clusters how well defined they are. What does the different scores (columns of the RMCluster output) mean and what do they tell you about the reliability of the clusters.

- Consider the location of the clusters (Start and End position in the alignment). If some clusters overlap try to explain why this may happened. How does this impact your conclusions which cluster(s) you trust most.

3. Integration of 3D modules in RNA secondary structure

   We applied RMDetect on the *H.sapiens* sequence of U4 Rfam family. RMDetect predicts two 3D modules:

Model	Model sequences	Start	End
KT	CCGAGG;CCAAUGAGG	40	26
TGA	UGAG;CGAG	30	41

   - Compare the predictions of RMDetect for a single sequence search with the predictions for multiple sequences (see task2).
   - Find the locations of detected RNA 3D modules in the RNAfold result (task 1)!
   - How well do the RNA 3D modules fit into the secondary structure? Discuss your observations!

4. Visualization and structure conservation of the kink-turn!

   - Search in PDB for the PDB identifiers 1E7K and 3CC2! These are crystal structures of which molecules and in which organisms?
   - Start PyMOL and fetch these two PDB files:
     `fetch 3CC2 1E7K, async=0`
   - Set the view of both molecule complexes to *Cartoon*! Can you identify the two molecule complexes (hint: turn them on/off)? Can you identify the RNAs (hint: shape of the cartoon)?
   - Both RNA molecules have at least one kink-turn module. Extract the kink-turns:
     `select kt-42, 3CC2 and (resi 1146-1155,1212-1217)`
     `select kt-u4, 1E7K and chain D and (resi 27-46)`
     Hide 1E7K and 3CC2 (hint: Hide → everything) and Show kt-42 and kt-u4 (hint: show → cartoon)!
     Color both kink-turns in your favorite colors! You can also show the nucleotide names (hint: label → residue name).
   - Now that we have selected both kink-turns we are going to align them:
     `align kt-42, kt-u4`
     How distant are the both modules? Look in the output of the PyMOL command line interface!
   - You would like to rotate your aligned kink-turns to inspect the similarity of both modules. To do that we center our view to kt-u4:
     `origin kt-u4`
   - Find both kink-turns in the kink-turn database `http://www.lifesci.dundee.ac.uk/groups/nasg/kturn/kturns_known.php`! Compare the primary sequences of the kink-turns! How does this relates to their similarity in tertiary structure (PyMOL view)?