Hands-On Six
RNA Secondary Structure Prediction

In this hands-on exercise, we will use publicly available programs to find the secondary structure of RNA sequences.

Problems 1 and 2 illustrate two ways of finding runs of complementary sequence in the RNA molecule:

5'-**AGCCAUUUUUUGGCU**-3'

Regions that can potentially base pair are displayed on a dot matrix.

Problem One

	A	**G**	**C**	**C**	**A**	**U**	**U**	**U**	**U**	**U**	**U**	**G**	**G**	**C**	**U**
A															
G															
C															
C															
A															
U															
U															
U															
U															
U															
U															
G															
G															
C															
U															

The table has the RNA molecule sequence both, across the top of the table and down the side.

- Place an X at each position where the bases across the top are capable of base pairing with those down the side (i.e. G/C and A/U base pairs).

Look for a diagonal row of X's running from the upper right of the matrix to the lower left.

a) What is the longest run of X's? ___________________________________.

b) Describe any symmetry in the matrix. ___________________________________.

c) What is the significance of this run? ___________________________________.

Problem Two

	A	G	C	C	A	U	U	U	U	U	U	G	G	C	U
A															
G															
C															
C															
A															
A															
A															
A															
A															
A															
U															
G															
G															
C															
U															

The table above shows the sequence of the RNA molecule across the top of the matrix and its reverse complement down the side.

- This time, place an X at each position where the bases across the top and down the side are the same (i.e., A/A, C/C, etc…).

Now look for a diagonal row of X's running from the upper left of the matrix to the lower right.

a) What is the longest run of X's? ______________________________ .

b) Describe any symmetry in the matrix. ______________________________ .

c) What is the significance of this run? ______________________________ .

Problem Three

A) We are going to get the complete genome of the potato spindle tuber viroid (PSTV) from NCBI. Its accession number is NC_002030.

- Go to http://www.ncbi.nlm.nih.gov/
- Choose “Nucleotide” from the drop-down menu.
- Type “NC_002030” in the window to the left of the blue “Search” button.
- Click on the blue “Search” button.
- From the new page, choose “FASTA” at the top of the page (on the left).
- Copy and save the sequence in a file. Do not close the NCBI window. We need it for part B.

B) We will get the reverse complement of the sequence.

- Open “Customize View” by clicking in the box to its right.
- Click in the “Show reverse complement” box.
- Click on “Update View” and save the sequence in the file you created in part A).

C) Use the DNA sequence dot matrix analysis feature at: http://arbl.cvmbs.colostate.edu/molkit/dnadot/index.html with various combinations of window and stringency (mismatch limit) options to compare these two sequences (hint: use a long window and quite relaxed stringency). The analysis is similar to that of Problem 2.

What type of secondary structure can be formed by the molecule? ________________________.
Hint: Use window size of 13 and match limit of 5.

Problem Four

In this problem you will go to Michael Zuker's RNA analysis package MFOLD (multiple foldings of RNA) at: http://mfold.rna.albany.edu/ to find the minimum energy structure of the RNA molecule of Problem 3.

- Click on “Mfold Web Server” from the menu at the top of the page.
- Click on “RNA Folding Form” under “Applications” on the left hand side of the page.
- Copy and paste the PSTV sequence (from Problem 4) into the input window of MFOLD and accept the default program parameters provided.
- Click on “Fold RNA” at the bottom of the page, and wait for the results.
- To view the output structures, scroll down to “View Individual Structures” and click on structures 1, 2, 3, etc….

Describe the first structure reported, find its energy, and write it down ______________________ .

Problem Five

Consider the following sequence:

**GGGGAATTAGCTCAAGCGGTAGAGCGCTTGCTTAGCATGCAAGAG
GTAGTGGGATCGATGCCCACATTCTCCA**

You are going to investigate and see what type of sequence it is.
Use at least 3 different methods/procedures to support your claim.
Hint: the following URL might be helpful (as a last resort) http://selab.janelia.org/tRNAscan-SE/