

BIO231/CS232 - Computational Biology I

Assignment AM1

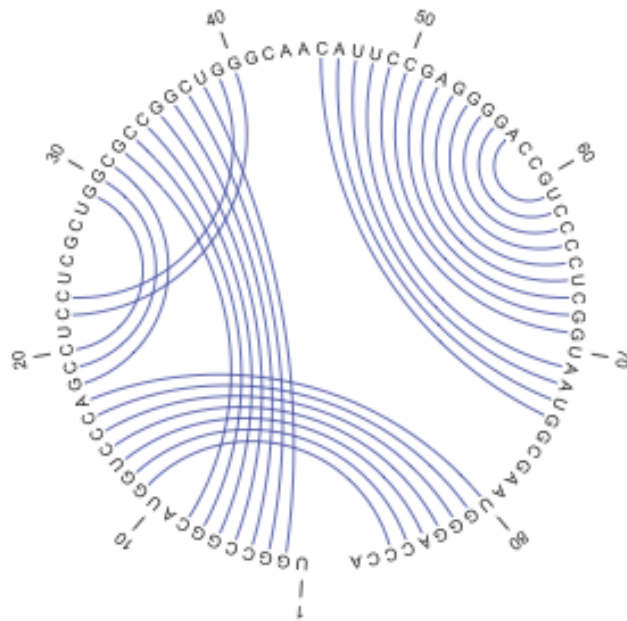
Assignment Due: April 1, 2022 5pm

Instructions

- Read the section on plagiarism in your student handbook. Plagiarised assignments will be referred to the DC.
- Assignments must be submitted in the “Assignment” section on LMS as a zip file containing a single word document for section 1 & 2, and individual files for codes. Code must be properly commented.
- Hard copy of the assignment must also be submitted to the TAs within the deadline.
- For programming questions, **DO NOT** use Biocoductor, Biopython or any other built-in libraries/packages.
- Assignments submitted on Dropbox, emailed assignments or late submissions will not be accepted.
- This is an individual assignment. Discussion is **NOT** allowed.

1 RNA Structure

1. Draw the RNA structure corresponding to the following circular representation. [5]



2. Using the energy values from Table 5.2 from your reading material, calculate the free energy for the hair-pin loop resulting from the folding of 5'-CCUUGGAACCACCAGGG-3'. [5]
3. Predict the secondary structure of 5'-UGCUCUAGUACGAGAGGACCGGAGUG-3' using Nussinov-Jacobson algorithm. [10]
4. Predict the secondary structure of 16S ribosomal RNA in humans using mfold tool. [5]

2 Proteins

1. Download PDB file (PDB ID: 5CHB) from Protein Data Bank (<http://rcsb.org>) and answer following questions. [5]
 - (a) How many amino acids are present in this protein structure?

- (b) Which method is used to determine this structure?
 - (c) What is the resolution of the structure? Is it good or bad?
 - (d) Use an online tool to generate Ramachandran plot for this protein.
 - (e) Discuss the output of Ramachandran plot.
2. Find two structures for each of the following domain classes from PDB and generate high quality images of each structure using an online tool. [5]
- α Domain
 - β Domain
 - $\alpha+\beta$ Domain
 - Multi-domain
 - Membrane protein

3 Programming

1. Write a program that takes DNA sequence as input and output six reading frames, and their amino acid sequence. [5]
2. Write a program that takes any PDB file as input and output the following details. [5]
 - Amino acid sequence
 - Number of hydrophobic residues
 - Number of protein chains
 - Number of helices
 - Number of sheets