CS425
Assignment 2 (Due Feb 23, 2017)

Computer Science Department
Colorado State University

February 15, 2017

1. RNA folding with a limited number of multi branch loops [35 pts].
Suppose $k \in \mathbb{Z}$ is a non-negative second input (in addition to the input sequence R) to the RNA folding algorithm. Devise a dynamic programming algorithm that finds a minimum free energy pseudoknot-free structure of R (according to the Turner model used in the McCaskill's algorithm) subject to the constraint that the structure contains at most k multi branch loops.

2. Predicting pseudoknots is hard or maybe not? [35 pts].
Give a polynomial time algorithm for prediction of a structure with maximum number of canonical base pairs for a given input sequence R. Note that pseudoknots are allowed. Hint: use maximum matching. *Remark:* Predicting pseudoknots according to the Nearest Neighbor energy model is hard. Essentially, the interdependency introduced by the Nearest Neighbor model renders the problem hard. In this problem, we show that predicting pseudoknots according to the base pair counting model is in P.

3. Hands-on experience [30 pts].
Download and compile UNAFold and VARNA from `http://mfold.rna.albany.edu/?q=DINAMelt/software` and `http://varna.lri.fr` respectively. Download OxyS and fhlA sequences from the course website and run UNAFold with the default parameters to predict their minimum free energy structures. Visualize the predicted structures using VARNA, and report the minimum free energies and structures. Does the predicted structure look like what we saw in class?

Upload your answer on Canvas in one zip file or tarball. Include all the code/scripts you have written in your submission.