Mitra Shokat
Bio 131
Spring 2017

Predicting the Secondary Structure of RNA

The structure of RNA differs from that of DNA in a few key ways. First, the nitrogenous base thymine (T) found in DNA is replaced by uracil (U) in RNA. Second, RNA is single-stranded while DNA is double-stranded. This single-strandedness allows for intramolecular base pairings between nitrogenous bases on the same strand of RNA. The pattern

up. I chose to test my program on this small sequence because it allowed me to compare the program's output to the "known" secondary structure of the strand of RNA (Figure 1).
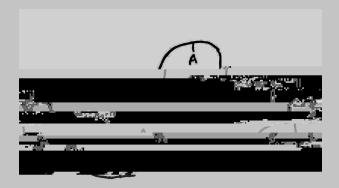

Figure 1. Predicted RNA secondary structure of test sequence.

After confirming that my program worked on a short sequence of RNA, I attempted to implement it on the RNA primary sequence of alanine tRNA ('AGGGAAAAUAGUUUAAUAAAAAUAUUUUACUUGCAGUAAAAAGUUAUUUCUAU AAUUUUUCUUU'). Again, by using a sequence with a known secondary structure, I was able to evaluate the efficacy of my program.

The basic approach for this algorithm is to create a dynamic programming table that compares the string of RNA to itself. The table is initialized with zeros on the main diagonal and on the diagonal to the left of the main one. Each index of the table is then filled in with a score. At each index i,j , there are four possible choices for determining the score, and thus the maximum of these scores is chosen for the table. At an index i,j , the goal is to determine the optimal secondary structure that contains the most base pairings. This is done by breaking down the sequence into subsequences, and breaking down those subsequences further into even smaller subsequences. The first choice for index i,j is to pair rna[i] and rna[j] and attach to best structure

ird
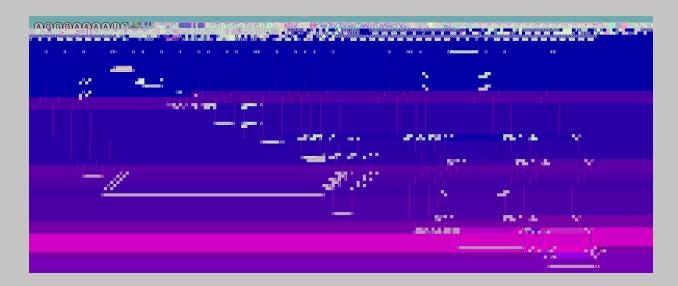
bine two

ne

the

cktracking

of the

h the

Figure 5. Output graphics of algorithm using alanine tRNA sequence as input. Red lines indicate base pairings.

Thus, it is evident that my implementation of the Nussinov algorithm, which is actually a slight variation of the original algorithm, still needs to be improved. One possible option for improving the scoring conditions would be to somehow weight consecutive base pairings, which are more favorable in RNA secondary structures.

References: