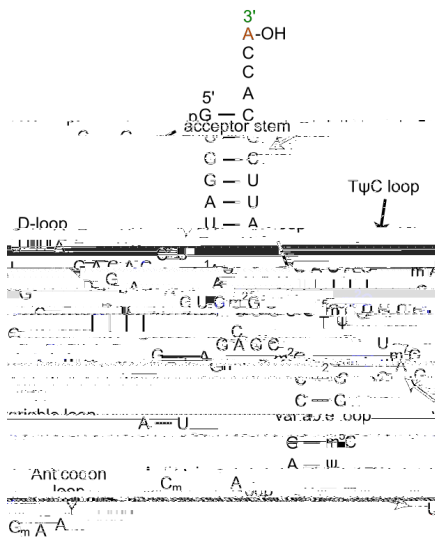


RNA Secondary Structure Prediction

Background: Though the primary structure of RNA, single stranded and linear such as in mRNA, is highly important for the functions of a cell, it accounts for only between 1%-5% of cellular RNA. The rest of the cellular RNA is composed of RNA in its secondary structure, the majority in the form of tRNA or rRNA. The secondary structure of RNA is arrived at through RNA molecules folding in upon themselves and forming bonds between their own complementary nucleotide bases. The structure of the resulting molecule determines its function



within the cell.

Figure 1. Secondary structure of tRNA^{Phe} from yeast.

The secondary structure of RNA is difficult to predict because the number of possible structures increases exponentially as the sequence length increases. It is generally accepted that the most stable possible structure is the most likely to occur, and that the most stable possible structure will be the one with the most bonded complementary base pairs.

Data Collection: Data was collected from the RNA STRAND database from the BETA Lab of the University of British Columbia. This includes the sequence of the RNA molecule and the predicted secondary structure of that sequence.

High Level Program Steps: Given a string of text the program will find all the possible ways of 'splitting' the string into two new ones to simulate the folding of RNA. It essentially cleaves the string at a given point and arranges the two resulting strings into the directions they would be had it really folded. For each split possible it calculates the number of aligning complementary bases and returns the split with the highest score. It then finds sections within the alignment that

are unpaired and performs the function again on the sequences of both sides of the unpaired section. Finally it performs the function again on any part of the sequence that extended past the alignment based on the length of the two strings.

Discussion of Results:

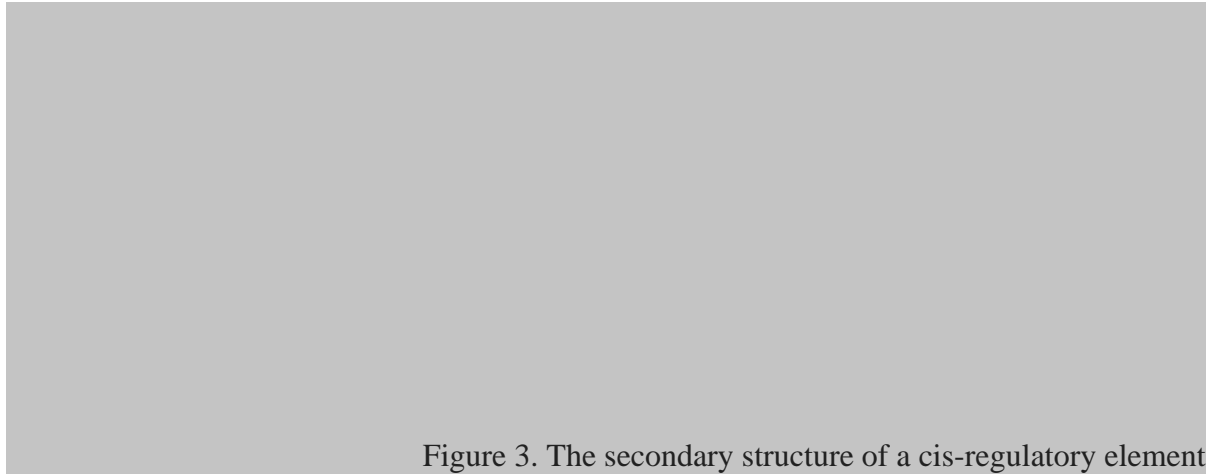


Figure 3. The secondary structure of a cis-regulatory element (Molecule ID: RFA_00640). It contains 36 paired bases, my program predicts 38. The structures



