

## BioE 131 HW3

Yifan Chen 10/23/2017

I worked Sara Xiuhui Gao

---

1. As a mathematical explanation, mutual information measures the amount of information that one random variable conveys about another. High mutual information indicates a large correlation between the two random variables. When we have the information about one of the variables, high mutual information means a large reduction of uncertainty in identifying the other. In the contrary, low mutual information between two random variables indicates they are not highly related, and zero means the two are independent of each other. In terms of a biological explanation, in the field of RNA or DNA sequence, the high mutual information between two positions indicates that the specific sites in the sequence always have a particular nucleotide combination where the sequence can establish a certain folding structure and later can be translated into a specific functional protein if it can be translated into proteins. As the structural Central Dogma states: "sequence determines structure determines functions." Thus, high mutual information between position  $i$  and  $j$  implies that position  $i$  and  $j$  may have some underlying interaction or binding connections vital to the structure and the function. Therefore, we can use high mutual information to predict structure and function of RNAs in this case.

2. The sequence of the tRNA

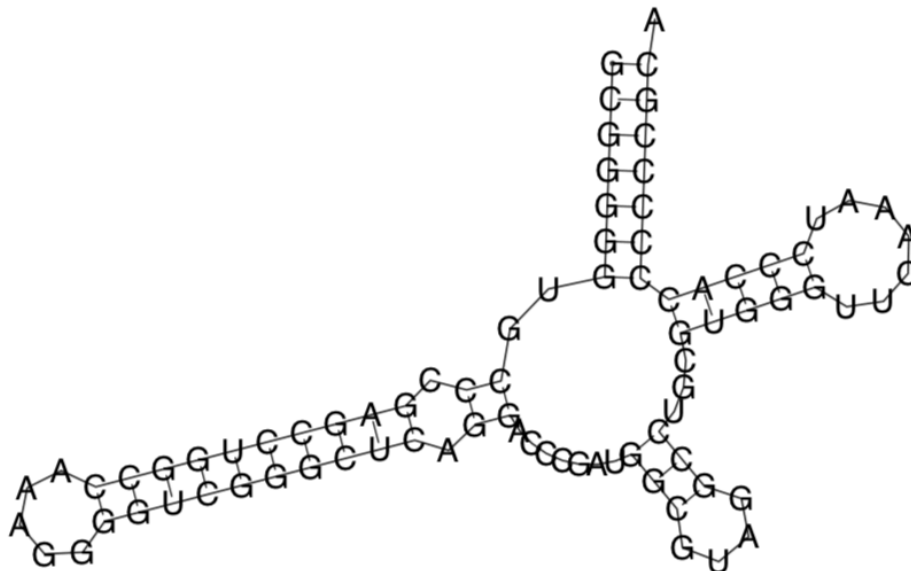
>AP000063.1/59179-59095

GCGGGGGUGCCCGAGCCUGGCCAAAGGGGUCGGGCUCAGGACCCGAUGGCGUAGGCCU

GCGUGGGUUCAAAUCCACCCC

CCGCA

The simulated the structure of AP0000631/59179-59095 by RNAfold, ViennaRNA.



3. Most of the “bottom 10” columns are the gap positions in the MSA. Since most of the sequences have gaps in those positions, the “bottom 10” columns are conserved with a relatively low entropy. With the help of ViennaRNA package, the results show that nucleotides in those positions do not have base pair with other nucleotides in the secondary structure and do not have an explicit correlation with other columns. No matter what nucleotides in other column positions, the “bottom 10” positions will remain conserved.

And the “top 50” pairing columns are most likely the conserved basepairs with a strong correlation to each other. ViennaRNA shows that such positions form hydrogen bonds to bind in pairs in almost all of the sequences. And they tend to fold into stems in secondary structures. Therefore, the columns with highest mutual information always form hydrogen bonds and pair with each other to fold stems in the secondary structure.

In the sequences, conserved positions might form base pairs in RNA, and have critical residues (such as active sites and binding pockets in both RNA and protein), functional domains in proteins, or protein-coding genes in genomes. By doing sequence alignment, we can easily find out the conserved elements, the structure and the function of RNA/ proteins regarding specific positions.