

## Solution: Measures to detect the effect of SNPs on RNA secondary structure

### 1. Structural effect on MFE structure

- (i) Case1 Hamming distance: 4 (there are four mismatches between the WT and MUT structure);  
Base pair distance: 9 (5 base pairs present in WT but not in MUT and 4 base pairs present in MUT but not in WT)
- Case2 Hamming distance: 4 (there are four mismatches between the WT and MUT structure);  
Base pair distance: 12 (6 base pairs present in WT but not in MUT and 6 base pairs present in MUT but not in WT)
- (ii) The hamming distance did show no difference between the impact of the SNP on the RNA secondary structure between both cases. However, the base pair distances did show a difference. The base pair distance is more suited for this comparison because it takes the base pairing partners into account whereas the hamming distance does not.

When you want to write a python script to calculate the base pair distance then, at first, you have to read the RNA secondary structure in dot-bracket notation. This can be done with the help of a stack (in python use the data type 'list') which allows to add or remove elements only from the top of the stack (last in first out). Each time you read an opening base pair '(' add its index on top of the stack. Each time you read a closing base pair ')' remove the top entry from the stack which is the index of the opening base of this base pair (the base pair partner).

### 2. Structural effect on ensemble structures

- (i) local region from position 156 to 236
- (ii) The hamming distance between both structures is 59. The MFE structure changes in the region from position 156 to 236 and this local region overlaps with the local region predicted by RNAsnp which has been predicted from base pair probabilities.
- (iii) The base pair in the wild type sequence with highest probability has a probability of 0.9136 at position (54,63) in the local region. In the mutant this base pair has probability 0.0348.
- (iv) The local region of altered RNA structures has 27 (18) basepairs of probability greater than 0.5 (0.8). In the wildtype this region has 18 (6) basepairs of probability greater than 0.5 (0.8). There is a consistent pattern that the number of base pairs in the mutant is higher than the wild-type. In particular for probability above 0.8, the number of base pairs in mutant is 3 times higher than the WT. This shows that the mutant gets stable structure in the affected region compared to WT, however, this information can not be deduced with the Hamming distance because the Hamming distance considers only the mismatches between the two structures being compared.