CSE 417: Homework 8

Name: Qingchuan Hou

Student ID: 2127437

UWNetID: qhou

Problem 1:

After get the OPT[] list from Nussinov's algorithm, we can use traceback algorithm to find the result list.

Assume an RNA sequence has n numbers. In the OPT list, we search the OPT[i,j] in sequence by j from 0 to n, and i from j to 0, When use the traceback, we will search the result in reverse order, by j from n to 0.

There are 3 situations in the traceback:

1. OPT[i,j] = 0. This is the basic case, which means that there are no matching combinations in the remaining RNA sequences
2. OPT[i, j] = OPT[i, j-1], this case means that RNA[j] does not belong to any pair, so we ignore RNA[j], and then search the sequence (i, j-1).
3. OPT[i, j] > OPT[i, j-1], this case indicates RNA[j] and belongs to one of the pair. But we need to search for RNA[t] combined with RNA[j], according to Nussinov's algorithm, we only need to find an index t in the sequence i to j that matches OPT[i, t-1] + OPT[t+1, j-1] + 1 = OPT[i, j]. Then again look for the remaining combinations in the sequences RNA[i] to RNA[t-1] and RNA[t+1] to RNA[j-1].

Pseudocode:

```
   Initall a zeros list A to store the result

 Find-Sulution[i, j]:

     if OPT[i, j] == 0:
        None

     elif OPT[i, j] == OPT[i, j-1]:
        Find-Sulution[i, j-1]

     else:
        set A[j] to -1       // means it is a member from a pair

        while t form i to j-4
           if OPT[i, t-1] + OPT[t+1, j-1] + 1 = OPT[i, j]
               set A[j] to -1            // means it is a member from pair (t,j)
               Break the while          // only need find one t for one solution
        Find-Sulution[i, t-1]
        Find-Sulution[t+1, j-1]

     return A
```

The number in list A show the result, number 1 and -1 is pairs.

Run Time:

In the traceback algorithm, every time find a member is j in a pair, it need let t go though the OPT[i, i] to OPT[i, j-4], There are two loop in the traceback algrithm, and each loop will go through evert remin number in list. Therefore, the total run time is: $O(n^2)$

The run-time of Nussinov OPT table construction is $O(n^3)$. In the Nussinov OPT table, we need to find the most suitable t for each OPT[i,j]. So there are 3 loops for each i, j, and t. In traceback, we don't need to search for t for each OPT[i,j], we just need to search for t as j decreases. So traceback has one less loop than Nussinov OPT。