CAP 6515 ASSIGNMENT #3

1 Solution to Question No: 1

1.1 Problem Statement

Formalizing the pseudocode for predicting the secondary structure of RNA sequence based on the maximum number of possible base pairs (Nussinov Model) such that no base-pair cross each other. Here we're considering only canonical base pairs (G – C, A – U, and G – U). So a base-pair is complementary if it belongs to the set, R = {(G,C), (C,G), (A,U), (U,A), (G,U), (U,G)}

1.2 Algorithm Description

Suppose a RNA sequence S with length n is given. We have to find the secondary structure of sequence S following the recursive equation given below:

$$M(i,j) = max \begin{cases} M(i,j-1), \\ M(i,k-1) + M(k+1,j-1) + \delta(k,j) \\ \qquad\qquad \text{for } i \leq k < j. \end{cases}$$

The pseudocode for predicting RNA secondary structure using Nussinov's model is given below. Here, M[1][n] is the maximum number of base pairs in the optimal base-paired structure for S[1...n]. The secondary structure can be found by tracing back through the array M. Another 2D array called K_backtrack is used to keep track of the positions of k for which the value of M[i][j] is maximum.

Algorithm 1 Nussinov Model: Maximum Possbile Base Pairs Calculation Algorithm

```
n ← length(S)                                   ▷ Length of RNA sequence
M[i,i] ← 0                                      ▷ Initialization
M[i,i-1] ← 0                                    ▷ Initialization
K_backtrack[i,j] ← 0                            ▷ Stores index of k for backtracking
for L ← 2 to n do
    for i ← 1 to n-L+1 do
        j ← L+i-1
        Temp_Max ← INT_MIN
        for k ← i to j-1 do
            if S[k] and S[j] complementary then
                δ(k, j) = 1
            else
                δ(k, j) = 0
            if k-1 = 0 then
                Temp ← M[k+1][j-1]+δ(k, j)
            else
                Temp ← M[i][k-1]+M[k+1][j-1]+δ(k, j)
            if Temp > Temp_Max then
                Temp_Max ← Temp
                K_index ← k
        M[i,j] ← max(M[i][j-1], Temp_Max)
        if Temp_Max > M[i][j-1] then
            K_backtrack[i,j] ← K_index          ▷ Storing the position of k for which the value of M[i][j] is maximum
Traceback(1, n)
return M[1][n]                                  ▷ Maximal number of base-pairs in the secondary structure of S
```

Algorithm 2 Nussinov Model: Traceback Algorithm

```
1:  Procedure Traceback(i, j)
2:  if j <= i then
3:      return
4:  else if M[i][j] = M[i][j-1] then
5:      Print(".")
6:      Traceback(i, j-1)
7:  else
8:      k ← K_backtrack[i,j]
9:      if S[k] and S[j] complementary then
10:         Print(")")
11:         Traceback(k+1, j-1)
12:         Print("(")
13:         Traceback(i, k-1)
14:     else
15:         Print(".")
16:         Traceback(k+1, j-1)
17:         Print(".")
18:         Traceback(i, k-1)
```