# STRUCTURAL BIOINFORMATICS LAB

Task II 2021

TA. Esraa Hamdi  TA Nourhan Mohammed

## Task Overview

In this task, you will implement the Nussinov algorithm for RNA secondary structure prediction.

## Background

### Nussinov Algorithm (Dynamic Programming Algorithm)

**Idea (biological):** Stacked base pairs of helical regions are considered to stabilize an RNA molecule.

→ maximize the number of base pairs.

**IN:** RNA sequence S

**OUT:** a non-crossing RNA structure P of S that maximizes

| P | (i.e. the number of base pairs in P).

- Nussinov considering (1) ij pair, (2) i being unpaired, (3) j being unpaired, and even (4) bifurcation:

$$S(i,j) = \max \begin{cases} S(i+1,j-1)+1 & [\text{if } i,j \text{ base pair}] \\ S(i+1,j) \\ S(i,j-1) \\ \max_{i<k<j} S(i,k) + S(k+1,j) \end{cases}$$

- Init: $\forall i = 1..|S|$: $S_{i,i} = 0; \forall i = 1..(|S|-1)$: $S_{i,i+1} = 0$
- Termination: $S_{1,|S|}$ = max. number of base pairs

***Note that the matrix is filled diagonally.***

As for the **traceback**, we start with the upper right cell, we travel through the optimal path and we record every diagonal (i,j) as a ( ) , the "(" on the lower index, ")" on the higher index, so if the cell is (3,5) :

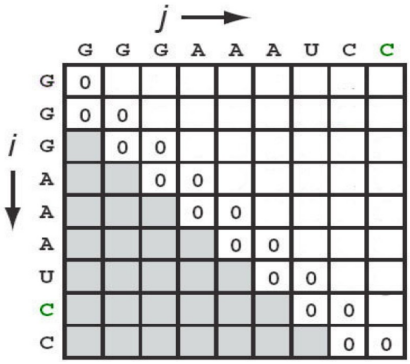
| | | | | | | |
|---|---|---|---|---|---|---|
| | | ( | | ) | | |

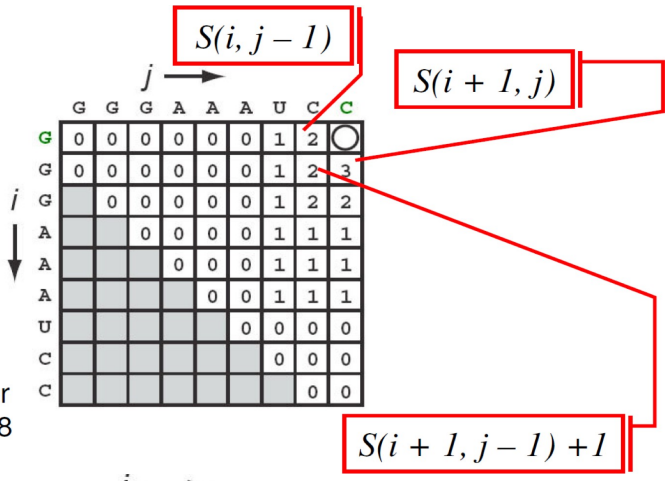
The rest of the empty cells would be dots.

Note that in case of bifurcation, you'd go to the 2 cells and record the diagonals as mentioned before.

We finish by reaching the diagonal of the matrix.

## Example :

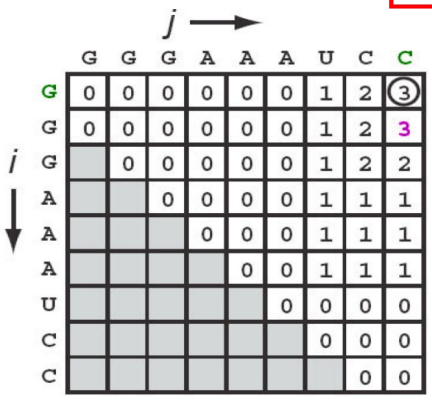


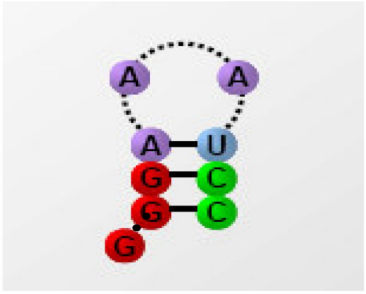




We still have to consider bifurcations for k=2,3,4,5,6,7,8

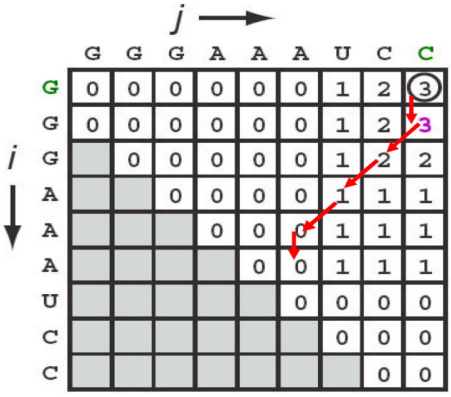
S(1,9)
= max { 2, 3, 2, 2 }
= 3.

Traceback to find the actual structure.




Nussinov algorithm gives the structure with maximum number of base pairings,
but does not always create viable secondary structures




which is the same as : .(((..)))

## Pseudocode for filling the matrix :

1. **Initialization:** fill the main diagonal and the diagonal just below it with zeros
2. Formally, the scoring matrix, M, is initialized:
    - M[i,i] = 0 for i = 1 to L (main diagonal)
    - M[i,i-1] = 0 for i = 2 to L (diagonal below main diagonal)
3. **Matrix Fill:** Starting with all subsequences of length 2, to length L

   **do**

   M[i,j] = max of the following :

    - M[i+1,j] (base i is hanging off by itself)
    - M[i,j-1] (base j is hanging off by itself)
    - M[i+1,j-1] + S(xi, xj) (bases i and j are paired; if xi = complement of xj, then S(xi, xj) = 1; otherwise it is 0)
    - M[i,j] = MAXi<k<j (M[i,k] + M[k+1,j]) (merging two substructures)

**Another test case :**

Initialization

| C | G | G | A | C | C | C | A | G | A | C | U | U | U | C |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |

| | |
|---|---|
| C | 1 |
| G | 2 |
| G | 3 |
| A | 4 |
| C | 5 |
| C | 6 |
| C | 7 |
| A | 8 |
| G | 9 |
| A | 10 |
| C | 11 |
| U | 12 |
| U | 13 |
| U | 14 |
| C | 15 |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | | | | | | | | | | | | | | |
| 0 | 0 | | | | | | | | | | | | | |
| | 0 | 0 | | | | | | | | | | | | |
| | | 0 | 0 | | | | | | | | | | | |
| | | | 0 | 0 | | | | | | | | | | |
| | | | | 0 | 0 | | | | | | | | | |
| | | | | | 0 | 0 | | | | | | | | |
| | | | | | | 0 | 0 | | | | | | | |
| | | | | | | | 0 | 0 | | | | | | |
| | | | | | | | | 0 | 0 | | | | | |
| | | | | | | | | | 0 | 0 | | | | |
| | | | | | | | | | | 0 | 0 | | | |
| | | | | | | | | | | | 0 | 0 | | |
| | | | | | | | | | | | | 0 | 0 | |
| | | | | | | | | | | | | | 0 | 0 |

| C | G | G | A | C | C | C | A | G | A | C | U | U | U | C |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |

Matrix Fill stage

Finished
length=15

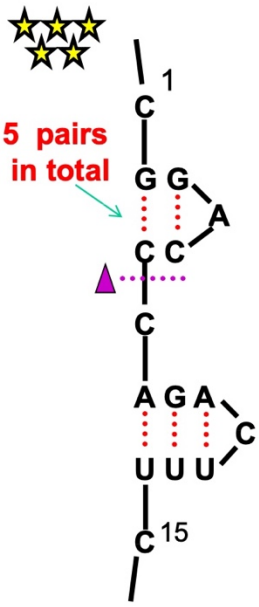
○ bifurcation occurred.

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 4 | 4 | 5 | 5 |
| 0 | 0 | 0 | 0 | 1 | 2 | 2 | 2 | 3 | 3 | 3 | 4 | 4 | 5 | 5 |
| | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 3 | 3 | 4 | 4 |
| | | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 2 | 3 | 3 | 3 |
| | | | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 2 | 2 | 3 | 3 |
| | | | | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 2 | 2 | 3 | 3 |
| | | | | | 0 | 0 | 0 | 1 | 1 | 1 | 2 | 2 | 3 | 3 |
| | | | | | | 0 | 0 | 0 | 0 | 1 | 2 | 2 | 3 | 3 |
| | | | | | | | 0 | 0 | 0 | 1 | 1 | 2 | 2 | 2 |
| | | | | | | | | 0 | 0 | 0 | 1 | 1 | 1 | 1 |
| | | | | | | | | | 0 | 0 | 0 | 0 | 0 | 0 |
| | | | | | | | | | | 0 | 0 | 0 | 0 | 0 |
| | | | | | | | | | | | 0 | 0 | 0 | 0 |
| | | | | | | | | | | | | 0 | 0 | 0 |
| | | | | | | | | | | | | | 0 | 0 |

| C | G | G | A | C | C | C | A | G | A | C | U | U | U | C |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |

Traceback stage

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| C | 1 | 0 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 4 | 4 | 5 | 5 |
| G | 2 | 0 | 0 | 0 | 0 | 1 | 2 | 2 | 2 | 3 | 3 | 3 | 4 | 4 | 5 | 5 |
| G | 3 | | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 3 | 3 | 4 | 4 |
| A | 4 | | | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 2 | 3 | 3 | 3 |
| C | 5 | | | | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 2 | 2 | 3 | 3 |
| C | 6 | | | | | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 2 | 2 | 3 | 3 |
| C | 7 | | | | | | 0 | 0 | 0 | 1 | 1 | 1 | 2 | 2 | 3 | 3 |
| A | 8 | | | | | | | 0 | 0 | 0 | 0 | 1 | 2 | 2 | 3 | 3 |
| G | 9 | | | | | | | | 0 | 0 | 0 | 1 | 1 | 2 | 2 | 2 |
| A | 10 | | | | | | | | | 0 | 0 | 0 | 1 | 1 | 1 | 1 |
| C | 11 | | | | | | | | | | 0 | 0 | 0 | 0 | 0 | 0 |
| U | 12 | | | | | | | | | | | 0 | 0 | 0 | 0 | 0 |
| U | 13 | | | | | | | | | | | | 0 | 0 | 0 | 0 |
| U | 14 | | | | | | | | | | | | | 0 | 0 | 0 |
| C | 15 | | | | | | | | | | | | | | 0 | 0 |




**You may ask any additional questions by email**

**Or on the assigned support time.**

**Good Luck ☺**