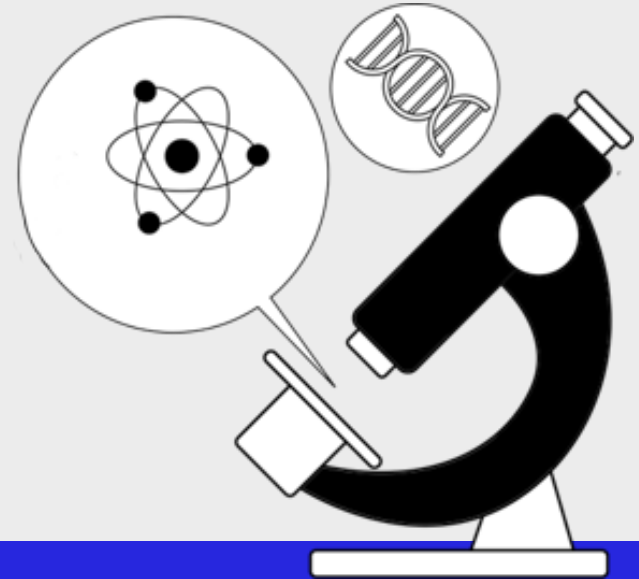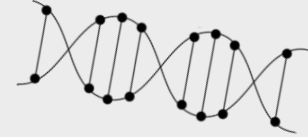# RNA structure

Saeedeh Akbari

Department of Computer Engineering

Sharif University of Technology

Fall 2023
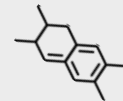
Table of contents

01

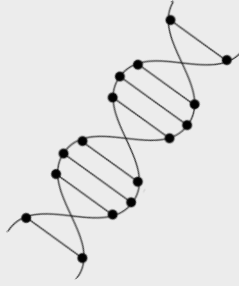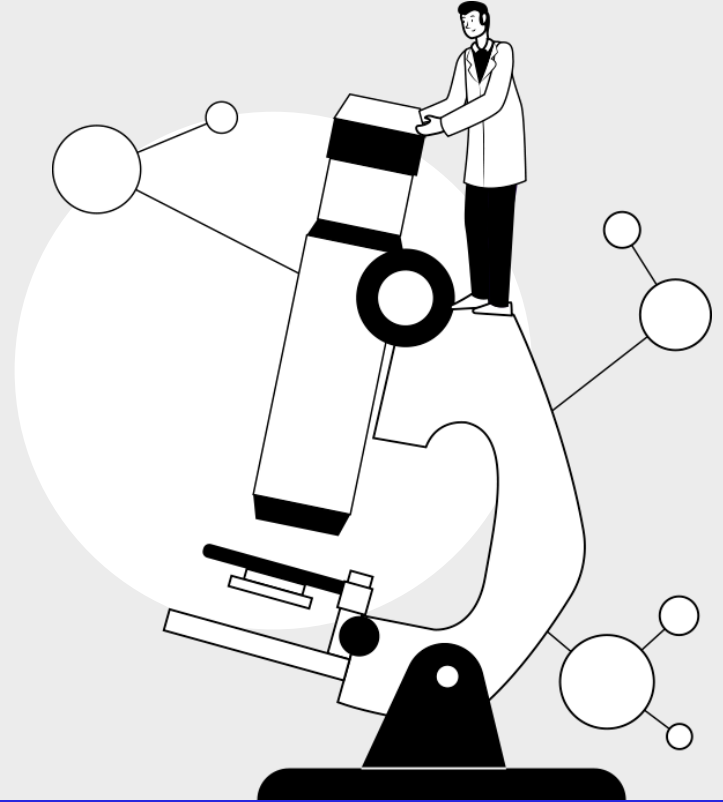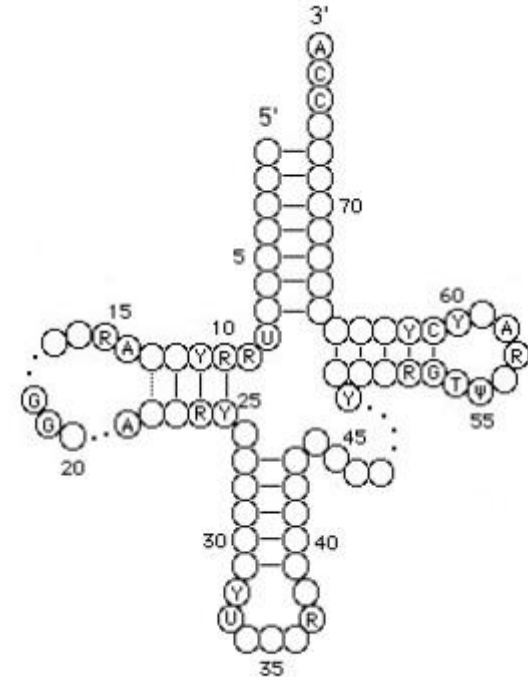# Introduction

Various types of RNA

- messenger RNA (mRNA)
- transfer RNA (tRNA)
- Ribosomal RNA (rRNA)
- small interfering RNA (siRNA)
- micro RNA (miRNA)
- small nuclear RNA (snRNA)
- small nucleolar RNA (snoRNA)
- guide RNA (gRNA)
- efference RNA(eRNA)

RNA Basics

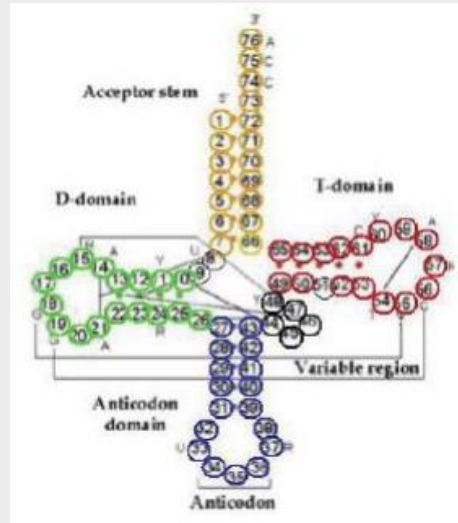- RNA bases: A, C, G, U
- Watson-Crick Pair
    - A-U (~ 2 kcal/mol)
    - G-C (~ 3 kcal/mol)
- Wobble pair
    - G-U (~ 1 kcal/mol)
- Non-Canonical pairs (modified suitably)
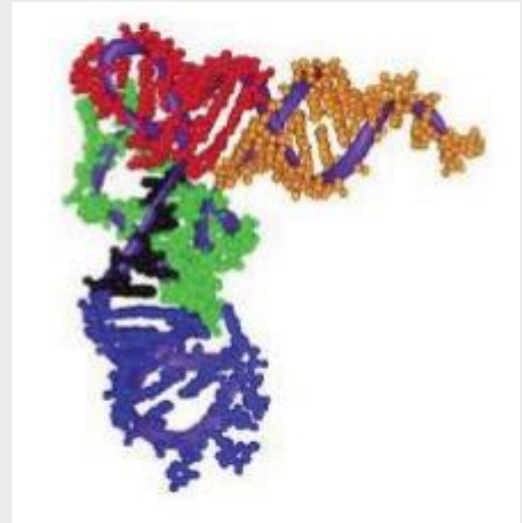- Bases can only pair with one other base

Secondary and Tertiary Structure

GCGGAUUUAGCUCAGUUGG
GAGAGCGCCAGACUGAAGA
UCUGGAGGUCCUGUGUUCG
AUCCACAGAAUUCGCACCA

Primary structure

Secondary structure

Tertiary structure

Secondary Structural Elements

a
Double-stranded segment (dsRNA)

Bulge — $i \geq 1$

Internal loop (if $i = j$, symmetric internal loop; if $i \neq j$, asymmetric internal loop) — $i \geq 1$, $j \geq 1$

Hairpin (stem-loop) — $3 \leq i \leq 8$

b
Two-stem junction (coaxial stack)

Pseudoknot

Kissing hairpins
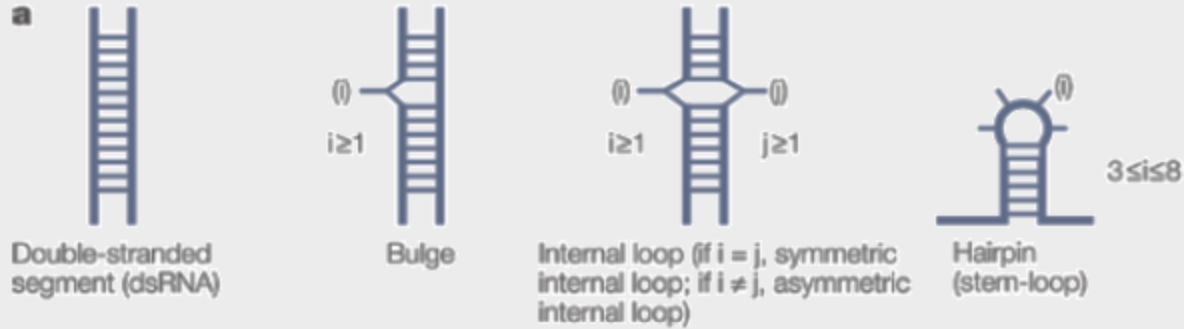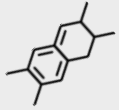
RNA Secondary structure/Motifs
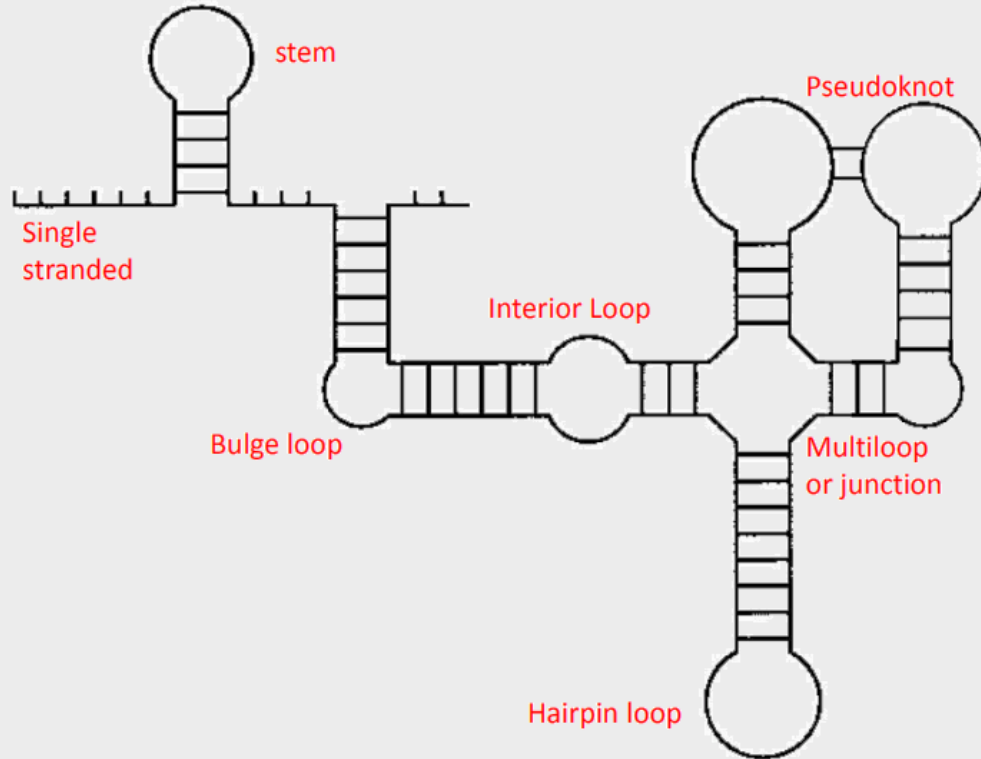
RNA Motifs Regulatory Effects

- Regulations of translations
- Processing of RNA
- Catalytic modification of other RNAs
- Transport and position in the cell
- Stability of RNA-transcript
- Expression of encoded proteins

Cells' mechanisms

Why predict structures?

- Current physical methods (X-Ray, NMR) are too expensive and time consuming

- Predict shape from sequence of bases

- Four basic structures: helices, loops, bulges and junctions

- Knowing the shape of a biomolecule is invaluable in drug design and understanding disease mechanisms

RNA secondary structure

- What makes RNA fold?

- Problem: given an RNA sequence, find the set of base pairs that is "correct" or "optimal"

  Maximize number of base pairs (Nussinov et al)

  Minimize energy (Zucker et al)

- Search problem: very high number of possible structures

- Algorithm: dynamic programming

  Cannot handle pseudoknots

Structure Representation

- Secondary structure described as a graph
- base pairs are described via pairs of indices (i, j), indicating links between base vertices

S={(1,13), (2,12), (3,11), (4,10)}
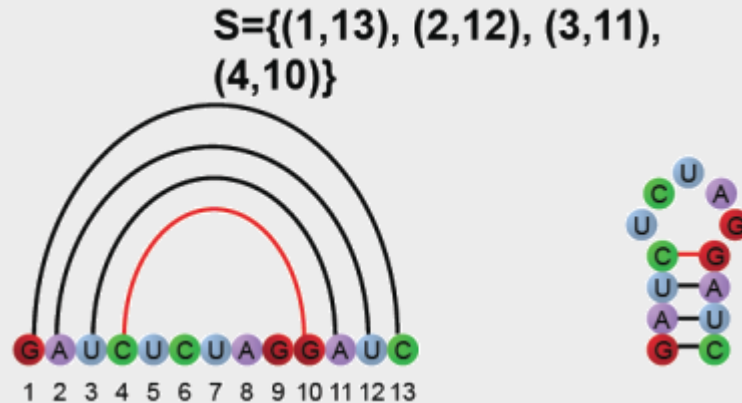
Definitions

- Sequence 5' (r_1 r_2 r_3 … r_n) 3' in {A, C, G, U/T}
- A Secondary Structure is a set of pairs i,j s.t.:
  1. $i < j - 4$
  2. If   i, j are two pairs with $i \leq i'$, then
     a) $i = i'$ & $j = j'$, or
     b) $j < i'$, or
     c) $i < i' < j' < j$

First pair precedes 2nd, or is nested within it. No "pseudoknots."
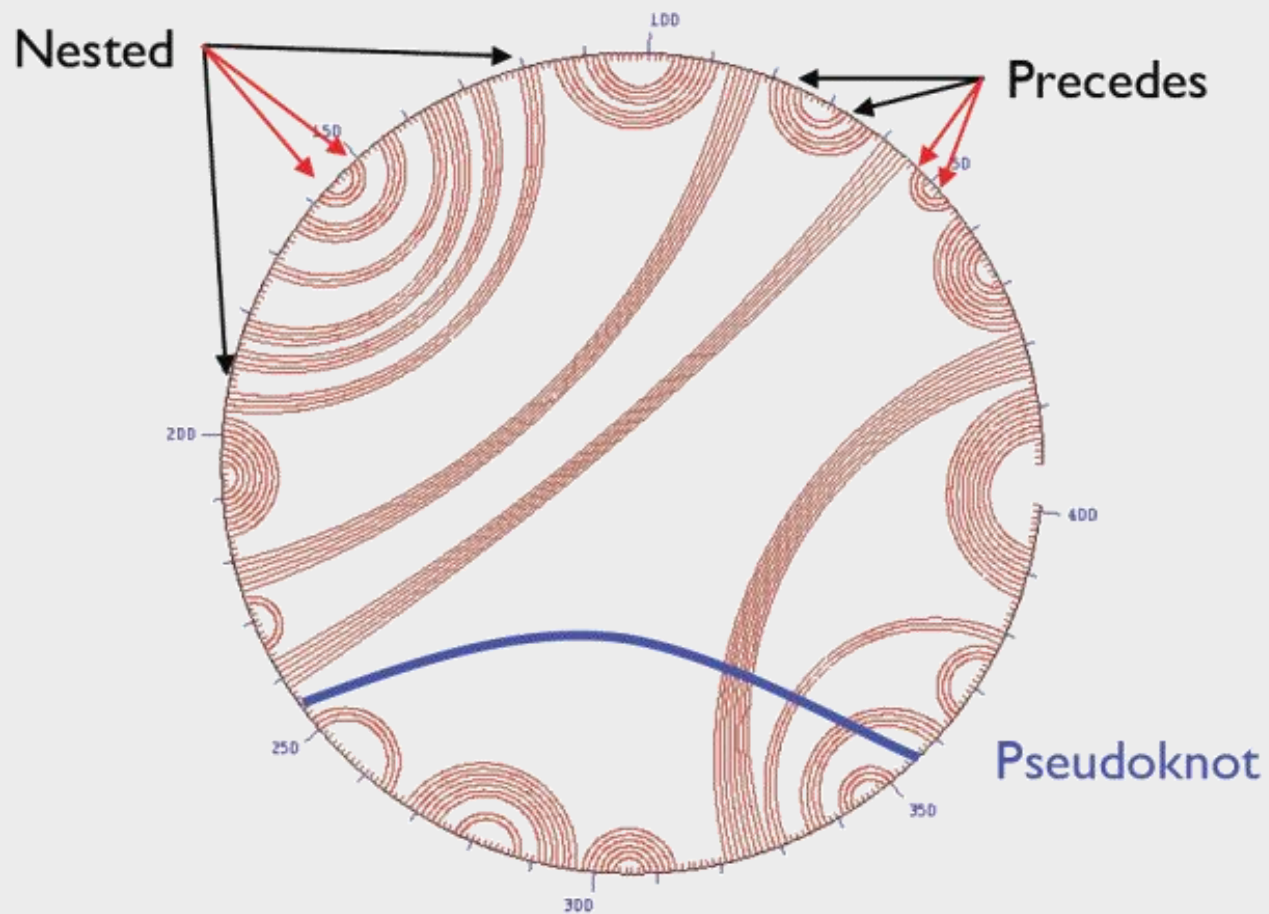
Approaches to Structure Prediction

- Maximum Pairing
  + works on single sequences
  + simple
  - too inaccurate

- Minimum Energy
  + works on single sequences
  - ignores pseudoknots
  - only finds "optimal" fold

- Partition Function
  + finds all folds
  - ignores pseudoknots

- Comparative sequence analysis
  + handles all pairings (incl. pseudoknots)
  - requires several (many?) aligned, appropriately diverged sequences

- Stochastic Context-free Grammars
  Roughly combines min energy & comparative, but no pseudoknots
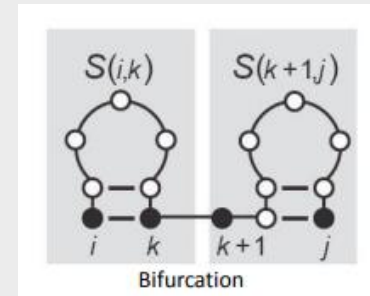
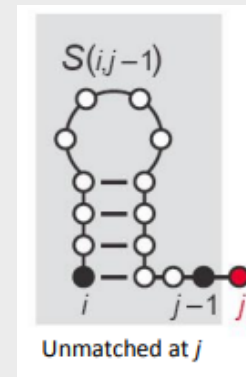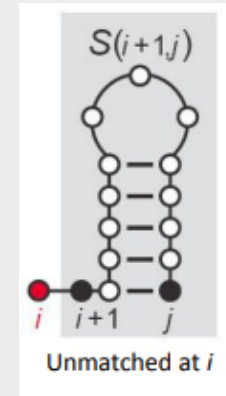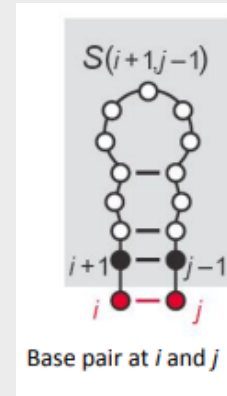- Physical experiments (x-ray crystalography, NMR)
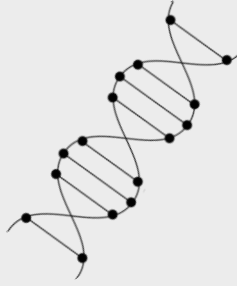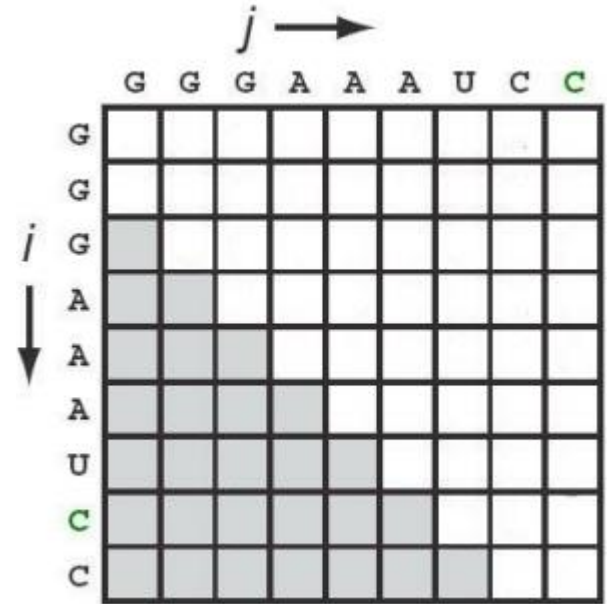
02

DP Method
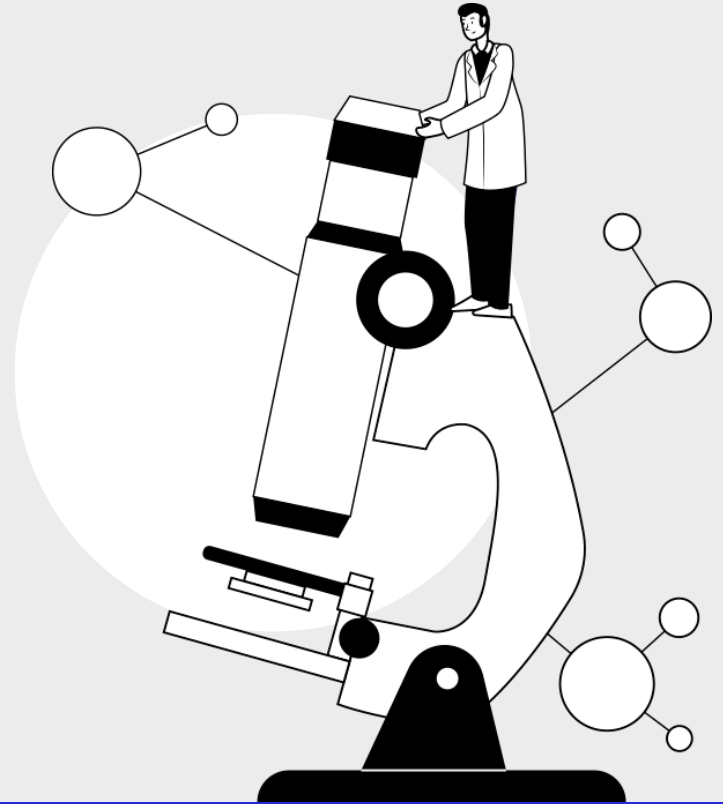
Base Pair Maximization: DP

- S(i, j) is the folding of the RNA subsequence of the strand from index i to index j which results in the highest number of base pairs.

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

S(i+1,j-1)

Base pair at *i* and *j*

S(i+1,j)

Unmatched at *i*

S(i,j−1)

Unmatched at *j*

S(i,k)   S(k+1,j)

Bifurcation

Base Pair Maximization: DP

- Alignment Method:
  - Align RNA strand to itself
  - Score increases for feasible base pairs
- Each score independent of overall structure
- Bifurcation adds extra dimension

Base Pair Maximization: DP

- Alignment Method:
    - Align RNA strand to itself
    - Score increases for feasible base pairs
- Each score independent of overall structure
- Bifurcation adds extra dimension

Base Pair Maximization: DP

- Alignment Method:
    - Align RNA strand to itself
    - Score increases for feasible base pairs
- Each score independent of overall structure
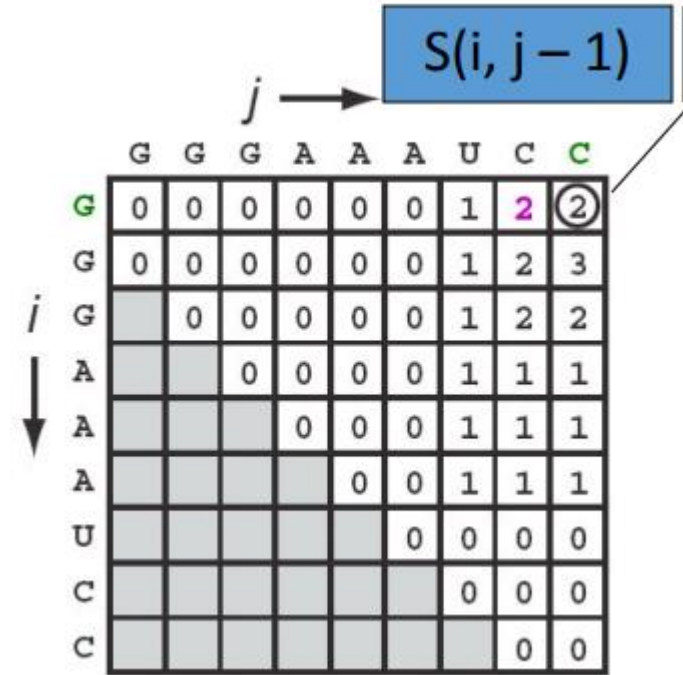- Bifurcation adds extra dimension

Base Pair Maximization: DP

- Alignment Method:
    - Align RNA strand to itself
    - Score increases for feasible base pairs
- Each score independent of overall structure
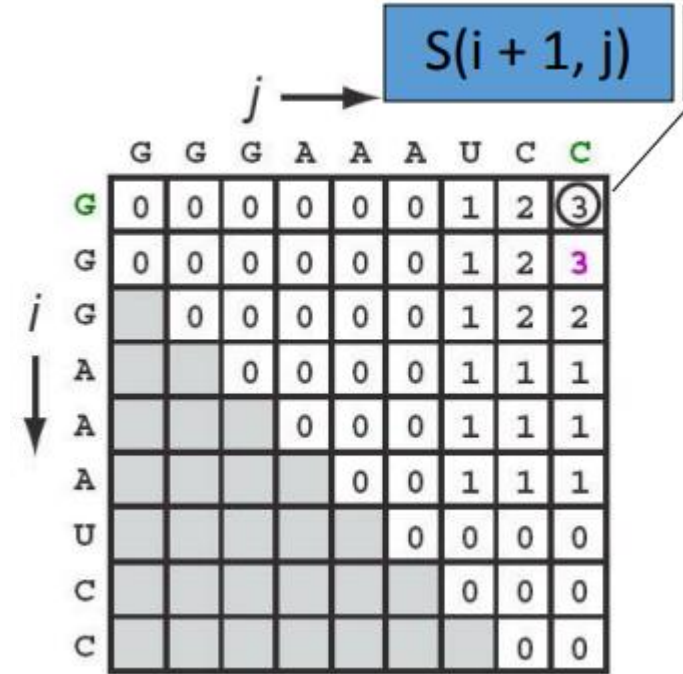- Bifurcation adds extra dimension

Base Pair Maximization: DP

- Alignment Method:
  - Align RNA strand to itself
  - Score increases for feasible base pairs
- Each score independent of overall structure
- Bifurcation adds extra dimension

Base Pair Maximization: DP

- Alignment Method:
    - Align RNA strand to itself
    - Score increases for feasible base pairs
- Each score independent of overall structure
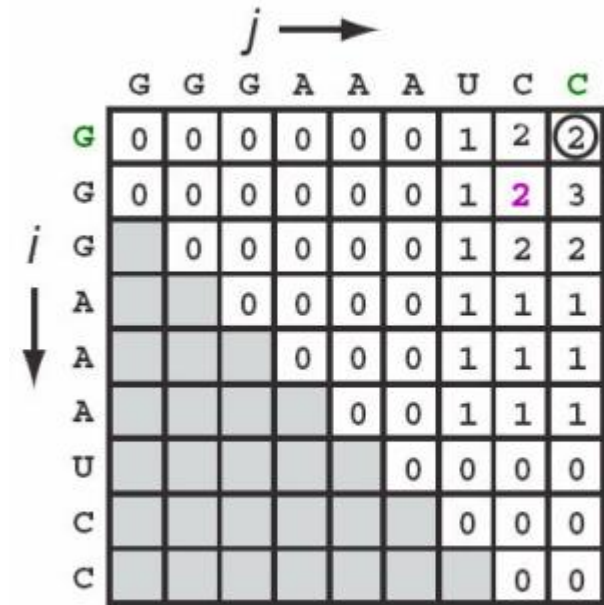- Bifurcation adds extra dimension

$$S(i + 1, j - 1) + 1$$
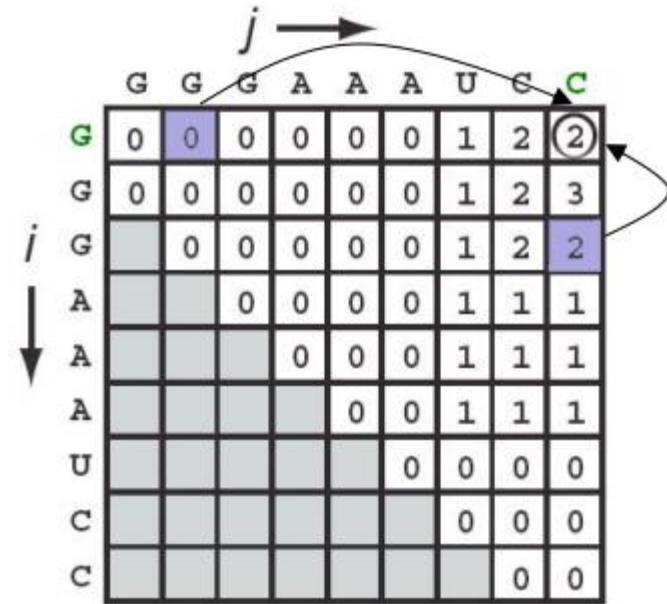
Base Pair Maximization: DP

- Alignment Method:
  - Align RNA strand to itself
  - Score increases for feasible base pairs
- Each score independent of overall structure
- Bifurcation adds extra dimension
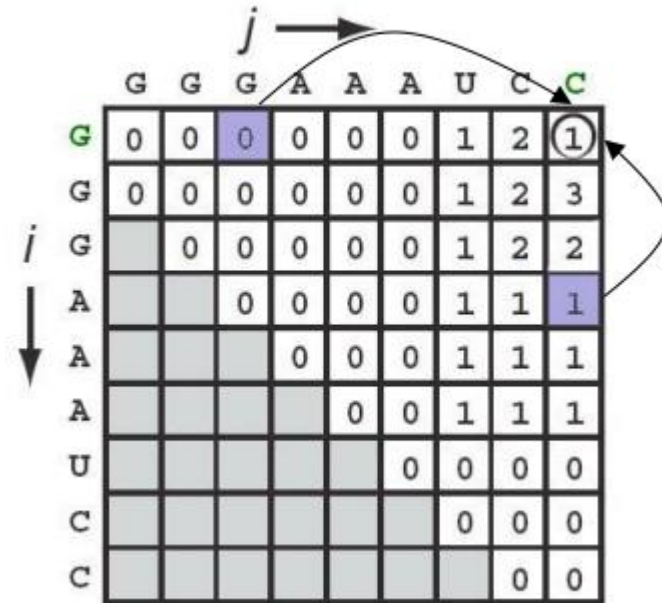
Base Pair Maximization: DP

- Alignment Method:
  - Align RNA strand to itself
  - Score increases for feasible base pairs
- Each score independent of overall structure
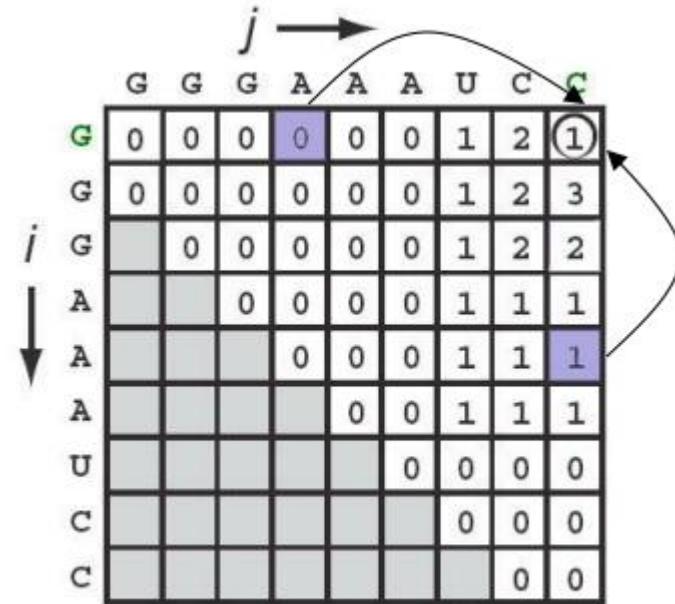- Bifurcation adds extra dimension

Base Pair Maximization: DP

- Alignment Method:
  - Align RNA strand to itself
  - Score increases for feasible base pairs
- Each score independent of overall structure
- Bifurcation adds extra dimension

Base Pair Maximization: Drawbacks

- Base pair maximization will not necessarily lead to the most stable structure.

- It may create structure with many interior loops or hairpins which are energetically unfavorable.

- Results comparable to aligning sequences with scattered matches not biologically reasonable.