

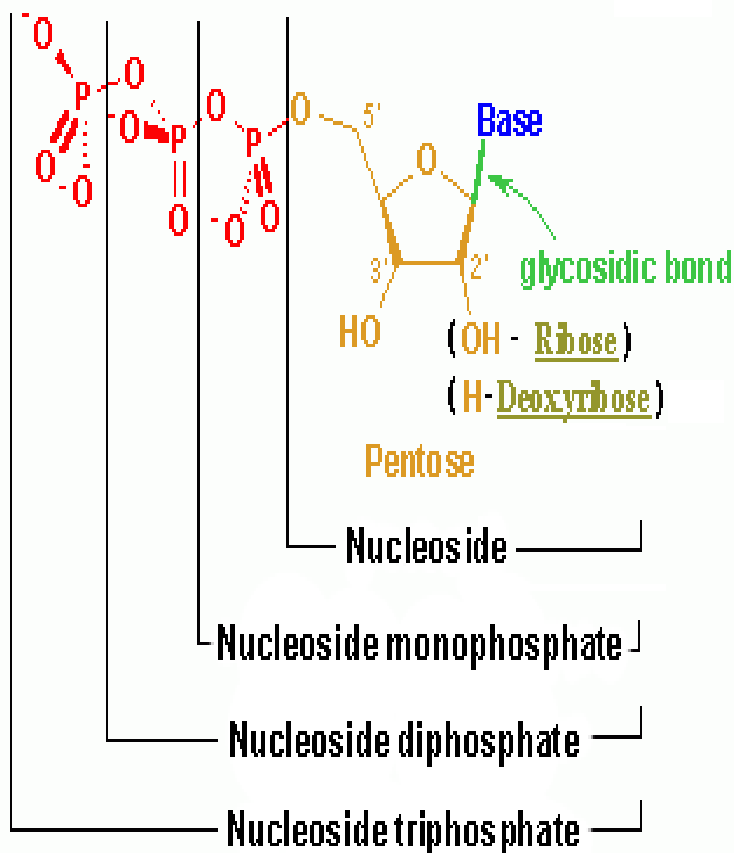
# RNA Secondary Structure Prediction

## What is RNA and where is it used

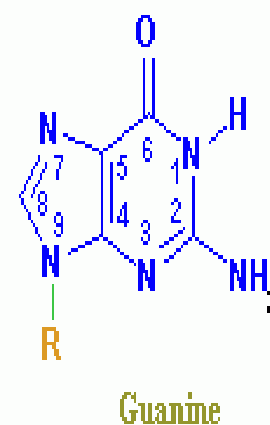
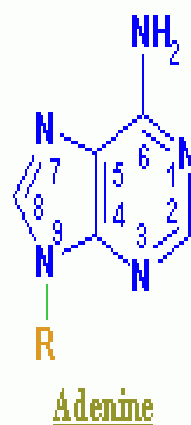
- Proteins and nucleic acids, like RNA(ribonucleic acid) and DNA(deoxyribonucleic acid), play an important role in reproducing and maintaining life:
- Proteins are important because they control processes like energy metabolism, intercellular communication and biosyntheses. They are synthesized using the genetic information which is stored in the DNA

- RNA molecules are used for the synthesis of proteins, they act as messengers.
- Both, DNA and RNA are composed of subunits, the so-called nucleotides or bases.
- There are only four different types of nucleotides in a molecule, but DNA and RNA do have other sets of nucleotides.
- Nucleotides consist of a nitrogen-containing base, a five-carbon sugar ring and a phosphate group .

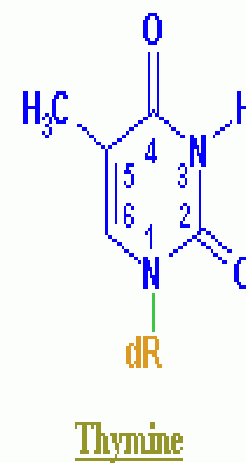
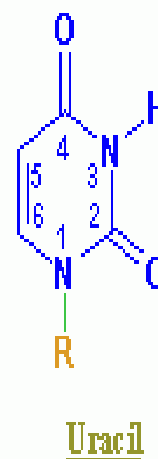
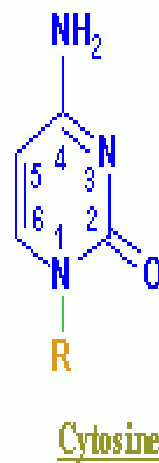
- The nucleotides are linked together by phosphodiester linkages through the hydroxyl group on the sugar on one nucleotide and the phosphate on the next one.
- As a result one can observe a strand with the so-called 5'-end, where  
a
- free phosphate group can be found, and the 3'-end with a free hydroxyl group.
- The nitrogenous base has the structure of a planar ring and is either a purine or a pyrimidine.



## Purines



## Pyrimidines



# The structure of RNA

- In RNA nucleotides the sugar which is used is ribose and therefore they are also called ribonucleotides.
  - The purine bases are also adenosine (A) and guanine (G), but the pyrimidine bases are cytosine (C) and uracil (U) .
  - RNA molecules are much smaller than DNA molecules, but they are also linear polymers
  - .Moreover they do not seem to have a regular three-dimensional structure and are mostly single stranded.
  - This makes them more flexible than DNA and they can also act also as enzymes
- cont...*

- The molecules often contain a very stable three dimensional structure with unpaired regions, which are very flexible.
- The wobble base-pairs make up an important factor for this flexibility.
- Beside the Watson-Crick base-pairs, A:U and G:C, is the wobble base-pair, G:U, one of the most common base-pairs in RNA molecules .
- But actually any of the bases can build a hydrogen bond with any other base. *cont...*

- Another difference between DNA and RNA is that double-stranded RNA builds up Alpha-helices while double-stranded DNA builds up Beta-helices .
- The difference becomes apparent observing the size of the major grooves of the helices:

---
- The major groove of the A-helix is rather narrow and deeper.
- This is due to ribose needing more space than deoxyribose .



# RNA Basics

- RNA bases A,C,G,U

- Canonical Base Pairs

- A-U



- G-C

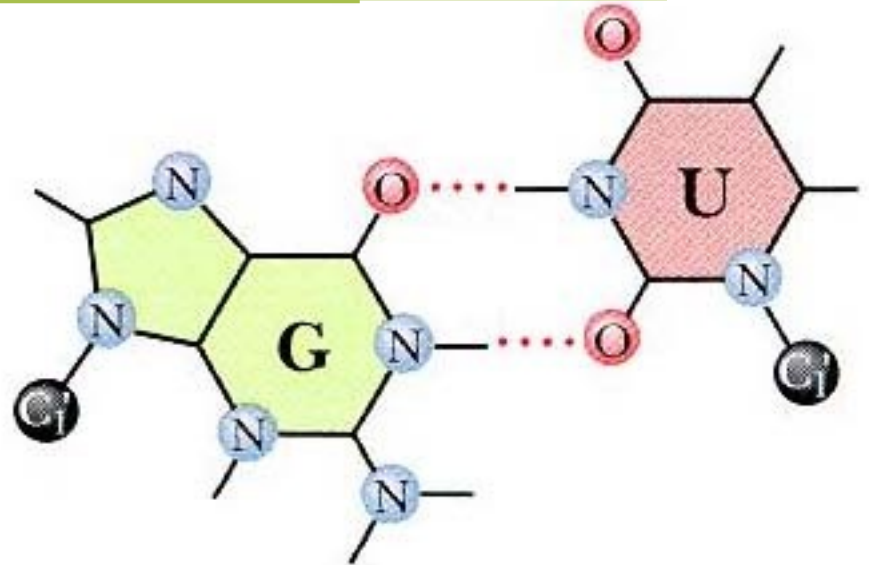


- G-U



“wobble” pairing

3 Hydrogen Bonds – more stable



DNA

RNA

Protein

## Primary, secondary and tertiary structures

- The primary structure of a molecule describes only the one-dimensional sequence of its components.
- The primary structure of RNA is almost identical to the primary structure of DNA, besides the components being A, C, G and U instead of T.
- The secondary structure of molecules is more complex than the primary structure and can be drawn in two-dimensional space .
- By definition, the local conformation of polymers is called secondary structure .
- For RNA and DNA the secondary structure is their base-paired structure. *cont...*

- RNA secondary structure is mainly composed of double stranded RNA regions formed by folding the single stranded RNA molecule back on itself.
- The tertiary structure is the overall three-dimensional structure of molecules .
- It is built on the interactions of the lower-order secondary structures .
- Helices are examples of RNA and DNA tertiary structures.
- pseudoknot is a tertiary structure of RNA.

## Importance of RNA secondary structure prediction

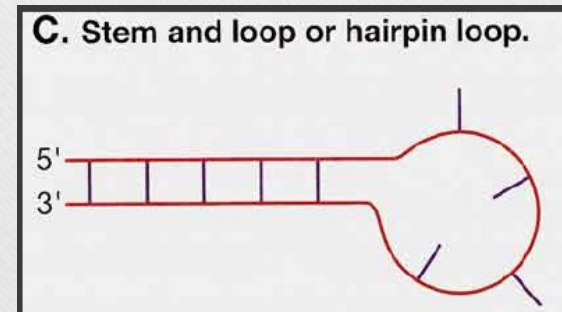
- Important aspect of the prediction of RNA secondary structure is that there are many sequences whose structures have not yet been experimentally determined and
- for which there are no homologues in the databases from which the structure could be derived. Hence it is a good idea to predict the structure.
- Moreover it has been shown that RNA secondary structure prediction has applications to the design of nucleic acid probes . *cont...*

- It is also used by molecular biologists to help predict conserved structural elements in non-coding regions of gene transcripts.
- Finally there is also an application in predicting structures that are conserved during evolution .

# Different secondary structure elements

## 1. Stem Loops (Hairpin Loops)

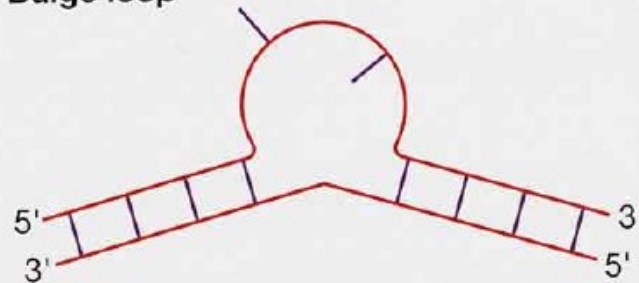
- ❖ is a lollipop shaped structure formed when a single stranded nucleic acid molecule loops back on itself to form a complementary double helix (stem) topped by a loop.
- ❖ stem loops are atleast 4 bases long.



## 2. Bulge Loops

- Bulge loops are commonly found in helical segments of cellular RNAs.
- Bulge loops occur when bases on one side of the structure cannot form base pairs and they cause bends in the helix.

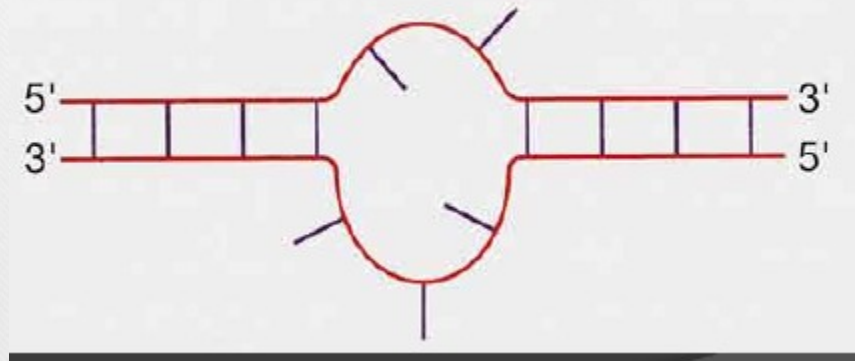
D. Bulge loop



## 3. Interior Loops

- Interior loops occur when bases on both sides of the structure cannot form base pairs.

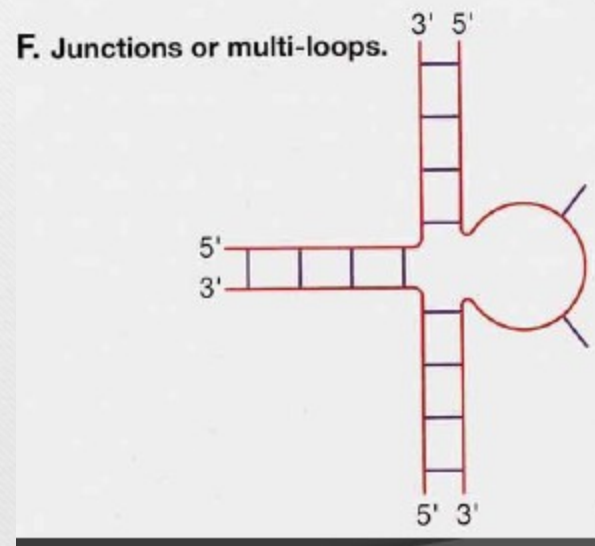
**E. Interior loop**





## 4. Junctions or Multiloops

- Junctions include two or more double stranded regions converging to form a closed structure

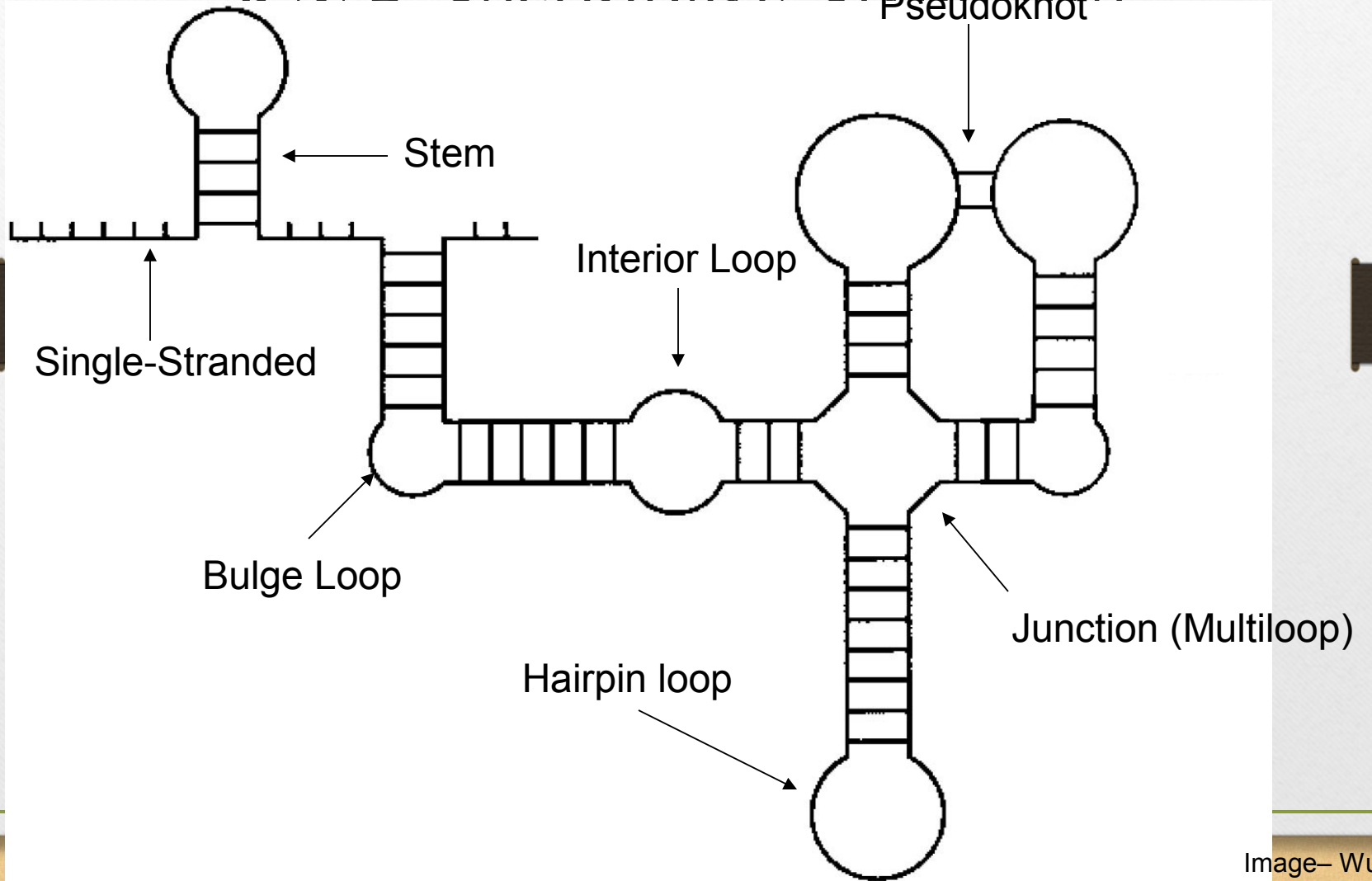


# Pseudoknots

- A pseudoknot is a tertiary structural element of RNA.
- It is formed by base-pairing between an already existing secondary structure loop and a free ending.
- Nucleotides within a hairpin loop form base pairs with nucleotides outside the stem .
- Hence basepairs occur that overlap each other in their sequence position .



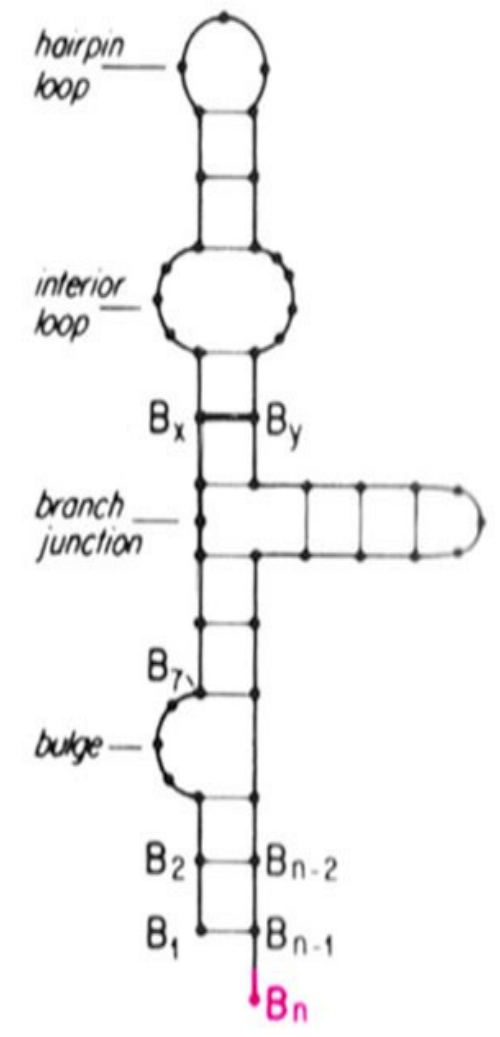
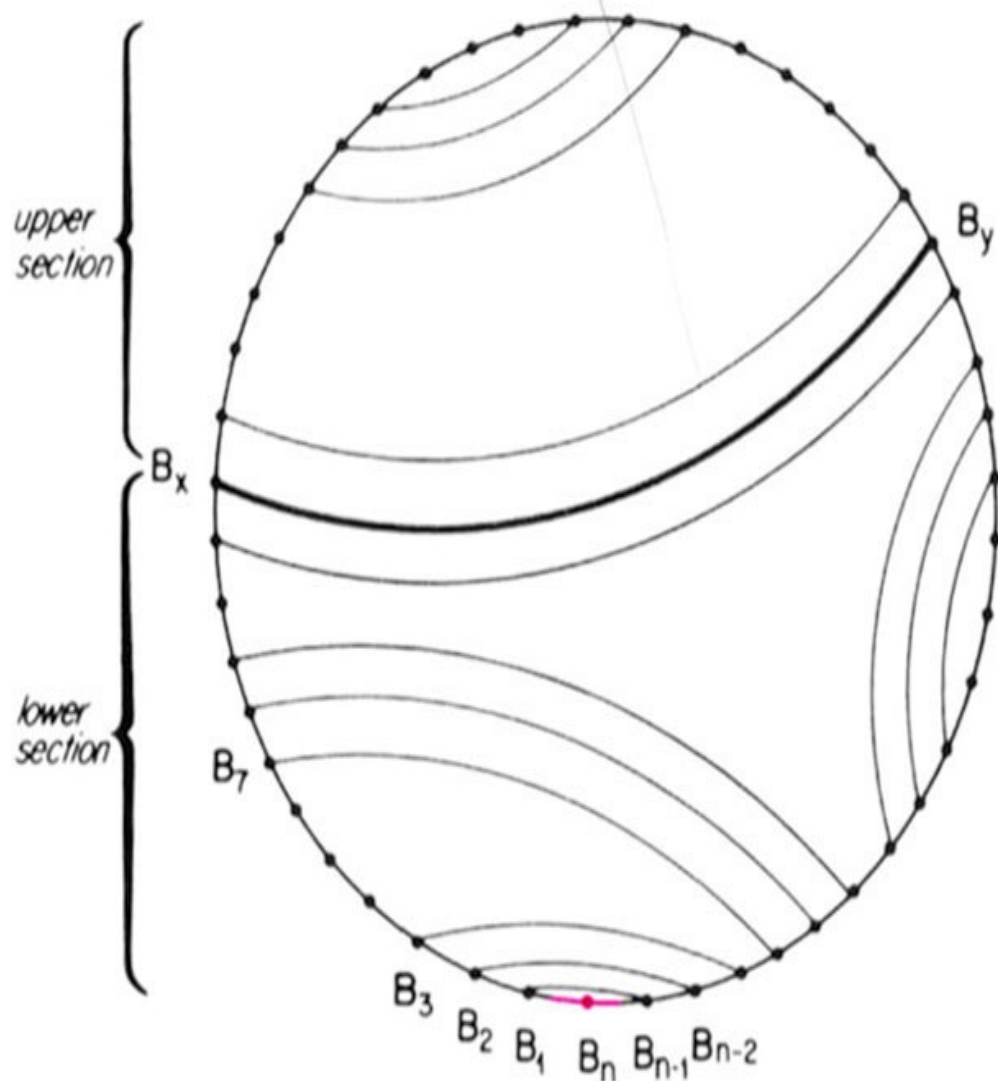
# DNA Secondary Structure



# ASSUMPTIONS IN RNA STRUCTURE PREDICTION

- The most likely structure is similar to the energetically most stable structure.
- The energy associated with any position in the structure is only influenced by local sequence and structure.
- The structure formed does not produce pseudoknots.

- One method of representing the base pairs of a secondary structure is to draw the structure in a circle.
  - An arc is drawn to represent each base pairing found in the structure.
- 
- If any of the arc cross, then a pseudoknot is present.



# RNA structure prediction methods

---

- Base Pair Maximization
- Energy Minimization

# Base Pairs Maximization

- This approach is to find the configuration with the greatest numbers of paired bases.
- Given a RNA sequence, determine the set of maximal base pairs(no base pair across each other)
- Align bases according to their ability to pair with each other gives an approach to determining the optimal structure



# Methods adopted

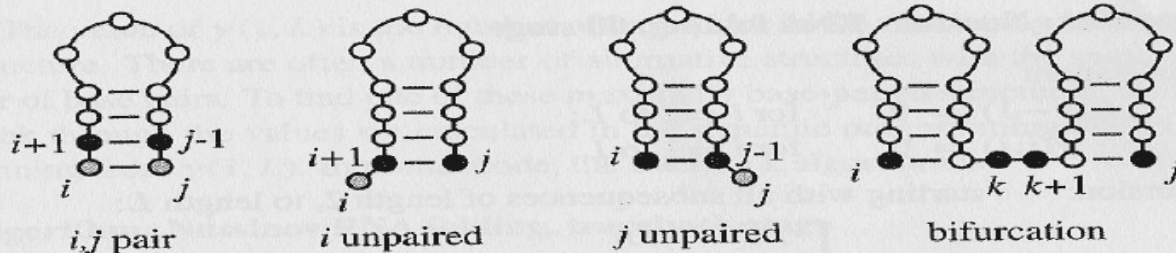
---

- Dynamic programming approach
- Nussinov Algorithm

# Nussinov Algorithm

Four ways to get the optimal structure between position  $i$  and  $j$  from the optimal substructure

1. Add  $i, j$  pair onto best structure found for subsequence  $i+1, j-1$
2. Add unpaired position  $i$  onto best structure for subsequence  $i+1, j$
3. Add unpaired position  $j$  onto best structure for subsequence  $i, j-1$
4. Combine two optimal structures  $i, k$  and  $k+1, j$



# Nussinov Algorithm

- compares a sequence against itself in a  $n \times n$  matrix
- Find the maximum of the scores for the four possible structures at a particular position.

# Base Pair Maximization - Drawbacks

---

- Base pair maximization will not necessarily lead to the most stable structure
  - May create structure with many interior loops or hairpins which are energetically unfavorable
- Comparable to aligning sequences with scattered matches – not biologically reasonable

# Energy Minimization

- Thermodynamic Stability

- Estimated using experimental techniques
- Theory : Most Stable is the Most likely

- No Pseudoknots due to algorithm limitations

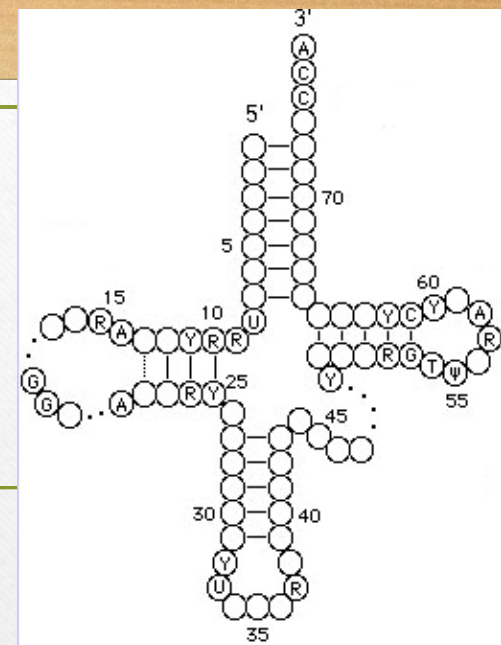
- Uses Dynamic Programming alignment technique

- Attempts to maximize the score taking into account thermodynamics

- Gaps represent some form of a loop

- The most widely used software that incorporates this minimum free energy algorithm is

MFOLD/RNAfold/ViennaRNA



# Energy Minimization Drawbacks

---

- Compute only one optimal structure
- Usual drawbacks of purely mathematical approaches
  - Similar difficulties in other algorithms
    - Protein structure
    - Exon finding