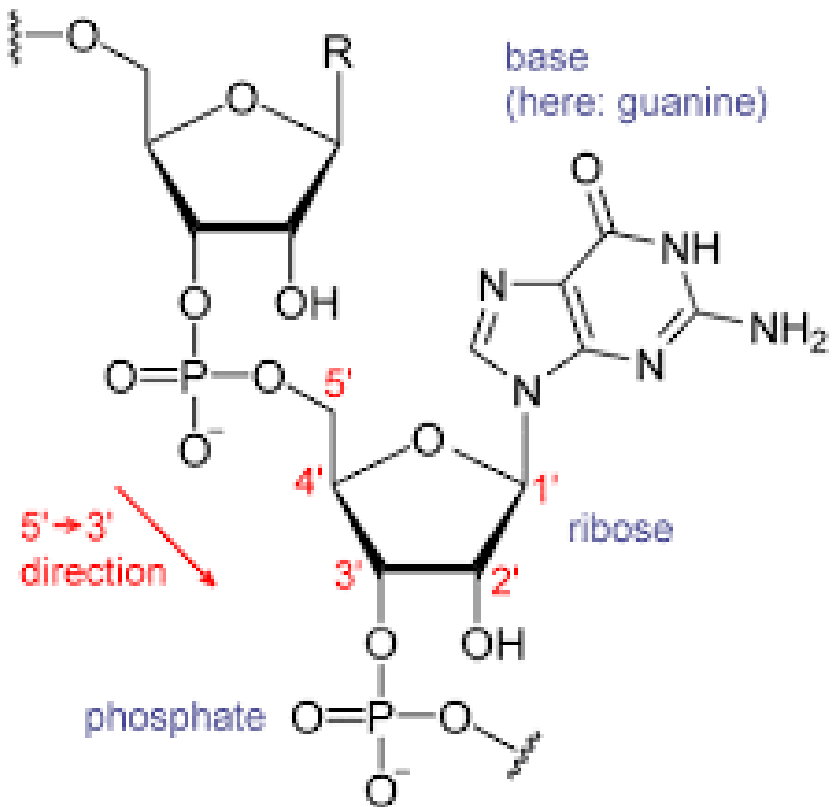


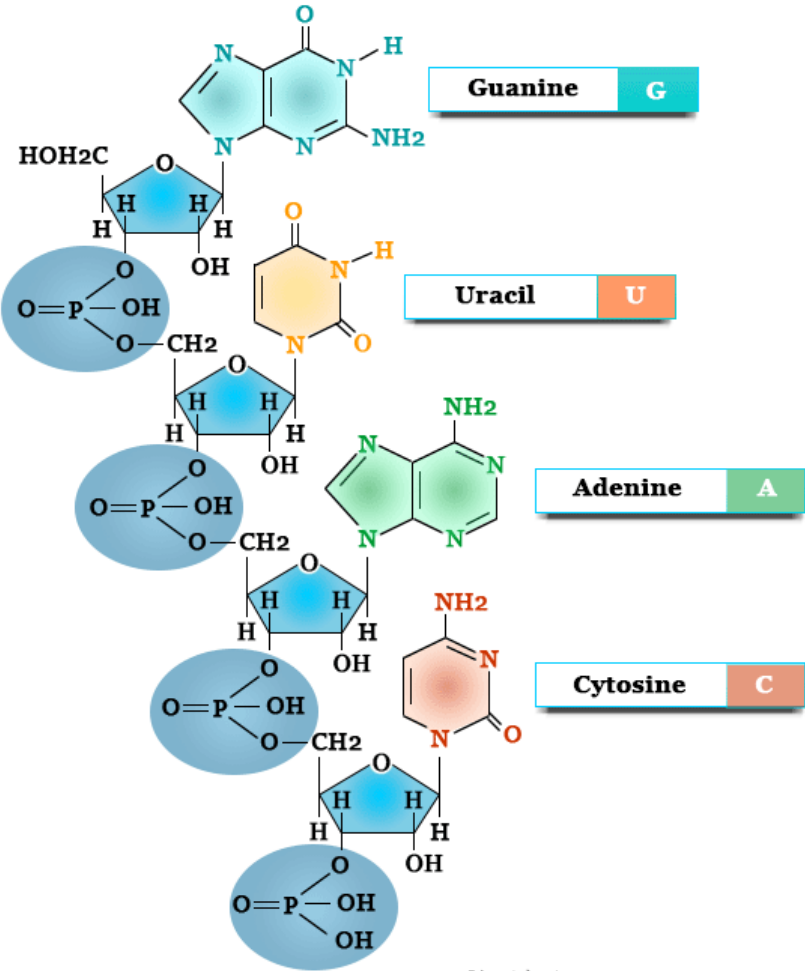
# RNA Structure Prediction Tools

Shilpa Deshpande Kaistha  
Department of Biotechnology  
School of Life Sciences & Biotechnology  
CSJM Univeristy, Kanpur

# RNA structure



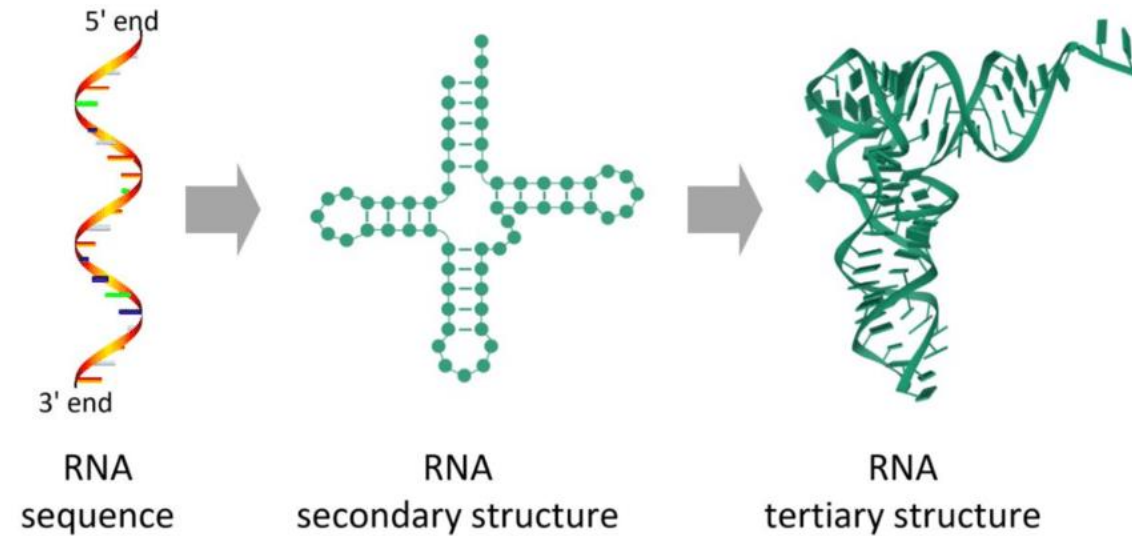
## Ribonucleic Acid (RNA)



Severo Ochoa won the 1959 Nobel Prize in Medicine after he discovered how RNA is synthesized

# RNA Structures

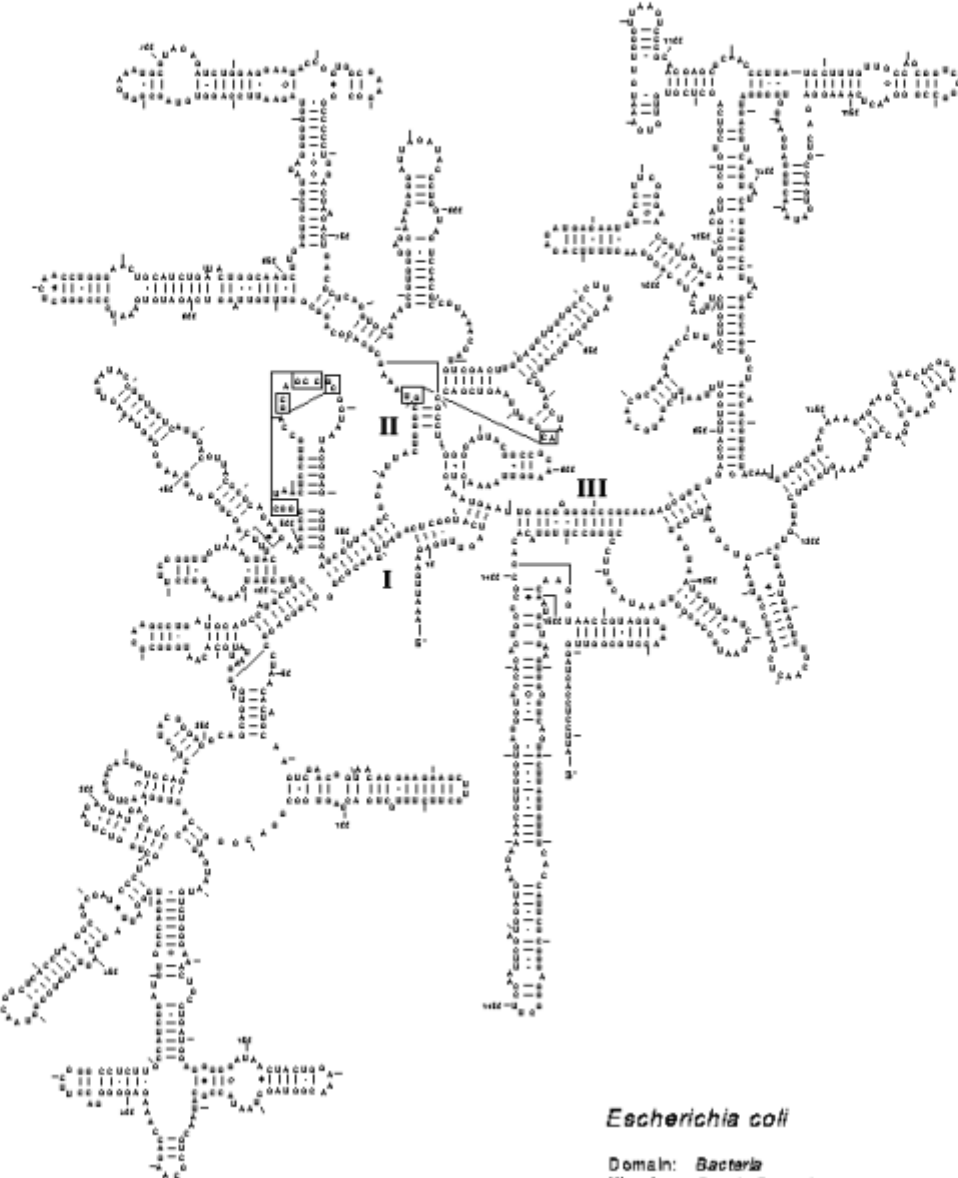
- Primary Structure: One dimensional sequence of complimentary to DNA
- Secondary Structure: Local conformation of polymer in 2D space due to base pairing
  - Double stranded regions formed by single strand RNA folding back on itself
- Tertiary Structure: 3 D structure of molecule formed by interactions of lower order secondary structures- Pseudoknots, helices are tertiary structures of RNA
- RNA secondary structure can be determined from atomic coordinates obtained from X-ray crystallography, nuclear magnetic resonance (NMR), or cryogenic electron microscopy- low output
- Computationally predicting the secondary structure of RNA is an active research area in computational biology and bioinformatics.



# Applications for RNA structure prediction tools

- Design nucleic acid probes
- Predict conserved residues in non coding regions of genome
- Predicting structures conserved during evolution for phylogenetic studies
- Useful for regions where no homology found in database

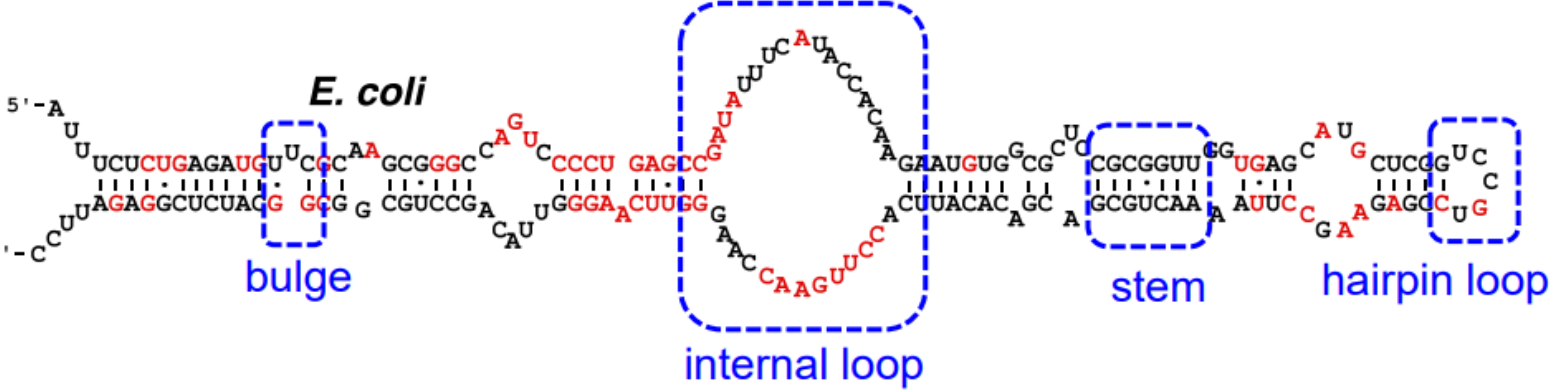
# Small Subunit Ribosomal RNA



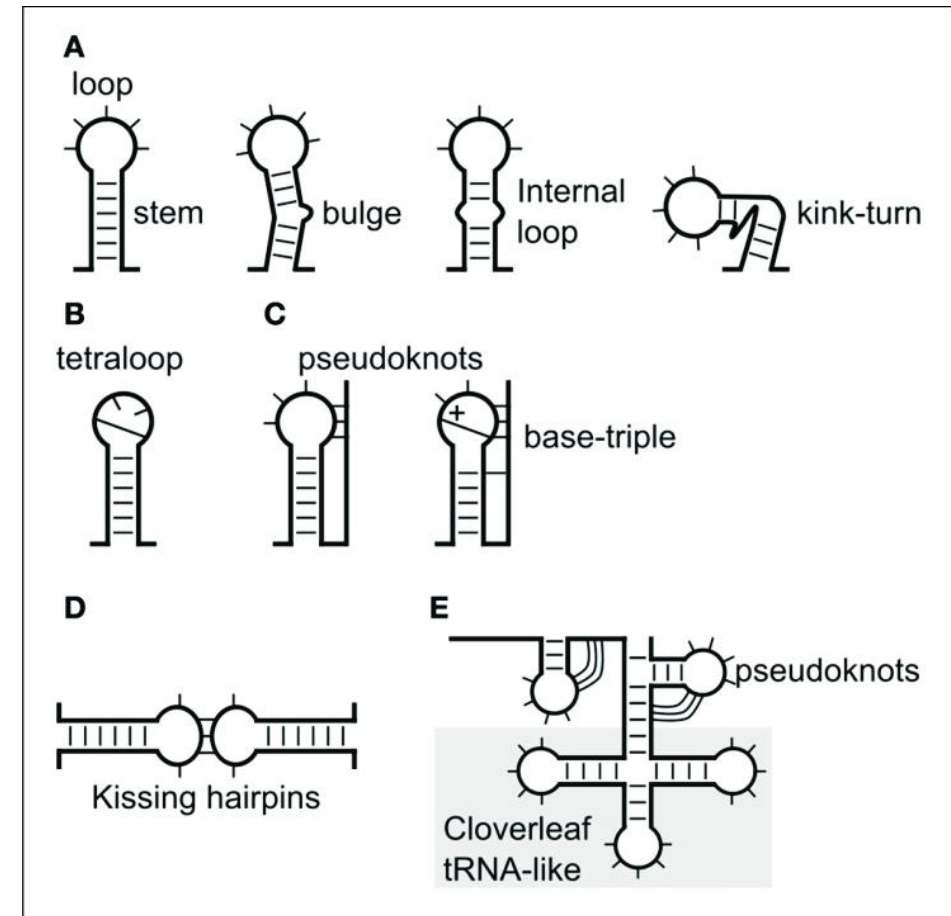
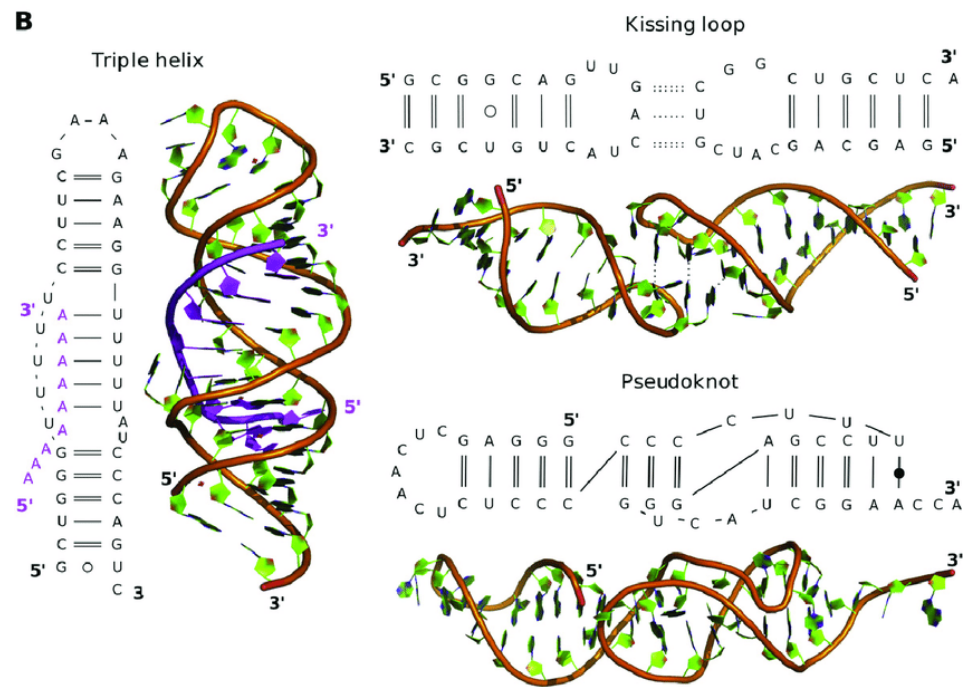
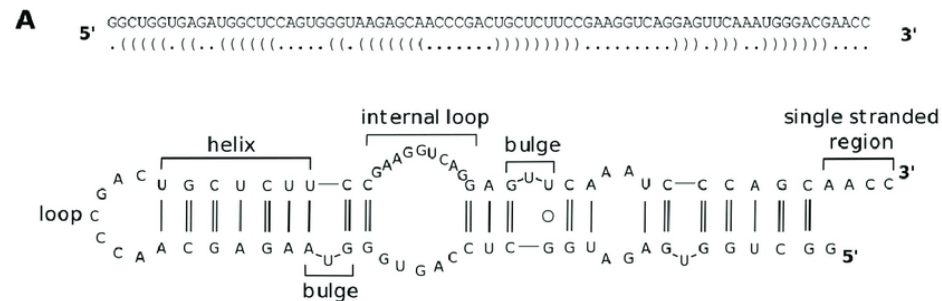
*Escherichia coli*

Domain: *Bacteria*  
Kingdom: *Proteobacteria*  
Order: *gamma*

# Secondary Structure Features



# Secondary RNA structures



# RNA structure prediction methods

- **Base Pair Maximization/Free energy minimization** based RNA structure prediction methods are usually applied to a single RNA sequence. The most energetically stable RNA secondary structure(s) that are composed of canonical G:C, A:U, and G:U base-pairs and organized into standard helices are predicted. Non-canonical base-pairs and base-pairs not in standard helices cannot be predicted at this time.
- **Comparitive analysis:**
  - 1) different, homologous RNA sequences are capable of folding into the same secondary and tertiary structure, and
  - 2) during the course of evolution, the secondary and tertiary structure of an RNA molecule remains mostly unchanged, while the primary structure can change significantly.



# Base Pair Maximization/ Free energy minimization model

- Approach find configuration with greatest number of paired bases
- Given RNA sequence, determine maximal base pairs
- Align bases according to ability to pair with each other for finding optimal structure
- Dynamic Programming
- Nussinov Algorithm: Ruth Nussinov et al. (1978) **computes for a given RNA sequence the maximal number of base pairs of any nested structure using dynamic programming**
- **Zuker Algorithm**
- Compares a sequence against itself in a  $n \times n$  matrix. Finds the maximum of scores for 4 possible structures at a possible position

<http://rna.informatik.uni-freiburg.de/Teaching/index.jsp?toolName=Nussinov>



# Free Energy Minimization/Single sequence prediction method

- search for thermodynamically stable states through free energy minimization.
- If the secondary structure contains only nested base pairing, the energy minimization can be efficiently solved through dynamic programming, such as those implemented in Vienna RNAfold, MFold (does not predict pseudoknots), RNAstructure and CONTRAfold.
- Faster implementations that try to improve the speed of dynamic programming include Rfold , Vienna RNAplfold, LocalFold and LinearFold .

# Mfold (biotools:mfold) ID Verified

<http://mfold.rna.albany.edu/?q=mfold/download-mfold>

[Structure prediction >](#) [Nucleic acid structure analysis >](#) [RNA >](#) [Protein secondary structure >](#) [Functional, regulatory and non-coding RNA >](#)

Web application  

Predict RNA secondary structure from sequence; does not predict pseudoknots - see PKNOTS.

[RNA secondary structure prediction >](#)

[Structure visualisation >](#)

[Protein secondary structure prediction >](#)

[RNA secondary structure alignment >](#)

## Credits & Support

Michael J. Muskus  
Primary contact | [✉ mmuskus at DOM.wustl.edu](mailto:mmuskus@DOM.wustl.edu) | [Link >](#)

## Documentation

<http://unafold.rna.albany.edu/?q=mfold/faq>

General

## Publication details

# Disadvantages of RNA prediction tools

- Does not lead to most stable structures. Corroborate with experimental techniques
- Create many structures with interior loops or hairpins that are energetically unfavourable
- pseudoknots, which include two stem–loop structures with half of one stem intercalating between the two halves of another stem are difficult to predict
- overfitting is a common risk for such highly parameterized models.
- RNA single sequence is free energy minimization using a dynamic programming approach
- Programs based on this technique include Mfold , RNAStructure and RNAFold

# Recent advances in RNA structure prediction tools

- Deep learning techniques subset of machine learning based on neural networks have started to emerge as an alternative approach to functional structure prediction problems including RNA secondary structure prediction problems
- learning-based methods benefit from making few assumptions, allowing pseudoknots, and accounting for tertiary interactions, noncanonical base pairing, or other previously unrecognized base-pairing constraints.
- eg **DeepDRU**, **U FOLD**

# RNA 3 D prediction tools

- The RNAComposer system offers a new user-friendly approach to the fully automated prediction of large RNA 3D structures. The method is based on the machine translation principle and operates on the RNA FRABASE database acting as the dictionary relating RNA secondary structure and tertiary structure elements

# Reference

- Zhao, Qi & Zhao, Zheng & Fan, Xiaoya & Yuan, Zhengwei & Mao, Qian & Yao, Yudong. (2020). Review of Machine-Learning Methods for RNA Secondary Structure Prediction
- Mathews DH, Turner DH, Zuker M. RNA secondary structure prediction. *Curr Protoc Nucleic Acid Chem.* 2007 Mar;Chapter 11:Unit 11.2
- Hajiaghayi, M., Condon, A. & Hoos, H.H. Analysis of energy-based algorithms for RNA secondary structure prediction. *BMC Bioinformatics* **13**, 22 (2012)
- Laiyi Fu, Yingxin Cao, Jie Wu, Qinke Peng, Qing Nie, Xiaohui Xie, UFold: fast and accurate RNA secondary structure prediction with deep learning, *Nucleic Acids Research*, Volume 50, Issue 3, 22 February 2022, Page e14, <https://doi.org/10.1093/nar/gkab1074>