Lecture 8
RNA Secondary Structure

Central Dogma

DNA
↓
RNA
↓
Protein

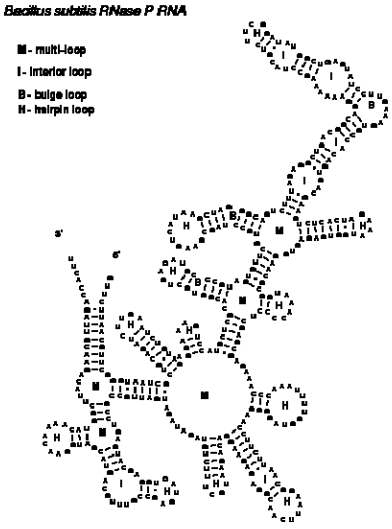
RNA

- In the central dodgma, we talk about mRNA coding for protein. There are also tRNA and rRNA that are also coded for by the DNA. The MicroRNAs (miRNA) are small (22 nucleotides) non-coding RNA gene products that seem to regulate translation
- The RNA has features in it sequence that gives it a structure based on the folding and pairing of bases within the strand.
- tRNA very highly conserved.

Secondary and Tertiary Structure

- Watson-Crick pairing of G/C and A/U
- G/C pairing give more energetic stability
- Double stranded regions and loop regions are the secondary structure elements
- Tertiary structure is the interaction between secondary structure elements

RNA Secondary Structure



http://www.bioinfo.rpi.edu/~zukerm/Bio-5495/RNAfold-html/img26.gif

---

Methods

- Sequence and base pairing patterns
  - there is conservation of sequence in RNA structures
- Energy minimization
  - find the energetically most stable structure
  - based on local sequence
  - no knots

History of Methods

- Energy calculations based on base pairings (Tinoco et al 1971)
- Free energy of regions based on the Boltzman function (Martinez, 1984). All possible structures are sampled using the Monte Carlo method. The selection of regions is based on weights determined by the free energy.

History of Methods

- Nussinov and Jacobson (1980) computed a scoring matrix to keep track of the maximum number of pairs. They then performed a traceback to get the optimal structure
- Zuker and Stiegler (1981) used dynamic programing and energy rules to get the energetically most favorable structure.

History of Methods

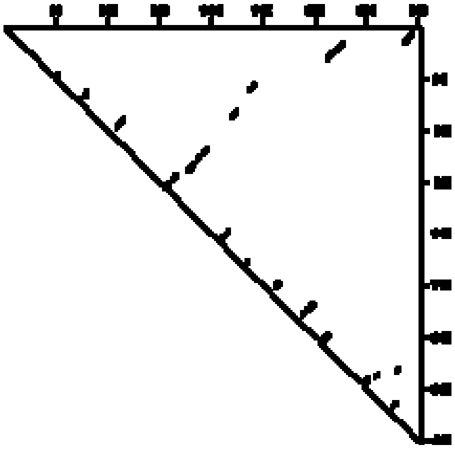
- Mfold is software developed by Zuker and co-workers. It is very computationally expensive and can be used on a maximum of about 1000 nucleotides.
- Genetic algorithm (Shapiro and Navetta, 1994) take into account both sequence and secondary structure.

Circle Plots

• Red lines connect complementary base pairs
• Bases numbered around outside of circle
• If not red lines intersect, then the structure is free of pseudo knots.

http://www.bioinfo.rpi.edu/~zukerm/Bio-5495/RNAfold-html/img28.gif

Dot Plots



- Dots are place where base pairings occur (self-complementarity).
- Diagonal lines indicate regions of complementary bases.

http://www.bioinfo.rpi.edu/~zukerm/Bio-5495/RNAfold-html/img32.gif

Energy Minimization

- Calculate energy based on some scoring matrix.
- Use dynamic programming to calculate the energetic most stable structure.
- Dynamics programming can be altered to also find suboptimal alignments as with mfold.

Evaluation of Methods

Energy minimization

- Advantages
  - Predict thermodynamically most stable structures
  - Reliable account for experimental and alignment data

- Disadvantages
  - Tertiary structures not predicted
  - Computationall expensive

Base Covariation

- Uses aligned sequences of the same class of RNA molecules from different species.
- Conserved regions are likely the base pairing regions.
- Arrange the likely paired regions with complementary regions. This can be done using various methods such as msa, hidden markov models, an ordered tree model, etc.

Hidden Markov Models

A machine learning approach that is trained on observed variation of alignament from msa including matches, substitutions, indels.

The HMM can then classify a new mRNA

Ordered tree model

Secondary structure elements are place at the node if a tree with the edges reflecting the relations between the nodes

Evaluation of Methods

Sequence base covariation methods rely on having a conserved pattern of base pairing

- Advantages
  - Computationally efficient
  - Predict both secondary and tertiary structure

- Disadvantages
  - Need information of sequence covariation – ie need to know structure of similar sequences

Context-free Grammers

- Use of formal language theory to develop a gramar to describe RNA secondary structure.
- Find groups of logically related words separated by intervening words.
- They can accommodate stochasticity i.e. variations in the sequence

RNA Genes

- RNA has various functions
- There are software developed to search for RNA genes in the genome.
- tRNAscan searched for tRNA looking at conserved bases and self-complementarity (to give correct structure).

RNA Structural Modeling

- Using RNA structural modeling a new model of phylogeny of the eukaryotes, prokaryotes, and achaea has been proposed
- RNAi are designed to inhibit specific genes
- Ribozymes are self-replicating RNA – RNA universe.

Exercises

Exercise

1. Find RNA sequence
2. Use software to find stucture

http://www.genebee.msu.su/services/rna2_reduced.html

http://www.bioinfo.rpi.edu/applications/mfold/