Overview

PART I (Secondary Structure)

- Introduction
  - Motivation.
  - Levels of RNA Structure?
  - Secondary Structure Prediction
  - Representation
- Software
  - Prediction (mfold, vienna, dynalign)
  - Rendering (Annotation s2s, others.)
- Everyones Favorite FIKA!

PART II (Rigid Block Model)

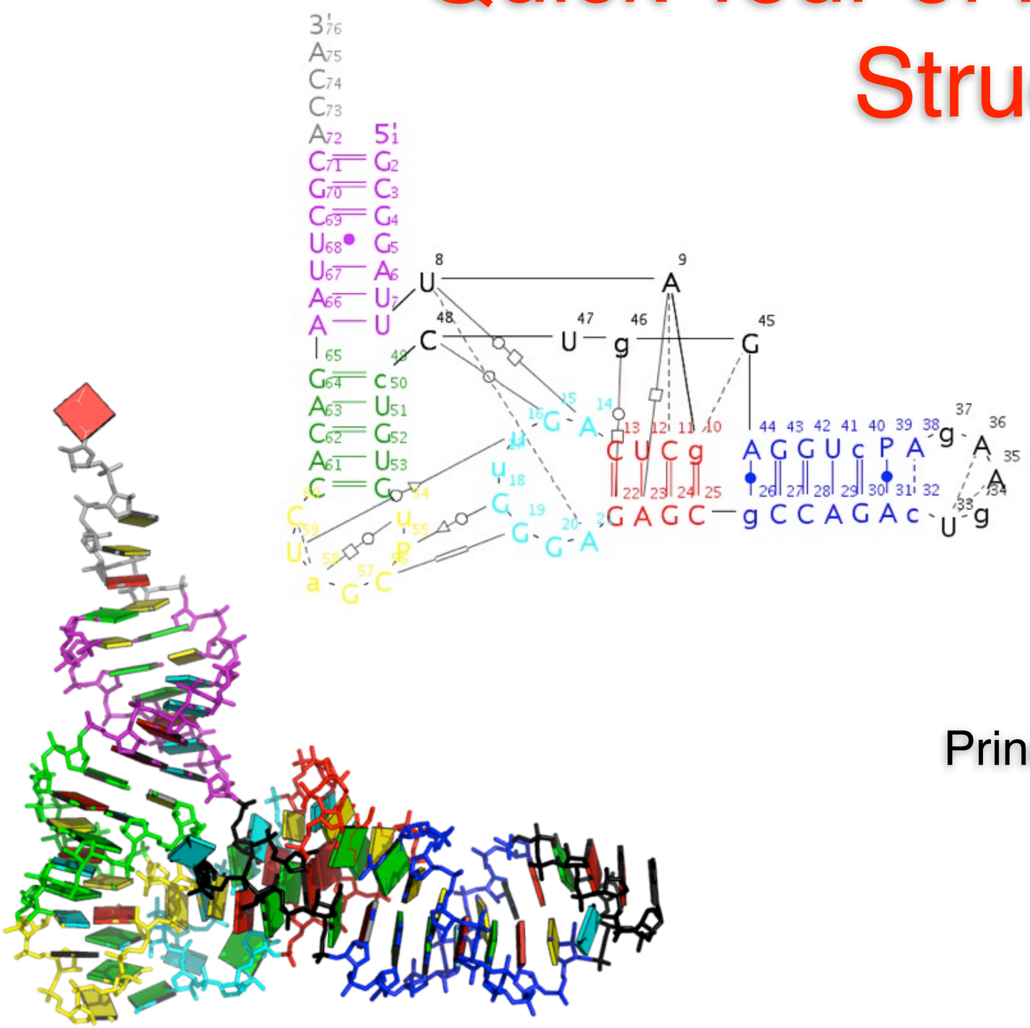
- Rigid Block Model for Nucleic Acids.

PART III (Practical Workshop-Tutorial)

- Practical Examples.
  - Using paper and pencil alone do Nussinov algorithm to some sequence. This gives you some structure, not necessarily minimal energy I guess.
  - Use computer to go to mfold, or vienna webserver to do prediction for same sequence. (maybe mfold is better because it gives more cartoons)
  - Make A-DNA, B-DNA, Z-DNA. Chromosomal DNA.
-

```
GCGGAUU UA gCUC AGuuGGGA GAGC gCCAGAc UgAAg APcUGGA GgUC cUGUG uPCGaUC CACAG AAUUCGC ACCA
1234567 89 0123 45678901 2345 6789012 34567 8901234 5678 90123 4567890 12345 6789012 3456
```

Quick Tour of RNA Secondary Structure Prediction



Mauricio Esguerra Neira

Lennart Nilsson Group

http://mesguerra.net

Principles of Nucleic Acid Structure

KI Doctoral Course 2430

Torsdag, Februari Två, 2012

13:00 - 16:00

GENOMICS

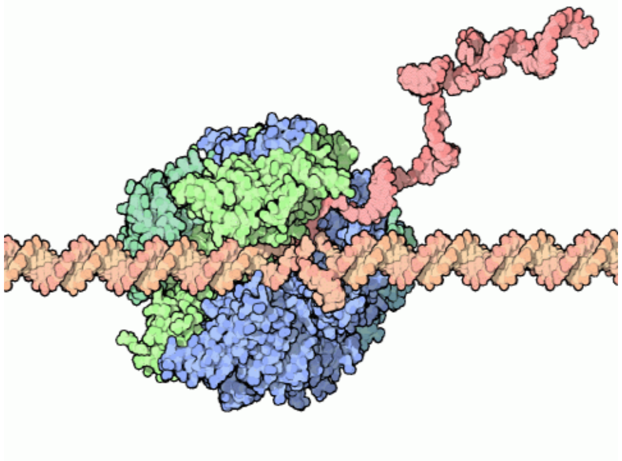
PROTEOMICS

POST-GENOMICS

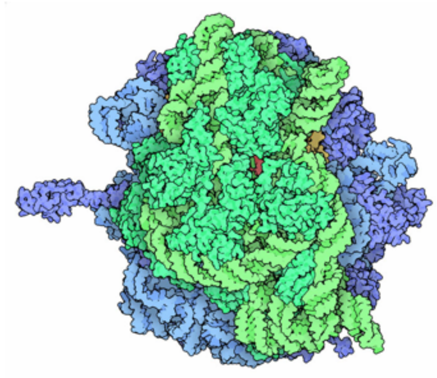
EPIGENETICS

RIBONOMICS

Biology’s Central Dogma



TRANSCRIPTION
RNA Pol II



TRANSLATION
Ribosome, aka rRNA

DNA ⇨ RNA ⇨ mRNA ⇨ Protein

Replication

antigene RNA (agRNA)

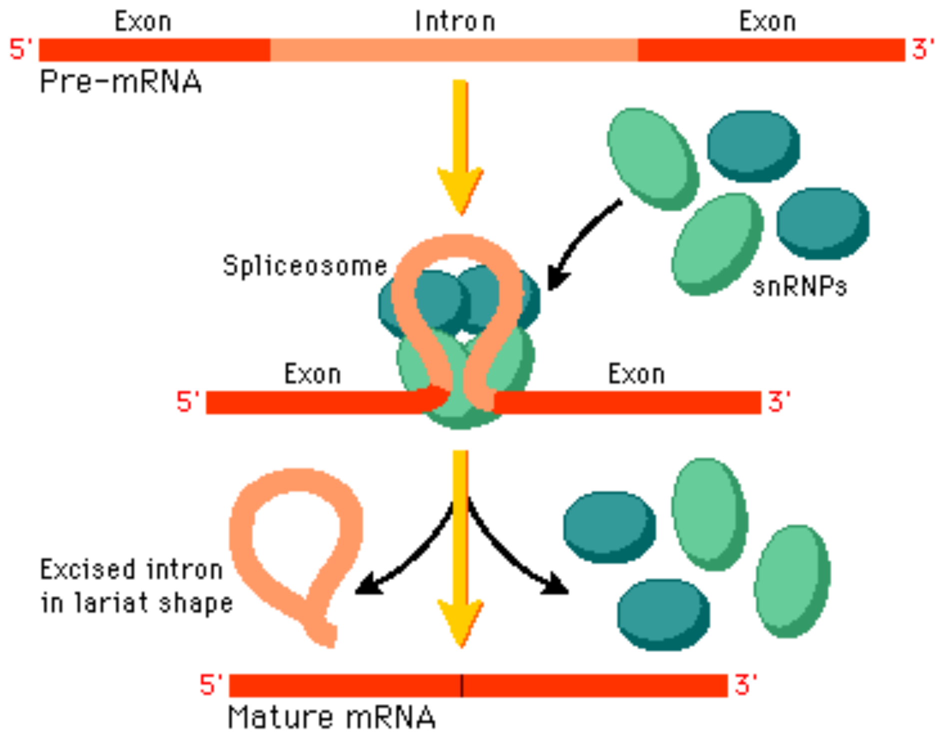
ncRNA (miRNA, siRNA, ...)

Genotype

Gene Expression (Activation, Silencing)

Phenotype

RNA is spliced by the spliceosome after transcription.



Non-coding RNA's

- RNApol I rRNA not 5S (~ 50% cell RNA)
- RNApol II precursors of <u>mRNA</u> and most <u>snRNA</u> and <u>microRNA</u>
- RNApol III 5S rRNA, tRNA and other small RNAs

- piwiRNA
- siRNA
-

The RNA Zoo

Structure Determines Function

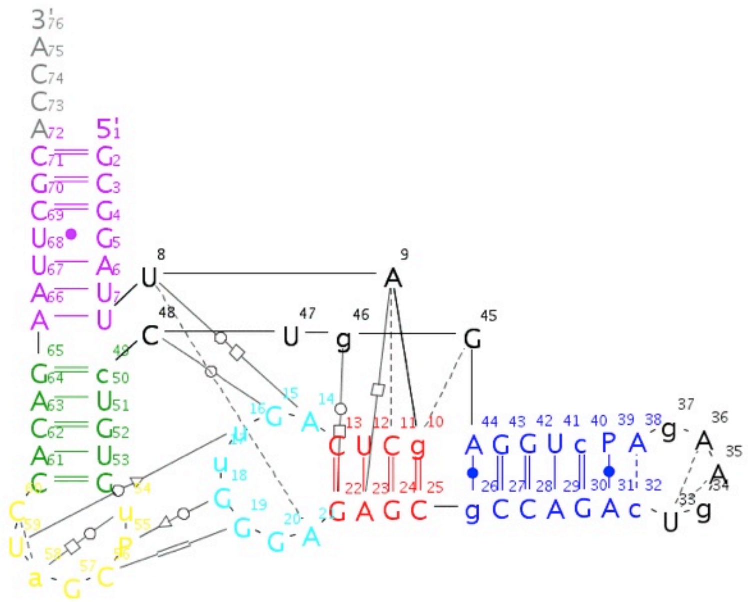
Primary Structure

Backbone covalent linked through phosphodiester group with
5’ to 3’ sense.

```
GCGGAUU UA gCUC AGuuGGGA GAGC gCCAGAc UgAAg APcUGGA GgUC cUGUG uPCGaUC CACAG AAUUCGC ACCA
1234567 89 0123 45678901 2345 6789012 34567 8901234 5678 90123 4567890 12345 6789012 3456
```

Secondary Structure

Hydrogen Bonded Base-Pairs



Two Kinda ThreeTypes of RNA Secondary Structure Prediction

- RNA Sequence Covariation (Gutell)
- RNA Free Energy Minimization (Tinoco-Uhlenbeck)
- RNA Base-Pair Maximization (Nussinov)

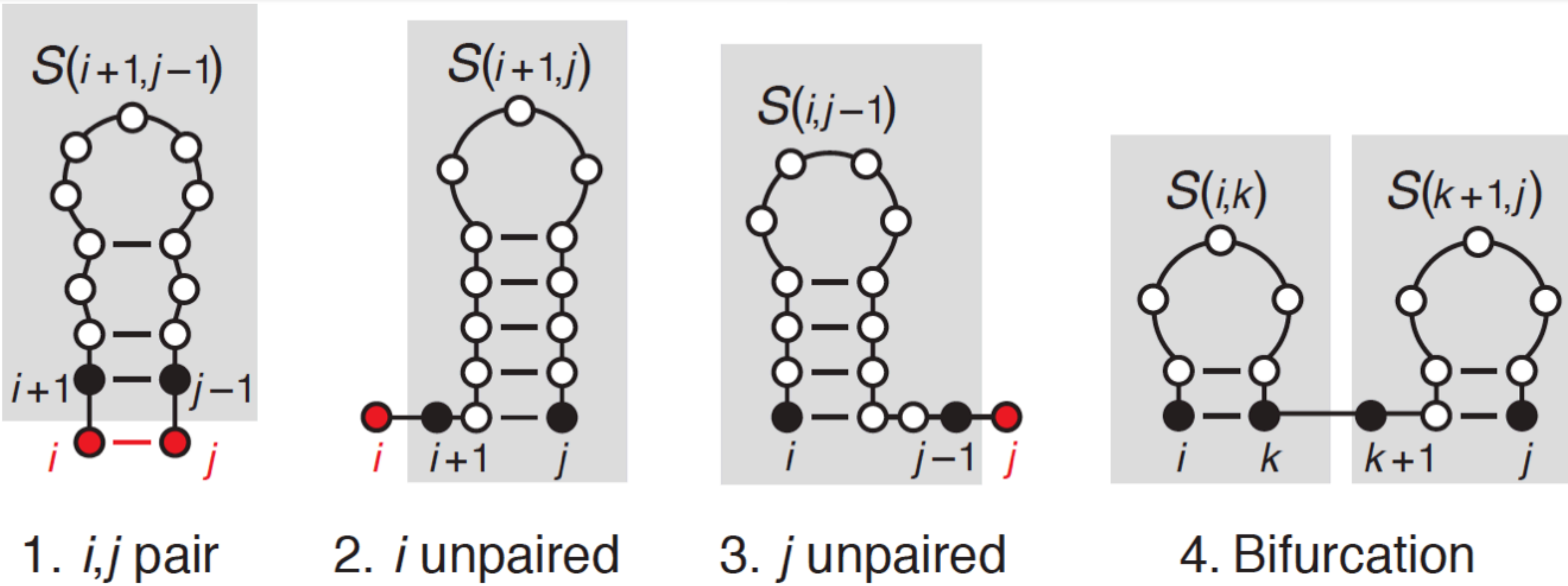
Sequence Covariation

- Main Idea is:
- Gutell quite successful in getting rRNA.

Delta G Minimization

- Tinoco-Uhlenbeck Postulate
- Dynamic Programming Nussinov Algorithm

Nussinov Algorithm (Maximum Number of Base-Pairs)



$$D(i,j) = \max \begin{Bmatrix} max_{i<k<j} D(i,k) + D(k+1,j) \\ D(i+1,j-1) + \ w(i,j) \\ D(i+1,j) \\ D(i,j-1) \end{Bmatrix}$$

Nussinov et al. *SIAM J. Appl. Math.* **35**, 68-82 (1978)

Initialization

$$D(i,j) = \max \begin{Bmatrix} D(i,i) = 0 & \forall i = 1..L \\ D(i,i-1) = 0 & \forall i = 2..L \end{Bmatrix}$$

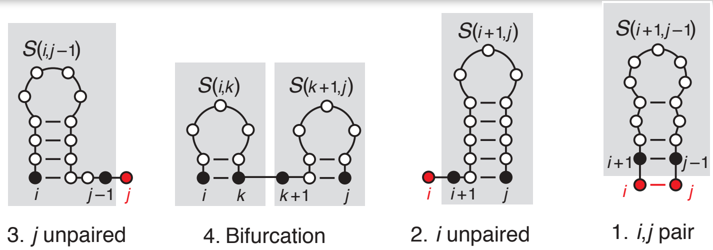
Recursion

$$D(i,j) = \max \begin{Bmatrix} D(i,k) + D(k+1,j) & where\ i \leq k < j \\ D(i+1,j-1) + \ w(i,j) & \end{Bmatrix}$$

Nussinov and Jacobson *PNAS* **77**, 6309-6313 (1980)

The Algorithm in Action (4 MNBP)

$$D(i,j) = \max \begin{cases} D(i,j-1) \\ D(i+1,j-1) + w(i,j) \\ D(i+1,j) \\ \max_{i<k<j} D(i,k) + D(k+1,j) \end{cases}$$



S. Eddy, *Nature Biotech.* **22**, 1457-1458 (2004)

D(1,1) = max
- i<k<j NOT
- D(2,0) + w(1,1) = 0
- D(2,1) = 0
- D(1,0) = 0

D(2,5) = max
- D(2,4) = 1
- D(3,4) + w(2,5) = 0
- D(3,5) = 1
- k={3,4}
  max{
  D(2,3)+D(4,5) = 2
  D(2,4)+D(5,5) = 1
  } = 2

	j	1	2	3	4	5	6	7	8	9	10
i		A	C	G	G	C	A	A	C	G	U
1	A	0	0	1	1	2	2	2	2	3	4
2	C	0	0	1	1	2	2	2	2	3	4
3	G		0	0	0	1	1	1	2	2	3
4	G			0	0	1	1	1	1	2	3
5	C				0	0	0	0	0	1	2
6	A					0	0	0	0	1	2
7	A						0	0	0	1	2
8	C							0	0	1	1
9	G								0	0	0
10	U									0	0

Backtracking to Structure (Counterclockwise)

$$D(i,j) = \begin{cases} D(i,j-1) \\ D(i+1,j) \\ D(i+1,j-1) + w(i,j) \\ for_{\ i<k<j}\ do\ D(i,k) + D(k+1,j) \end{cases}$$

D(1,10) =

- D(1,9) = 3 ≠ 4 ✖
- D(2,10) = 4 = 4 ✓
- D(2,9) + w(1,10) = 4 = 4 ✓

k={2,3,4,5,6,7,8,9}
D(1,2)+D(3,10) = 3
D(1,3)+D(4,10) = 4 ✓
D(1,4)+D(5,10) = 3
D(1,5)+D(6,10) = 4 ✓
D(1,6)+D(7,10) = 4 ✓
D(1,7)+D(8,10) = 3
D(1,8)+D(9,10) = 2
D(1,9)+D(10,10) = 3

	j	1	2	3	4	5	6	7	8	9	10
i		A	C	G	G	C	A	A	C	G	U
1	A	0	0	1	1	2	2	2	2	3	4
2	C	0	0	1	1	2	2	2	2	3	4
3	G		0	0	0	1	1	1	2	2	3
4	G			0	0	1	1	1	1	2	3
5	C				0	0	0	0	0	1	2
6	A					0	0	0	0	1	2
7	A						0	0	0	1	2
8	C							0	0	1	1
9	G								0	0	0
10	U									0	0

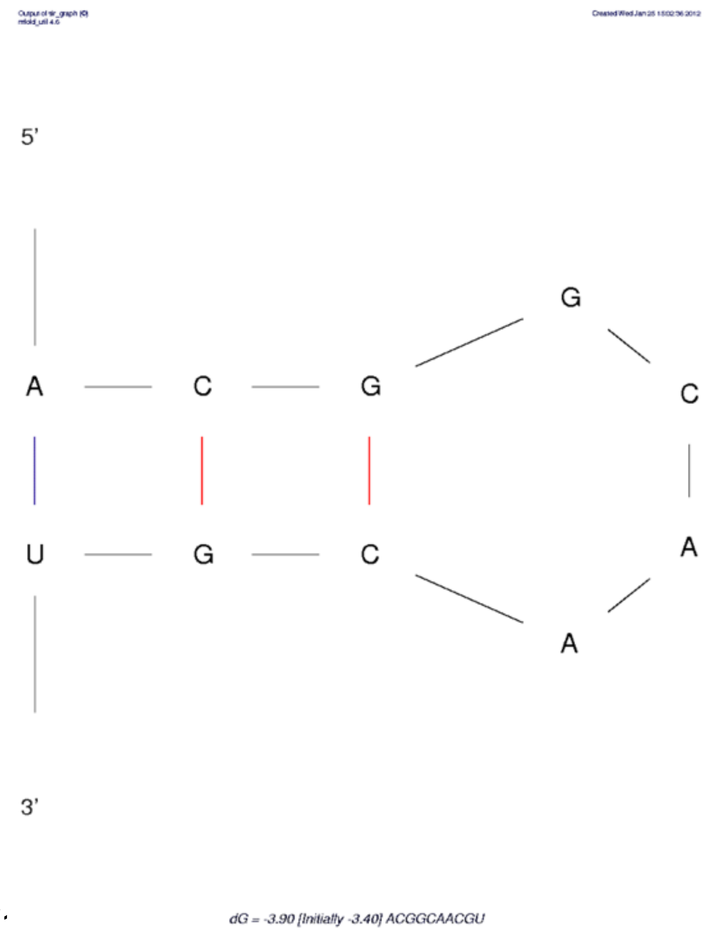
One Possible Structure With 4 Base-Pairs

ACGGCAACGU
(((()..)))
but |j-1| > 1
(((....)))

	j	1	2	3	4	5	6	7	8	9	10
i		A	C	G	G	C	A	A	C	G	U
1	A	0	0	1	1	2	2	2	2	3	4
2	C	0	0	1	1	2	2	2	2	3	4
3	G		0	0	0	1	1	1	2	2	3
4	G			0	0	1	1	1	1	2	3
5	C				0	0	0	0	0	1	2
6	A					0	0	0	0	1	2
7	A						0	0	0	1	2
8	C							0	0	1	1
9	G								0	0	0
10	U									0	0

One Possible Structure With 4 Base-Pairs

ACGGCAACGU
((((()..)))
but |j-1| > 1
(((....)))



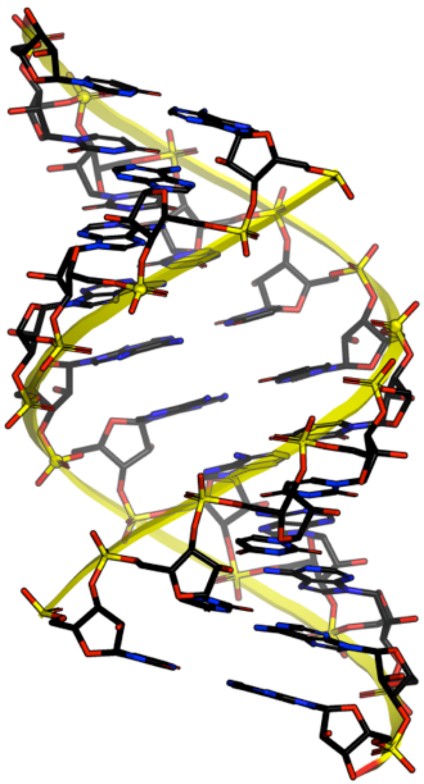
	j	1	2	3	4	5	6	7	8	9	10
i		A	C	G	G	C	A	A	C	G	U
1	A	0	0	1	1	2	2	2	2	3	4
2	C	0	0	1	1	2	2	2	2	3	4
3	G		0	0	0	1	1	1	2	2	3
4	G			0	0	1	1	1	1	2	3
5	C				0	0	0	0	0	1	2
6	A					0	0	0	0	1	2
7	A						0	0	0	1	2
8	C							0	0	1	1
9	G								0	0	0
10	U									0	0

Nussinov algorithm - Ultrastudio.org

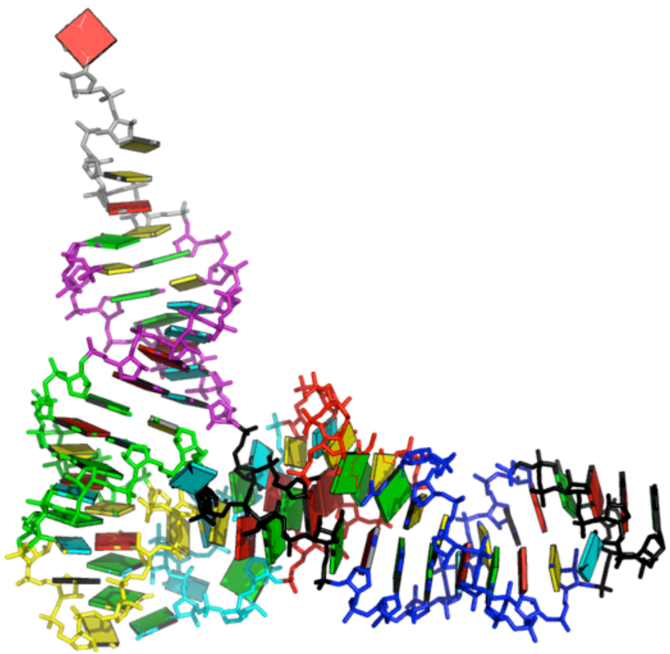
Visualization of Secondary Structure

- Standard
- Circular
- Dot Plot

Tertiary Structure



A-RNA



Nobel Prizes and RNA

Classic Papers on Secondary Structure

- Classics
  - Tinoco - Uhlenbeck Nature 1971
- Books
  - Tinoco’s Book
  - Bioinformatics Book Chapter by Mathews
- Online
- Recent Trends
- People

RNA Secondary Structure Prediction Groups

- Doug Turner University of Rochester (Turner Rules)
- David Mathews University of Rochester (Dynalign)
- Michael Zuker Rensselaer Polytechnic Institute (mfold)
- Ivo Hoffacker Wien (vienna)
- Peter Stadler Leipzig (vienna)
- Francois Major Montreal (mc-sym)
- Tamar Schlick NYU (graph-grammars)
- Ruth Nussinov Tel Aviv (1978 dynamic algorithm)
- Nacho Tinoco (1971 free-energy minimization)