

-old homepage-

Post Doctoral Researcher
Department of Biosciences & Nutrition
Karolinska Institutet
Phone number: (+46)-08-524-81079 (lab)
(+46)-073-678-5334 (cell)
e-mail: [mauricio.esguerra at ki dot se](mailto:mauricio.esguerra@ki.se)

<http://mesguerra.net>

All class materials at:

<https://github.com/esguerra/ki2430>

Why Study RNA?

1. To **understand** what RNA does.
2. To **predict** what RNA can do.
3. To **control** RNA.

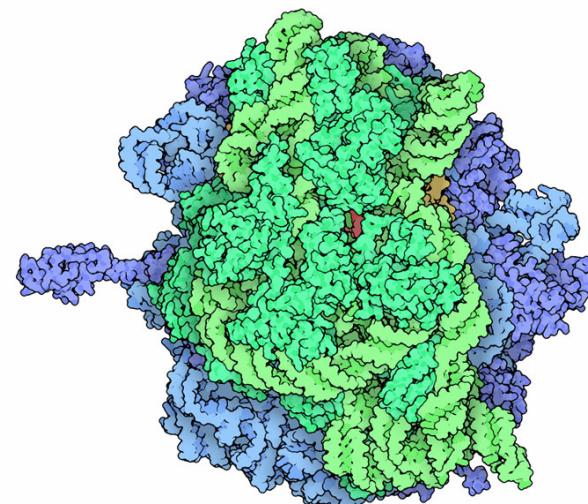
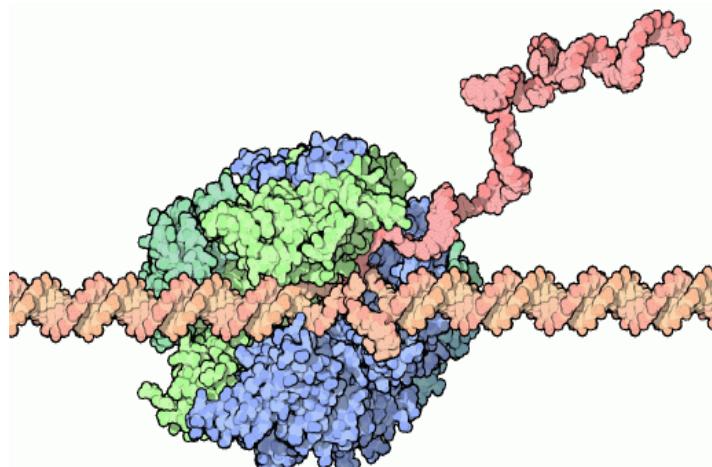
Borrowed from:

Ignacio Tinoco @ MSRI June 2000

<http://www.msri.org/realvideo/ln/msri/2000/molbio/tinoco/1/banner/01.html>

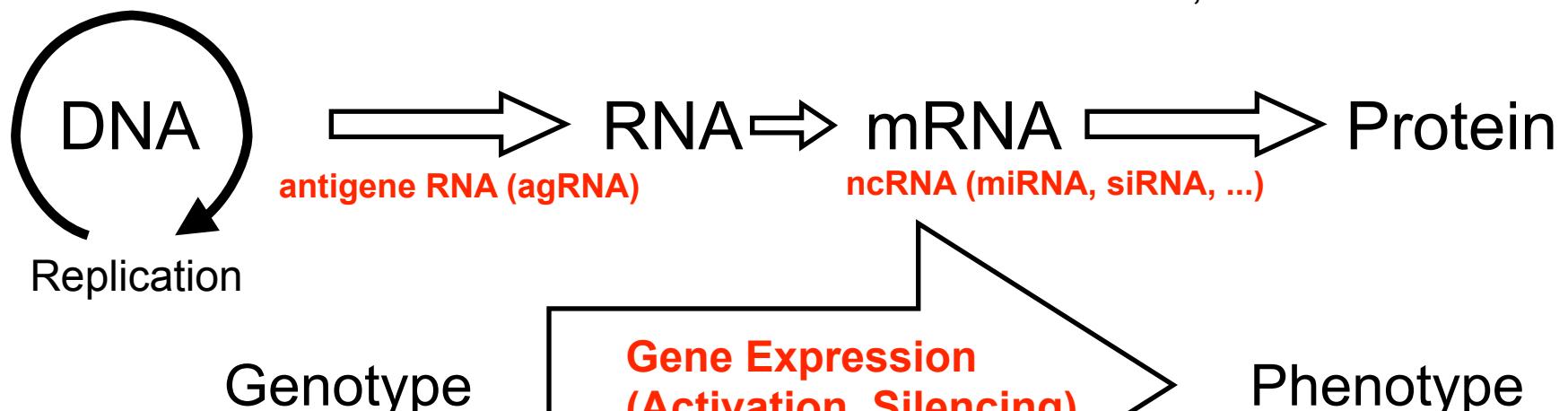
Highly Recommended!

Biology's Central Dogma and Molecular Machines

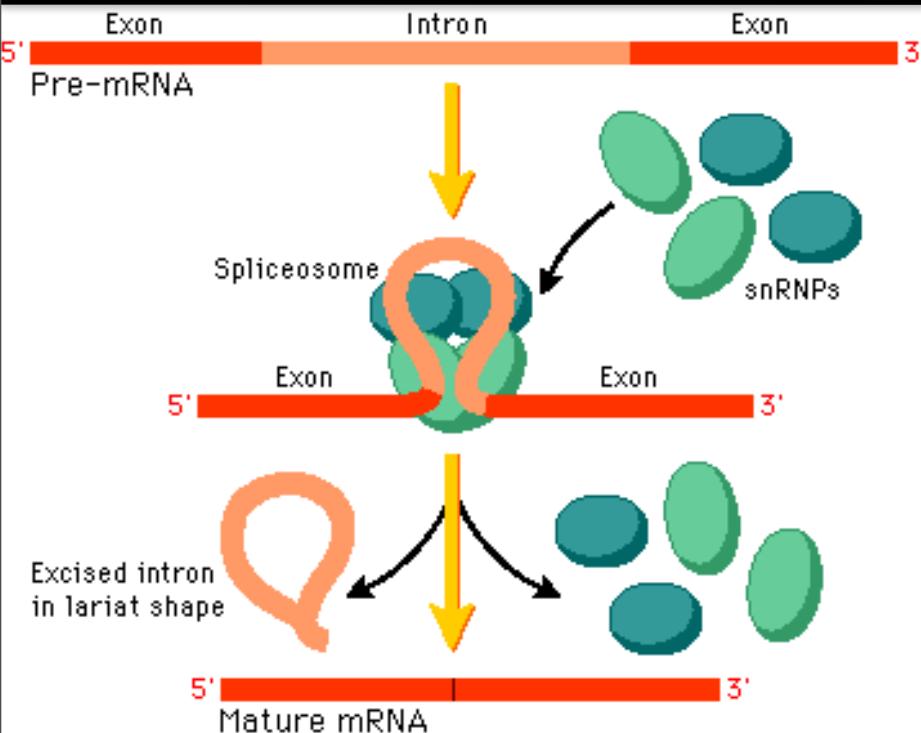


TRANSCRIPTION
RNA Pol II

TRANSLATION
Ribosome, aka rRNA



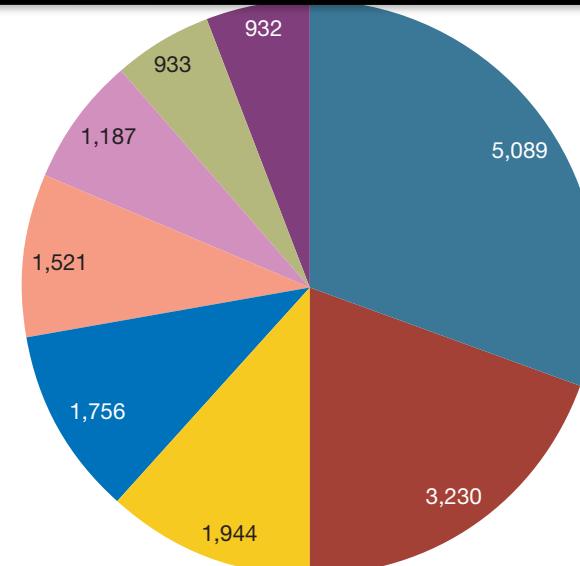
RNA is Spliced by the Spliceosome After Transcription.



Protein GENES
RNA GENES

<http://genome.ucsc.edu/ENCODE/>

<http://www.gencodegenes.org/>



J. Harrow, Havana project, preliminary Gencode 7 data

Total noncoding RNA genes = 16,592

■ lincRNA ■ snoRNA
■ Antisense ■ miscRNA
■ snRNA ■ Uncategorized processed transcript
■ microRNA ■ Other

Long noncoding RNAs are just one of many noncoding transcripts being annotated. lincRNA, long intergenic noncoding RNA; snRNA, small nuclear RNA; snoRNA, small nucleolar RNA; and miscRNA, miscellaneous RNA.

~3200 Million base-pairs (Humans)
< 2% makes mature mRNA (protein)
~50% transcribed ncRNA.

Some Common Non-coding RNA's

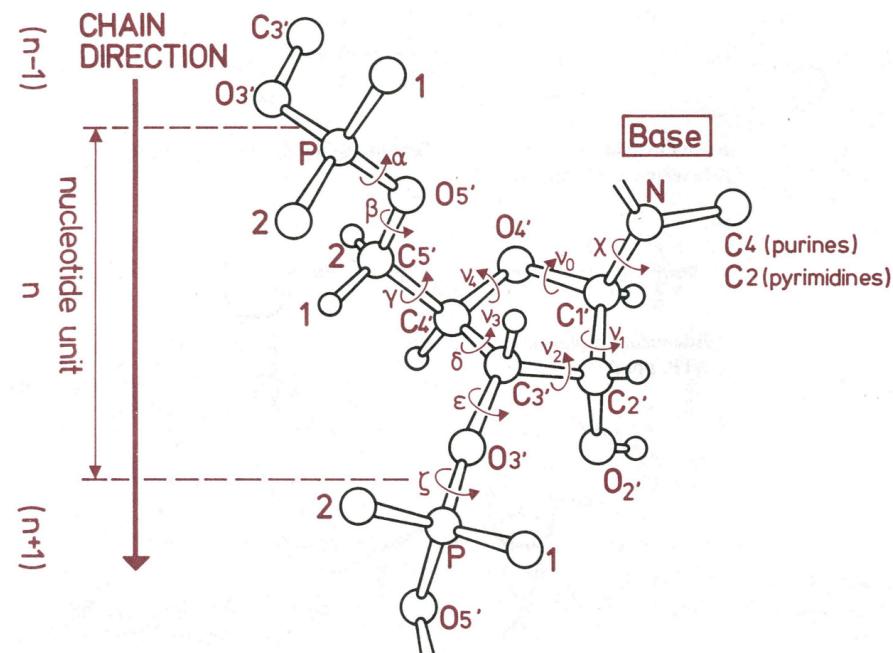
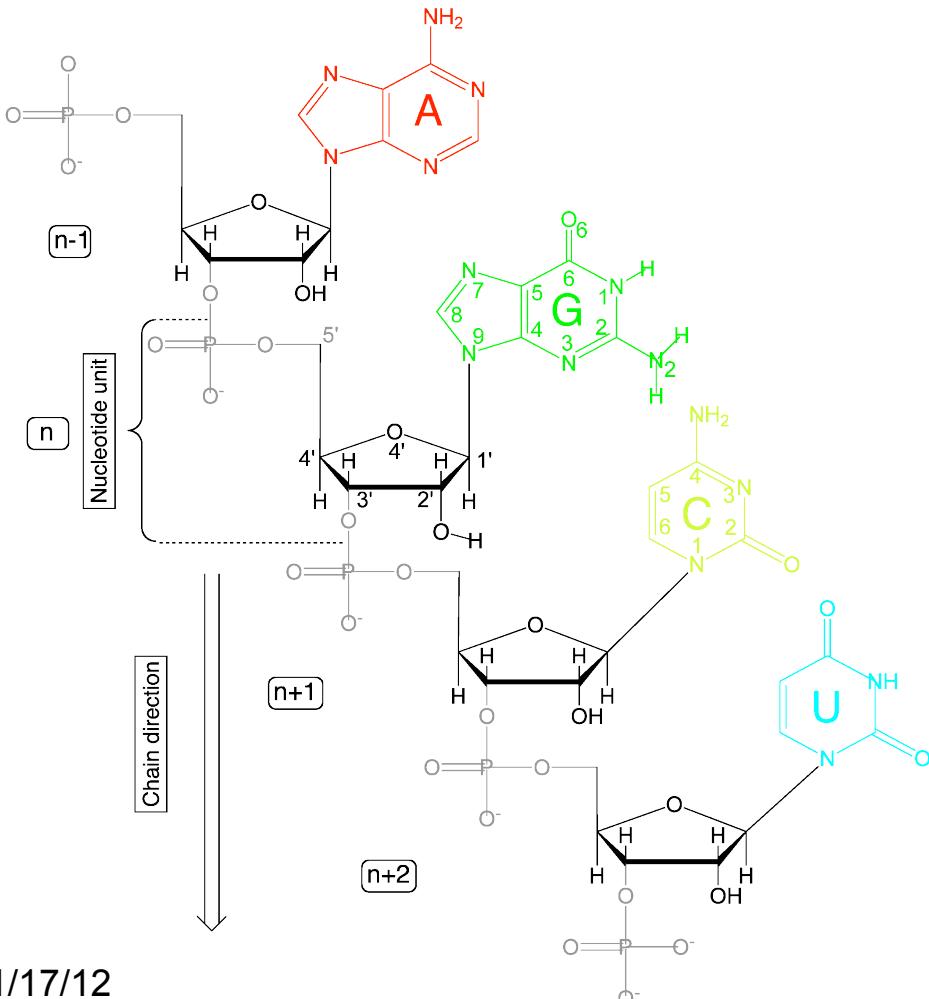
- RNAPol I rRNA not 5S (~ 50% cell RNA)
- RNAPol II precursors of mRNA and most snRNA and microRNA (non-coding-RNA genes)
- RNAPol III 5S rRNA, tRNA and other small RNAs
- piwiRNA, siRNA, etc.
- http://rfam.sanger.ac.uk/family/browse/with_structure#A

Structure Determines Function

RNA Structure Can Be Understood at Levels.

Primary Structure (Sequence)

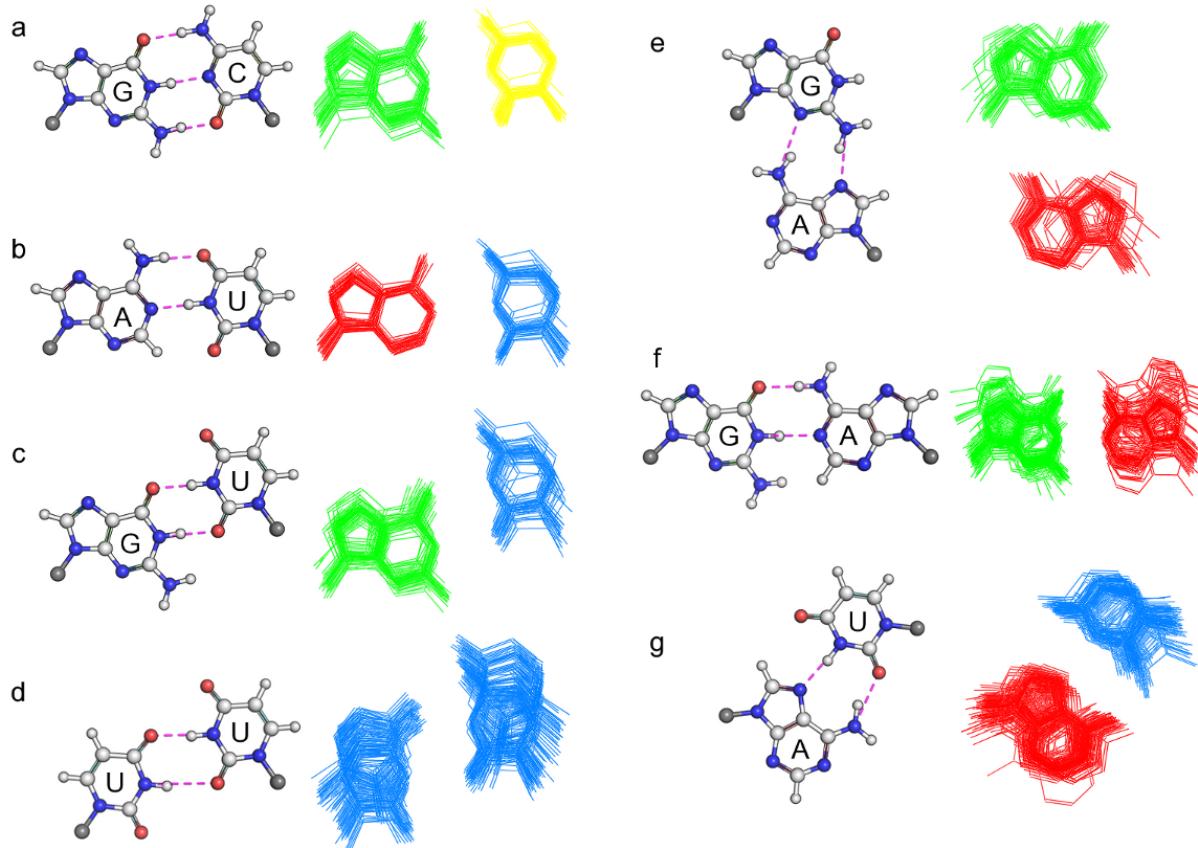
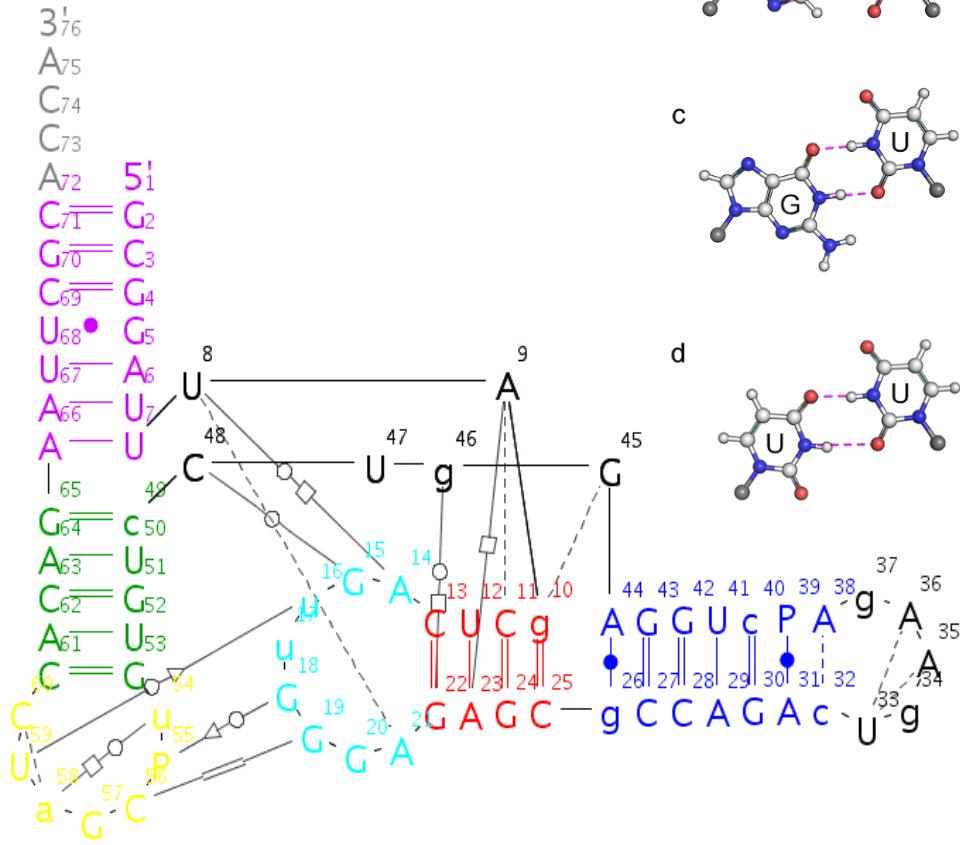
GC₂GGAUU UA gCUC AGGuuGGGA GAGC gCCAGAc UgAAg APCUGGA GgUC cUGUG uPCGaUC CACAG AAUUCGC ACCA
1234567 89 0123 45678901 2345 6789012 34567 8901234 5678 90123 4567890 12345 6789012 3456



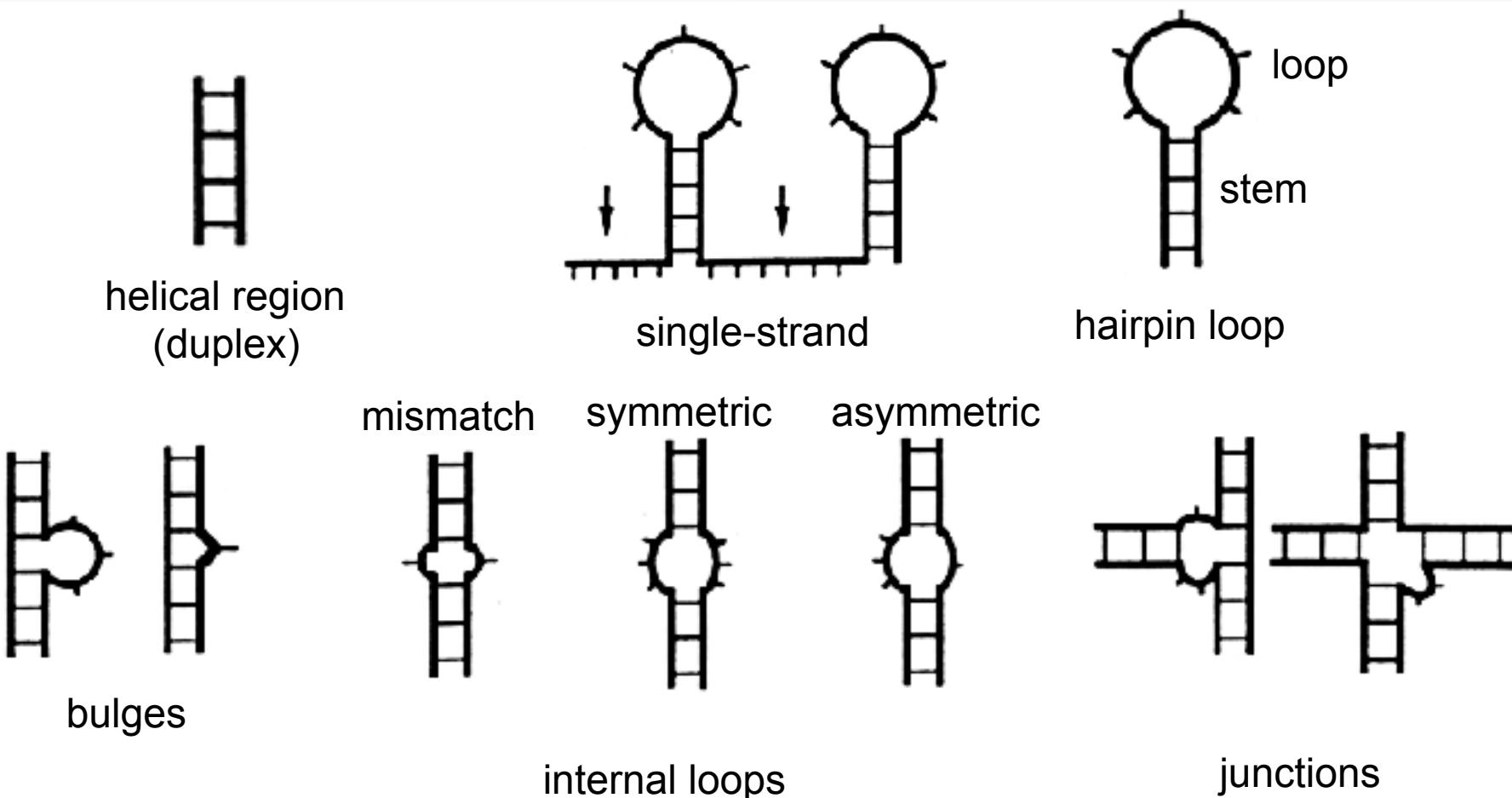
Backbone covalently linked through phosphodiester groups with 5' to 3' sense.

Secondary Structure (Base-pairing)

1. Sequence
2. Network Hydrogen Bonded Base-Pairs (Canonical and Non-Canonical)



RNA Secondary Structure “Elements”

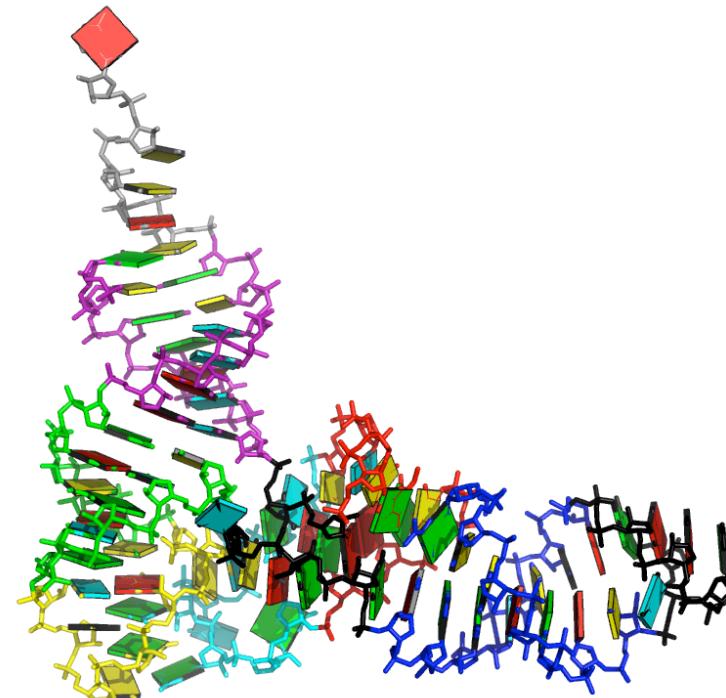
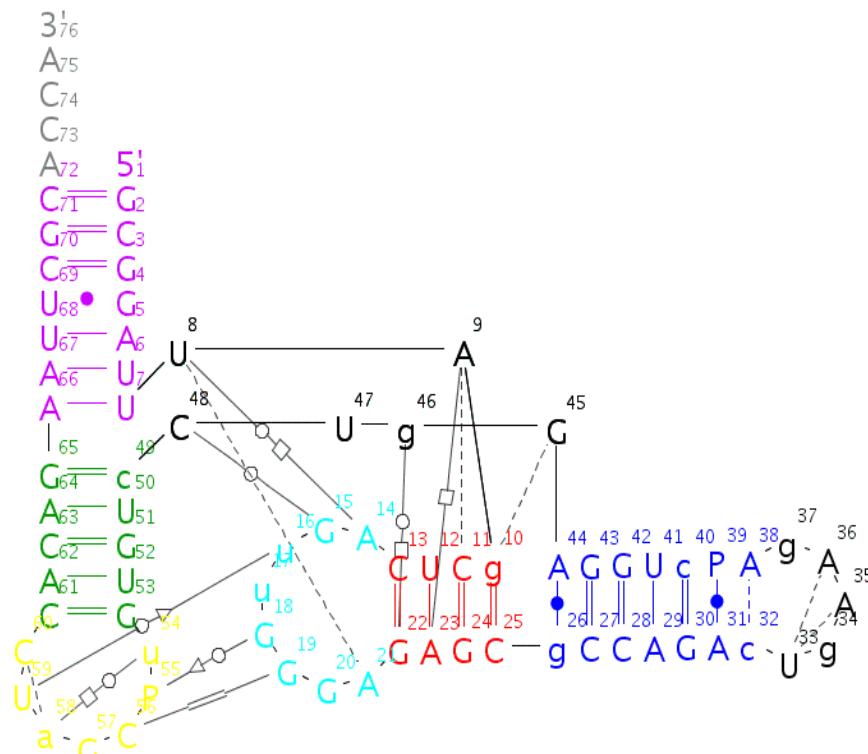


SCOR: Structural Classification of RNA
<http://scor.berkeley.edu>

Rfam: RNA Families
<http://rfam.sanger.ac.uk>

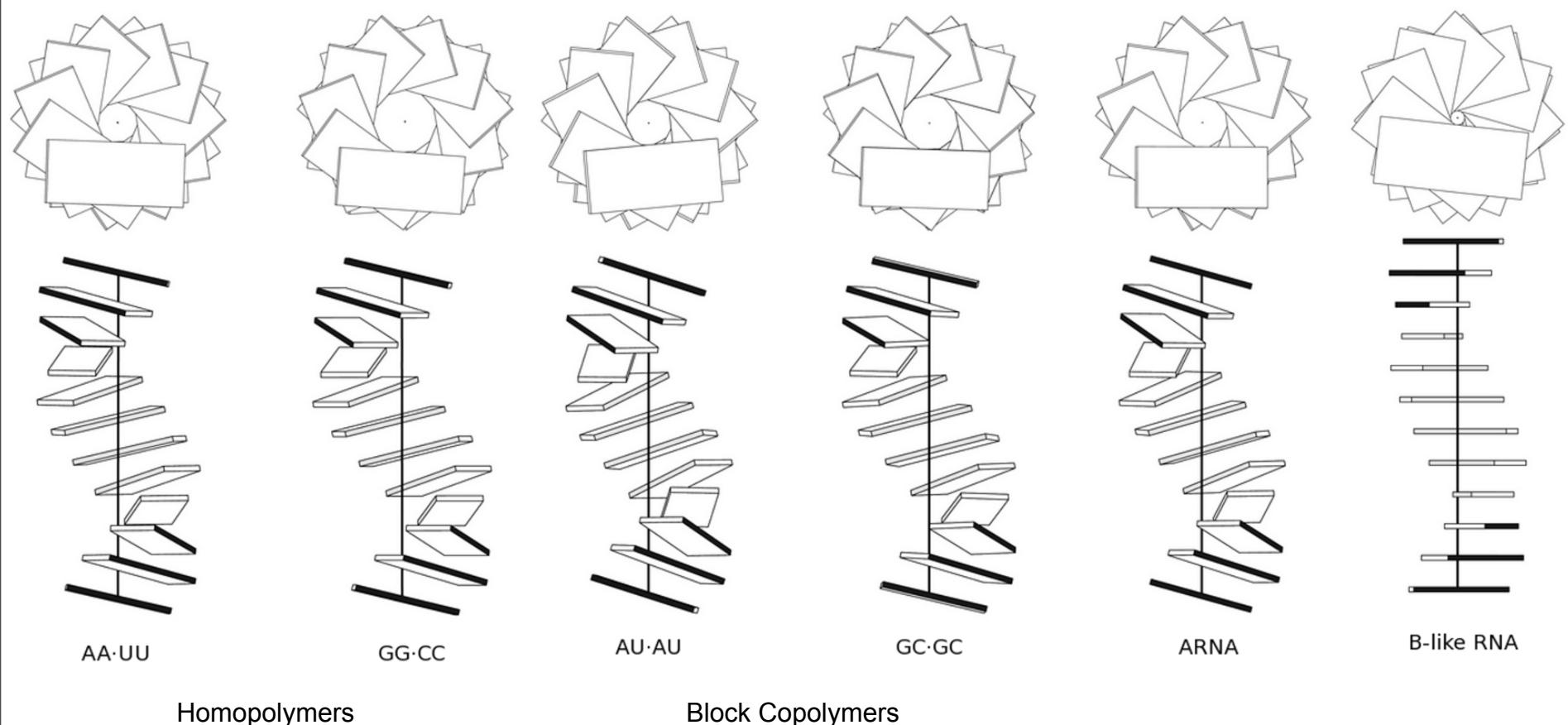
3D Structure of RNA (The Folding Problem)

GC₆GGAUU 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



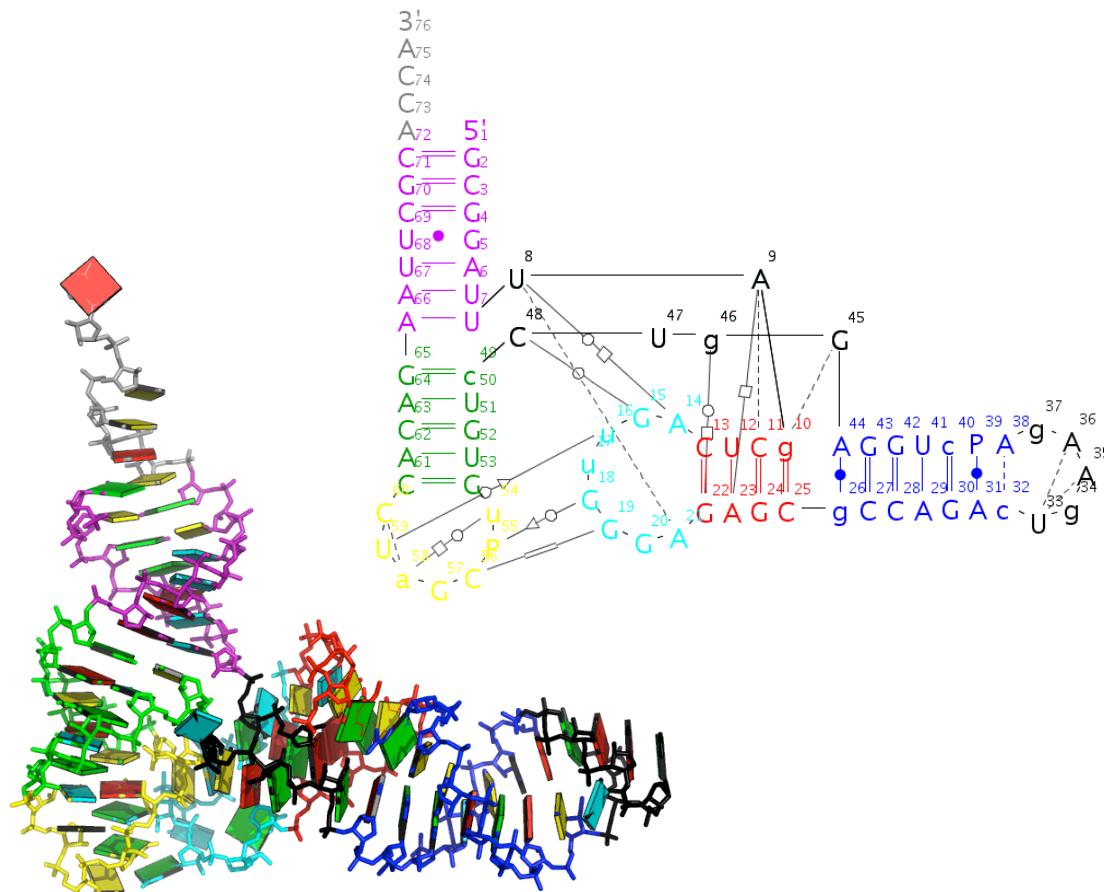
1. Sequence
2. Network Hydrogen Bonded Base-Pairs (Canonical and Non-Canonical)
3. Network of so-called “Tertiary Interactions” (Secondary-Motif interactions), e.g. ion-mediated, base-backbone, inter-helical base-pairs.

3D Structure (Subtle Conformational Variation of Helical Regions)



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



Two, Kinda Three, Types 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 to find helical regions.

“...helix is considered to exist when:

- (i) it can be formed in at least two different 16S rRNAs (involving homologous segments in the molecule in all cases), and
- (ii) a canonical base pairing covariance can be demonstrated (i.e., a Watson-Crick pairing in one case being replaced by a different Watson-Crick pairing in some other case) for at least two pairs in the helix”

TABLE 1. Helix 9-13/21-25^a

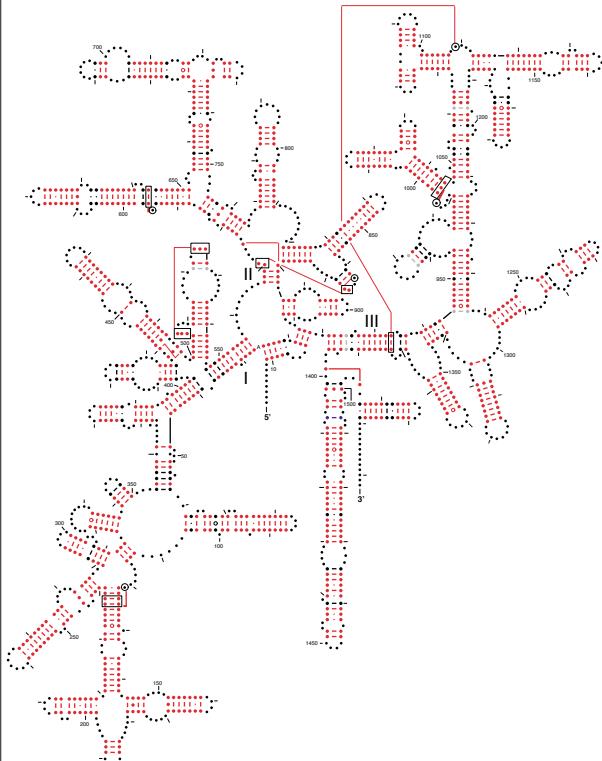
Organism/ organelle	Sequence	
	10	20
<i>E. coli</i> G A A [G A G U U] U G A U C A U [G G C U C] A
<i>H. volcanii</i>	pA U [U C C G G U] U G A U C C U [G C C G G A]
<i>X. laevis</i>	pU A C [C U G G U] U G A U C C U [G C C A G] U
Human mitochondria	pA A U [A G G U U] U G G U C C U [A G C C U] U

- Robin Gutell quite successful in getting rRNA.

Secondary Structure of E.Coli Ribosome

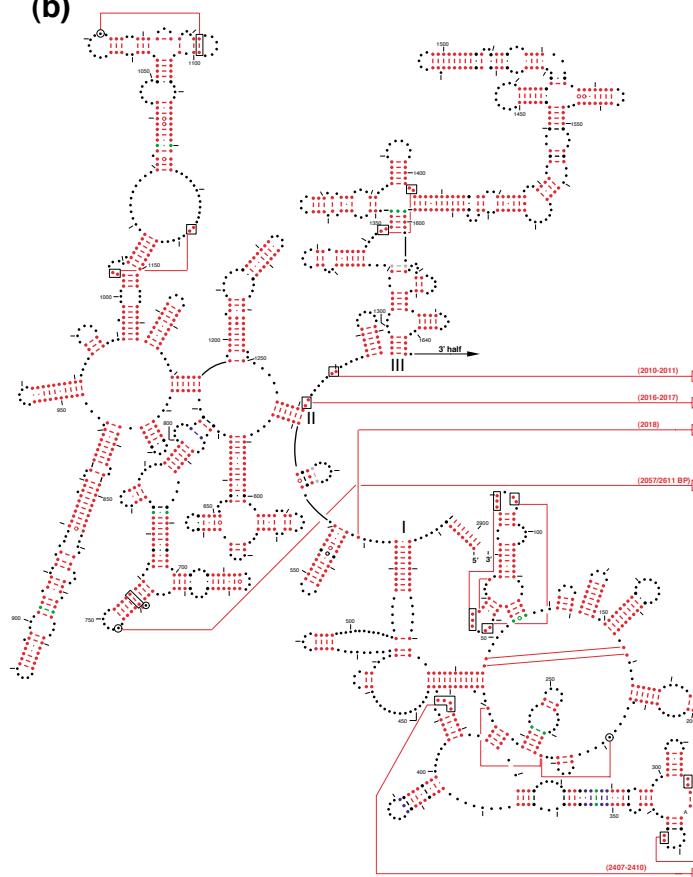
(looks like airport map)

(a)



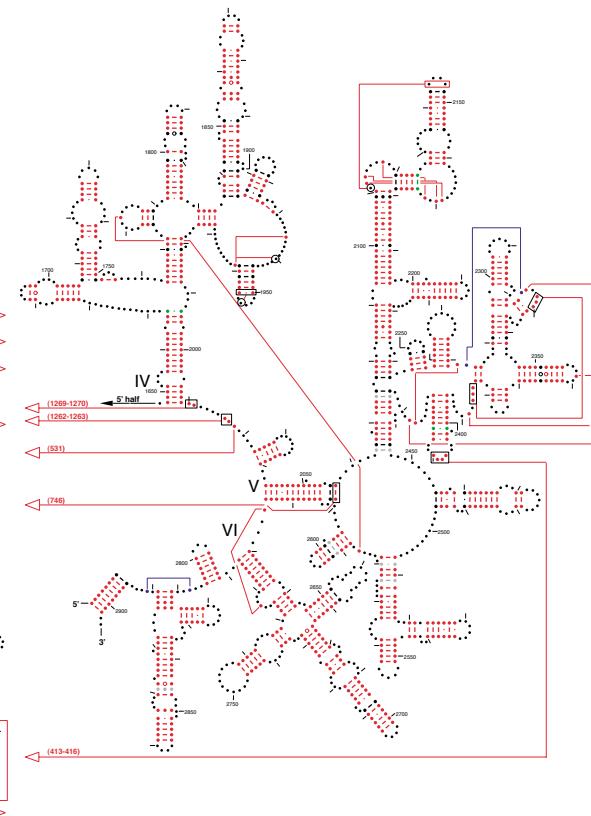
16S

(b)



23S - 5'

(c)



23S - 3'

<http://www.rna.icmb.utexas.edu/SAE/2A/RNA2DMap/version2/Main.php>

Note

Delta G Minimization

Tinoco-Uhlenbeck Postulate:

Assume: The energy of each base-pair is independent of all of the other pairs and the loop structure.

Consequence: Total free energy is the sum of all base-pair free energies.

Turner Rules:

Assume: The energy of secondary structure motifs is independent of all other secondary structure motifs.

Consequence: Total free energy is the sum of all secondary structure motifs.

“The secondary structure which an RNA molecule takes up at equilibrium in any solution is that which minimizes the free energy of the solution”

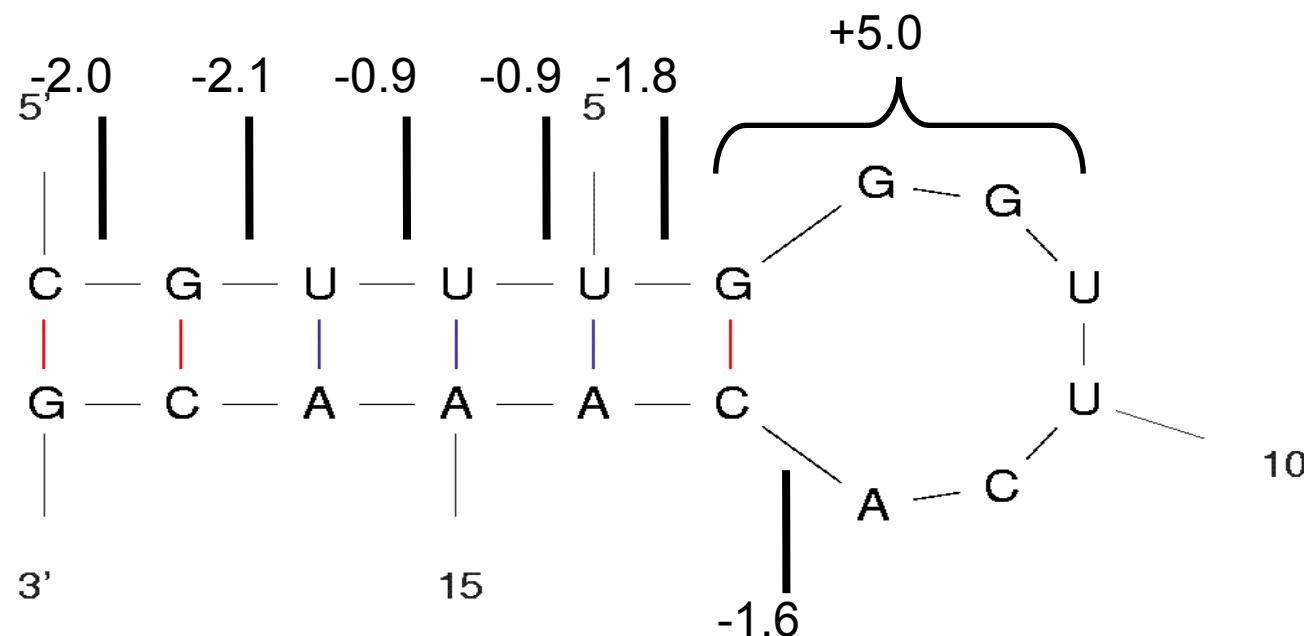
Tinoco, Uhlenbeck, and Levine (Nature V.230, 362, 1971)

Nearest Neighbor Model

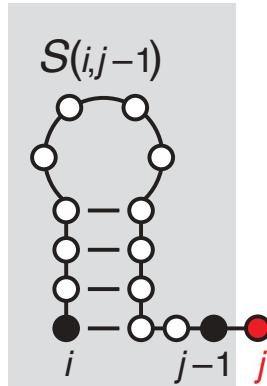
$$\Delta G^o_{stem} = \Delta G^o \begin{bmatrix} C & G \\ G & C \end{bmatrix} + \Delta G^o \begin{bmatrix} G & U \\ C & A \end{bmatrix} + 2\Delta G^o \begin{bmatrix} U & U \\ A & A \end{bmatrix} + \Delta G^o \begin{bmatrix} U & G \\ A & C \end{bmatrix} = -7.7 \text{kcal/mol}$$

$$\Delta G^o_{loop} = \Delta G^o_{6-nucleotide} + \Delta G^o_{non-canonical} \begin{bmatrix} G & G \\ C & A \end{bmatrix} = 3.4 \text{kcal/mol}$$

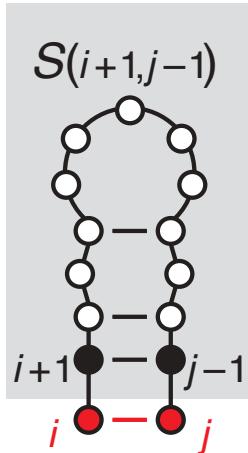
$$\Delta G^o_{total} = \Delta G^o_{stem} + \Delta G^o_{hairpin loop} = -4.3 \text{kcal/mol}$$



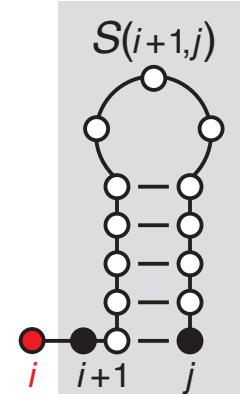
Nussinov Algorithm (Maximum Number of Base-Pairs)



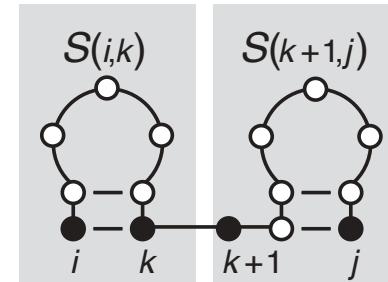
3. j unpaired



1. i,j pair



2. i unpaired



4. Bifurcation

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

Initialization

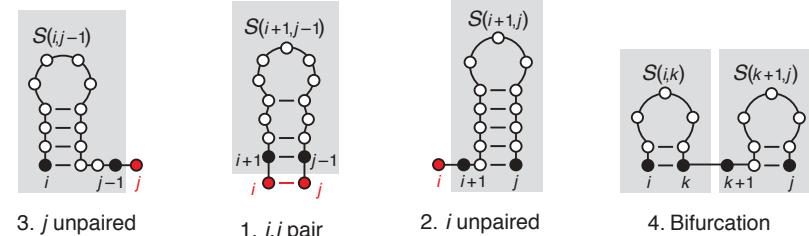
$$D(i,j) = \max \left\{ \begin{array}{ll} D(i,i) = 0 & \forall i = 1..L \\ D(i,i-1) = 0 & \forall i = 2..L \end{array} \right\}$$

Recursion

$$D(i,j) = \max \left\{ \begin{array}{ll} D(i,k) + D(k+1,j) & \text{where } i \leq k < j \\ D(i+1,j-1) + w(i,j) & \end{array} \right\}$$

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)

	$i < k < j$ NOT $D(2,0) + w(1,1) = 0$ $D(2,1) = 0$ $D(1,0) = 0$
$D(1,1) = \max$	$D(2,4) = 1$
$D(2,5) = \max$	$D(3,4) + w(2,5) = 0$ $D(3,5) = 1$ $k=\{3,4\}$ $\max\{$ $D(2,3)+D(4,5) = 2$ $D(2,4)+D(5,5) = 1$ $\} = 2$

	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	0

Backtracking to Structure (Counterclockwise)

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

$$D(1, 10) = \boxed{\begin{array}{ll} D(1, 9) = 3 \neq 4 & \times \\ D(2, 10) = 4 = 4 & \checkmark \end{array}}$$

$$\boxed{D(2, 9) + w(1, 10) = 4 = 4} \quad \checkmark$$

$$k = \{2, 3, 4, 5, 6, 7, 8, 9\}$$

$$D(1, 2) + D(3, 10) = 3$$

$$D(1, 3) + D(4, 10) = 4 \quad \checkmark$$

$$D(1, 4) + D(5, 10) = 3$$

$$D(1, 5) + D(6, 10) = 4 \quad \checkmark$$

$$D(1, 6) + D(7, 10) = 4 \quad \checkmark$$

$$D(1, 7) + D(8, 10) = 3$$

$$D(1, 8) + D(9, 10) = 2$$

$$D(1, 9) + D(10, 10) = 3$$

	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	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	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

Output of sr_graph (0)
mfold util 4.6

Created Wed Jan 25 15:02:36 2012

5'

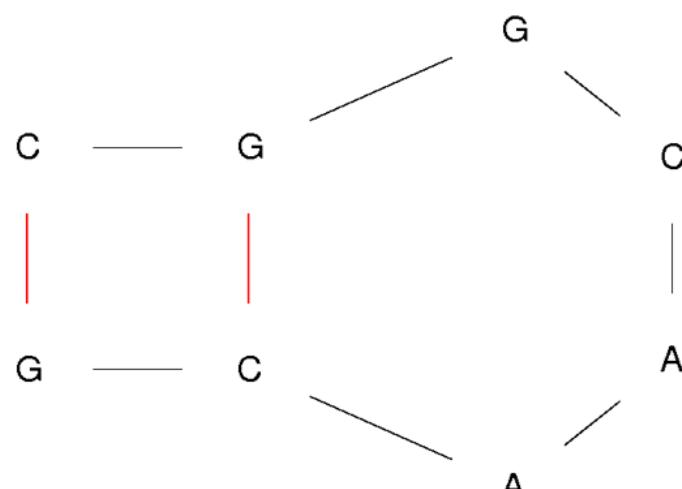
A

U

3'

ACGGCAACGU

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



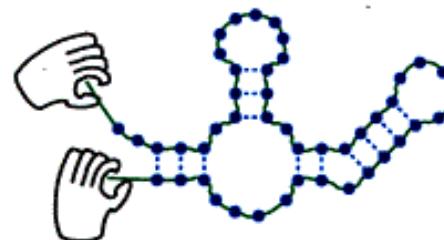
	j	1	2	3	4	5	6	7	8	9	10
i		A	C	G	G	C	A	A	C	G	U
1	A	0	0	1	1	2	2	2	2	3	4
2	C	0	0	1	1	2	2	2	2	3	4
3	G		0	0	0	1	1	1	2	2	3
4	G			0	0	1	1	1	1	2	3
5	C				0	0	0	0	0	1	2
6	A					0	0	0	0	1	2
7	A						0	0	0	1	2
8	C							0	0	1	1
9	G								0	0	0
10	U									0	0

Online Java Tool Using the Nussinov Algorithm

[Nussinov algorithm - Ultrastudio.org](#)

Visualization of Secondary Structure

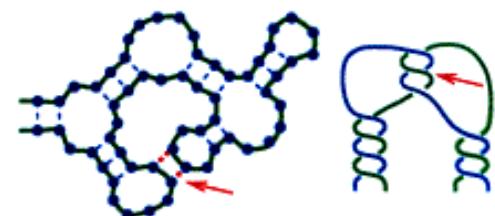
- Planar graph



- Arch diagram



- Pseudo-knots neglected
⇒ arches do not cross



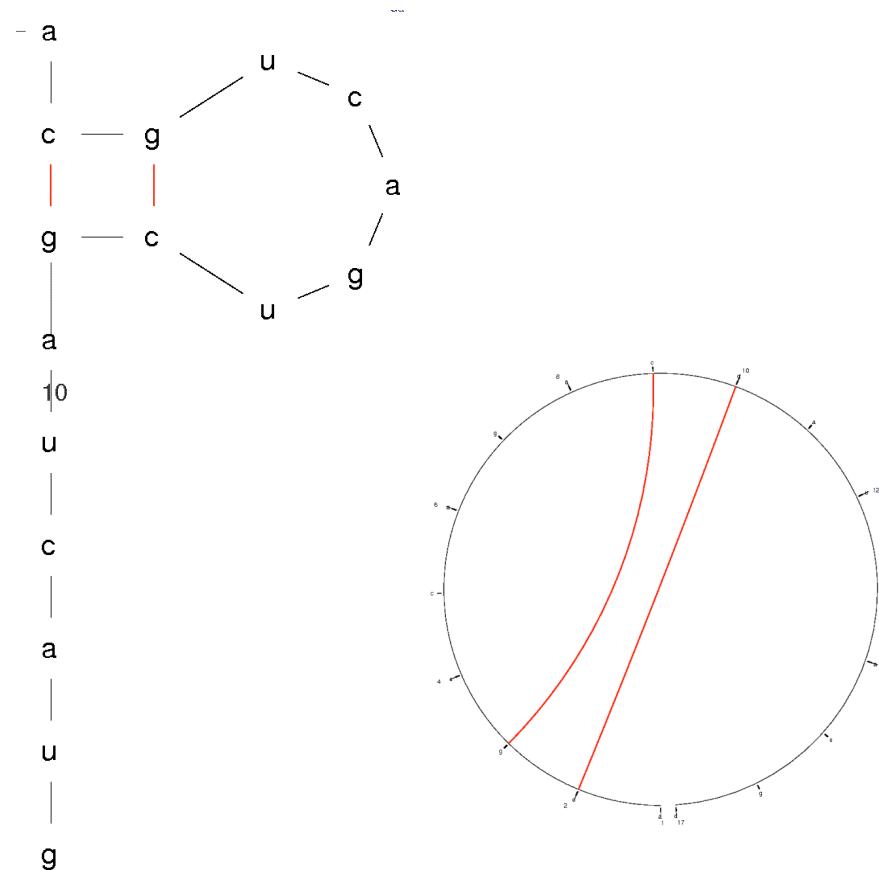
- Grammatically correct string of parentheses
...(((.((((.....))))((((((.....))))..))....)))

From Ralf Bunschuh at:

[http://online.itp.ucsb.edu/
online/infobio01/bunschuh1/](http://online.itp.ucsb.edu/online/infobio01/bunschuh1/)

More Visualization

- Standard
- Circular
- Dot Plot

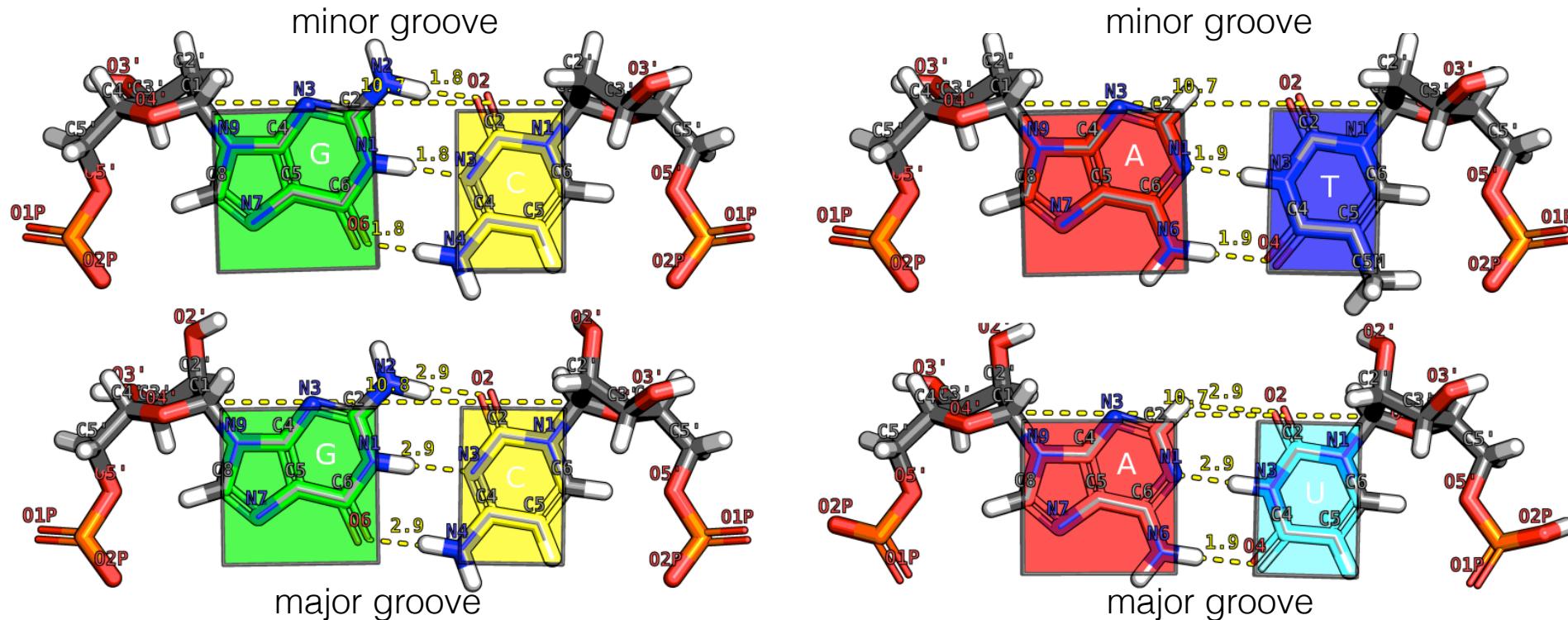


<http://mfold.rna.albany.edu/?q=mfold/RNA-Folding-Form>



RNA 3D STRUCTURE

NDB/PBD Color Convention

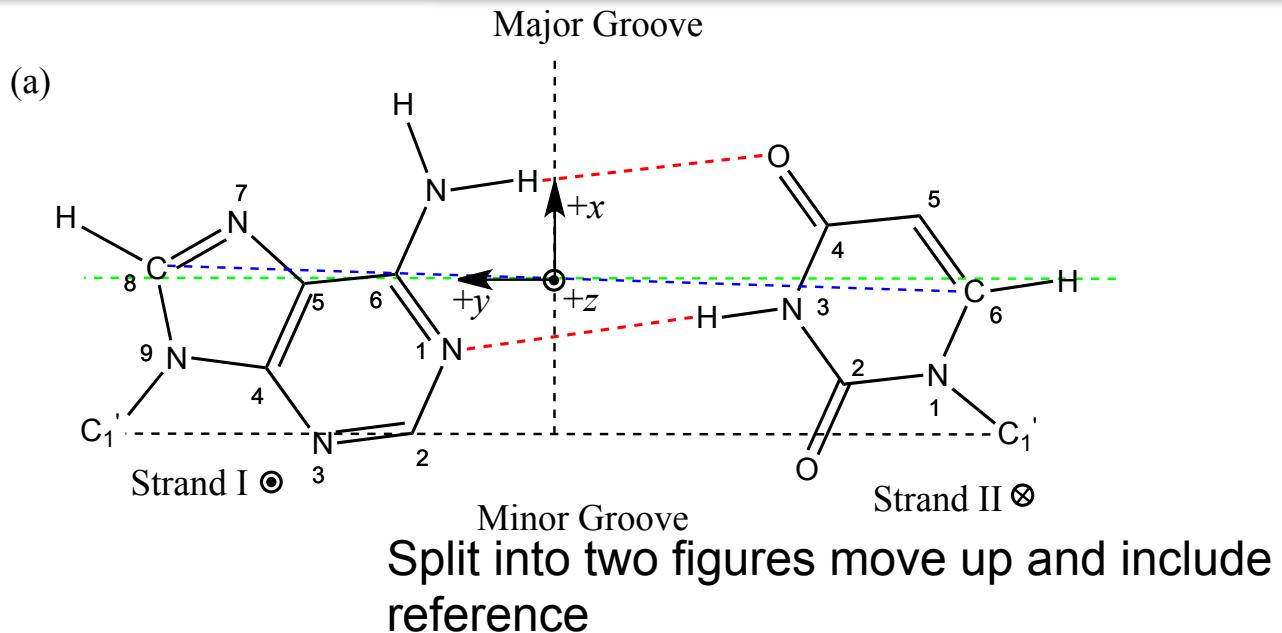


Purines aka **R** = Guanine (green) and Adenine (red)

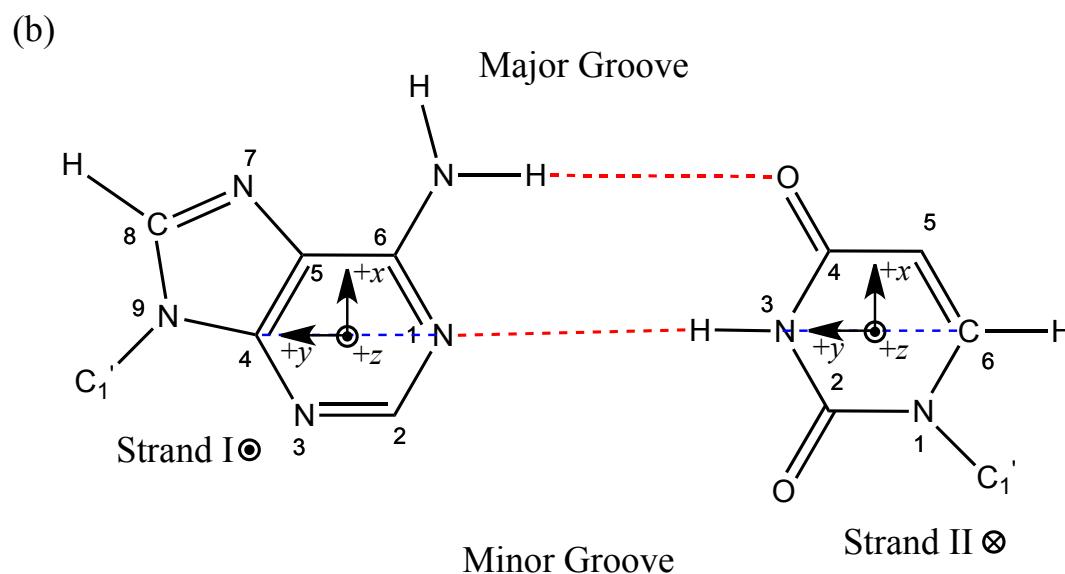
Pyrimidines aka **Y** = Cytosine (yellow), Thymine (blue) and Uracil (cyan)

Standard Reference Frame

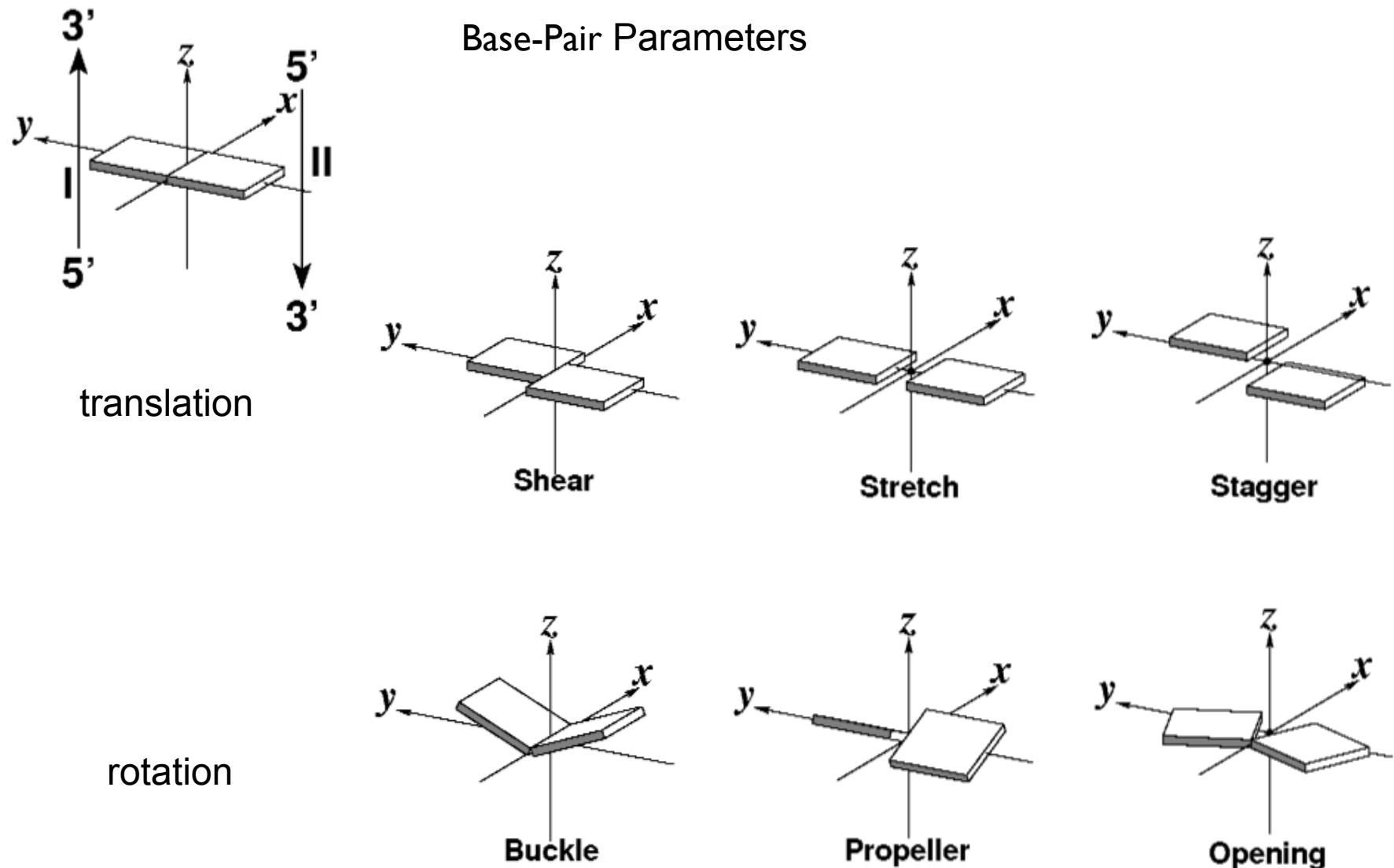
new
3DNA, Curves+
1999-2001



old
SCHNAAp
1997

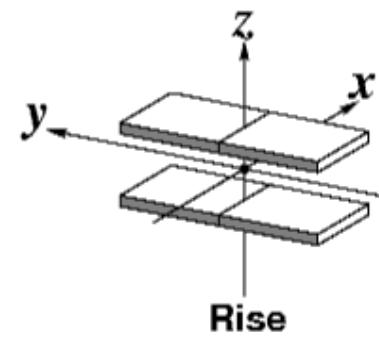
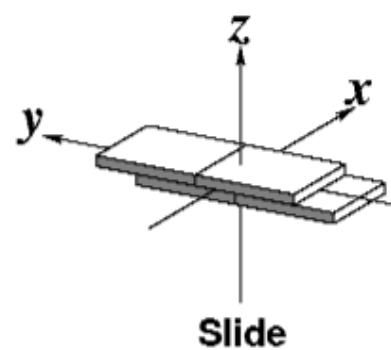
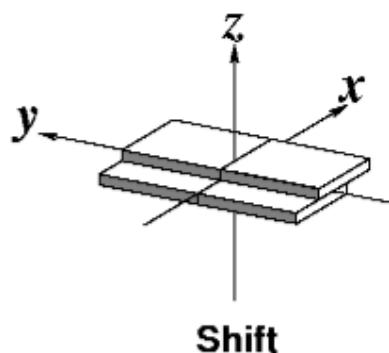


Calladine-Drew Rigid Block Models (Base-Pairs)

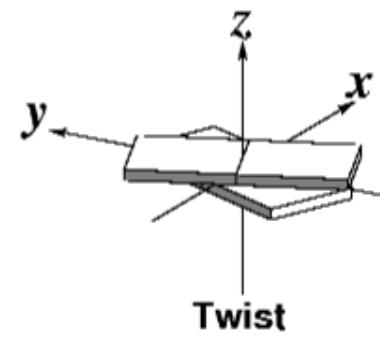
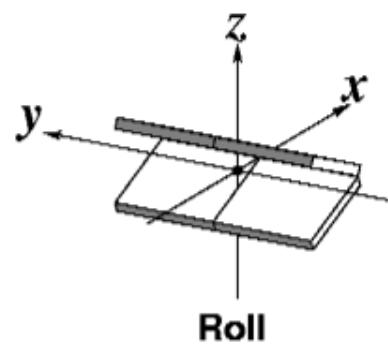
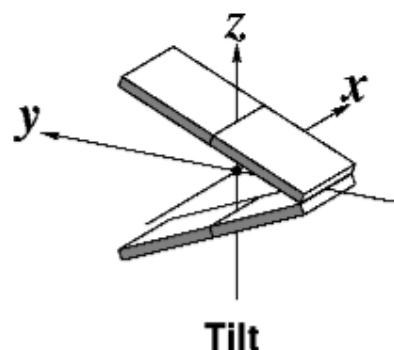


Rigid Block Model Base-Pair Steps

Base-Pair-Step Parameters



translation



rotation

Software to Compute Rigid-Body Parameters

NEW

- 3DNA (Xiang-Jun Lu @ Columbia, Bussemaker lab.)
 - <http://w3dna.rutgers.edu>
 - <http://rutchem.rutgers.edu/~xiangjun/3DNA>
- 3dnaV2, UNIX preferred (linux, OS-X, freebsd, cygwin
“yikes!”)
- Curves+ (Richard Lavery @ Universite d' Lyon)

OLD

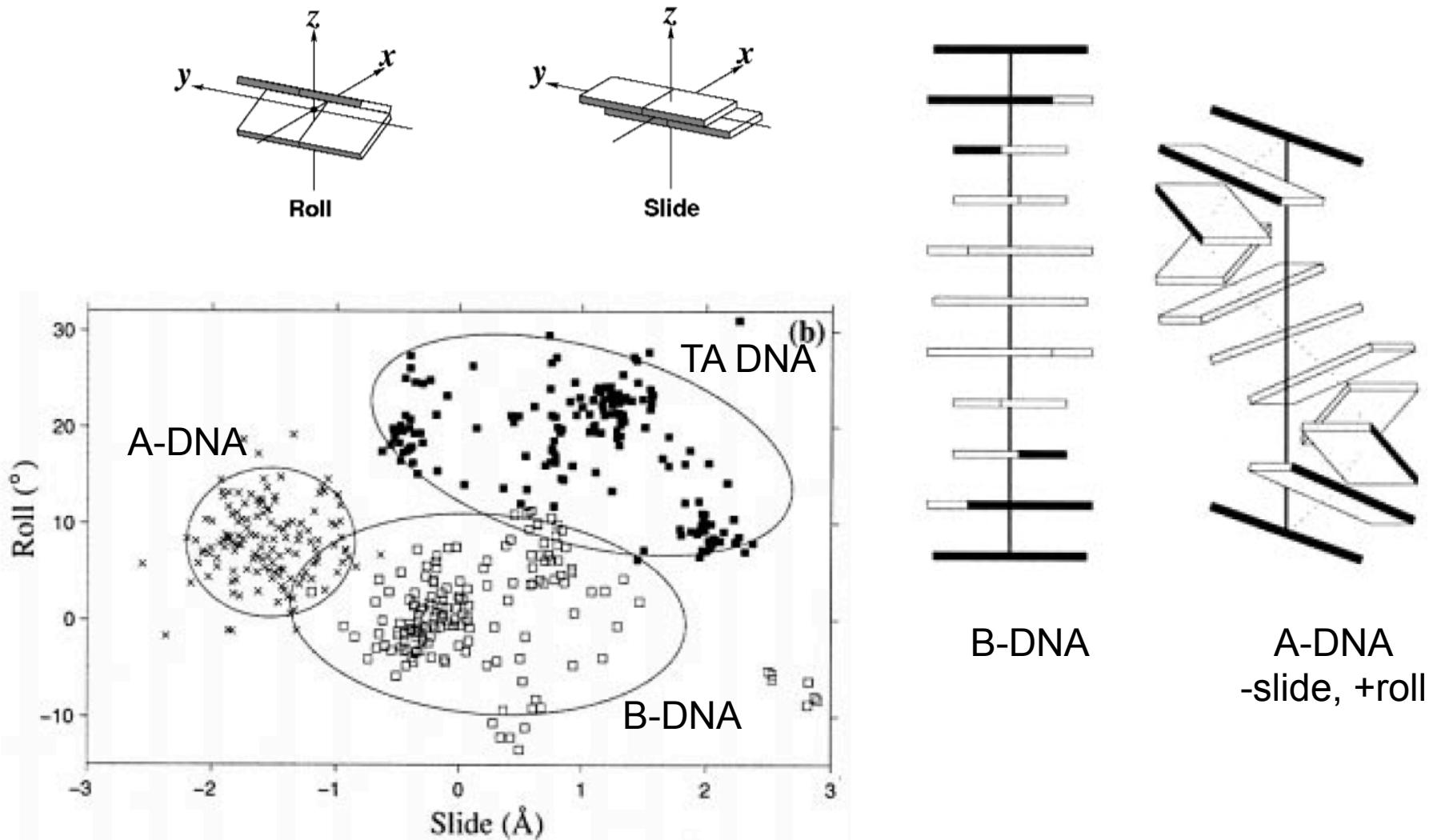
- FREEHELIX, RNA, SCHNAaP, SCHNArP, compDNA,
NUPARM

What's a base-pair in 3DNA? (In misc_3dna.pra)

- The distance between the origins of the two bases (as defined by their standard reference frames) must be less than certain limit (15.0 Å by default) - otherwise, they would be too far away to be called a pair.
- The vertical separation (i.e., stagger) between the two bases must be less than certain limit (2.5 Å by default) - otherwise, they would be stacking instead of pairing.
- The angle between the two base z-axes (i.e., their normal vectors) is less than a cut-off (65.0° by default).
- There is at least one pair of nitrogen/oxygen base atoms that are within a H-bonding cut off distance (4.0 Å by default).

Please Show Us What They Look Like.

Step-Parameters Make Analysis Easier



Base-Pair Databases

- <http://bps.rutgers.edu> for RNA.
- <http://rnasteps.rutgers.edu> for RNA Steps.
- <http://3dnascapes.rutgers.edu> for DNA.

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)

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!

Overview

PART II (Rigid Block Model for Nucleic Acids)

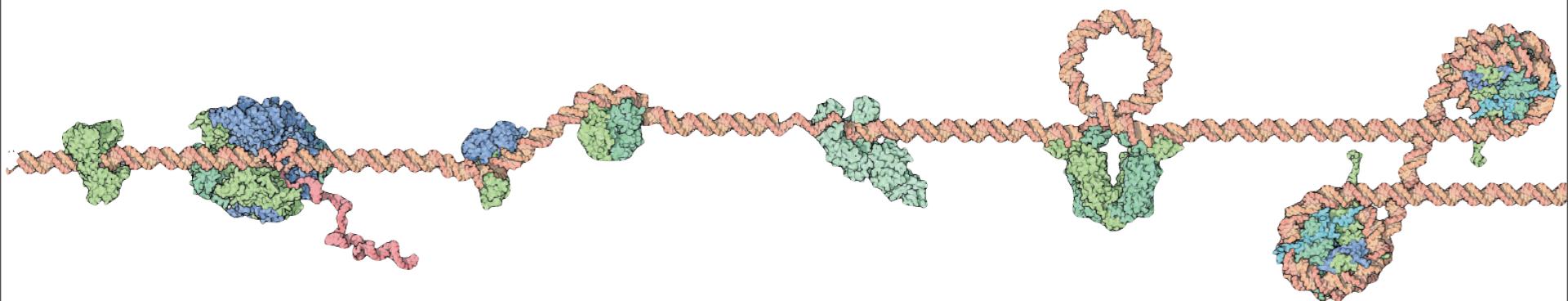
- NDB/PDB Color Conventions.
- 1991 Tsukuba - Standard Reference Frame.
- Calladine-Drew Rigid Block Model.
- Software (3DNA(ansi-C), Curves+(fortran))

Overview

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 B-DNA, A-RNA, Z-DNA.
 - Chromosomal DNA. (download luger's - get rid of protein - analyze - rebuild.)
 - Download the ribosome of s. cerevisiae. latest at 3.0
 - (What do the conformation of helices look like in ribosome?)

Not Only RNA Pol II Reads DNA.



Sidenote on Ribosomal RNA “News”

There used to be three high resolution x-ray structures of the ribosome up to 2011. Now there are four and the new one is of an Eukaryote. So now the high resolution ribosomal structures are:

Thermus Thermophilus (1hnz - 3.30) Rama.

Haloharcula Marismortui (1jj2 - 2.40) Steitz.

Deinococcus Radiodurans (2zjr - 2.91) Ada Yonath.

Saccharomyces Cerevisiae.(3u5d - 3.00) - Yusupov and Yusupova