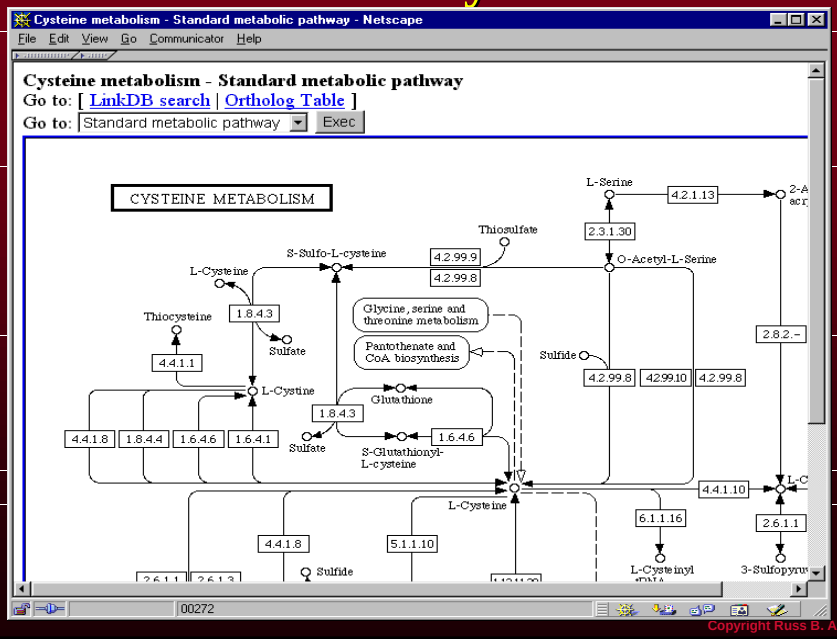
Genetics Networks (revisited)

Russ B. Altman
MIS 214/CS 274

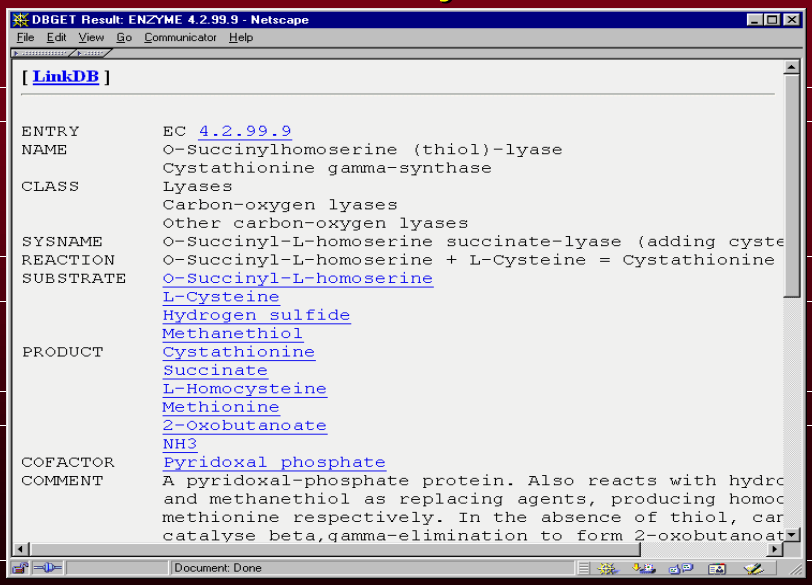
What is a genetic network?

- Individual genes have a *function* (e.g. transforming a substance or binding to a substance)
- Sets of functions when sequenced can produce *pathways* (e.g. output of one transformation is the input to another)
- Sets of pathways, as they interact with other pathways, create a *genetic network* of interactions.
- The emergent properties of these networks constitute the "observables" when we study cells.

KEGG Pathway Database

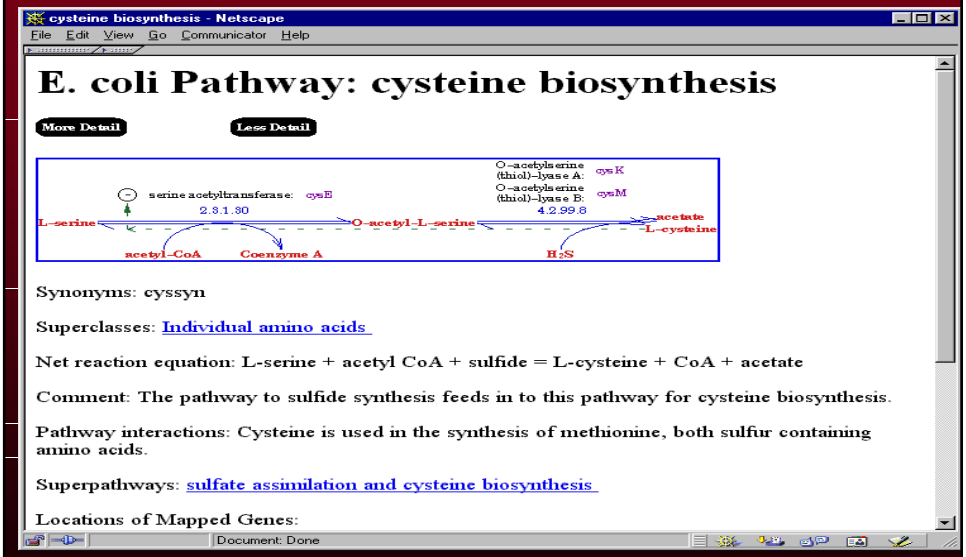


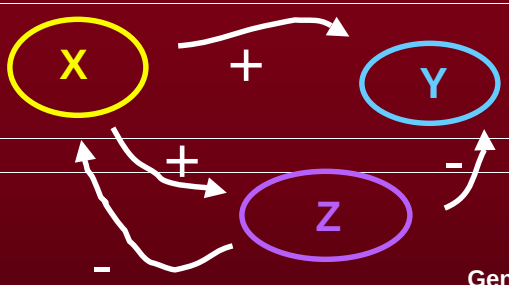
KEGG Pathway Database



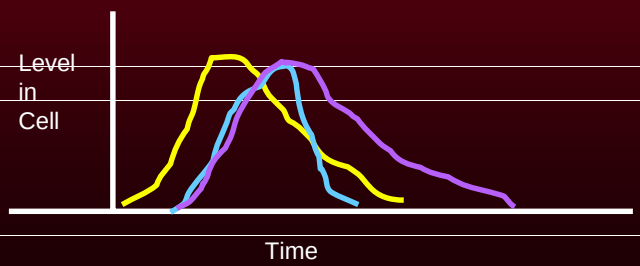
EcoCYC Metaboli Database

http://ecocyc.pangeasystems.com:1555/server.html





Gene Functions: X, Y, Z
~~Pathways: X-Y, X-Z-Y, X-Z-X~~
Networks: XYZ



Studying Gene Networks

Until recently, difficult because

(1) Data about interactions and timing of expression was very difficult to collect.
(2) No good methods to analyze these networks

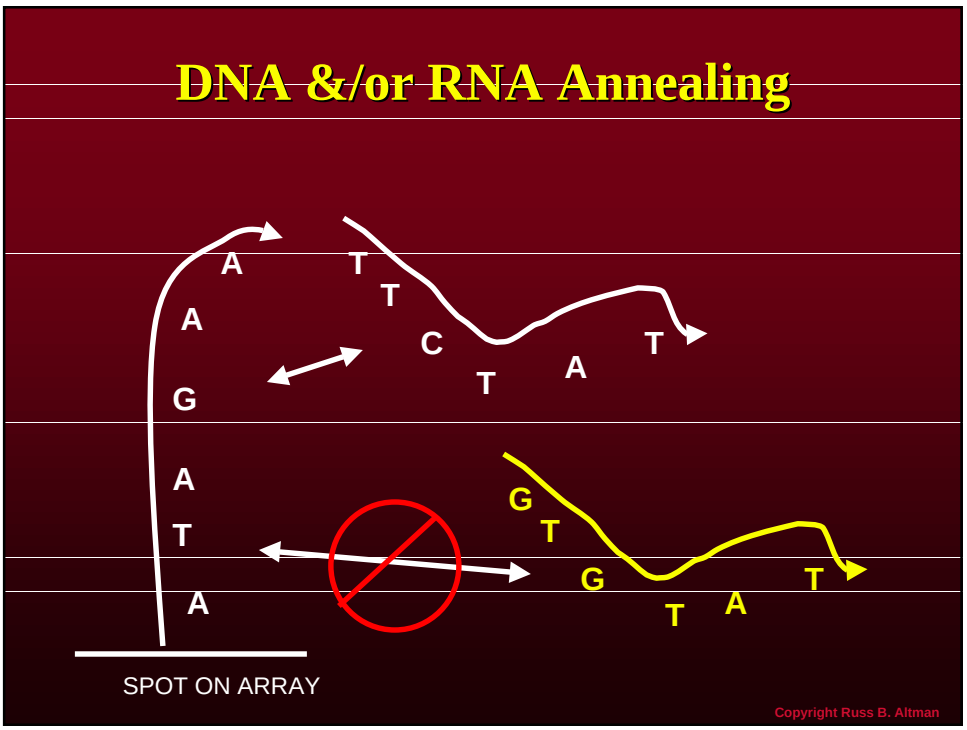
But now:

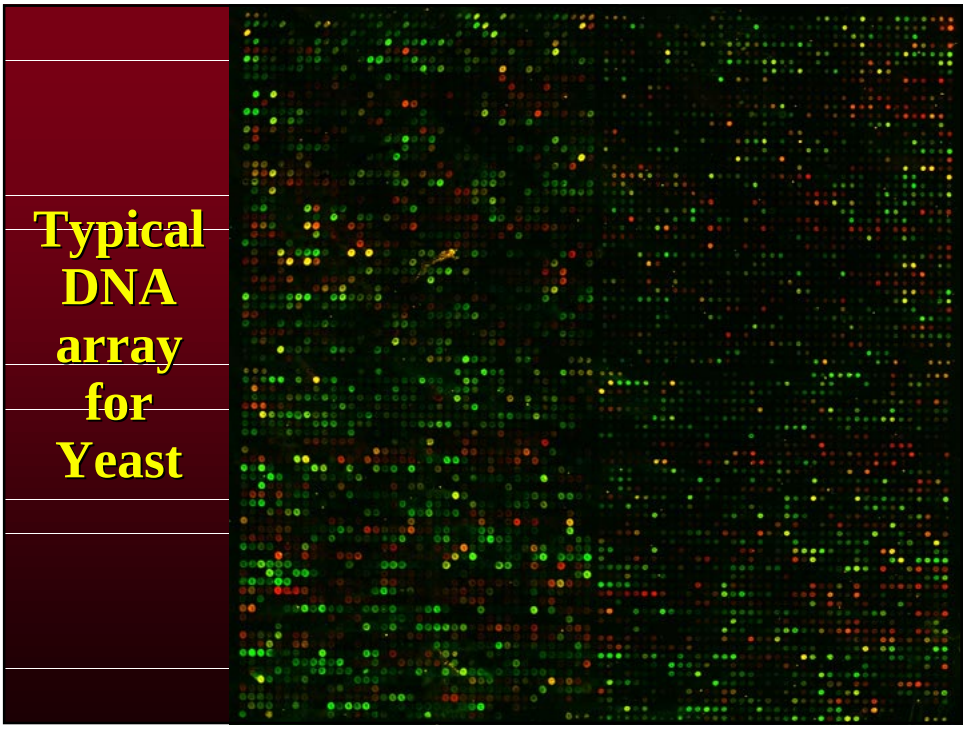
(1) Methods for measuring the "network" are available.
(2) Progress made in thinking about computing on these network.

Availability of Relevant Data

Gene Arrays

(1) Fix pieces of DNA of known sequence on a 2-dimensional array (e.g. all 6000 yeast genes)
(2) Gather samples from cells under two conditions (e.g. starved vs. not starved)
(3) Label the mRNA of the two samples with fluorescent dyes of different colors (green/red)
(4) Mix the samples and pour on the 2D array. mRNA that is complementary (in Watson-Crick sense) will anneal to pieces of DNA.
(5) Measure fluorescence to see if levels of mRNA in the one sample is same, more, less than the other.


DNA &/or RNA Annealing
A
T
T
A
C
T
A
T
G
A
T
A
G
T
G
T
A
T
SPOT ON ARRAY
Copyright Russ B. Altman


Typical DNA array for Yeast

What does this tell you?

mRNA is a proxy for how much protein is available in the cell.

DNA array gives a snapshot of the level of mRNA expression (and thus protein) in the cell at a particular time.

Can take multiple snapshots to watch the evolution of availability of mRNA over time, or in response to stimuli.

For example: in response to starvation, evolution of cancer, normal vs. disease, etc...

What to do with array data?

Some obvious choices…

1. Cluster genes based on similar gene expression = a new metric for "similarity" separate from sequence measures.

2. Try to infer genetic interactions based on timing and co-regulation of genes.

Cluster genes based on patterns of expression.

Cluster analysis = unsupervised learning

1. **Define a distance metric between two gene expression patterns (e.g. Euclidean distance = SQRT ((x1-x2)**2 + (y1-y2)**2 …))**

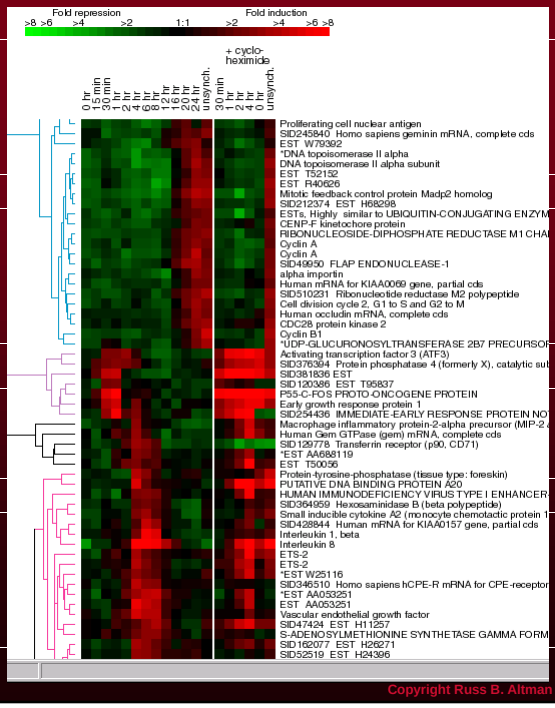
EITHER:

2. **Create a distance matrix of all pairwise distances, and then find closest pairs, and combine them (like phylogenetic trees). O(N**2)**

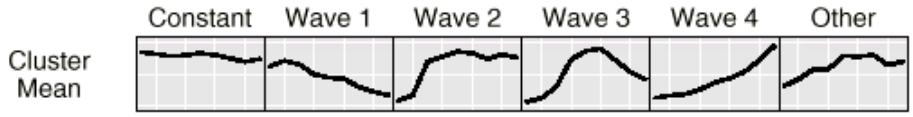
OR

2. **Randomly generate "seed" points of the same dimensionality of the data, and map data points to closest seed. O(N).**

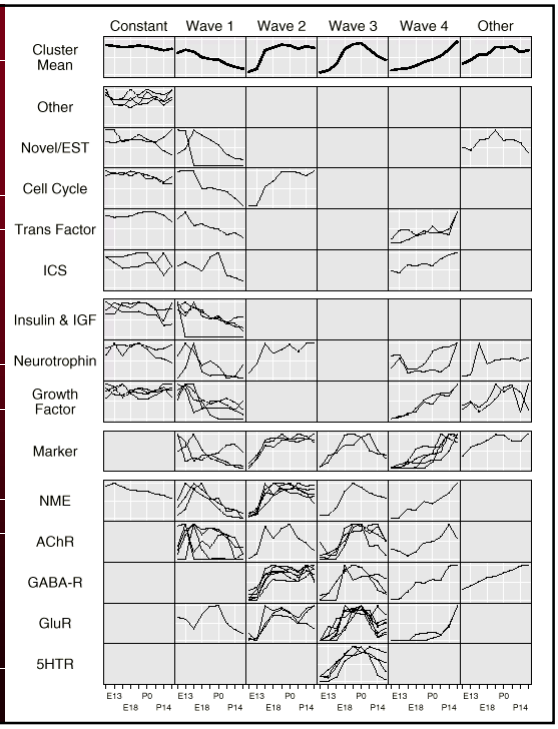
Can build trees from cluster analysis, groups genes by common patterns of expression.

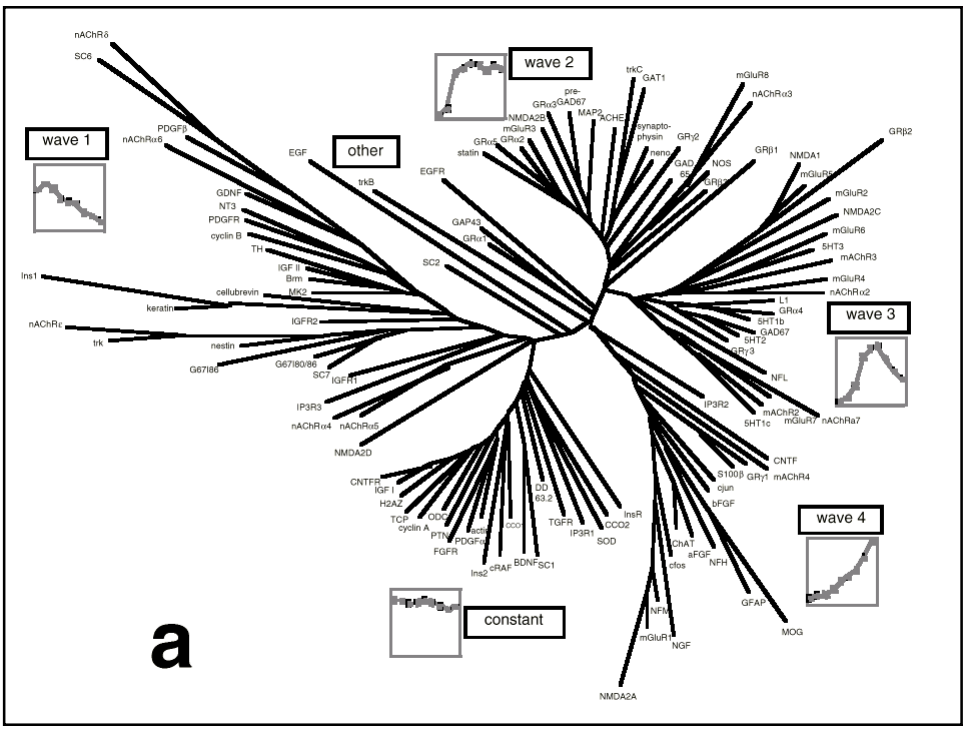


Average of clustered wave forms



Typical "wave forms" observed (note: not lots of bumps)





Imagine other array technologies

Protein chips to assess interaction of proteins (lay down proteins, and then label others, and look for binding events).

Reconstructing Genetic Network

Hard problem.

Given N genes, there are an exponential number of connections between the genes.

Relationships are not generally +/- but are continuous valued (e.g. concentration of molecule varies smoothely.

Must use knowledge about expected function and membership in pathways to prune the list of possible network interactions.

Simplification: Boolean Network

1. All genes are either in on state or off state.

2. State of a gene at time T determined by the logical combination of states of regulatory genes at time T-1.

3. Can propogate the states over time to simulate the evolution of the network.

Very simplified, but still useful for seeing the kinds of properties that can emerge from these networks.

Sample network

A’ = B
B’ = A or C
C’ = (A and B) or (B and C) or (A and C)
equivalent to:

INPUT			OUTPUT		
A	B	C	A’	B’	C’
0	0	0	0	0	0
0	0	1	0	1	0
0	1	0	1	0	0
0	1	1	1	1	1
1	0	0	0	1	0
1	0	1	0	1	1
1	1	0	1	1	1
1	1	1	1	1	1

Some sample state transitions

1 0 0 - 0 1 0 - 1 0 0 - 0 1 0 - 1 0 0 …

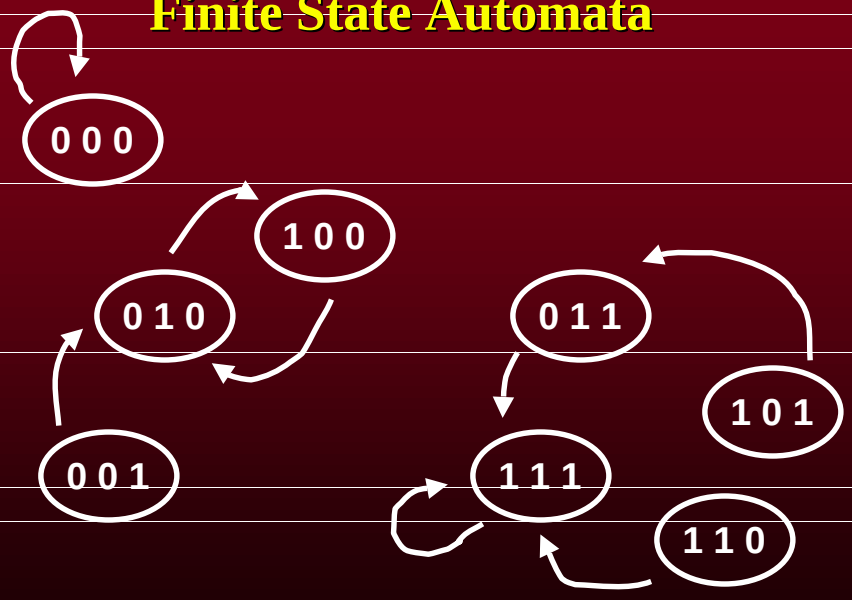
1 0 1 - 0 1 1 - 1 1 1 - 1 1 1 - 1 1 1 …

0 0 0 - 0 0 0 - 0 0 0 …

0 0 1 - 0 1 0 - 1 0 0 - 0 1 0 - 1 0 0 …

0 1 0 - 1 0 0 - 0 1 0 - 1 0 0 - 0 1 0 …

Finite State Automata



Things to notice

Attractors are can be *static* (only one node) or *dynamic* (sequence of nodes that repeat).

Different equilibriums are possible, even for very simple network. These can correspond to:

disease state
resting state
perturbed state

Transitions from one attractor to another require external events (starvation, food supply, heat, etc…) to create mutation in state, and move to different attractor.

Correspondences

Genetic Network	Boolean Network
Genotype/DNA	Wiring and rules
Gene	Element of state
Expression pattern	State
Development	Trajectory
Mature cell	Attractor

Old Premise (from Somogyi et al)

~~Gene for every function, function for every gene.~~

Complete reduction of organism into genes

~~Determination of protein structures/activities~~

Mapping molecular gene product interactions

Assembly of database of molecular mechanisms

Synthesis is sum-of-parts computer model.

New Premise (from Somogyi et al)

Gene function distributed across parallel network
Identify genes and genetic network elements
Determine states of network (expression patterns)
Map out alternative trajectories/attractors

Reverse engineer the network
- parallel trajectories suggest shared input
- temporal links determined by wave shapes

RNA Structure Computations

Russ B. Altman
MIS 214/ CS 274

Biological roles RNA

0. Like DNA, RNA contains 4 subunits (AUGC). It is less stable than DNA, so is not a storage media.

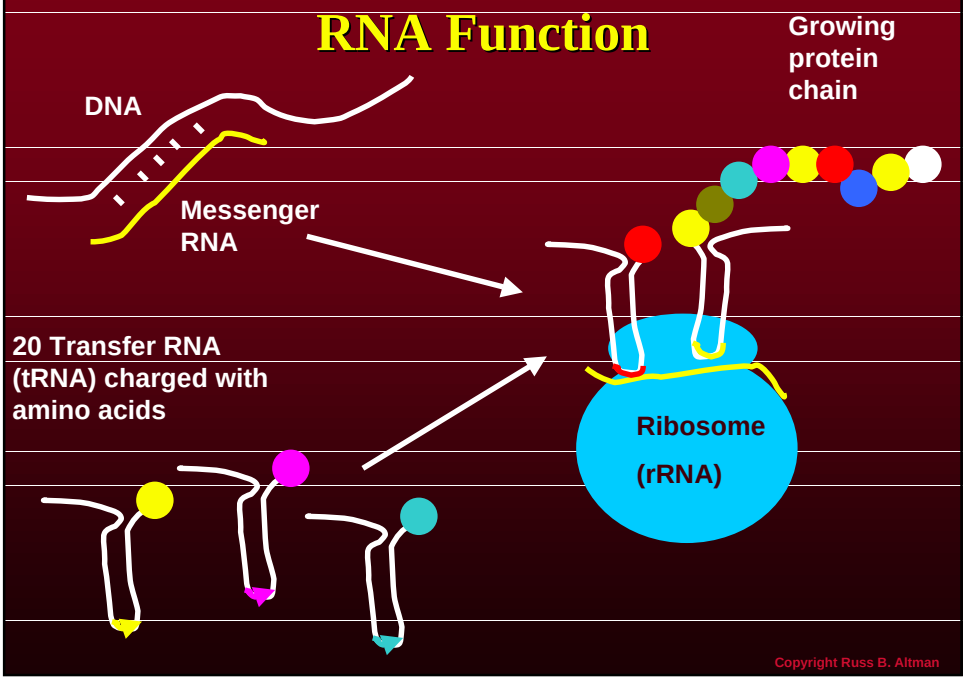
1. the DNA code a gene is copied into messenger RNA (mRNA)

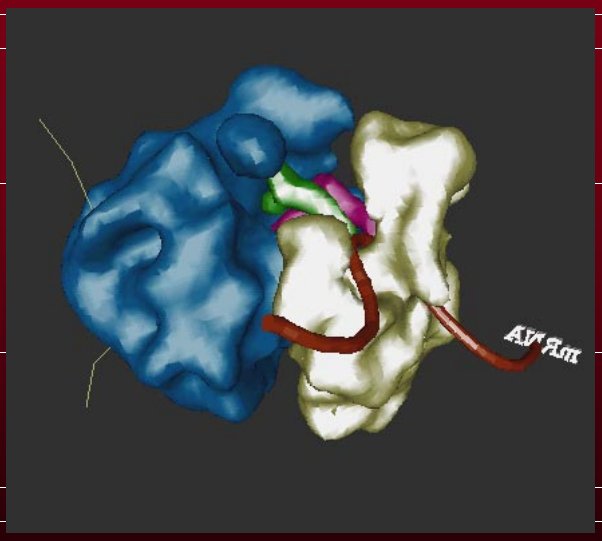
2. mRNA is the version of the genetic code translated at the ribosome.

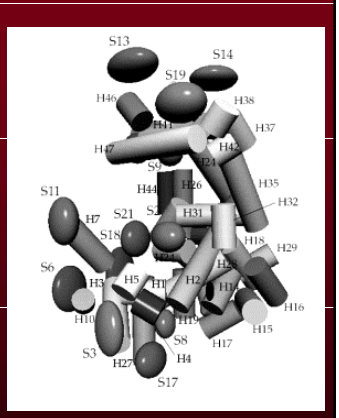
3. the ribosome is made up RNA (ribosomal RNA or rRNA)

4. The individual amino acids are brought to the ribosome, as it reads the mRNA by molecules called transfer RNAs (tRNA)

RNA Function

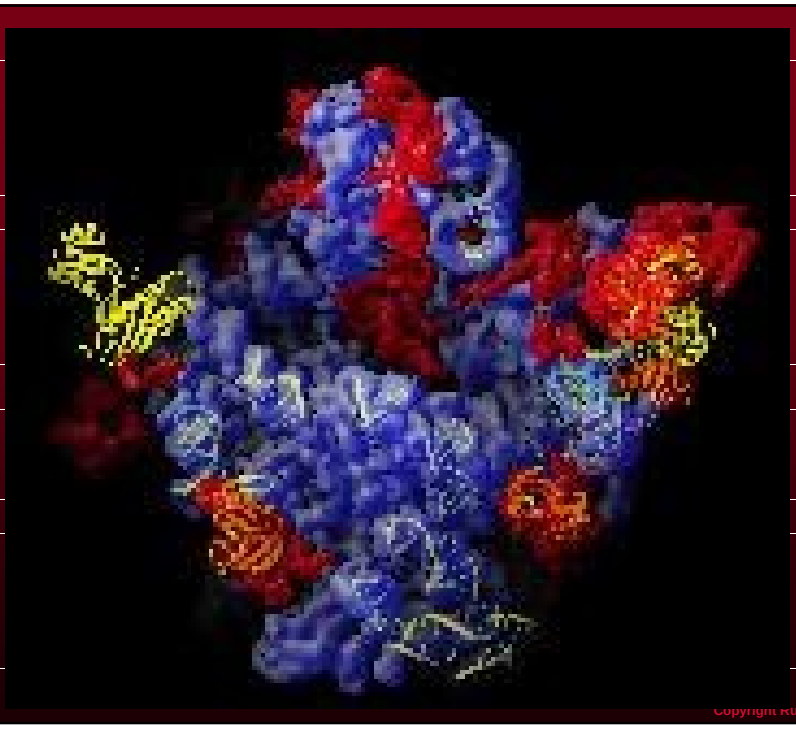






http://www.wadsworth.org/BMS/

http://www-smi.stanford.edu/projects/helix/ribo.html



Insert tRNA pix here.

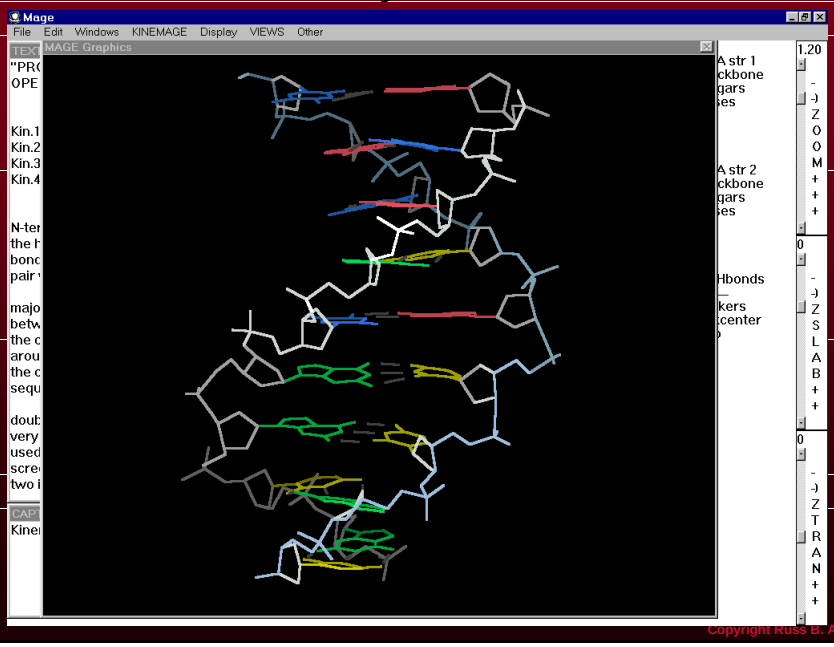
RNA has 3D structure

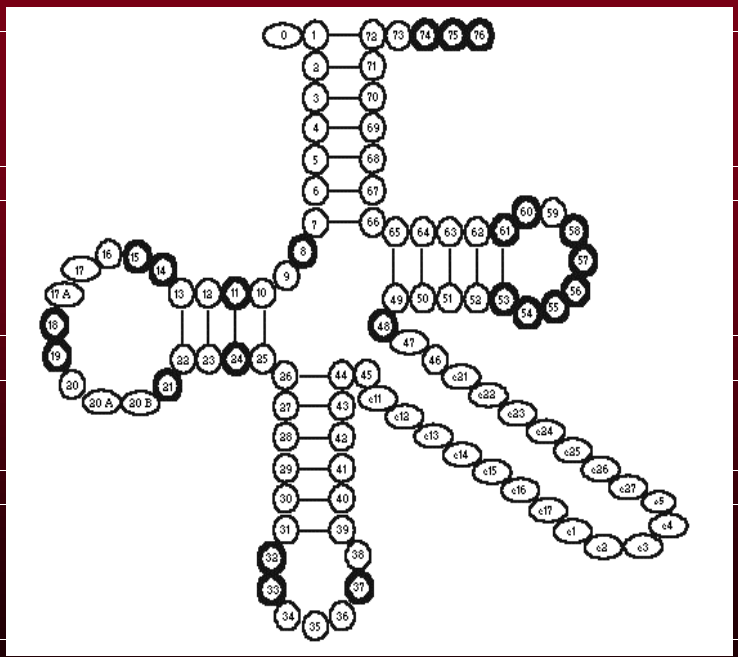
AUUCGGCGACGAAU

Primary Structure

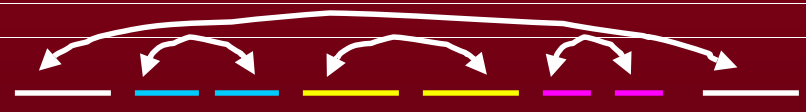
Secondary Structure

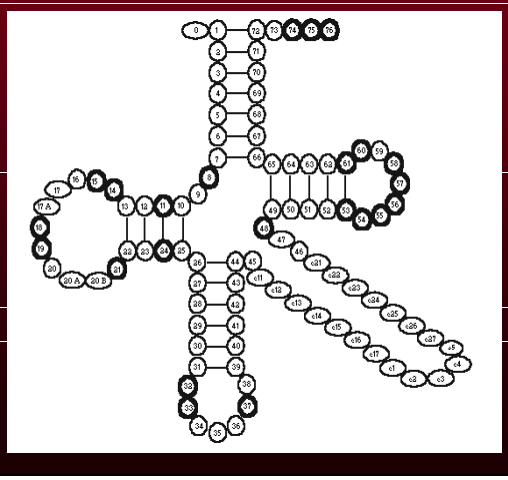
Tertiary Structure





RNA folding is usually nested





RNA can have pseudoknots

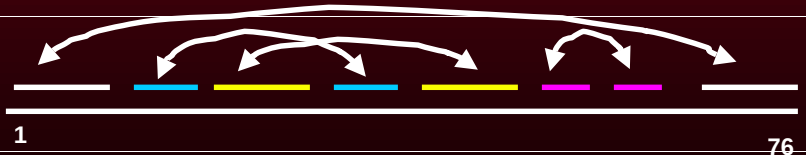
Normal base pairing (for base pairs i-j and i'-j'):

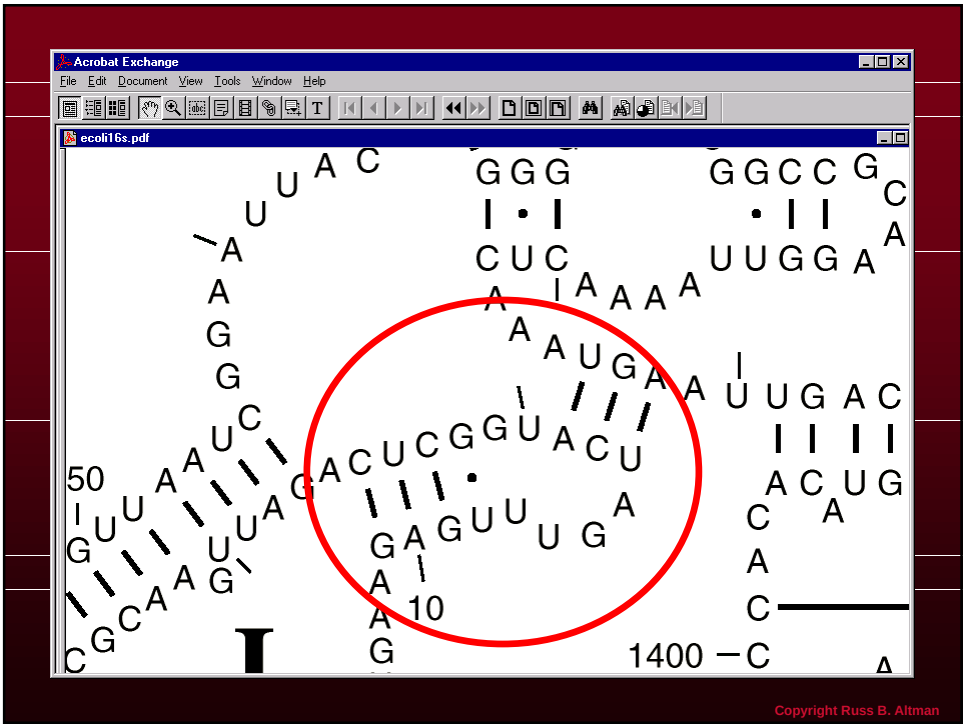
i < j < i' < j' or i < i' < j' < j

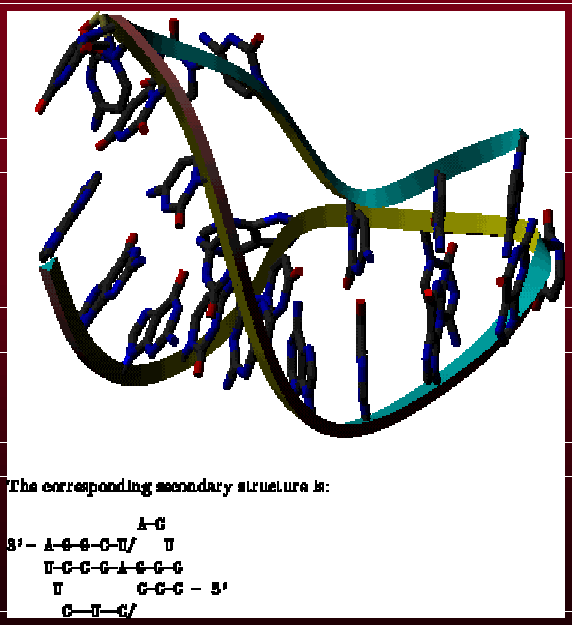
Pseudoknots

i < i' < j < j'

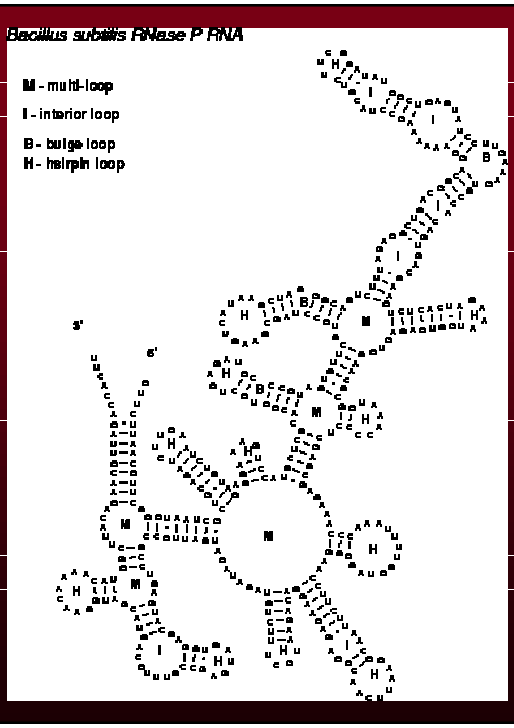
Yellow and blue helices form a pseudoknot.







The corresponding secondary structure is:



Show sample RNA folds

Show sample RNA structures

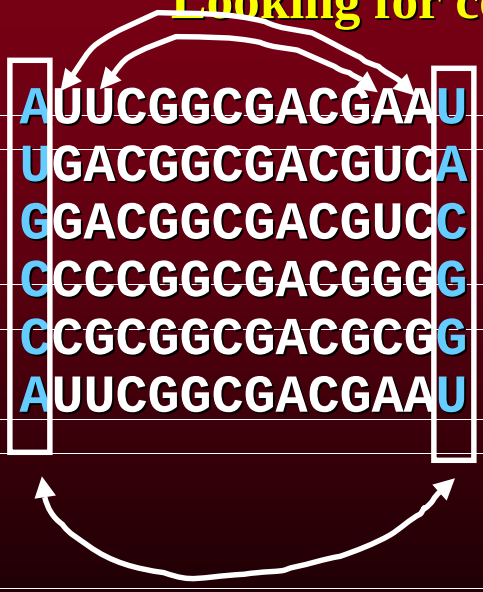
Predicting RNA Secondary Structure

1. Based on phylogenetic comparisons

--look at columns of a multiple alignment
--find WC or nonCanonical relationships that are preserved despite mutations.

2. Based on energetics
-- individual contributions to energy are additive
--assumption that optimal secondary structure is that which has lowest energy.

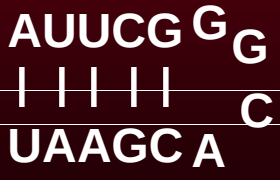
Looking for covariation



The consistent co-variation of the two columns in a Watson-Crick manner indicates that there is some sort of relationship between those two positions in the secondary structure.

Covariation is taken as a gold standard for other prediction methods.

AUUCGGCGACGAAU **Primary Structure**



Secondary Structure

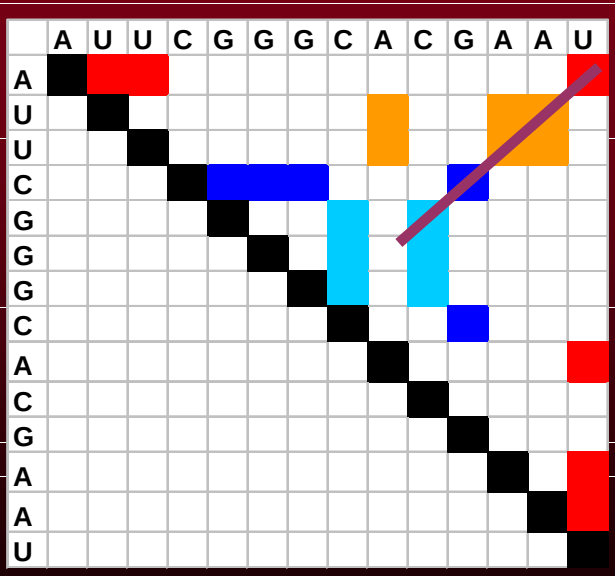
RNA Folding energetics

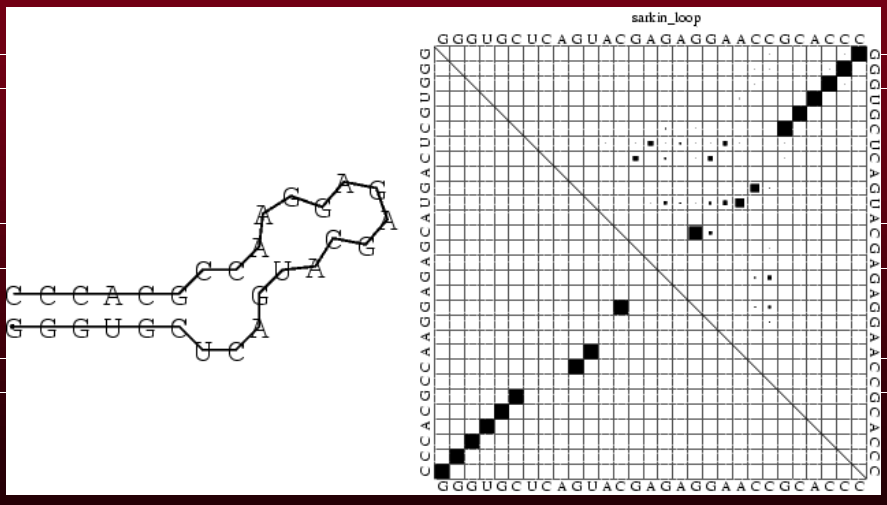
Based on the observation that the stability of an RNA fold can be decomposed into the contributions of individual energies.

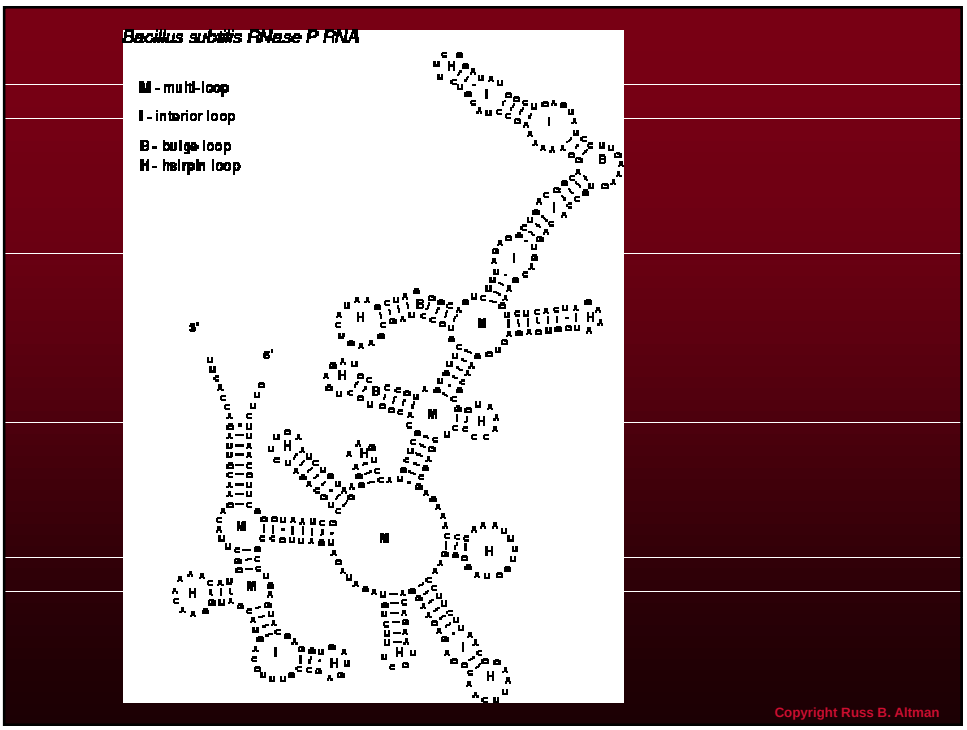
Favorable contributions from:

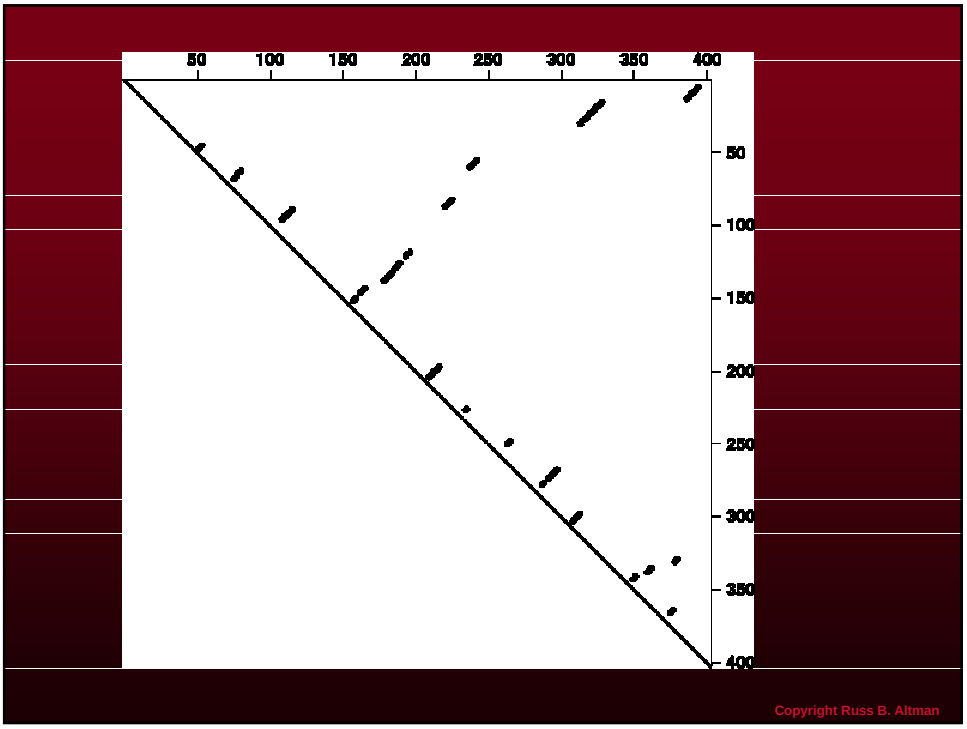
1. **Hydrogen bonds of base-pairs**
2. **Favorable "stacking" interactions of bases (parallel planes of rings)**
3. **Some Ad hoc basepairs created in irregular structures (such as loops of 4 = tetraloops)**

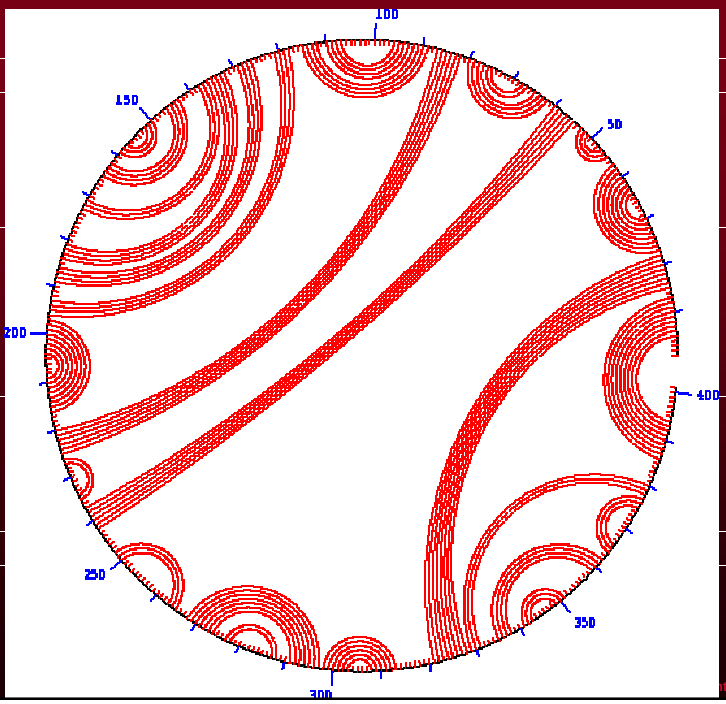
Dot plot approach (look for diagonals)






Bacillus subtilis RNase P RNA
M - multi-loop
I - interior loop
B - bulge loop
H - hairpin loop
5'
3'
Copyright Russ B. Altman


50
100
150
200
250
300
350
400
Copyright Russ B. Altman



Simple energies

G -- C = -3 (most favorable Watson-Crick)
A -- U = -2
G -- U = -1 (note: non-WC = non-canonical)

more elaborate schemes developed by Turner et al.

(http://www.ibc.wustl.edu/~zuker/cgi-bin/efiles.cgi?T=37#STACK)

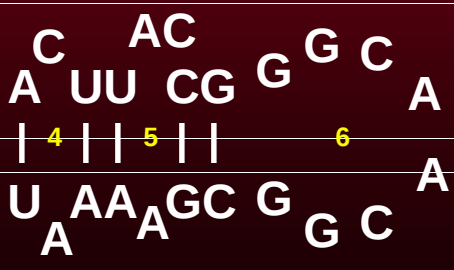
5′ --> 3′
UX
AY
3′ <-- 5′

	A	C	G	U
A	0.4	0.4	0.4	-1.1
C	0.4	0.4	-2.3	0.4
G	0.4	-1.8	0.4	-0.8
U	-0.9	0.4	-1.1	0.4

RNA Folding Energies

Unfavorable contributions from:

4. Symmetric bulges in helices
5. Asymmetric bulges in helices
6. Increasing size of loop at end of helix
7. Multiple branches from a single loop (next slide)



RNA Folding Algorithm

Use dynamic programming approach.

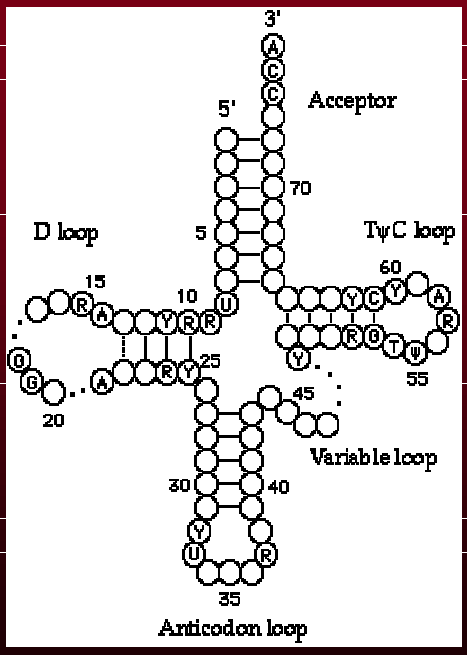
See Michael Zuker's home page for comprehensive description of state-of-the-art algorithms.

http://alice.wustl.edu/~zuker/Bio-5495/RNAfold-html/ = OLD

http://www.ibc.wustl.edu/~zuker/rna/

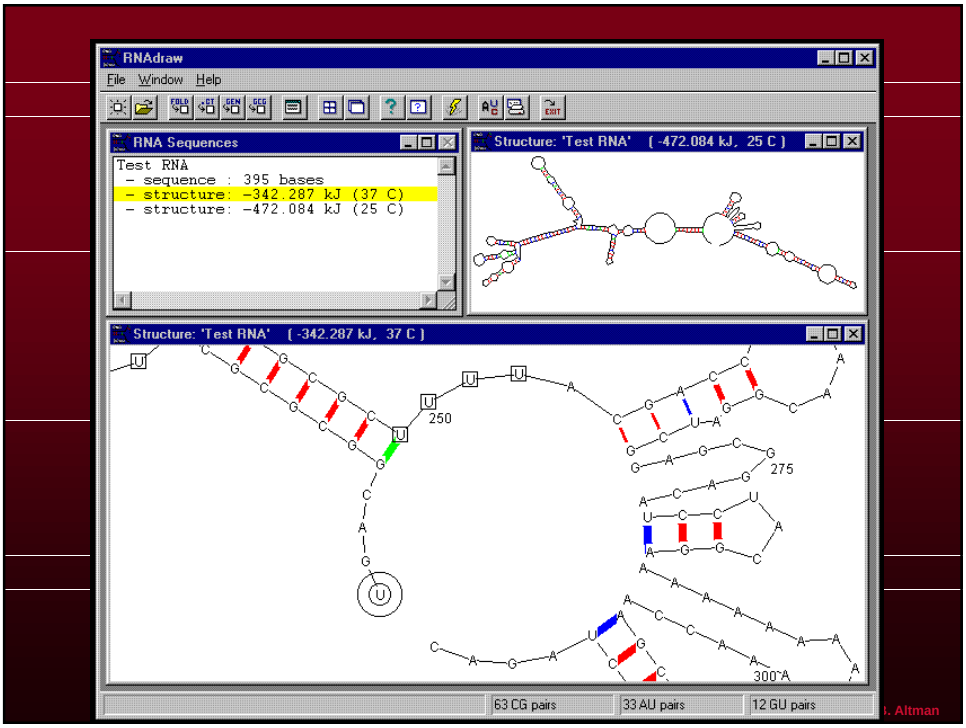
Recursions available at:

http://www.ibc.wustl.edu/~zuker/Bio-5495/RNAfold-html/node3.html



The MFOLD Program

http://www.ibc.wustl.edu/~zuker/rna/form1.shtml


RNAdraw
File Window Help
RNA Sequences
Test RNA
- sequence : 395 bases
- structure: -342.287 kJ (37 C)
- structure: -472.084 kJ (25 C)
Structure: 'Test RNA' (-472.084 kJ, 25 C)
Structure: 'Test RNA' (-342.287 kJ, 37 C)
250
275
300
63 CG pairs
33 AU pairs
12 GU pairs