

Introduction to RNA structure and folding

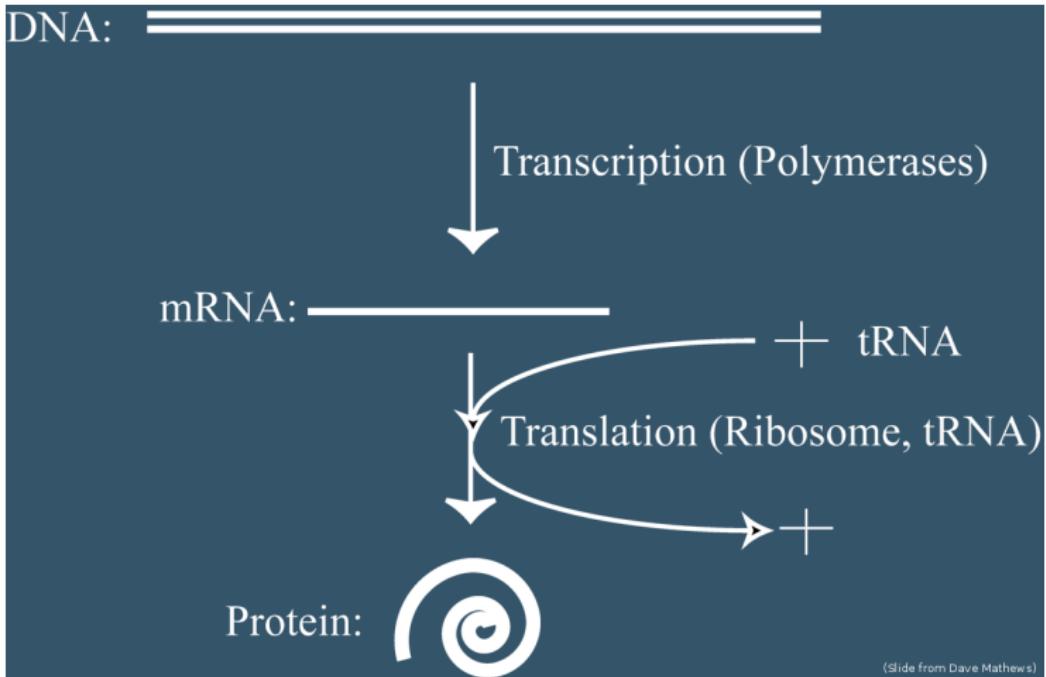
Jan Gorodkin, Center for non-coding RNA in Technology and Health,
University of Copenhagen, Denmark

<http://rth.dk>

Content

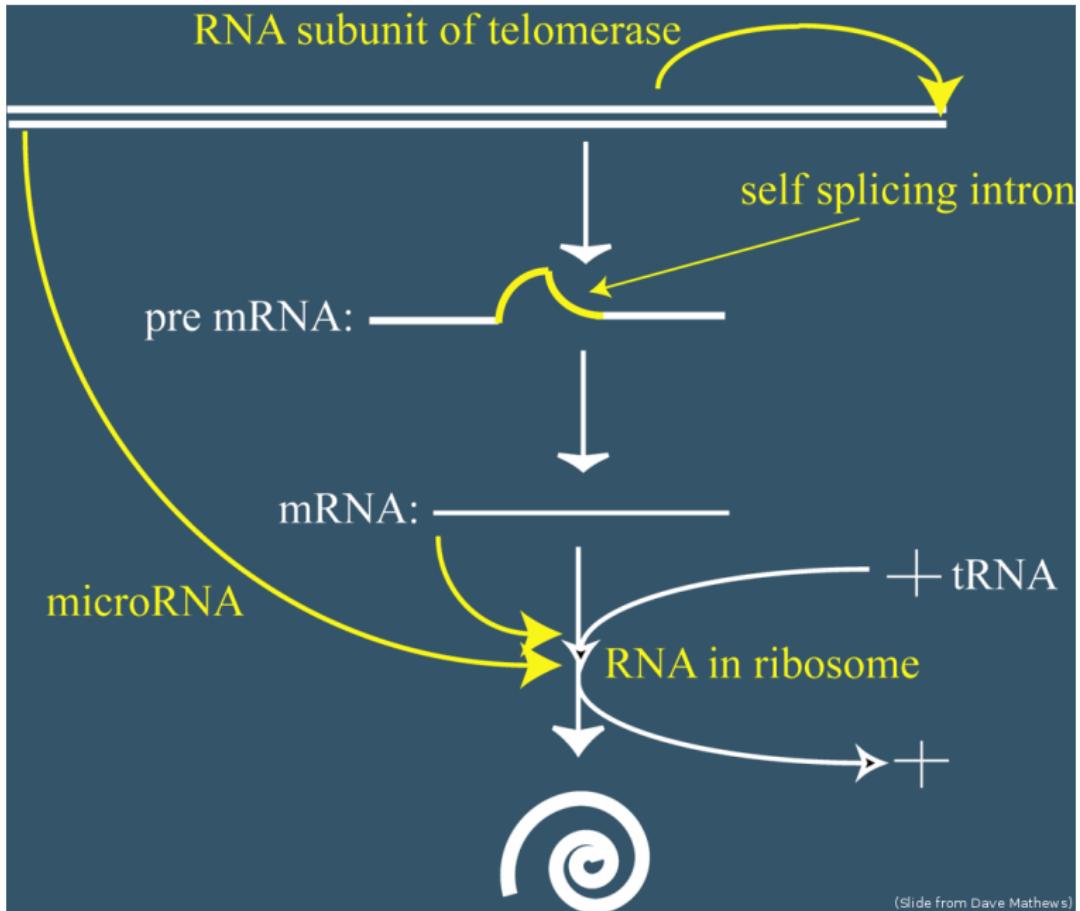
- The world of ncRNAs.
- Concepts of RNA structure.
- Towards the modern RNA world.
- Summary.

Central dogma of Biology



(Slide from Dave Mathews)

RNA is an active player



RNA structure and function; where is it

- Non-coding RNA genes: Functional active non protein coding transcripts.
- 5' and 3' UTRs (UnTranslated Regions) of mRNAs.
- Retroviruses (eg. HIV). General structure (RNA genome):

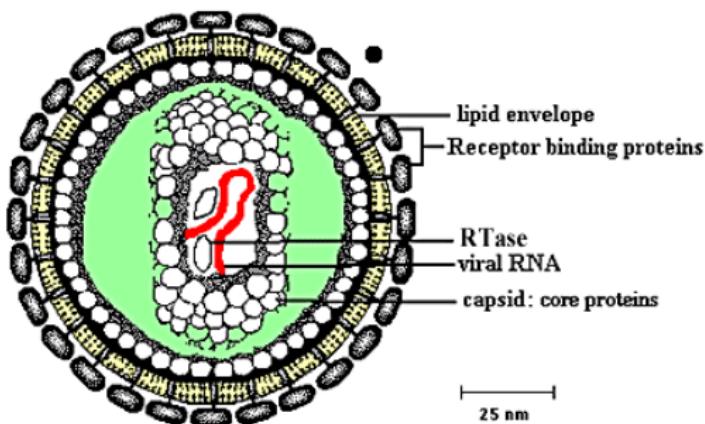


Diagram of a Retrovirus

(From <http://alyhoops3903.tripod.com/sitebuildercontent/sitebuilderpictures/diagram.gif>)

Room for RNA function

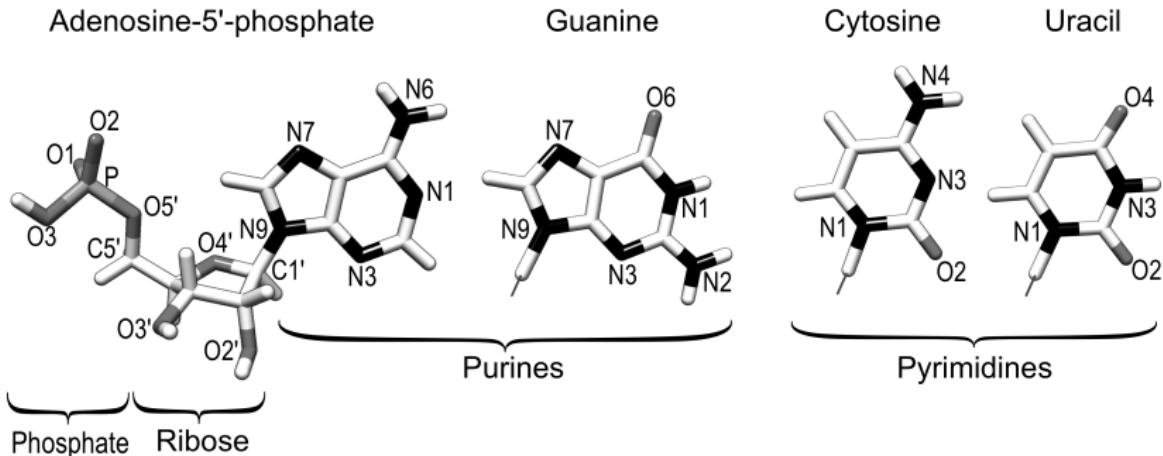
- ~ 1.2% of the human genome encodes protein.
- RNA catalytic and store information (oprise to the "RNA World")
- >80% of human disease-associated loci relates to non-coding DNA.
- Vary dramatic in size from 22nt (microRNA) to 17Kb (Xist RNA).

Concepts of RNA structure

RNA (Ribonucleic acid) is a polymer of ribonucleosides (sugar ribose bound to the bases):

- purines A (adenine); G (guanine)
- pyrimidines C (cytosine); U (uracil)

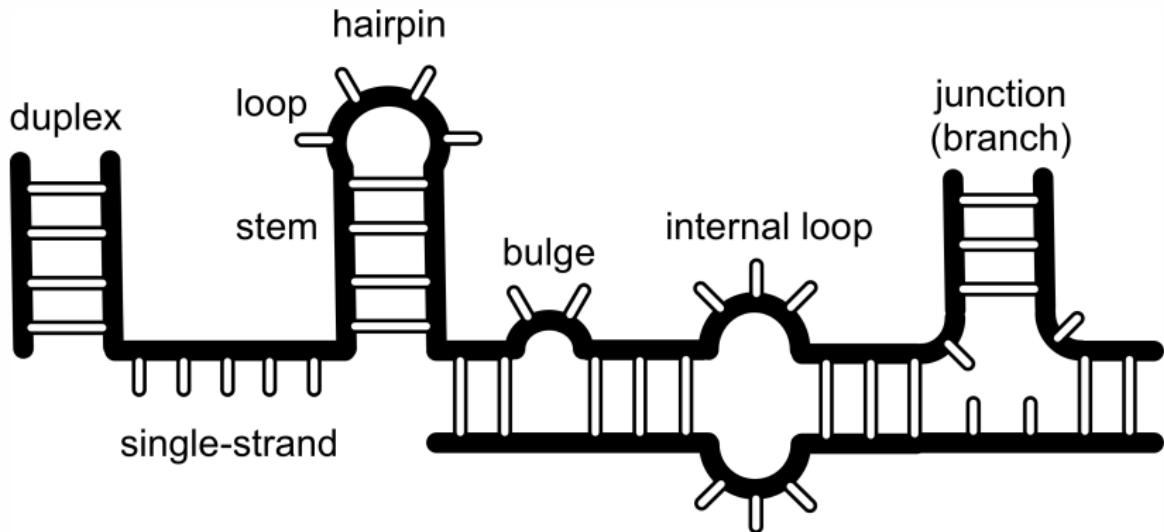
Nucleotide (nucleoside with a phosphate group).



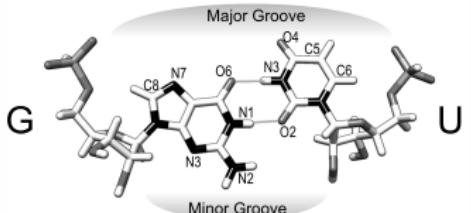
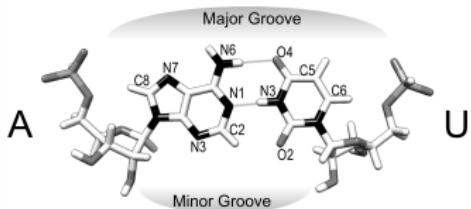
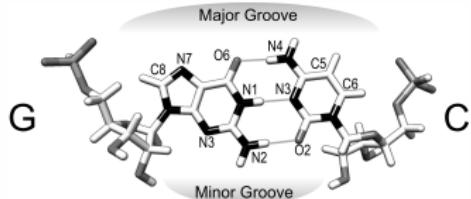
Nucleotides form RNA which can fold into structure.

Concepts of RNA structure

- RNA secondary structure, a base pair representation of the 3D structure.
- Different types of structure elements.
- Long range base pairs from tertiary interactions.



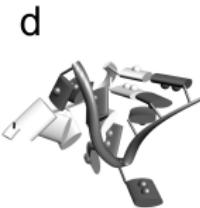
Concepts of RNA structure



- "Regular" base pairs
Watson-Crick: A-U, G-C and G-U Wobble (these are modelled in the standard energy folding algorithm)
- The other edges (sugar and and Hoogsteen) can also form base pairs.

Concepts of RNA structure

Towards tertiary structure



(a) Helix stacking

(b) Kissing hairpin loops

(c) Kink-turn

(d) Pseudoknot

(e) Tetraloop

(f) Tetraloop receptor complex

Concepts of RNA structure

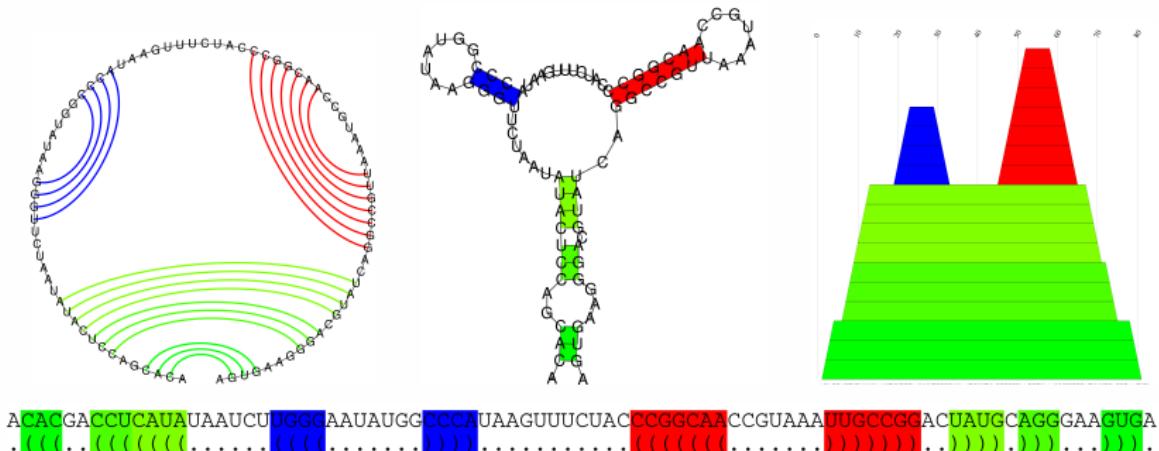
- Small molecules ions contribute significantly to folding of the RNA.
- RNA-targeting drugs The novel appearing RNAs holds great potential for this. (The tmRNA are bound by tmRNP; could be targets for novel antibiotic compounds*.)
- Proteins RNA holds the potential to be bound by many proteins. Structure could be involved.

* C. Zwieb, Principle of RNA structure, In press.

RNA structure representations

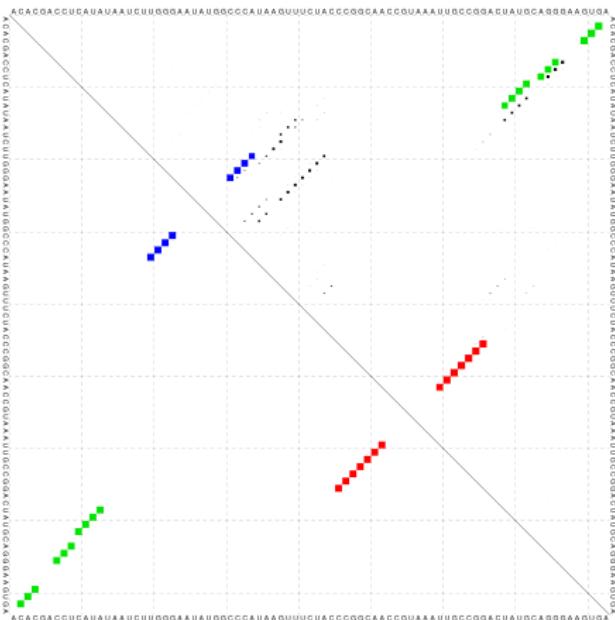
Example for the purine riboswitch, Rfam family RF00167.

- Circle plot (the "arcs" show base pairs).
- Secondary structure diagram.
- Mountain plot.
- Paranthesis representation.



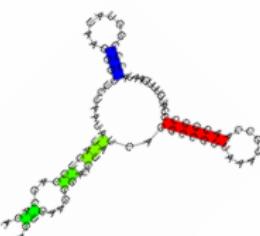
Representing ensembles of structures

The purine riboswitch, Rfam family RF00167.



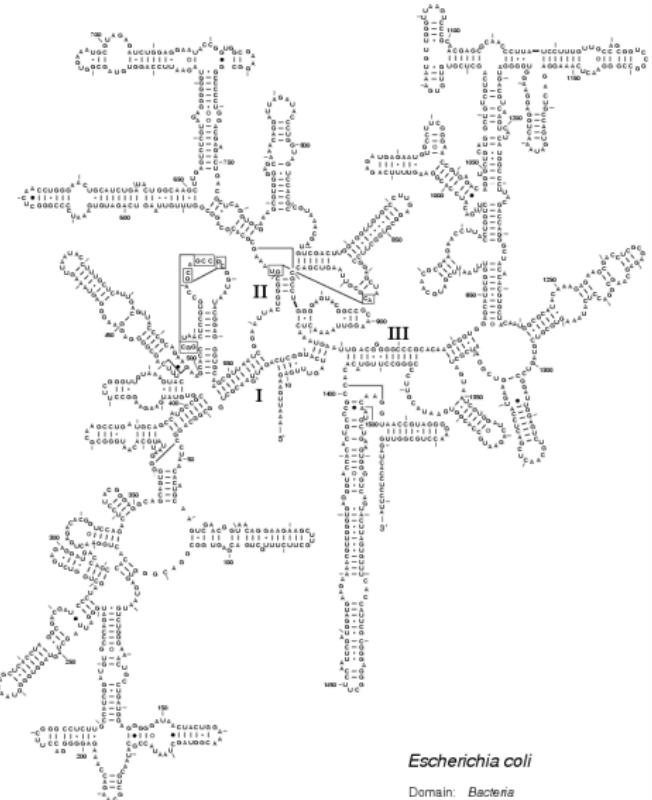
Ensembles of structures (thermodynamic equilibrium) are best represented by base pair probabilities.

A pair (i, j) with probability p is represented by a square in row i and column j with area p .



rRNA

Secondary Structure: small subunit ribosomal RNA



Escherichia coli

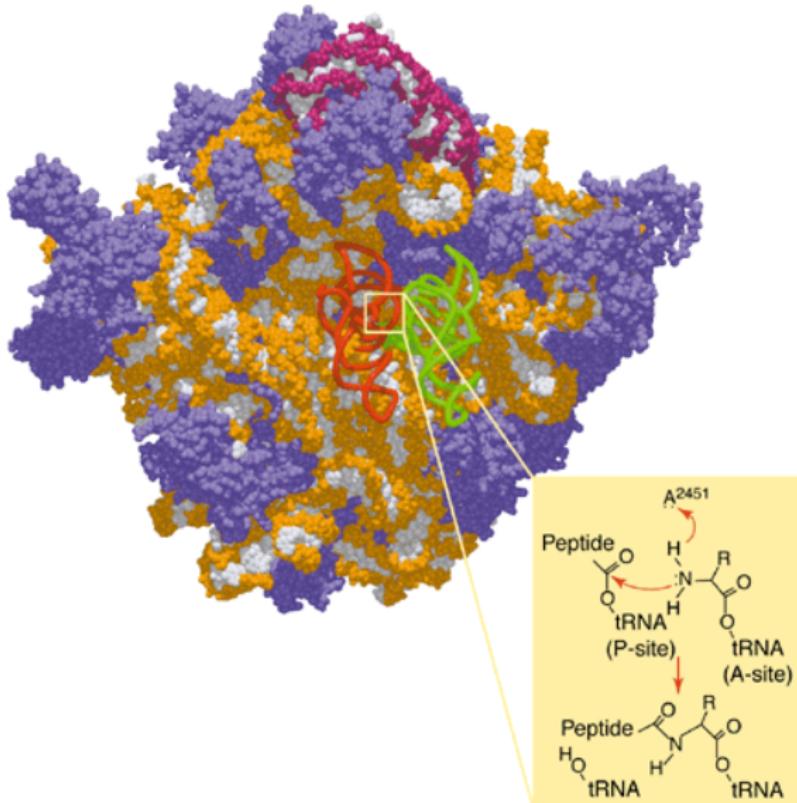
Domain: Bacteria
Kingdom: Purple Bacteria
Order: gamma

July 3, 1995 v4.0

- Responsible for proteins.
- Ribosome is a ribozyme.
- Two subunits (large / small).
- Interacts with tRNA and mRNA.
- Translation can be regulated.

rRNA

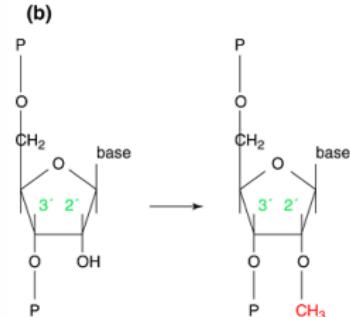
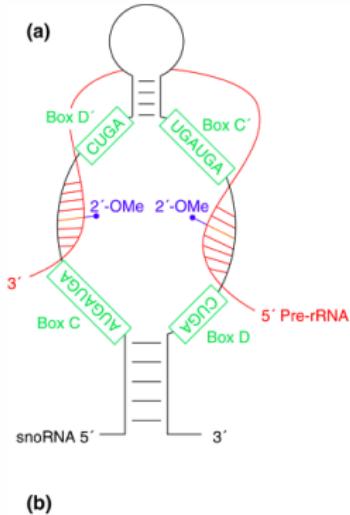
Three-dimensional structure resolved:



snoRNA

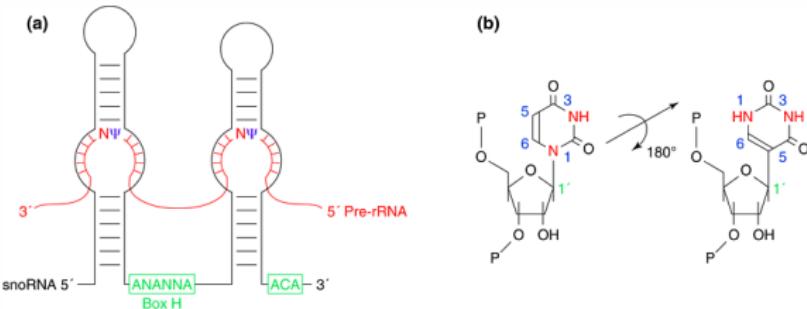
Small nucleolar RNA (snoRNA) 60–300 nts.

- snoRNA genes origin.
- Ribosome synthesis in nucleoli.
- rRNA maturation.
- 2' O-methylation. (box C+D)
- Pseudo-uridine formation (Ψ) (box H+ACA).
- Ribosome conserved. Modifications sites not!
- Site-site specific.
- Associated with proteins snoRNPs.
- Serve as a guide.
- H/ACA also in the Cajal body.



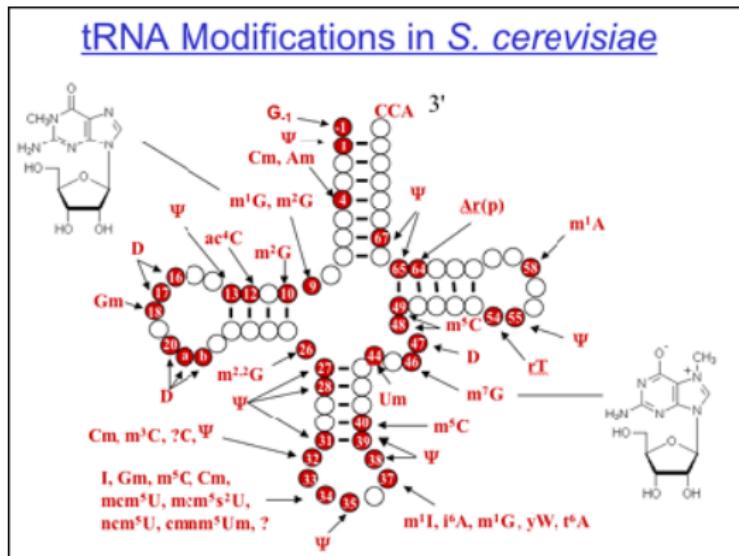
snoRNA

- HMM/SCFG for snoRNA search (Lowe&Eddy).
- Similar type of RNAs found in archaea using computational models.
- 18 experimentally found sequences were used to train a model, which then suggested about 200 candidates in archaea (Omer et al).
- Hüttenhofer (+coworkers): EST screen resulted in 133 new snoRNA, some which showed brain-specific expression.
- Can target other RNAs.
- Most come from introns, some from introns of imprinted ncRNA genes.



Concepts of RNA structure

- pre-tRNA has its 5' end cleaved by RNaseP.
 - Some tRNAs have introns.
 - tRNA is heavily modified.

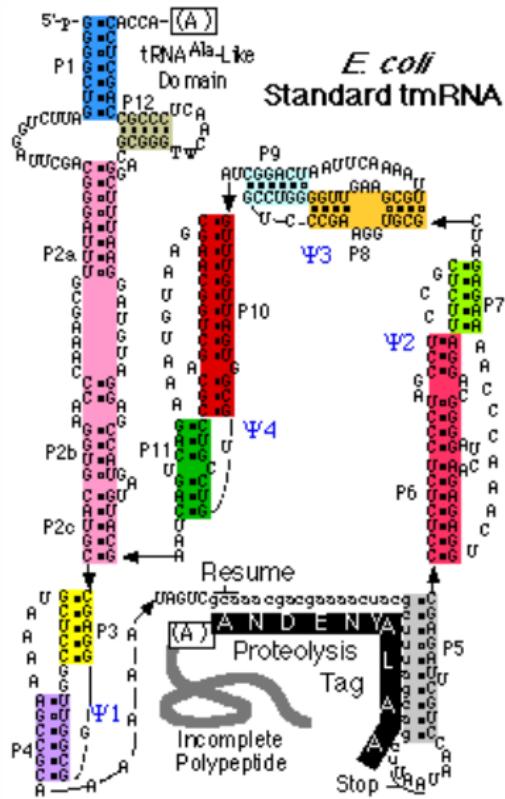


(<http://dbb.urmc.rochester.edu/labs/phizicky/index.html>)

tmRNA

- Previously named 10S RNA.
 - Combined properties of tRNA and mRNA.
 - Found in bacteria and plastid genomes.
 - Function: rescue ribosomes that have reached the 3' end of the mRNA without a stop codon.
 - Structural database:

<http://psyche.uthct.edu/dbs/tmRDB/tmRDB.html>

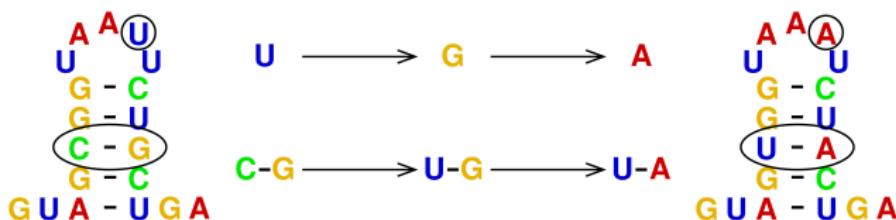


Concepts of RNA structure

RNA structure:



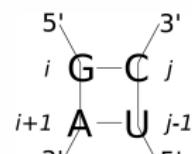
The structure is preserved. Not necessarily the sequence:



Stacked base pairs

- Major source of stabilizing energy
- All 21 combinations measured, accuracy at least 0.1 kcal/mol
- Include the hydrogen bonding energy of pair formation
- Energies of tandem G–U pairs depend on context, i.e. violate the nearest-neighbor model

		(i+1,j-1)					
		CG	GC	GU	UG	AU	UA
(i,j)	CG	-2.4	-3.3	-2.1	-1.4	-2.1	-2.1
	GC	-3.3	-3.4	-2.5	-1.5	-2.2	-2.4
	GU	-2.1	-2.5	1.3	-0.5	-1.4	-1.3
	UG	-1.4	-1.5	-0.5	0.3	-0.6	-1.0
	AU	-2.1	-2.2	-1.4	-0.6	-1.1	-0.9
	UA	-2.1	-2.4	-1.3	-1.0	-0.9	-1.3


-2.2 kcal/mol

ncRNA gene search by folding single sequences

- Search for stable structures by sliding a window of some length and fold it will in general not work.
- Real RNAs are largely impossible to distinguish from random RNAs.*
- For homologous sequences the structure can diverge and structure rather than sequence is the search object.

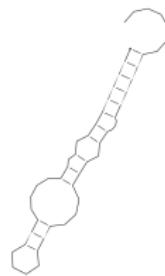
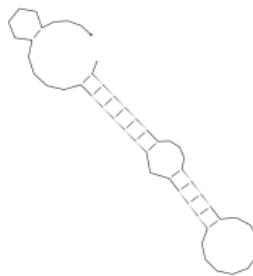
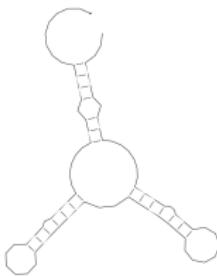
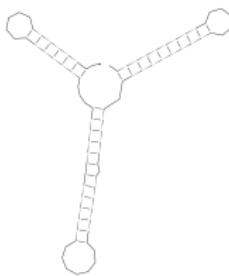
CACGAGCGATCCCATCGCTACTACGCCGACGTACGA



* Rivas & Eddy, Bioinformatics, 2000.

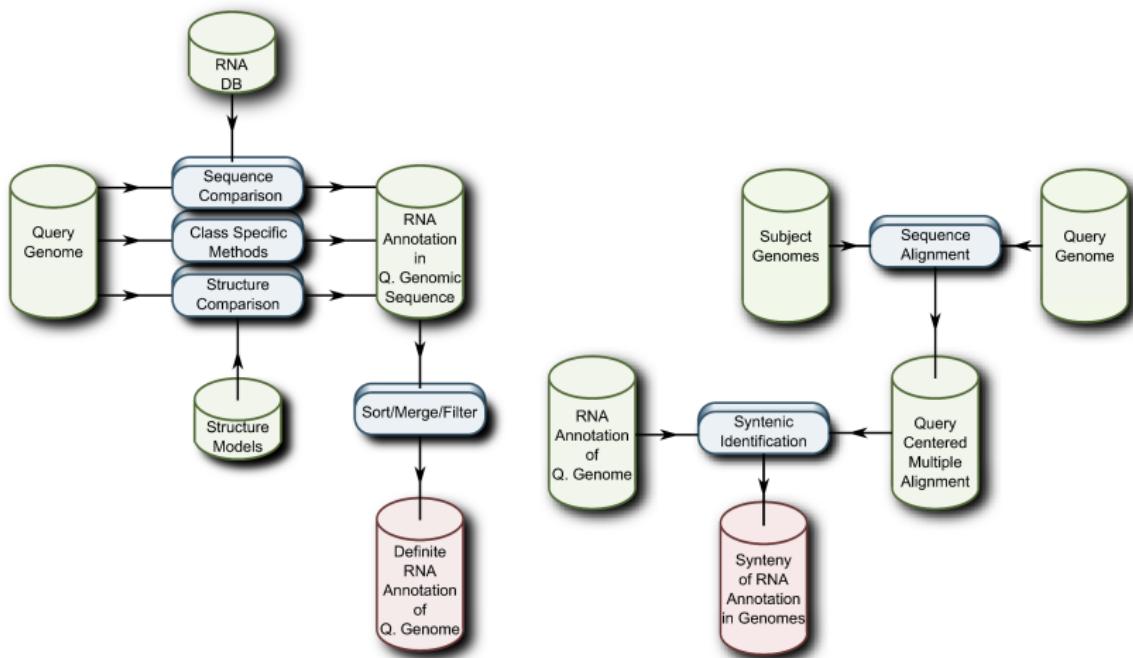
ncRNA gene search by folding single sequences

Predicted RNA structures for two ncRNAs (DsrA and DlcF) and two sequences with random nucleotide content of the same length. Can you tell them apart?



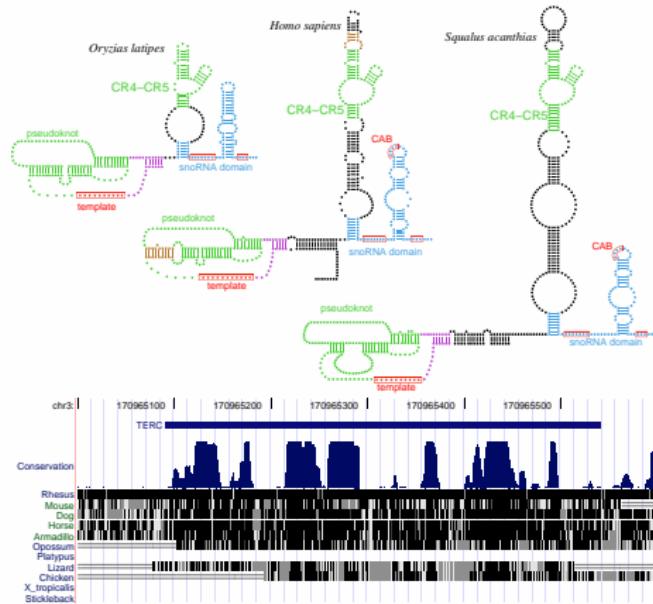
Strategies for RNA structure in homology search

- Homology search: use existing knowledge (database, e.g., Rfam).
- Strategy: model structured RNAs. Infer match of sequence to model.



Non-coding RNAs by homology search?

Major issue due to structural variation of ncRNAs:

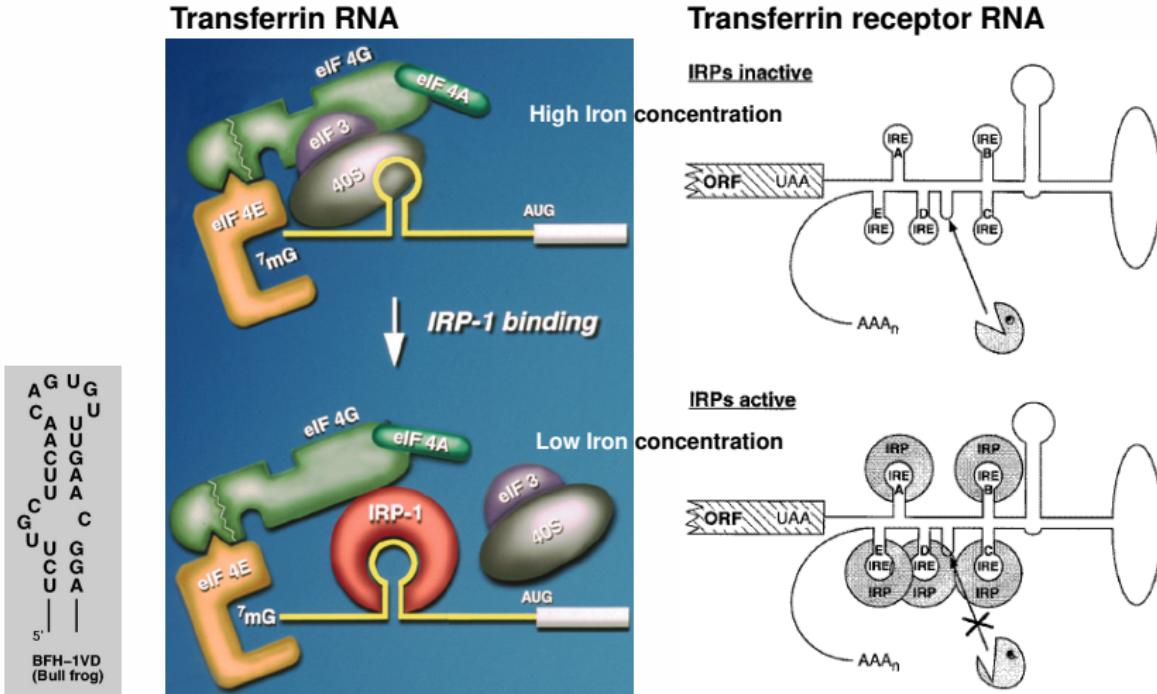


Most recent version of Infernal do find more remote homologs than BLAST*, but the main problem remain: Finding homologous ncRNAs is not a routine task.

* (Menzel, Gorodkin, Stadler, RNA, 2009)

Iron Responsive Elements (IRE)

In vertebrates.



(Muckenthaler *et al.*, 1998; Hentze & Kühn, 1996)

The RNA structure motif search problem

Randomized ferritin 5'UTR regions with an IRE-like element.

```
>seq_shuf AJ251148.1
GCATAA TTTCTTCTGTAAACAAGTCTTACAGTGGCATGTGACCGTTAACGGCTAAAAT
GTTCTCATTAAGGACTTAAATTTCGATTGACTGATTCTCTTACAAATTTCATAAT
GCAGTCACGTAGTTACAACTCCTCAAGGCTGGAAATTCCGTTCAACAGTAAGGCCGTG
ATTTAAAGAAGGTGAATTGGTTCGAGATCTAATTGCTTACATGTCTCACTGTGACAA
GTCTTATTTGAGATATTGTTAA
>seq_shuf M60170.1
CTGGCCCGCACCGCGCTTCGACCGCCCCCGTCCGCCCCCCCCTTCGTCGCCTCTTGCG
CCTCGACGAGTTCCGCTCGAACCGTCCTTCGCAAGAGTTCCGGCAGCCCAGAACCCAC
GTGGATGCCCATTCACGAGTAGTGGGTATTCCGTCGACGCCAGCGCTCGCC
>seq_shuf Y15629.1
ATGAGAAAAGTTCTCAAAGCTGAAAGCAGCTCTCTAGTCTTTGTCGCAATTACGTCCAC
AAGCGCCAATTGTTGAGTGATCTCACACAATTACGAGACACAAAGGCCTATAAAAACCT
TTTTTCGCAAAAAAATGGACTTTGCAAAAATGTATTCTGATGCACGTGCCATCAAGTACG
TCCACTATGTGAAGATCCTCAAAGAGTTGAGCAATGTTCCATTCAACTTTATTAACAA
GCGTCAAAGTTAGCCTCTATATCTTCCGGTCATCGACACGTCAATAGATCGCCTCAATT
GAGCAAAAACGAGAGAAAAAGAGGGCGCAGG
>seq_shuf D15071.1
GCCCGGCCGCGACGCCACCCGACGGCTCCCTCCGGCAGTGCGGCCTGGCCAGTGAGCT
GGGCCGCTCAGGAGCCCCAACTCGCTCGCTCCGCCAGTCGCGGCACAGGCC
```

Can you see a common structural motif?

The RNA structure motif search problem

Randomized ferritin 5'UTR regions with an IRE-like element.

```
>seq shuf AJ251148.1
GCATAATTTCTTCTTGTAAACAAGUCUUAcAGUGGcaugugaCCGUUUUAGGCTAAAAT
GTTCTCATTAAGGACTTAAATTTCGATTTGACTGATTCTCTTACAAATTTCATAAT
GCAGTCACGTAGTTACAATCCTCTCAAGGCTGGAAATTCCGTTCAACAGTAAGGCCGTG
ATTAAAGAAGGTGAATTGGTTCGAGATCTAATTGCTTACATGTCTACTGTGACAA
GTCTTATTTGAGATATGTTAA
>seq shuf M60170.1
CTGGCCCGCACCGCGCTTCGACCGCGCCCCCGTCGCCCGCCCTTCGTCGCCTTTGCG
CCTCGACGAGTTCGCTCGAACCGTCTTCGCAAGAGTTCCCGGCAGCCCAGAACCCAC
GTGGAUGCcCCAUUcacgaguAGUGGGUAUUCGTCCGACGCCAGCGCTCGCC
>seq shuf Y15629.1
ATGAGAAAAGTTCTCAAAGCTGAAAGCAGCTCTCTTAGTCTTTGTCGCAATTACGTCCAC
AAGCGCCAATTGTTGAGTGATCTCACACAATTACGAGACACAAAGGCCTATAAAAACCT
TTTTCGCAAAAAATGGACTTTGCAAAATGUAUUcUGAUUGcaccqugcCAUCAAGUACG
TCCACTATGTGAAGATCCTCAAAGAGTTGAGCAAATGTTCCATTCAACTTTATTAAACA
GCGTCAAAGTTAGCCTCTATATCTTCCGGTCATCGACACGTCAATAGATCGCCTCAATT
GAGCAAAACGAGAGAAAAAGAGGGCGCAGG
>seq shuf D15071.1
GCCCGGCCGCGACGCCACCCGACGGCTCCCTCCGGCAGTGC GG Uc CUGG Cca q u g a G C U
GGGCGCGTCAGGAGCCCCAACTCGCTCGCTCCGCCAGTCGGGCACAGGCC
```

Here it is!

The RNA structure motif search problem

Randomized ferritin 5'UTR regions with an IRE-like element.
Inferring structural alignment:

```
UCUUUAcAGUGGcaugugaCCGUUUUAAGG  
GAUGCcCCAUUcacgaguAGUGGGGUAUU  
GUUUUcUGAUGcacgugcCAUCAACUAC  
GCCGUccUGGCa-a-gugaGCUGGGCCGC  
((((.((((.....))))))))))
```

No sequence conservation.

The RNA World

- The phrase "RNA World" was first proposed in 1986 by Walter Gilbert.
- Early ideas in the 1960s by Alexander Rich and Carl Woese.
- RNA carried out information storage and catalytic roles.
- No evidence, it will never appear.
- But, test if conditions in RNA are present.
- Discovery of Catalytic RNAs (ribozymes) (Cech, Altman, Bartell, Szostak, Wilson ...)
- These include self splicing intron and RNase P RNA, ribozymes that can make single nucleotides, and the ribosome.

Towards a Modern RNA World

- RNAs with highly specialized function does not seem to fit into an RNA world.
- Examples: snoRNA and micro RNA (miRNA).
- Not all RNA seem to originate from old times.
- Non-coding RNA (ncRNA) might be greatly involved in regulation.

Summary

- RNA structure formed by base pair interactions.
- The strcuture is often involved in function.
- Some RNAs holds catalytic propeties.
- RNA structure is a feature of many noncoding RNAs.
- Homology search is exploits the RNA structure, but it remains a hard problem.

The Nussinov algorithm for RNA folding

Jan Gorodkin, Center for non-coding RNA in Technology and Health,
University of Copenhagen, Denmark

<http://rth.dk/>

Content

- RNA secondary structure
- RNA folding.
- The Nussinov algorithm for RNA folding.
- From single sequences to multiple sequences.

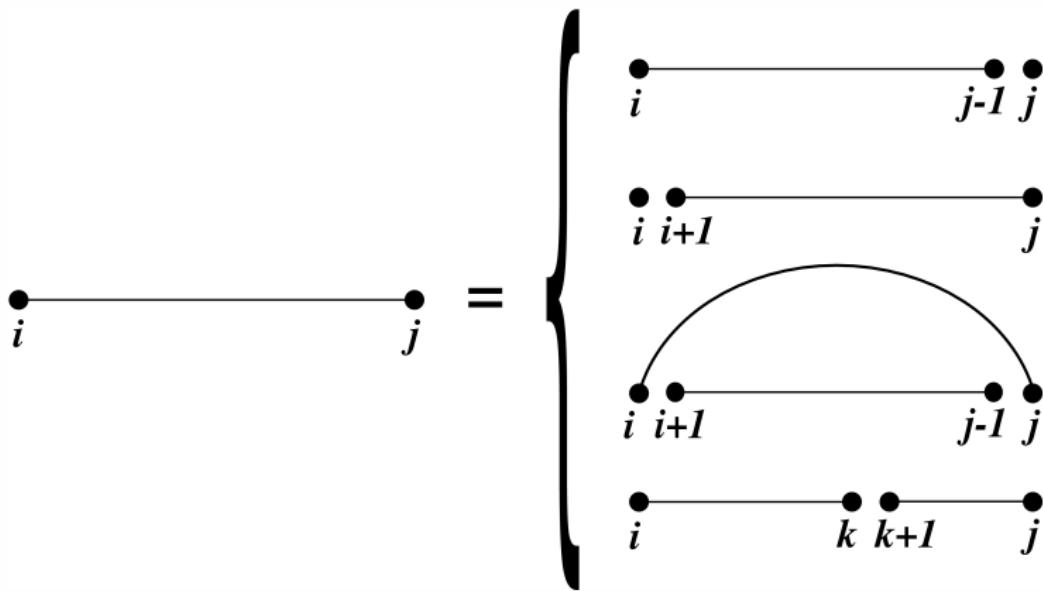
Towards RNA folding

- A first approach for folding is to find the maximum number of possible base pair. [Strategy introduced by Nussinov and Jacobson]
- This is in general an incomplete model, but concepts are the same in the more advanced models.
- Refresher: we use dynamical programming. What is the advantage of dynamical programming.

Basic RNA folding algorithm

Concept:

- Max number of base pairs found on a subsequence.
- Extend to either side, or merge two substructures.



Basic RNA folding algorithm

Maximum number of basepairs:

Subsequence: (x_i, \dots, x_j)

Recursion:

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

with

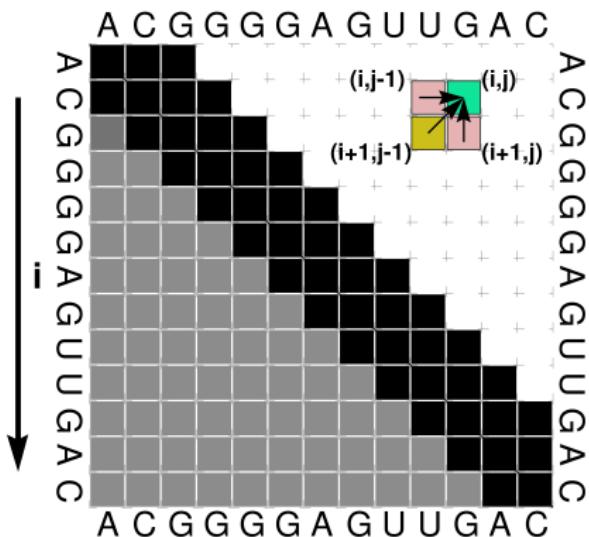
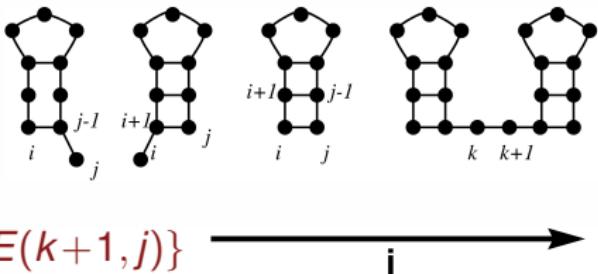
$$s(i, j) = \max \begin{cases} 1 & \text{if } x[i] \text{ and } x[j] \text{ pair} \\ 0 & \text{otherwise} \end{cases}$$

Time complexity: $O(N^3)$.

Memory complexity: $O(N^2)$.

No pseudoknots.

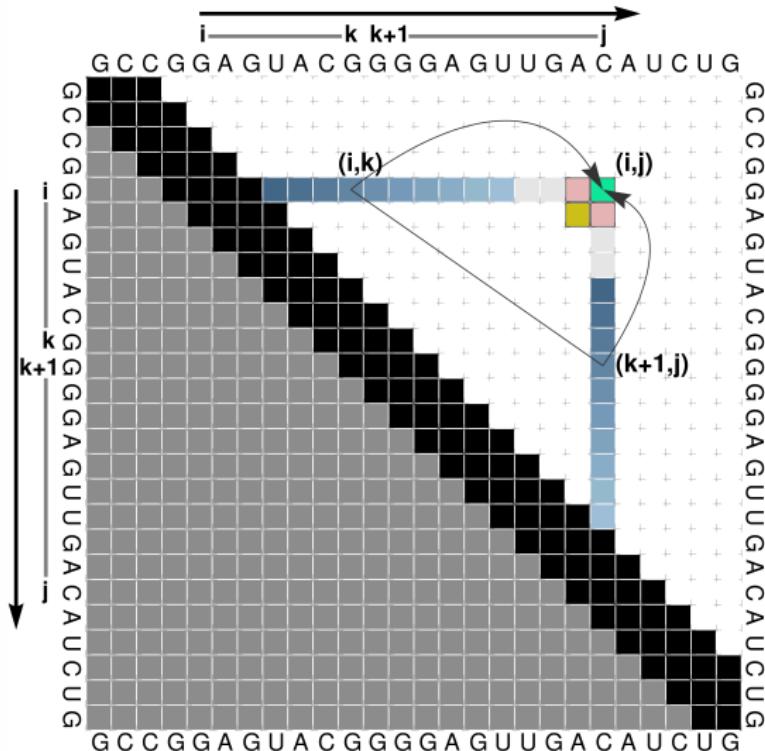
Energy folding: more complicated.



Basic RNA folding algorithm

Computing the bifurcating terms

Combine all terms (i, k) and $(k+1, j)$ for $i < k < j$.



Basic RNA folding algorithm: example

Folding the sequence: **AAACUUUCCCAGGG**

	A	A	A	C	U	U	U	C	C	C	A	G	G	G
A	0	0	0	0	1	2	3	3	3	3	3	4	5	6
A	0	0	0	0	1	2	2	2	2	2	3	3	4	5
A	X	0	0	0	1	1	1	1	1	1	2	3	3	4
C	X	X	0	0	0	0	0	0	0	0	1	2	3	4
U	X	X	X	0	0	0	0	0	0	0	1	2	3	3
U	X	X	X	X	0	0	0	0	0	0	1	2	2	3
U	X	X	X	X	X	0	0	0	0	0	1	1	2	3
C	X	X	X	X	X	X	0	0	0	0	0	1	2	3
C	X	X	X	X	X	X	X	0	0	0	0	1	2	2
C	X	X	X	X	X	X	X	X	0	0	0	1	1	1
A	X	X	X	X	X	X	X	X	X	0	0	0	0	0
G	X	X	X	X	X	X	X	X	X	X	0	0	0	0
G	X	X	X	X	X	X	X	X	X	X	0	0	0	0
G	X	X	X	X	X	X	X	X	X	X	X	0	0	0

Basic RNA folding algorithm: example

Folding the sequence: **AAACUUUCCCAGGG**

	A	A	A	C	U	U	U	C	C	C	A	G	G	G
A	0	0	0	0	1	2	3	3	3	3	3	4	5	6
A	0	0	0	0	1	2	2	2	2	2	3	3	4	5
A	X	0	0	0	1	1	1	1	1	1	2	3	3	4
C	X	X	0	0	0	0	0	0	0	0	1	2	3	4
U	X	X	X	0	0	0	0	0	0	0	1	2	3	3
U	X	X	X	X	0	0	0	0	0	0	1	2	2	3
U	X	X	X	X	X	0	0	0	0	0	1	1	2	3
C	X	X	X	X	X	X	0	0	0	0	0	1	2	3
C	X	X	X	X	X	X	X	0	0	0	0	1	2	2
C	X	X	X	X	X	X	X	X	0	0	0	1	1	1
A	X	X	X	X	X	X	X	X	X	0	0	0	0	0
G	X	X	X	X	X	X	X	X	X	X	0	0	0	0
G	X	X	X	X	X	X	X	X	X	X	0	0	0	0
G	X	X	X	X	X	X	X	X	X	X	X	0	0	0

Basic RNA folding algorithm: example

Folding the sequence: **AAACUUUCCCAGGG**

	A	A	A	C	U	U	U	C	C	C	A	G	G	G
A	0	0	0	0	1	2	3	3	3	3	3	4	5	6
A	0	0	0	0	1	2	2	2	2	2	3	3	4	5
A	X	0	0	0	1	1	1	1	1	1	2	3	3	4
C	X	X	0	0	0	0	0	0	0	0	1	2	3	4
U	X	X	X	0	0	0	0	0	0	0	1	2	3	3
U	X	X	X	X	0	0	0	0	0	0	1	2	2	3
U	X	X	X	X	X	0	0	0	0	0	1	1	2	3
C	X	X	X	X	X	X	0	0	0	0	0	1	2	3
C	X	X	X	X	X	X	X	0	0	0	0	1	2	2
C	X	X	X	X	X	X	X	X	0	0	0	1	1	1
A	X	X	X	X	X	X	X	X	X	0	0	0	0	0
G	X	X	X	X	X	X	X	X	X	X	0	0	0	0
G	X	X	X	X	X	X	X	X	X	X	0	0	0	0
G	X	X	X	X	X	X	X	X	X	X	X	0	0	0

Basic RNA folding algorithm: example

Folding the sequence: **AAA**CUUUCCC**AGGG**

	A	A	A	C	U	U	U	C	C	C	A	G	G	G
A	0	0	0	0	1	2	3	3	3	3	3	4	5	6
A	0	0	0	0	1	2	2	2	2	2	3	3	4	5
A	X	0	0	0	1	1	1	1	1	1	2	3	3	4
C	X	X	0	0	0	0	0	0	0	0	1	2	3	4
U	X	X	X	0	0	0	0	0	0	0	1	2	3	3
U	X	X	X	X	0	0	0	0	0	0	1	2	2	3
U	X	X	X	X	X	0	0	0	0	0	1	1	2	3
C	X	X	X	X	X	X	0	0	0	0	0	1	2	3
C	X	X	X	X	X	X	X	0	0	0	0	1	2	2
C	X	X	X	X	X	X	X	X	0	0	0	1	1	1
A	X	X	X	X	X	X	X	X	X	0	0	0	0	0
G	X	X	X	X	X	X	X	X	X	X	0	0	0	0
G	X	X	X	X	X	X	X	X	X	X	0	0	0	0
G	X	X	X	X	X	X	X	X	X	X	X	0	0	0

A-U C-G

Basic RNA folding algorithm: example

Folding the sequence: **AAA**CUUUCCC**AGGG**

	A	A	A	C	U	U	U	C	C	C	A	G	G	G
A	0	0	0	0	1	2	3	3	3	3	3	4	5	6
A	0	0	0	0	1	2	2	2	2	2	3	3	4	5
A	X	0	0	0	1	1	1	1	1	1	2	3	3	4
C	X	X	0	0	0	0	0	0	0	0	1	2	3	4
U	X	X	X	0	0	0	0	0	0	0	1	2	3	3
U	X	X	X	X	0	0	0	0	0	0	1	2	2	3
U	X	X	X	X	X	0	0	0	0	0	1	1	2	3
C	X	X	X	X	X	X	0	0	0	0	0	1	2	3
C	X	X	X	X	X	X	X	0	0	0	0	1	2	2
C	X	X	X	X	X	X	X	X	0	0	0	1	2	1
A	X	X	X	X	X	X	X	X	X	0	0	0	0	0
G	X	X	X	X	X	X	X	X	X	X	0	0	0	0
G	X	X	X	X	X	X	X	X	X	X	0	0	0	0
G	X	X	X	X	X	X	X	X	X	X	X	0	0	0

A-U C-G
A-U C-G

Basic RNA folding algorithm: example

Folding the sequence: **AAA**CUUUCCC**AGGG**

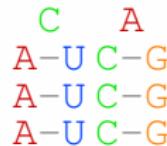
	A	A	A	C	U	U	U	C	C	C	A	G	G	G
A	0	0	0	0	1	2	3	3	3	3	3	4	5	6
A	0	0	0	0	1	2	2	2	2	2	3	3	4	5
A	X	0	0	0	1	1	1	1	1	1	2	3	3	4
C	X	X	0	0	0	0	0	0	0	0	1	2	3	4
U	X	X	X	0	0	0	0	0	0	0	1	2	3	3
U	X	X	X	X	0	0	0	0	0	0	1	2	2	3
U	X	X	X	X	X	0	0	0	0	0	1	1	2	3
C	X	X	X	X	X	X	0	0	0	0	0	1	2	3
C	X	X	X	X	X	X	X	0	0	0	0	1	2	2
C	X	X	X	X	X	X	X	X	0	0	0	1	1	1
A	X	X	X	X	X	X	X	X	X	0	0	0	0	0
G	X	X	X	X	X	X	X	X	X	X	0	0	0	0
G	X	X	X	X	X	X	X	X	X	X	0	0	0	0
G	X	X	X	X	X	X	X	X	X	X	X	0	0	0

A-U C-G
A-U C-G
A-U C-G

Basic RNA folding algorithm: example

Folding the sequence: **AAA**CUUUCCC**AGGG**

	A	A	A	C	U	U	U	C	C	C	A	G	G	G
A	0	0	0	0	1	2	3	3	3	3	3	4	5	6
A	0	0	0	0	1	2	2	2	2	2	3	3	4	5
A	X	0	0	0	1	1	1	1	1	1	2	3	3	4
C	X	X	0	0	0	0	0	0	0	0	1	2	3	4
U	X	X	X	0	0	0	0	0	0	0	1	2	3	3
U	X	X	X	X	0	0	0	0	0	0	1	2	2	3
U	X	X	X	X	X	0	0	0	0	0	1	1	2	3
C	X	X	X	X	X	X	0	0	0	0	0	1	2	3
C	X	X	X	X	X	X	X	0	0	0	0	1	2	2
C	X	X	X	X	X	X	X	X	0	0	0	1	1	1
A	X	X	X	X	X	X	X	X	X	0	0	0	0	0
G	X	X	X	X	X	X	X	X	X	X	0	0	0	0
G	X	X	X	X	X	X	X	X	X	X	0	0	0	0
G	X	X	X	X	X	X	X	X	X	X	X	0	0	0



Basic RNA folding algorithm (Nussionov-Jacobson)

Pseudo-code (As in Durbin et al. 1998):

```
Push (1,L) onto stack      (Start with this coordinate: i=1, j=L)

Recursion: Repeat until stack is empty
    -pop (i,j);
    if i >=j continue;

    else if M(i+1,j)           = M(i,j)   then push (i+1,j);

    else if M(i,j-1)           = M(i,j)   then push (i,j-1);

    else if M(i+1,j-1) + d[i,j] = M(i,j)   then
        record base i,j;
        push (i+1,j-1);

    else for k = i+1 to j-1:  if M(i,k) + M(k+1,j) = M(i,j) then
        push (k+1,j);
        push (i,k);
        break;
```

From maximizing to minimizing

- Energy parameters are negative reflecting contributions to the free energy.
- Thus one can compute the free energy (kcal/mol) by minimizing over the different contributions.
- The Nussinov algorithm can in the first approach be converted.

Ambiguity in Nussinov algorithm

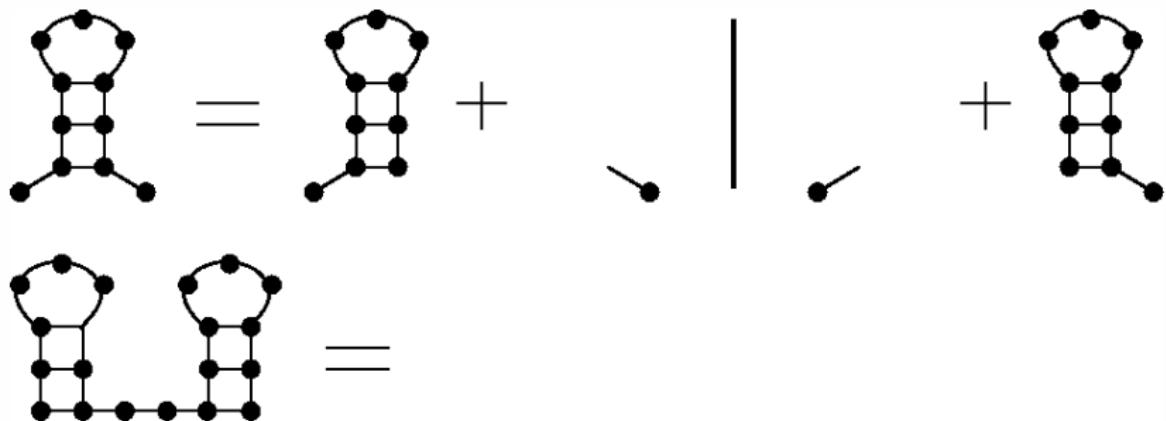
Ambiguity:

The algorithm contains redundancy and is ambiguous as some cells can be computed in more than one way.

Ambiguity in Nussinov algorithm

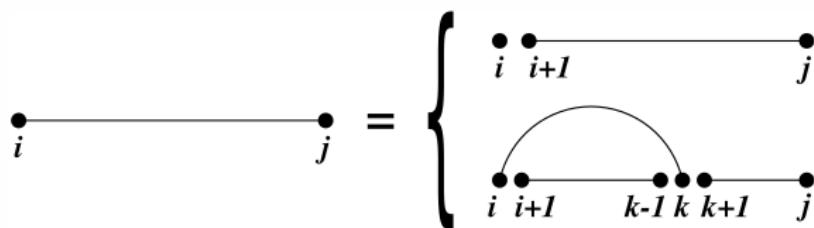
Ambiguity:

The algorithm contains redundancy and is ambiguous as some cells can be computed in more than one way.



Decomposing the Nussinov algorithm

Decomposing the Nussinov algortihm



$$E(i,j) = \max \left\{ \begin{array}{l} E(i+1,j) \\ \max_{i+3 \leq k \leq j \text{ s.t. } s(i,k)=1} \{ 1 + E(i+1, k-1) + E(k+1, j) \} \end{array} \right\}$$

How can this scoring scheme be changed to take different base pairings into account?

Make the scoring more realistic

Replace β_{ij} with $\beta_{x[i]x[j]}$.

$$\begin{aligned}\beta(C, G) &= \beta(G, C) = 3 & \beta(A, U) &= \beta(U, A) = 2 \\ \beta(G, U) &= \beta(U, G) = 1\end{aligned}$$

$$E_{ij} = E_{i+1,j} + E_{i+1,k-1} + E_{k+1,j} + \beta(x_i, x_k)$$

.	((.	(.	.	.)))	
A	G	C	A	C	A	C	A	G	G	C	
0	0	0	0	0	0	3	3	3	6	9	A
	0	0	0	0	0	3	3	3	6	9	G
	0	0	0	0	0	0	0	3	6	6	C
	0	0	0	0	0	0	3	3	3	3	A
	0	0	0	0	0	3	3	3	3	3	C
	0	0	0	0	0	0	0	0	0	0	A
		0	0	0	0	0	0	0	0	0	C
		0	0	0	0	0	0	0	0	0	A
			0	0	0	0	0	0	0	0	G
				0	0	0	0	0	0	0	G
					0	0	0	0	0	0	C

Make the scoring more realistic

For the full algorithm cost for structural elements are introduced, such as

$$E_{\text{loop}} = E_{\text{mismatch}} + E_{\text{size}} + E_{\text{special}}$$