COMP364

Introduction to RNA secondary structure prediction

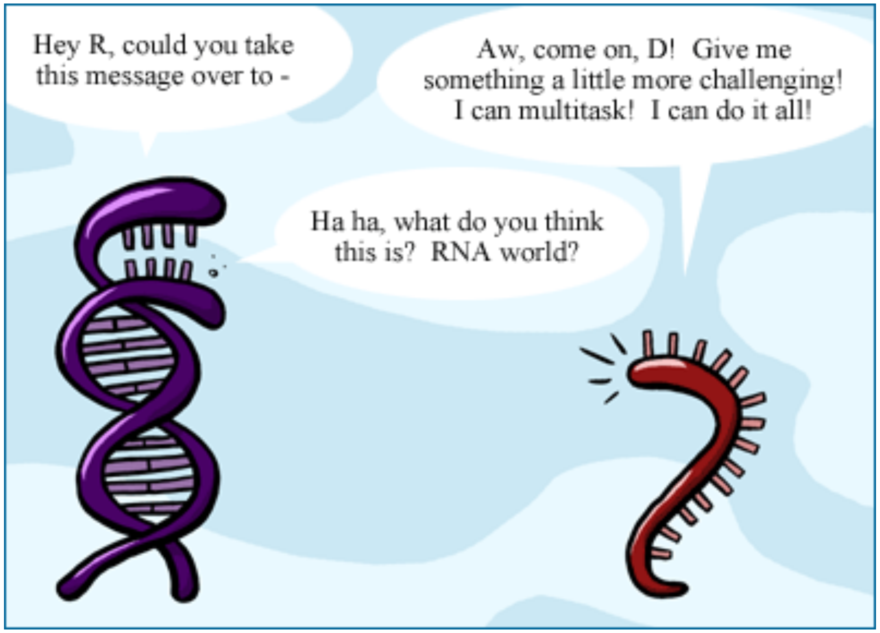
Jérôme Waldispühl
School of Computer Science, McGill

RNA world

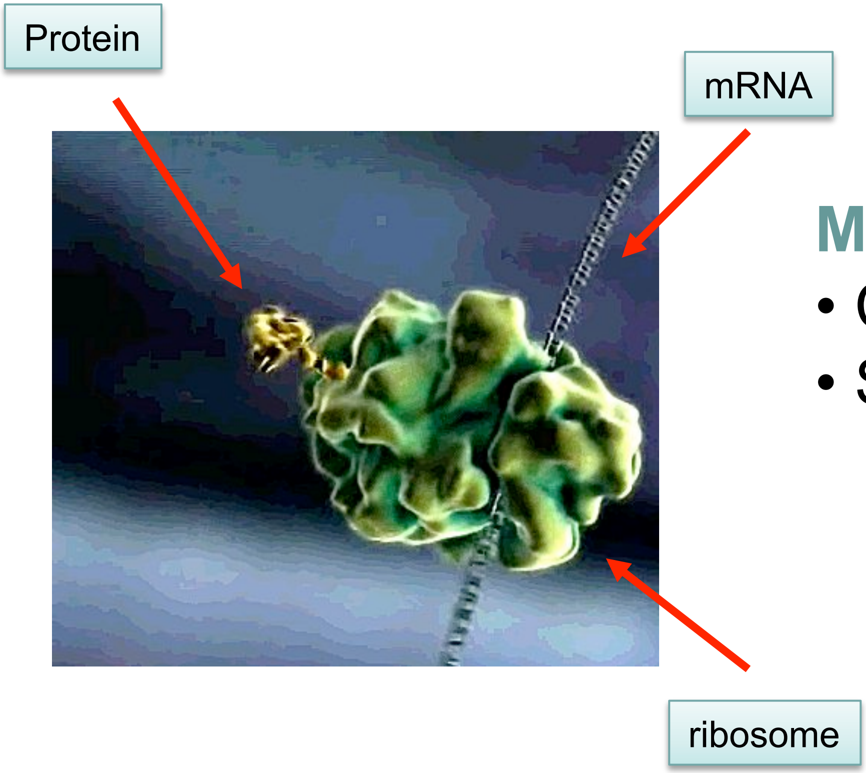
In prebiotic world, RNA thought to have filled two distinct roles:

1. an information carrying role because of RNA's ability (in principle) to self-replicate,
2. a catalytic role, because of RNA's ability to form complicated 3D shapes.

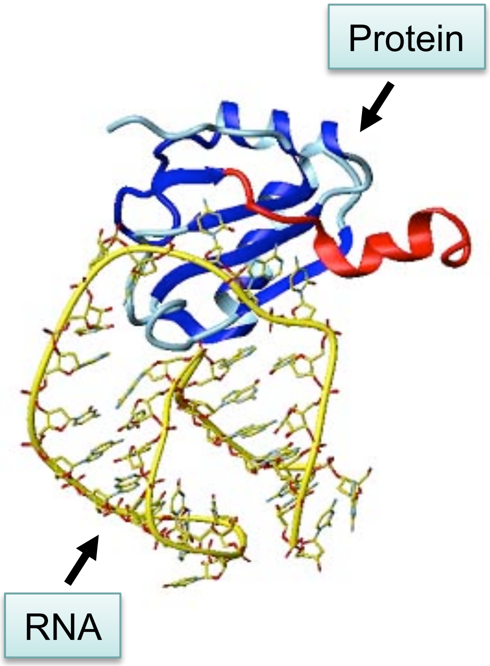
Over time, DNA replaced RNA in Its first role, while proteins replaced RNA in its second role.



RNA classification



Messenger RNA:
- Carry genetic information,
- Structure less important.



Non-coding RNA:
- Functional,
- Structure is important.

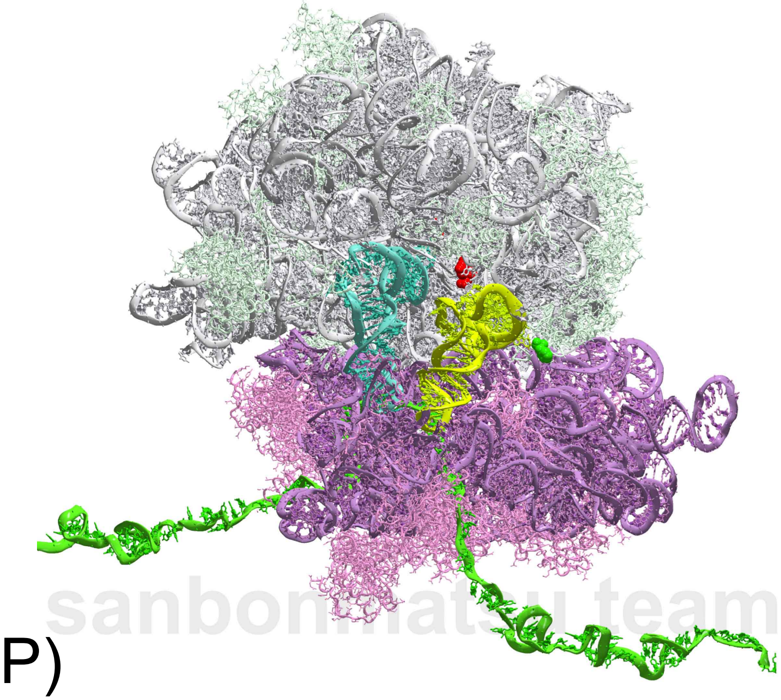
Cellular functions of RNA

Genetic Functions:

- Messenger RNA
- Viroids
- Transfer RNA

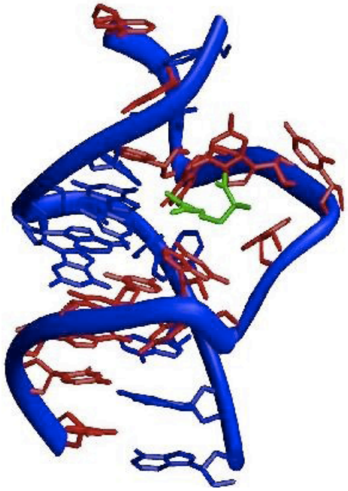
Enzymatic functions:

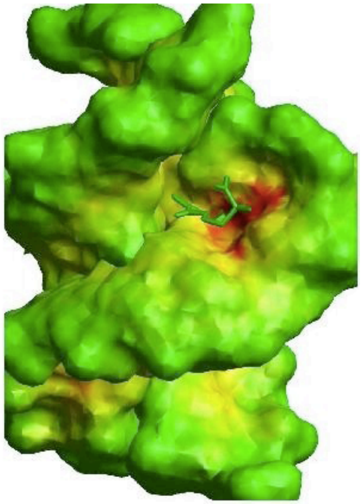
- Splicing (snRNA)
- RNA Maturation (ribonuclease P)
- Ribosomic RNA
- Guide RNA (snoRNA)

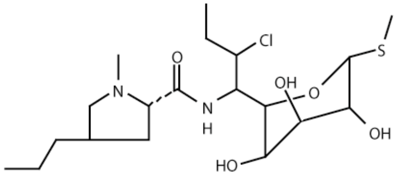


RNA structure and function

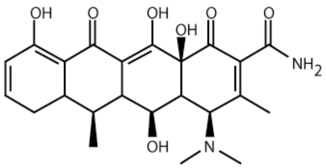
- RNAs have a 3D structure,
- This 3D structure allow complex functions,
- The variety of RNA structures allow the specific recognition of a wide range of ligands,
- Some molecules target these RNA structures (antibiotics, antimitotics, antiviruses):



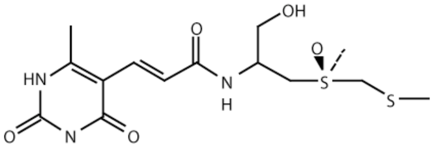




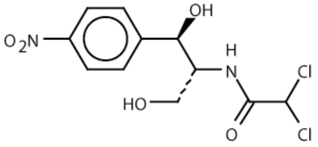
Linezolid



Doxycyclin

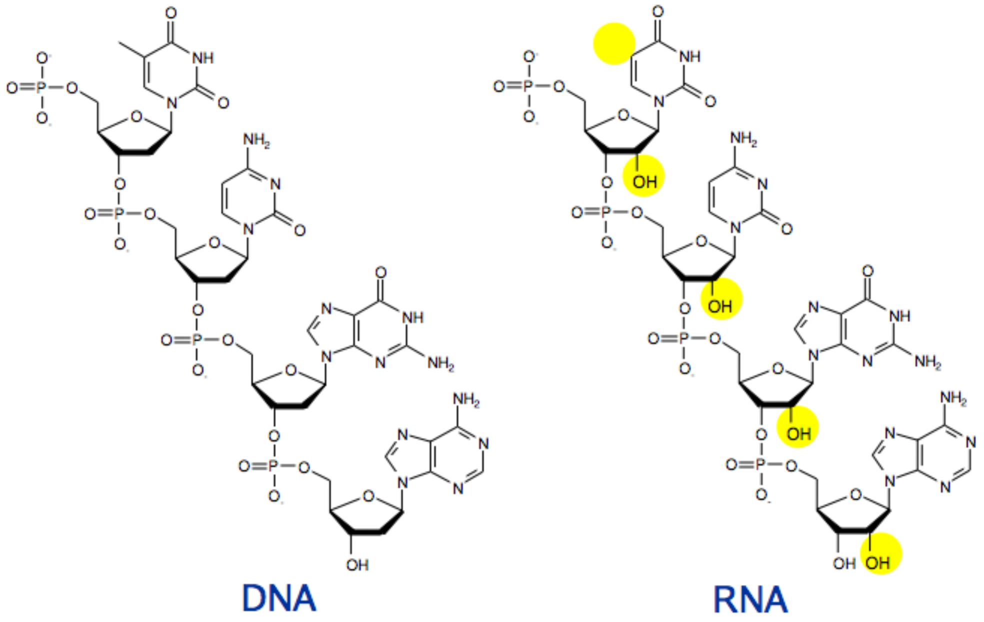


Sparsomycin



Chloramphenicol

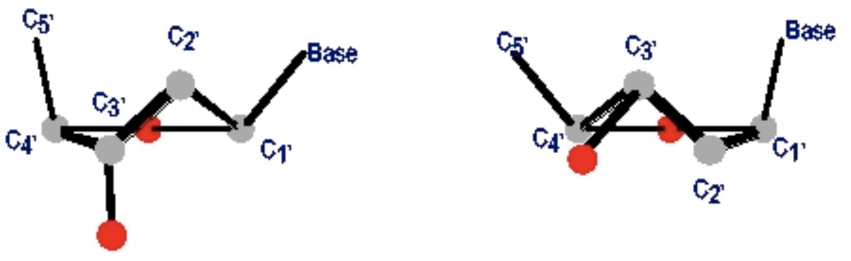
RNA vs DNA: Chemical nature



- 2’-OH group attached to sugar (instead of 2’-H): *more polar*
- Substitution of thymine by uracile = suppression of group 5-CH3

Small modifications => big effects

RNA vs DNA: Modification of the local and global geometry



C2'endo

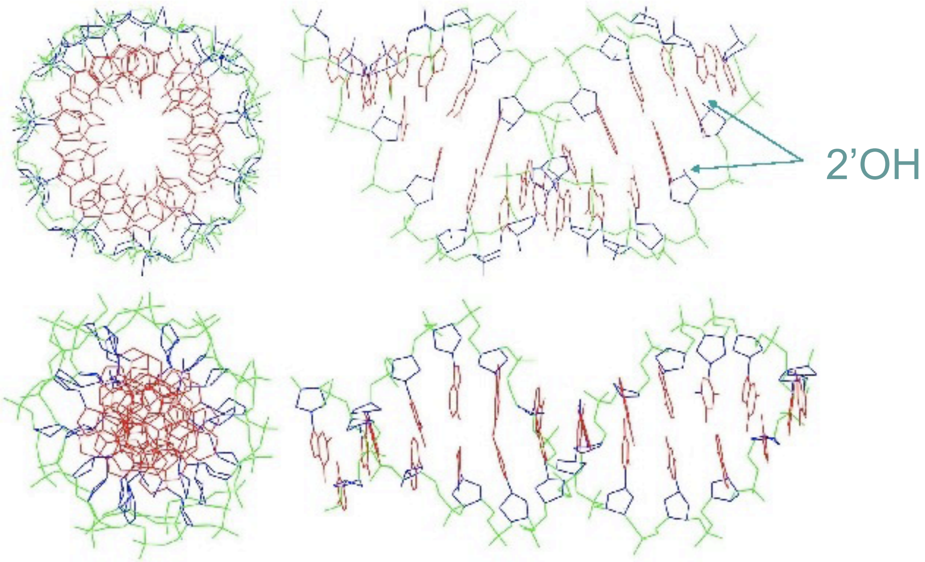
C3'endo

:Local conformation

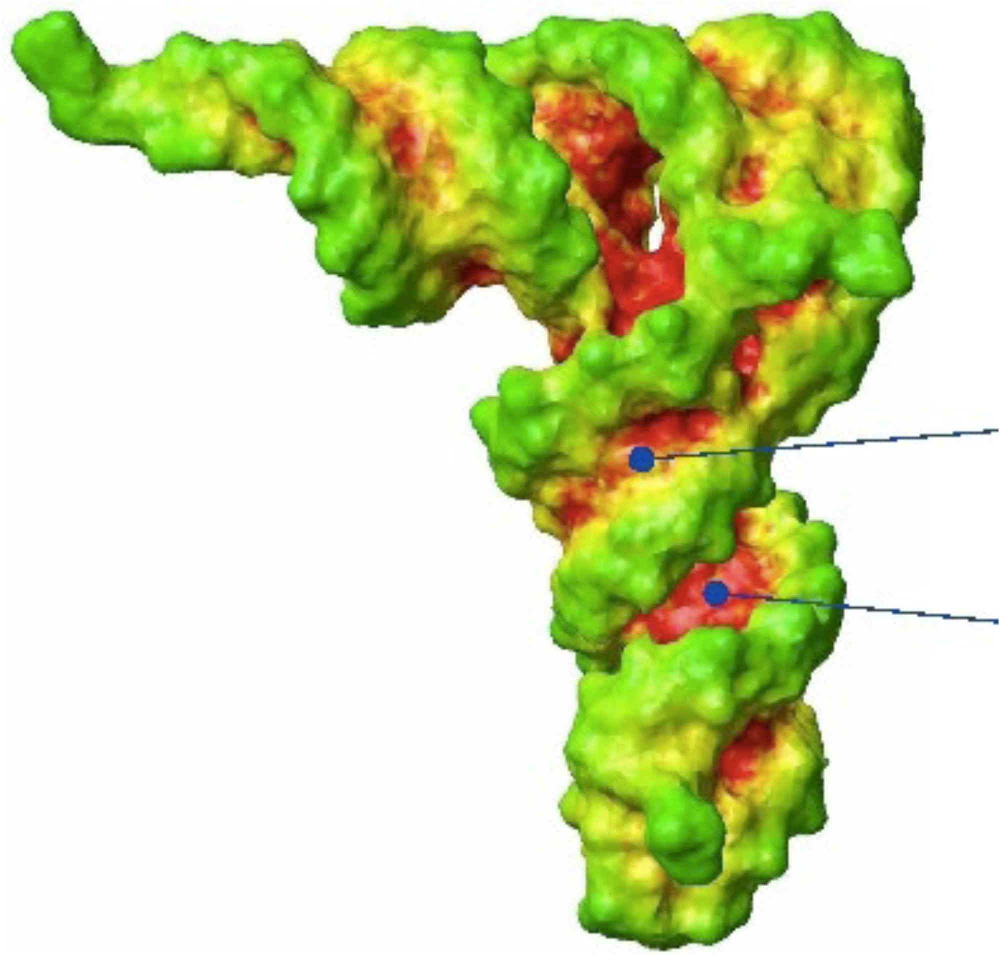
Global conformation:

RNA favorite:

DNA favorite:



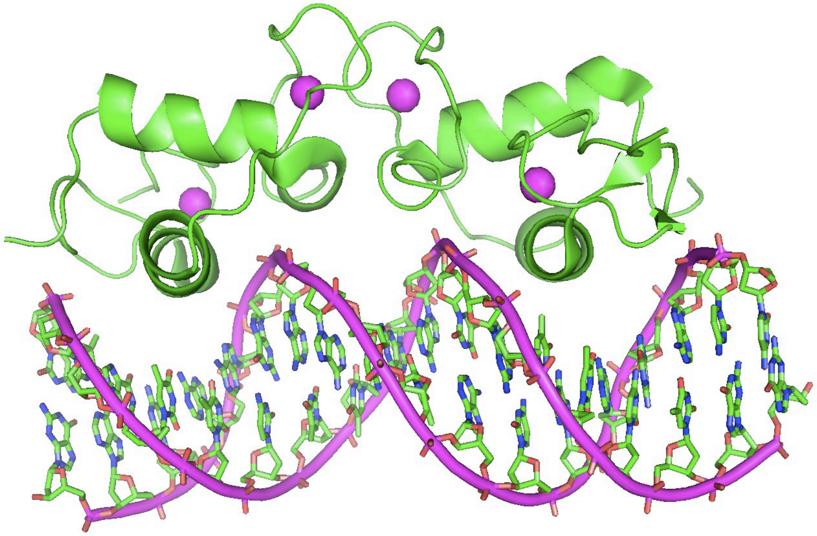
RNA vs DNA: Consequence of the modification of the geometry



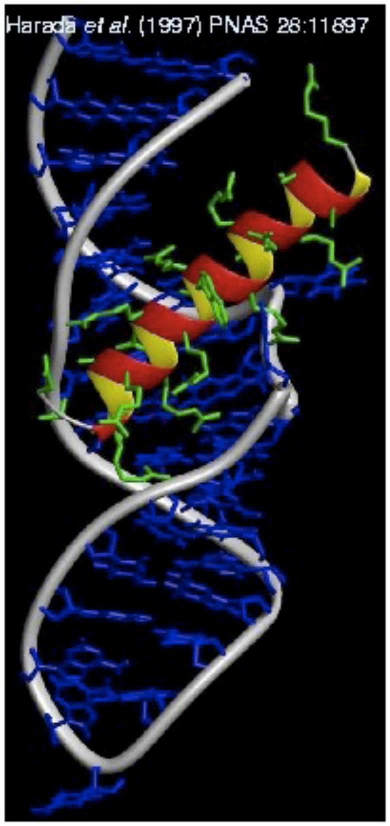
Small furrow is flat

Big furrow is deep

RNA vs DNA: RNA-Protein and DNA-Protein interactions are different



DNA-Protein: Secondary structure elements insert in big furrow



Harada *et al.* (1997) PNAS 28:11697

Protein binds to an irregularity of the helix

RNA-Protein interaction are more specific. Usually using less structured regions.

RNA vs DNA: Last (?) differences

- RNA is a short linear molecule
  DNA long ≠ RNA short

- RNA are usually single stranded
  ADN double stranded ≠ ARN single stranded

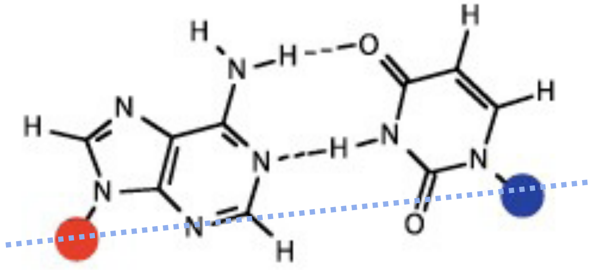
- « turnover » relatively fast
  ADN stable ≠ ARN versatile

Base pairing in RNAs

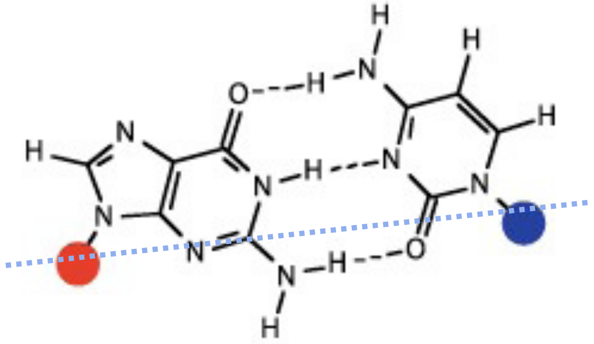
- As in DNA, bases can interact through hydrogen bonds.

- Beside the two canonical base-pairs, RNA structure allows “Wooble” base-pairs.

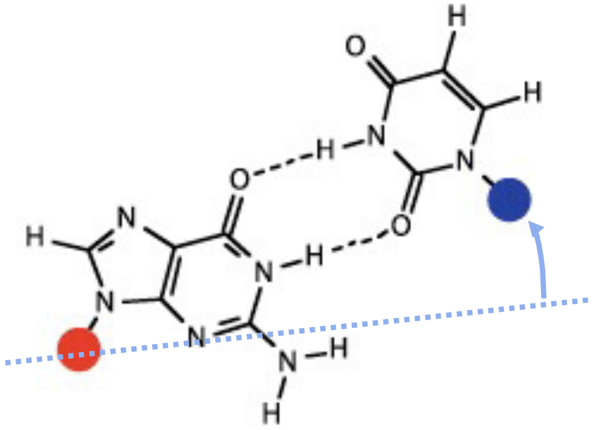
- A-U and G-C are “isosterus” while G-U induce a distortion of the backbone.



Paire A-U



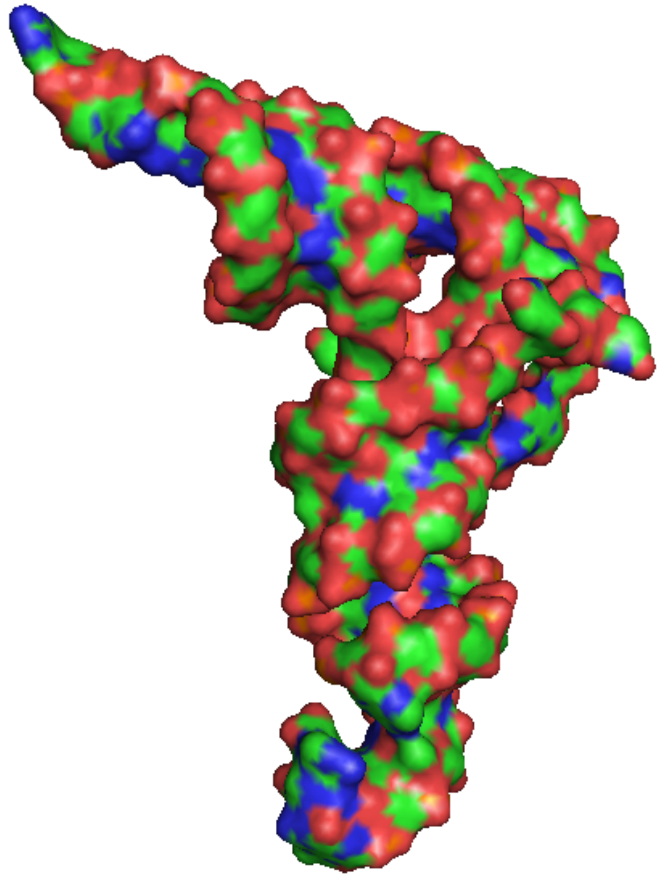
Paire G-C

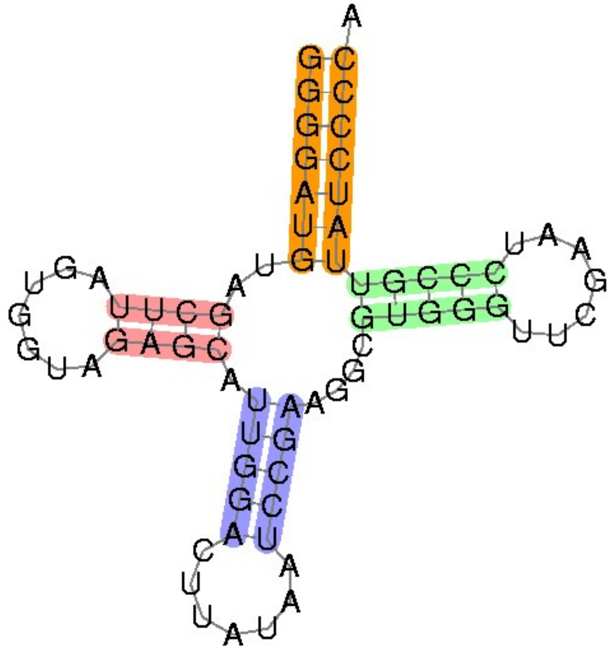


Paire "bancale" G-U

RNA secondary structure

The **secondary structure** is the ensemble of base-pairs of the structure.





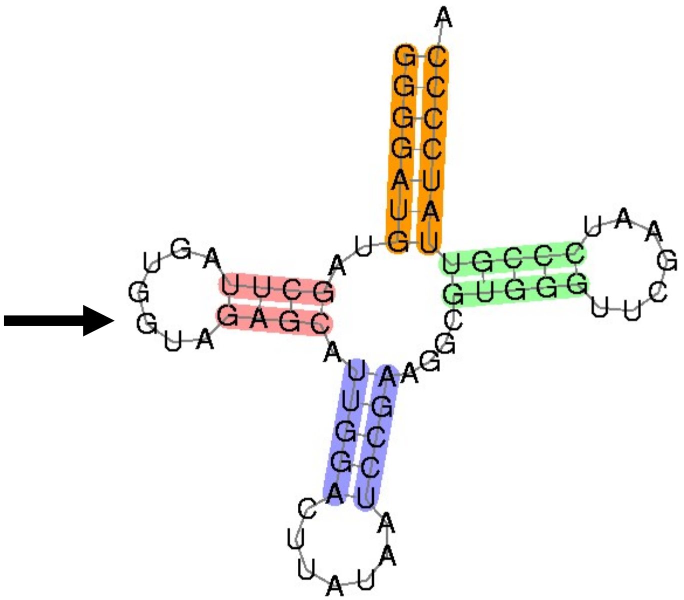
RNA secondary structure

Central assumption: RNA secondary structure forms before the tertiary structure.

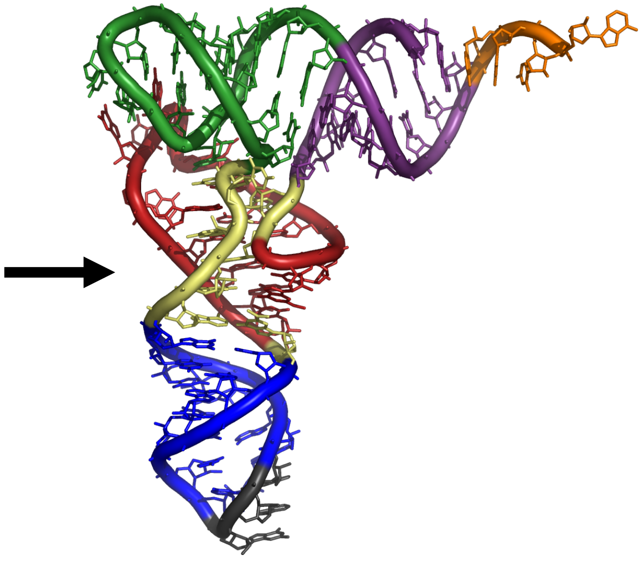
Primary structure

cgcggggttgatataatataaaaaataat
aaataataataataataattatcatcatt
tccgacccatattataataatacgggttg
gaaatatagatataatatttattatattga
tataatacatatatataagttagaggaaa
tgttgtttaaaggttaaactgttagattgc
aaatctacacatttagagttcgattctctt
catttcttatatatatactacccacgcg

Secondary structure



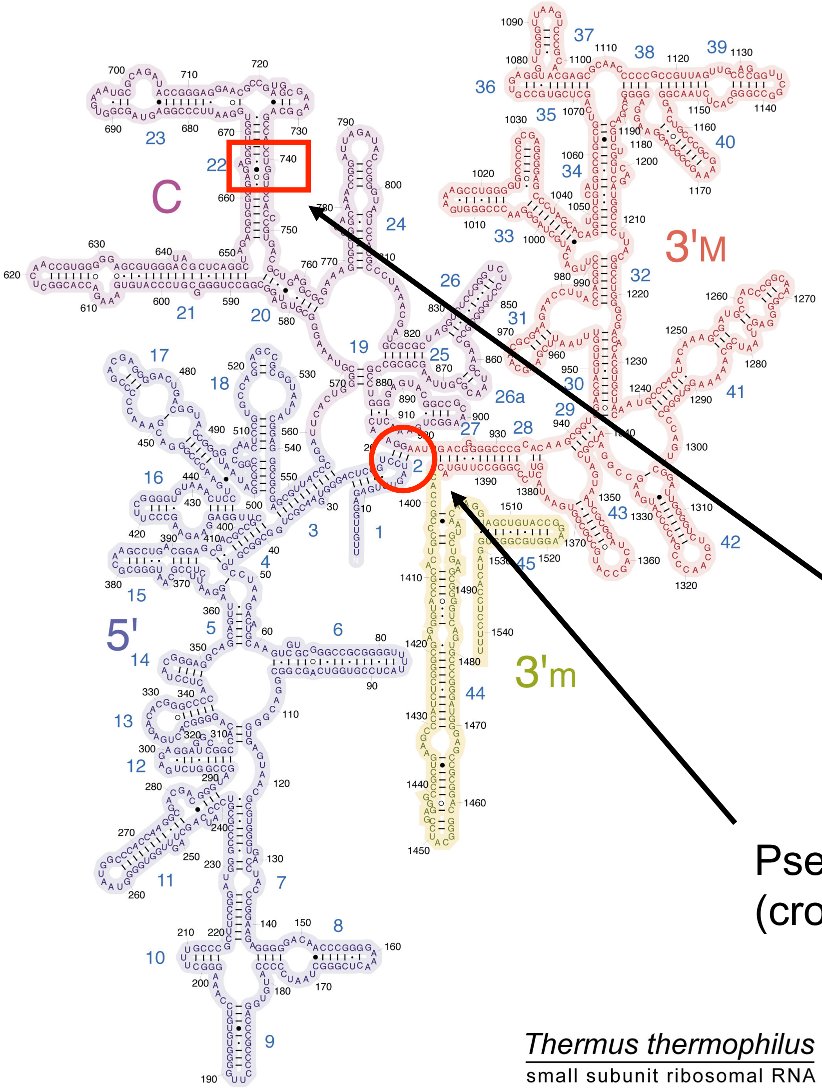
Tertiary structure



Secondary structure prediction is an important step toward 3D structure prediction.

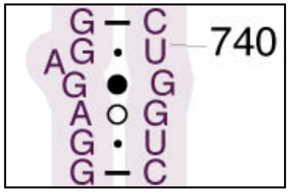
RNA secondary structure

The secondary structure can be very complex. Usually most of it can be drawn on a plane.



Thermus thermophilus
small subunit ribosomal RNA

Few "irregularities" remain.

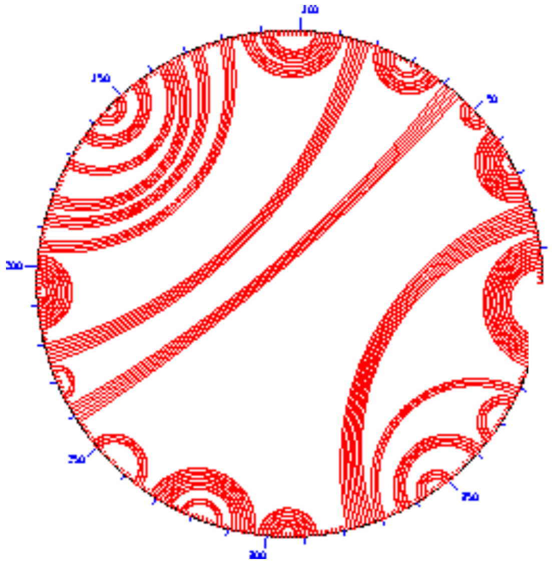


Non-canonical base-pairs

Pseudo-knot
(crossing interaction)

Base triplets
(Not on the picture)

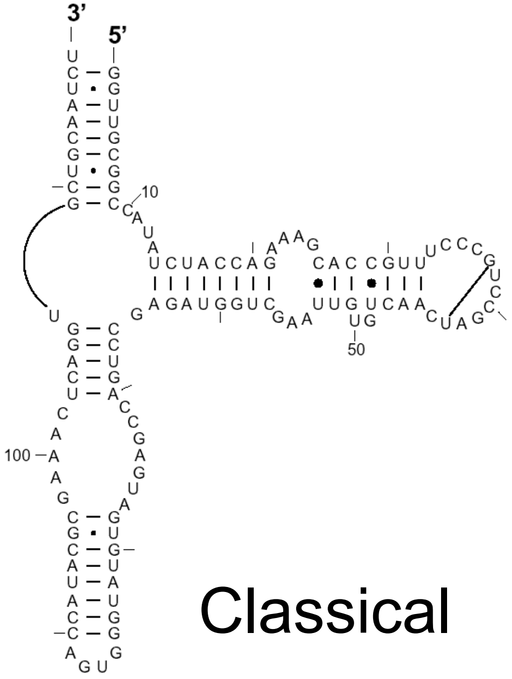
RNA secondary structure representations



Circular

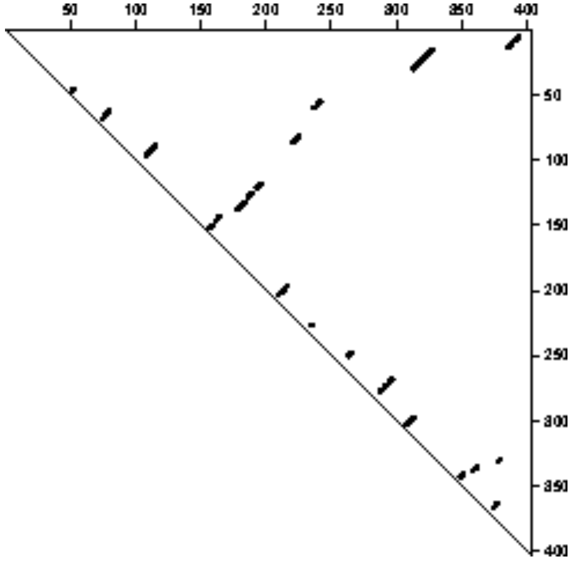
Brackets

..(((((((.(((..((...))))))))...(((....))))).)))))



Classical

Dot plot



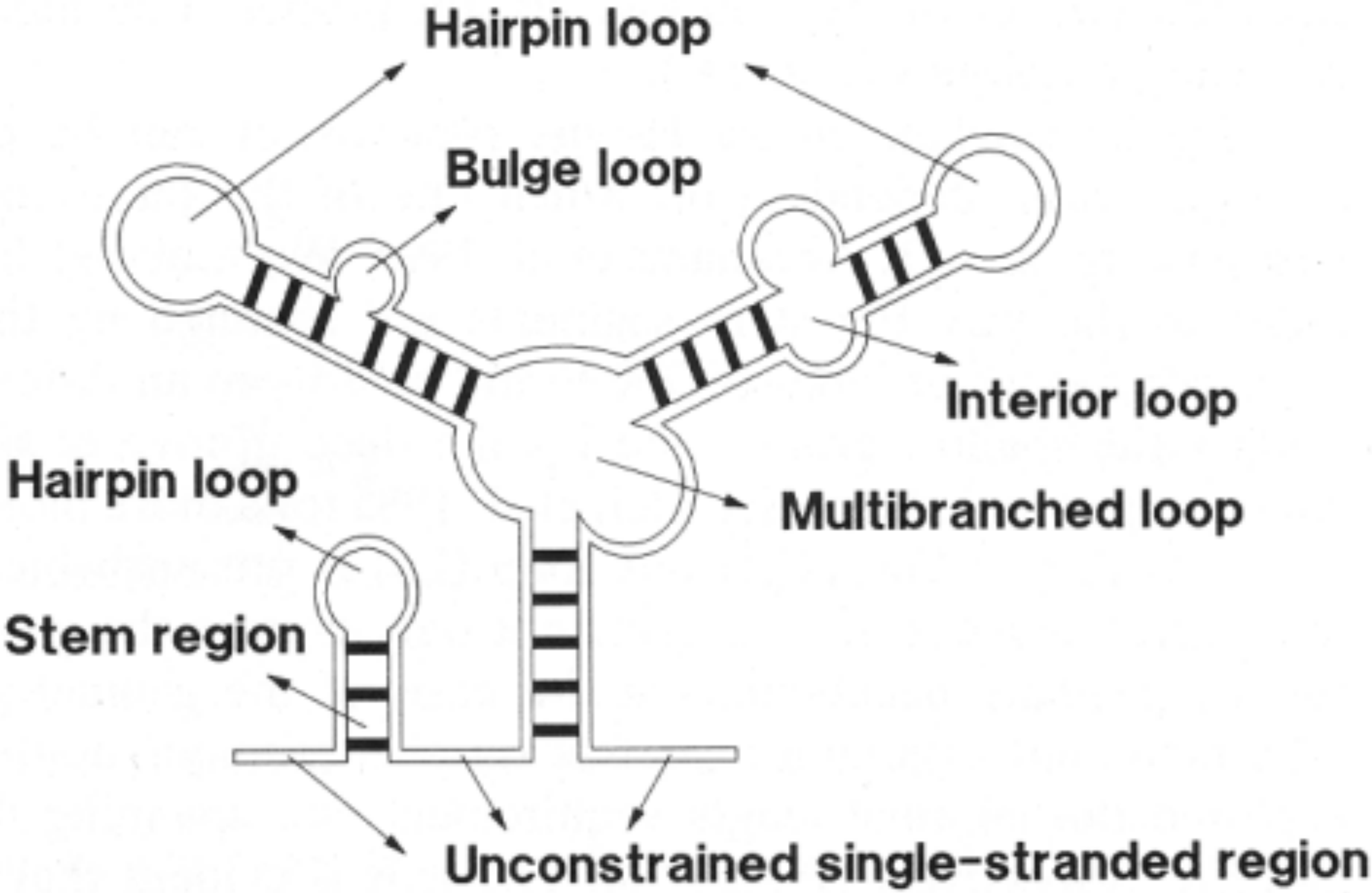
RNA secondary structure prediction using comparative methods

The secondary structure can be predicted from the alignment of homologous sequences. Base-pairs are identified through compensatory mutations.

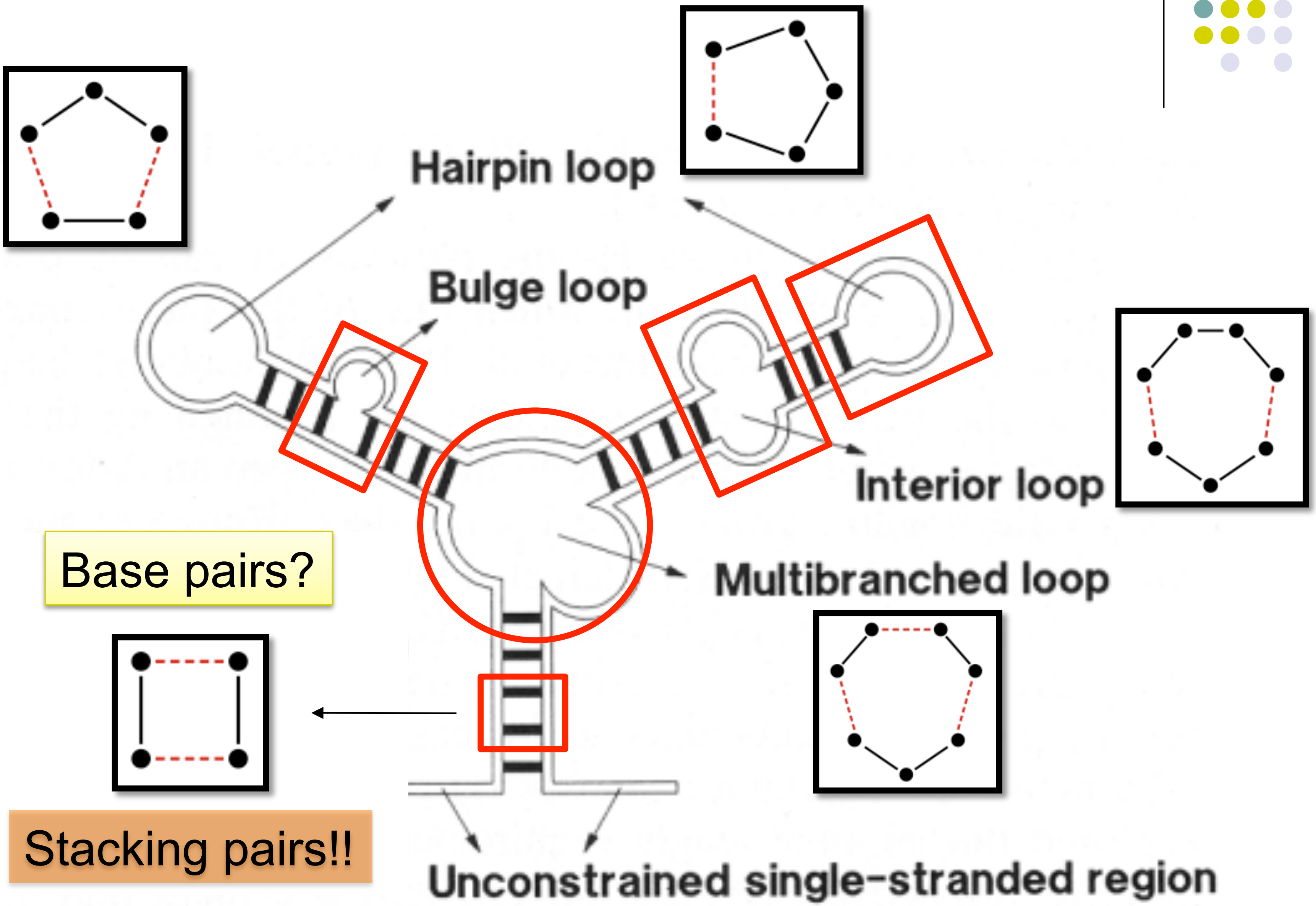
```
AJ617357.1/475-507    Car.Enc.  ACGGUCACAAACACUCAAUCAACUGUGGGCCGU
M88547.1/564-596      Car.Men.  ACGGUCACAAACACCCAAUCAACCGUUGGUCGU
U33047.1/505-537      Car.The.  UCGGCCACAAACACACAAUCUACUGUUGGUCGG
X56019.1/1572-1604    Car.The.  UCGGCCACAAACACACAGUCUACUGUUGGCCGG
AJ617361.1/475-507    Car.Enc.  ACGGUCACAAACACUCAAUCAACUGUGGGCCGU
M20562.1/1573-1605    Car.The.  UCGGCCACAAACACACAGUCUACUGUUGGCCGG
AF030574.1/505-537    Car.The.  UCGGCCACAAACACACAAUCUACCGUUGGUCGA
AJ617358.1/475-507    Car.Enc.  ACGGUCACAAACACUCAAUCAACUGUGGGCCGU
SS_cons                         <<<<<<<...<<<..........>>>>>>>>>>
```

97% of the base pairs predicted by comparative analysis in rRNAs have been confirmed later in the crystal structure.

RNA secondary structure elements



Loop decomposition

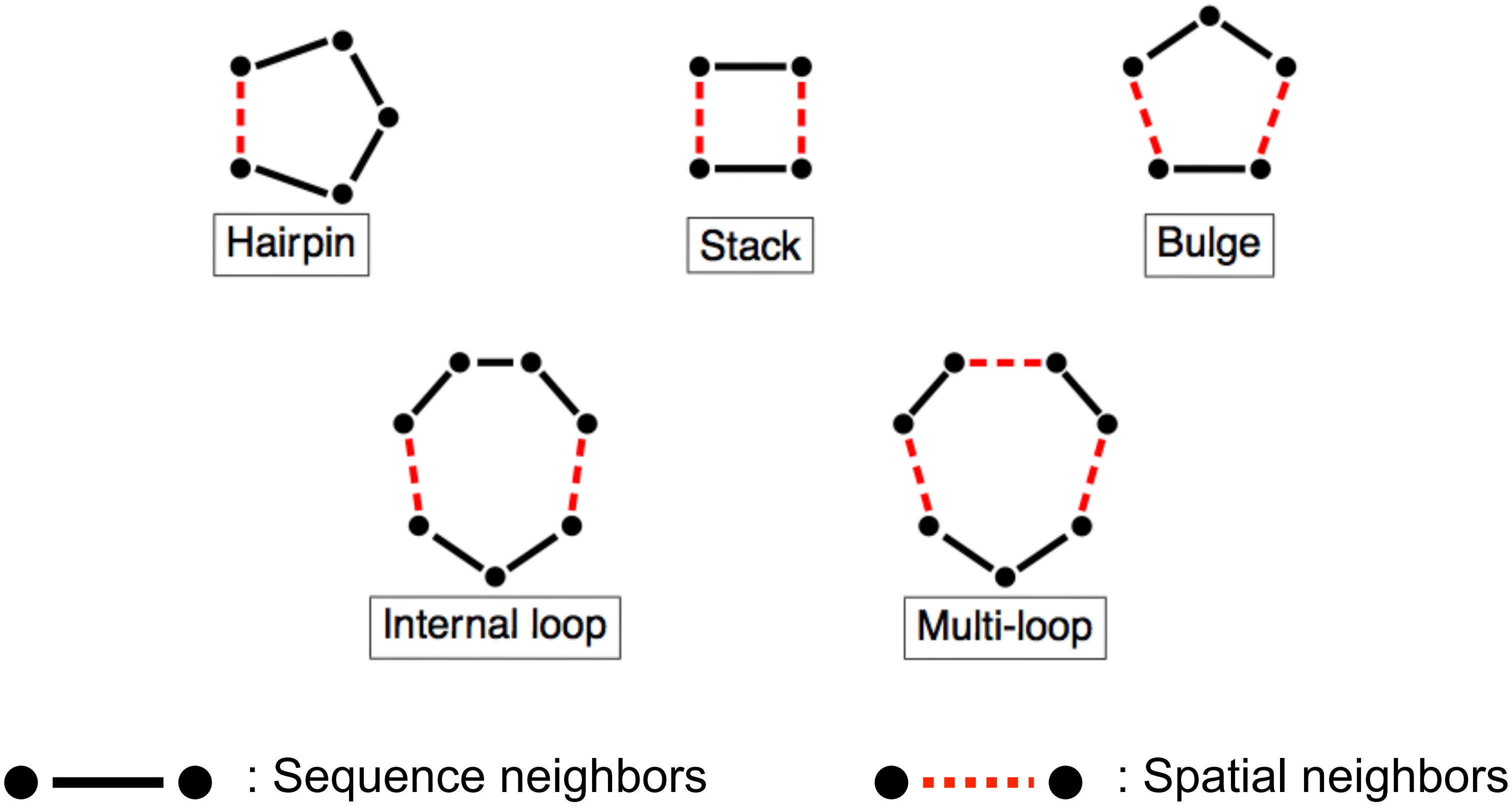


Base pairs?

Stacking pairs!!

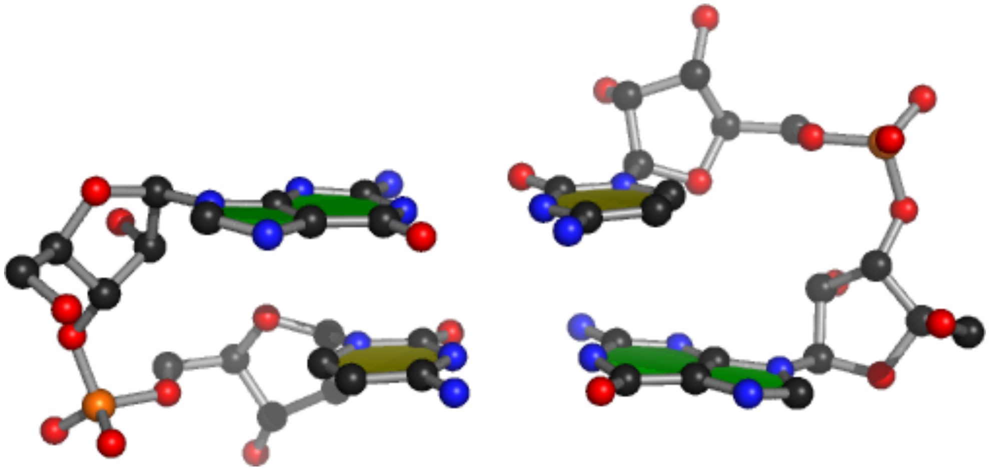
RNA secondary structure description

A secondary structure can be decomposed in a sequence of loops:



Stacking base pairs

Base stacking interactions between the pi orbitals of the bases' aromatic rings contribute to stability. GC stacking interactions with adjacent bases tend to be more favorable.



Note: Stacking energy are orientated.

5’ - CG - 3’
3’ - GC - 5’

≠

5’ - GC - 3’
3’ - CG - 5’

Vienna RNA package

Web server & source code: http://rna.tbi.univie.ac.at/

Suite of tools to predict and analyze the secondary structure of RNA. Here, we will overview:

- RNAfold: The main program that is used to predict the secondary structure from sequence data alone.
- RNAdistance: A program to compare secondary structures.

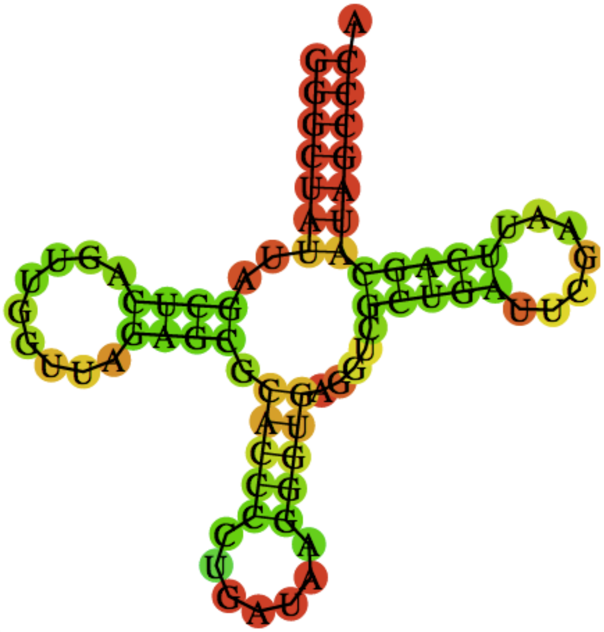
RNAfold

Command line: `RNAfold -p -d2 --noLP < test_sequenc.fa`

Text output:

```
GGGCUAUUAGCUCAGUUGGUUAGAGCGCACCCUGAUAAGGGUGAGGUCGCUGAUUCGAAUUCAGCAUAGCCCA
(((((((..((((.........)))).(((((.......))))).....(((((.......))))))))))))).
```

Graphic Output:



RNAdistance

Base pair distance:

```
$ RNAdistance -DP
Input structure;  @ to quit
....,....1....,....2
(((..(((...))))))
..((((((...))))))
P: 6
```

Number of base pairs you need to remove and add to go from one structure to the other.

WARNING: Use only with structures with same length.

RNA distance implements many other metrics.