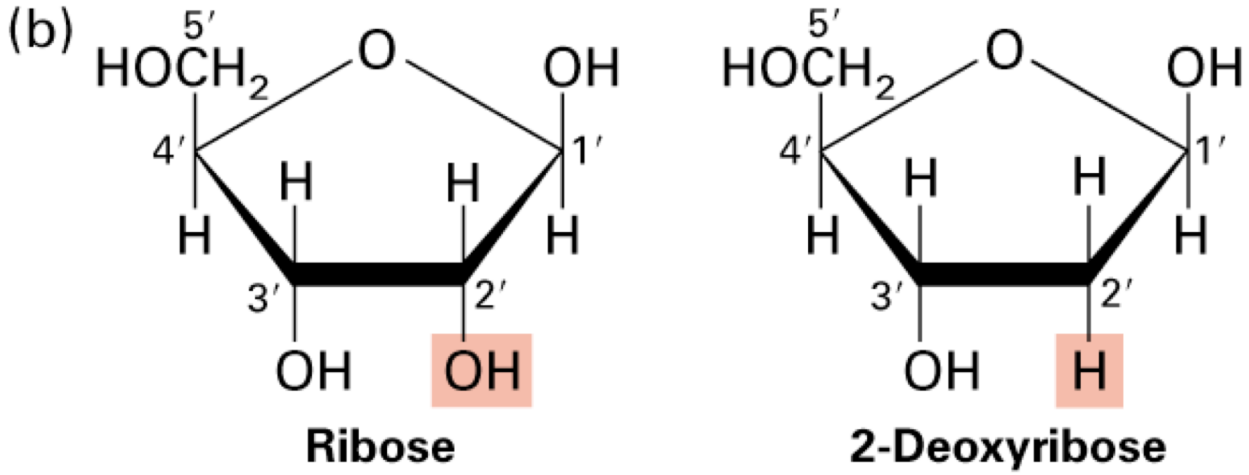
RNA world

Conformations of RNA

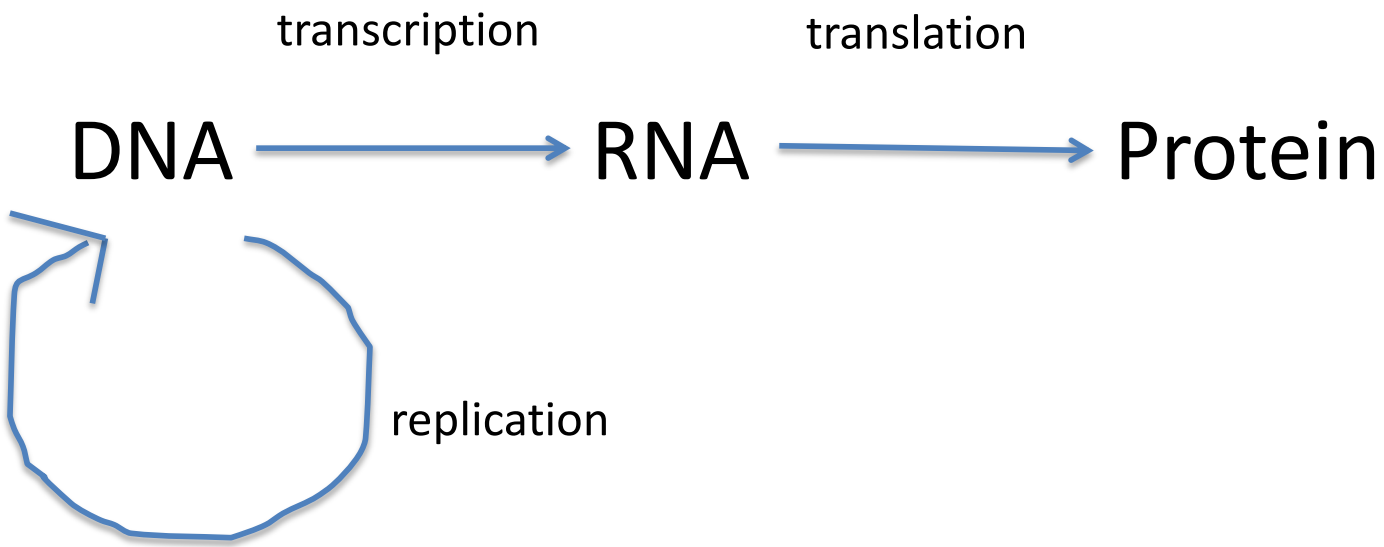
- Primary structure of RNA similar to DNA



- RNA, like DNA, can be single or double stranded, linear or circular.
- Unlike DNA, RNA can exhibit different foldings
- Different folds permit the RNAs to carry out specific functions in the cell

Central dogma

The flow of genetic information

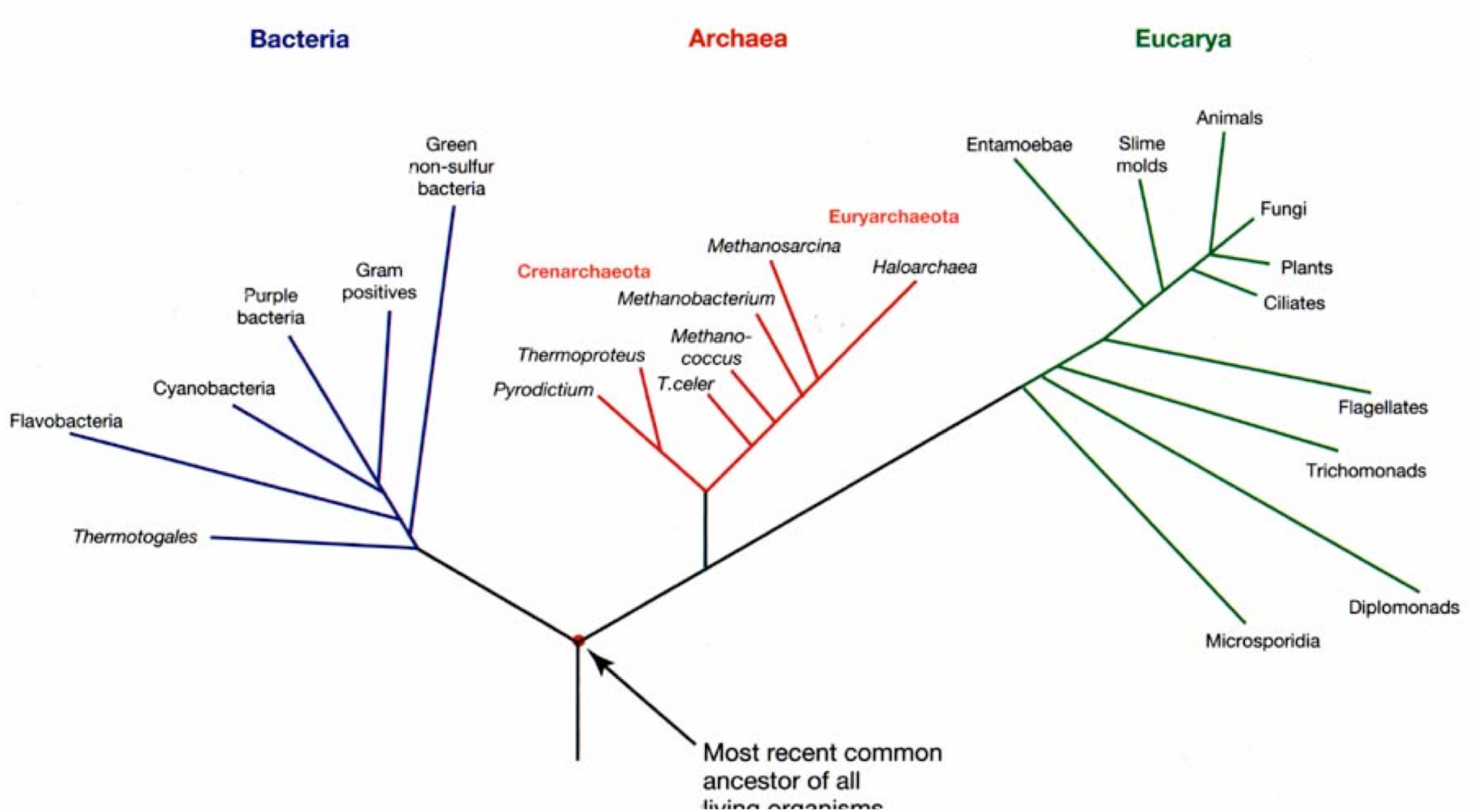


RNA

"Three" different types of RNA

- mRNA - messenger RNA, specifies order of amino acids during protein synthesis
- tRNA - transfer RNA, during translation mRNA information is interpreted by tRNA
- rRNA – ribosomal RNA, combined with proteins aids tRNA in translation

Small subunit 18S rRNA



then:

- Discovery of catalytic RNA (ribonuclease P, self-splicing introns, hepatitis delta virus, …)
- Discovery of other roles of RNA
  - ncRNAs - functional RNA molecules (RNA other than mRNA)

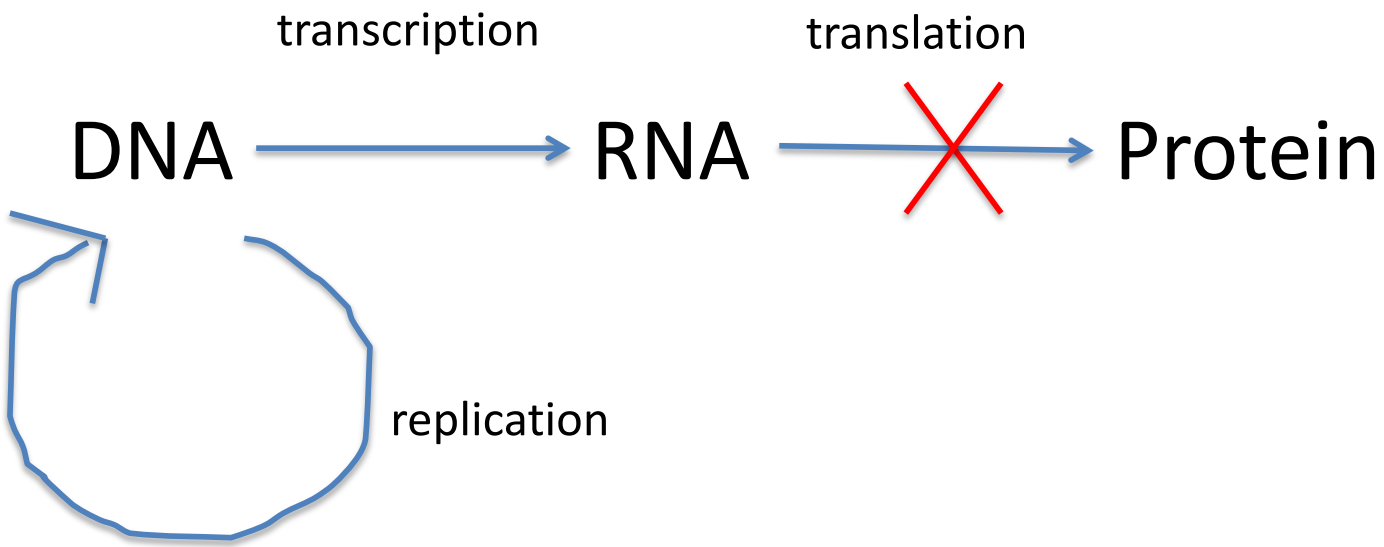
    Genomic dark matter
    - Ignored by gene prediction methods
    - Not in EnsEMBL
    - Computational complexity
    - ~10% of human gene count?
  - RNA interference (siRNA, miRNA, tiny-noncoding RNA, small modulatory RNA, ..
  - cofactor RNA (telomerases,..)
  - …. to be discovered

Local RNA structures in untranslated regions (UTR)

- have known roles in regulation of gene expression :
  - mRNA stabilization
    - 5’ UTR elements in bacteria reduce mRNA degradation
    - 3’ UTR elements in eukaryotes control mRNA degradation
  - mRNA translation
    - Control and Rate of translation
    - IRES (viruses)
  - mRNA localization
    - Transport – development
  - mRNA processing
    - Splicing of introns (alternative)
- In the coding regions, redundancy of the genetic code leaves (some) room for RNA sec. structure on top

Central dogma

The flow of genetic information



Properties of RNA molecules

- Assemble in double-stranded helices like DNA

  Carry GENETIC INFORMATION like DNA

- Fold in complex tertiary architectures like proteins

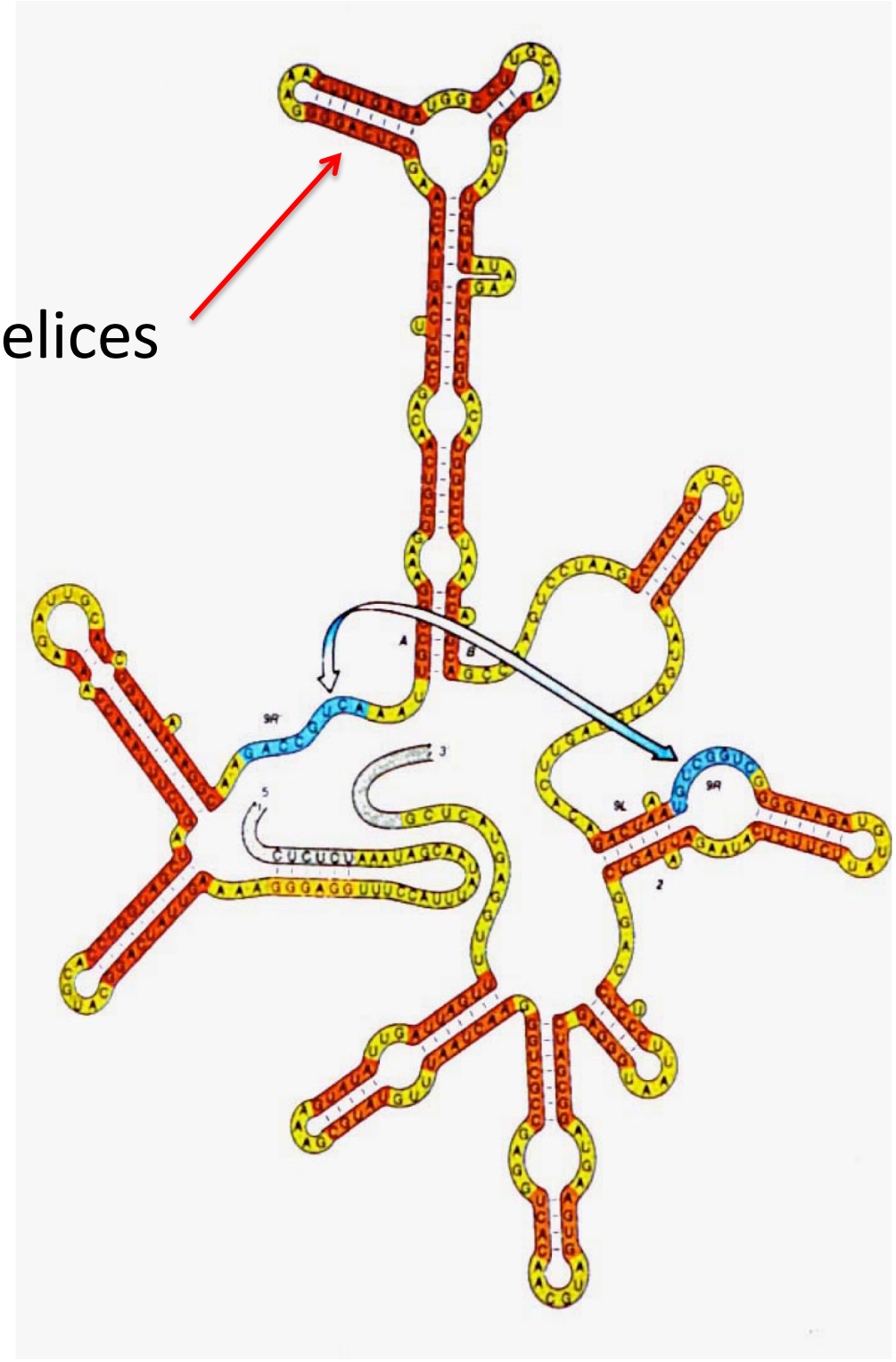
  Perform CHEMICAL CATALYSIS like proteins

Biological sequence analysis

- Proteins – “easy”
- RNAs - hard

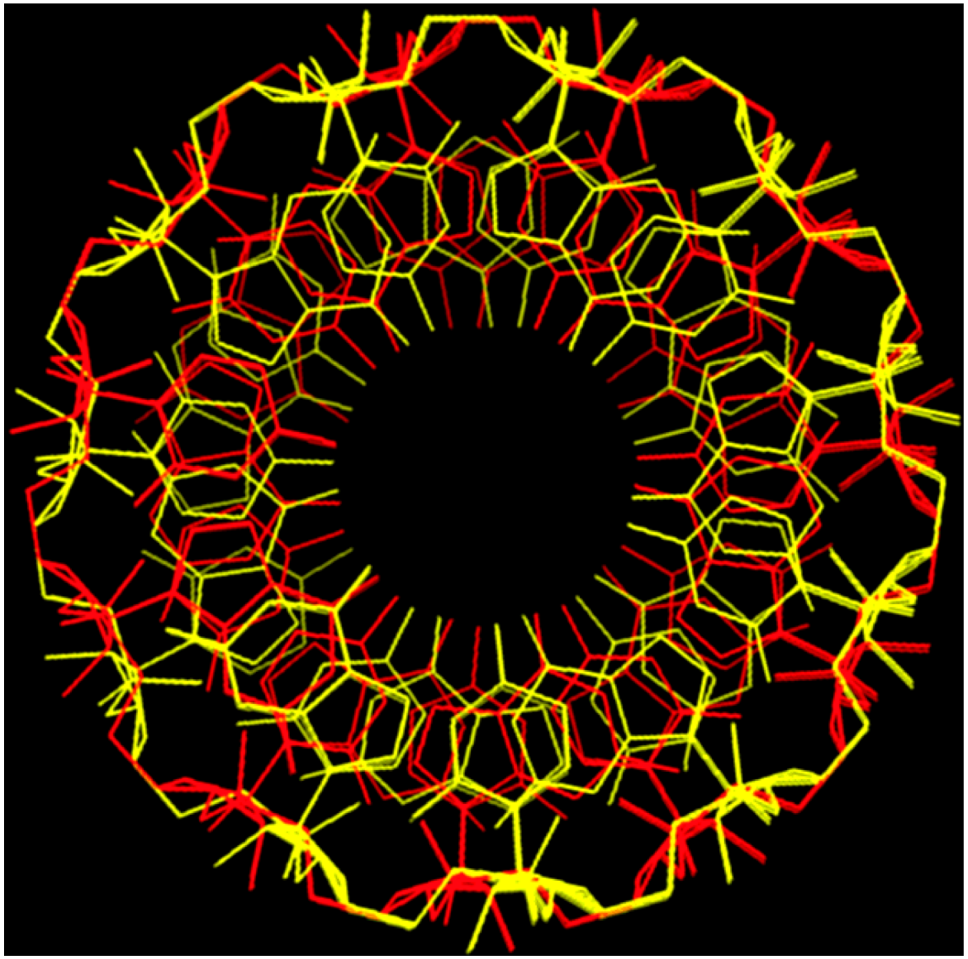
2D structures

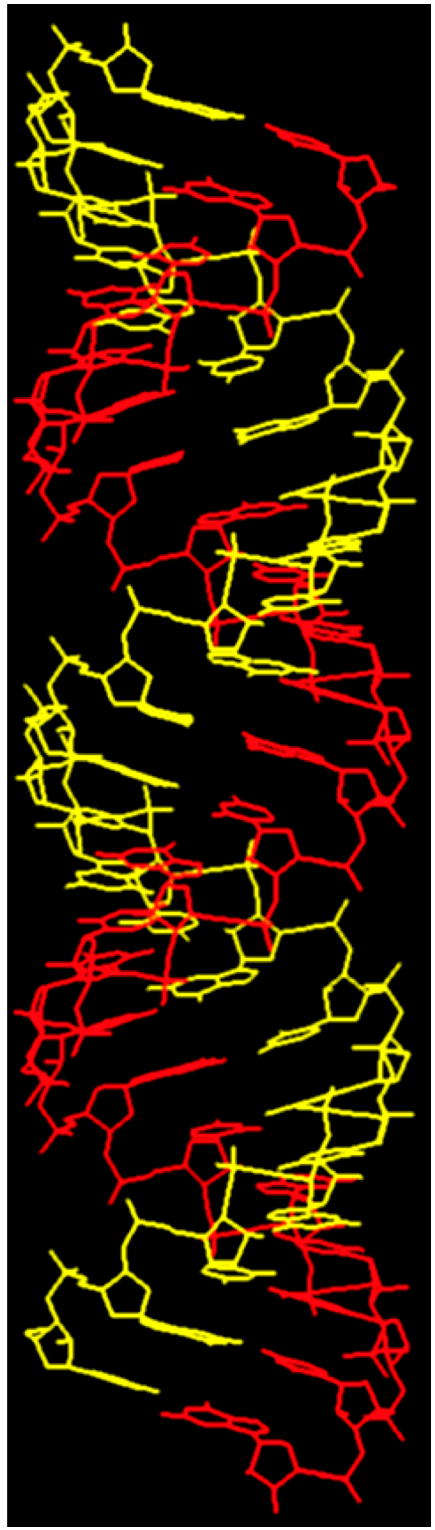
Watson-Crick base paired helices



Main building block

the RNA double helix held together by Watson-Crick pairs





2D structures

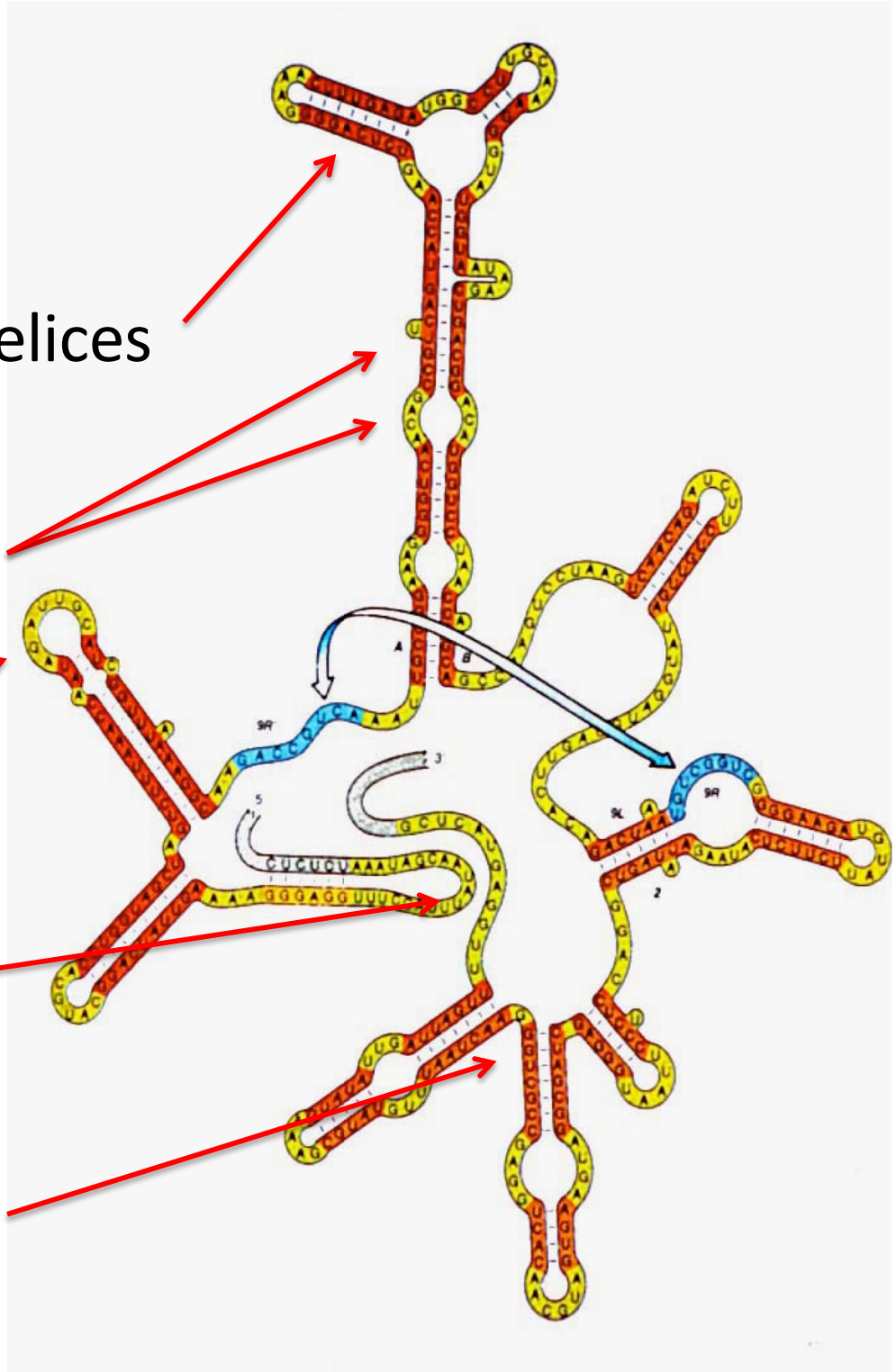
Watson-Crick base paired helices

Internal loops (symmetric, Asymmetric, bulge)

Hairpin loops

Single-strands junctions

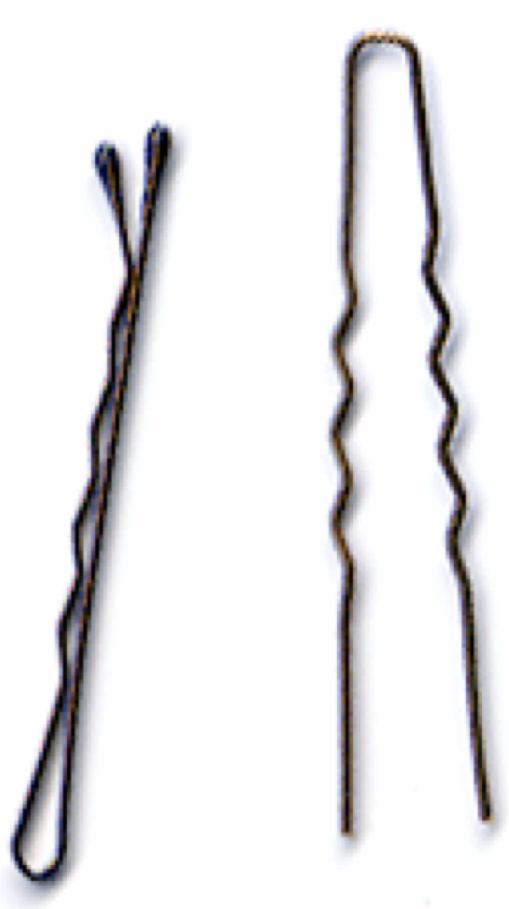
Multi-branched loops from which three or more stems radiate



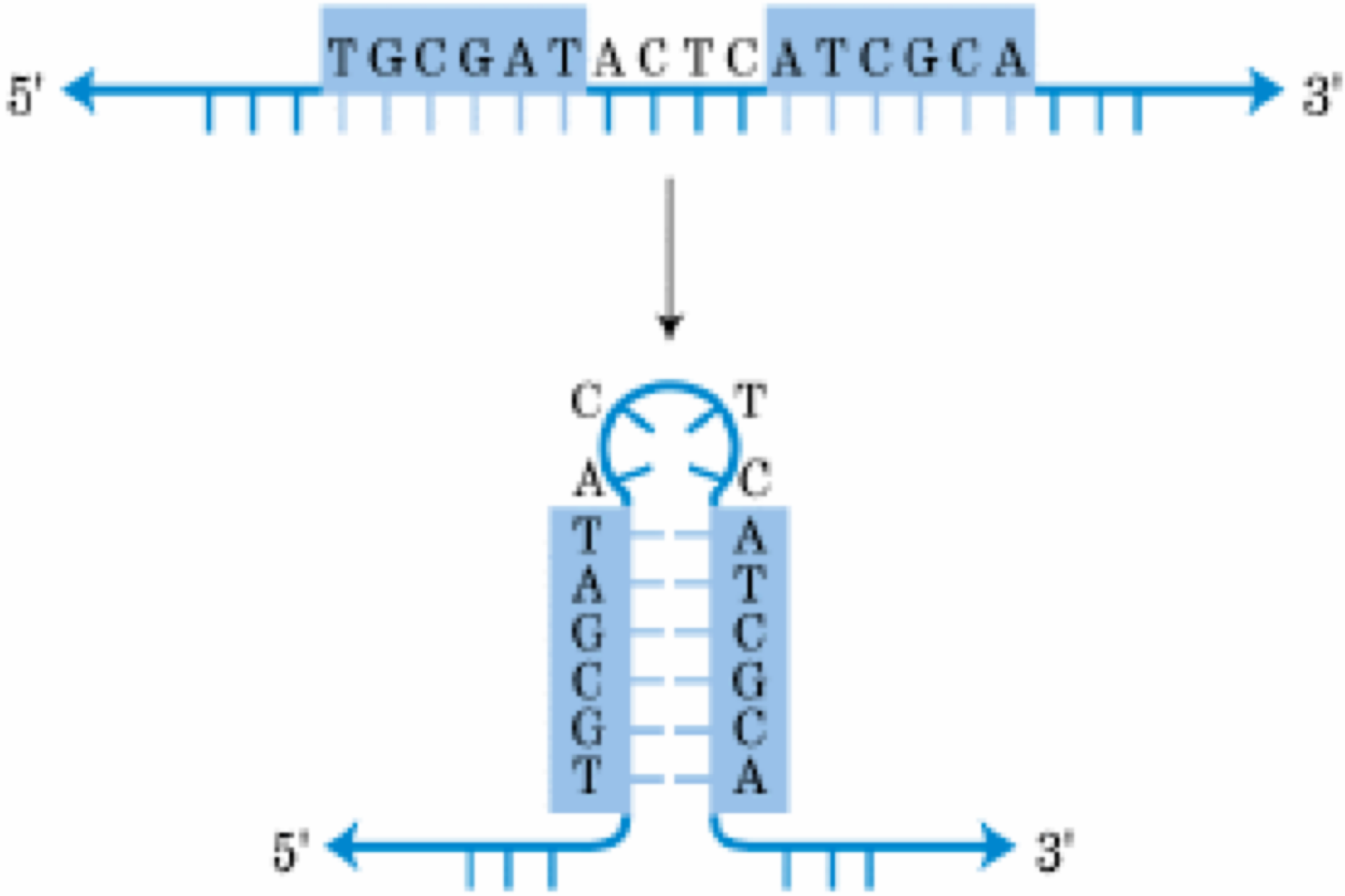
Hairpins

DNA strands with self-complementary base sequences have the potential to form hairpin structures. Formed only with a single DNA (or RNA) strand.

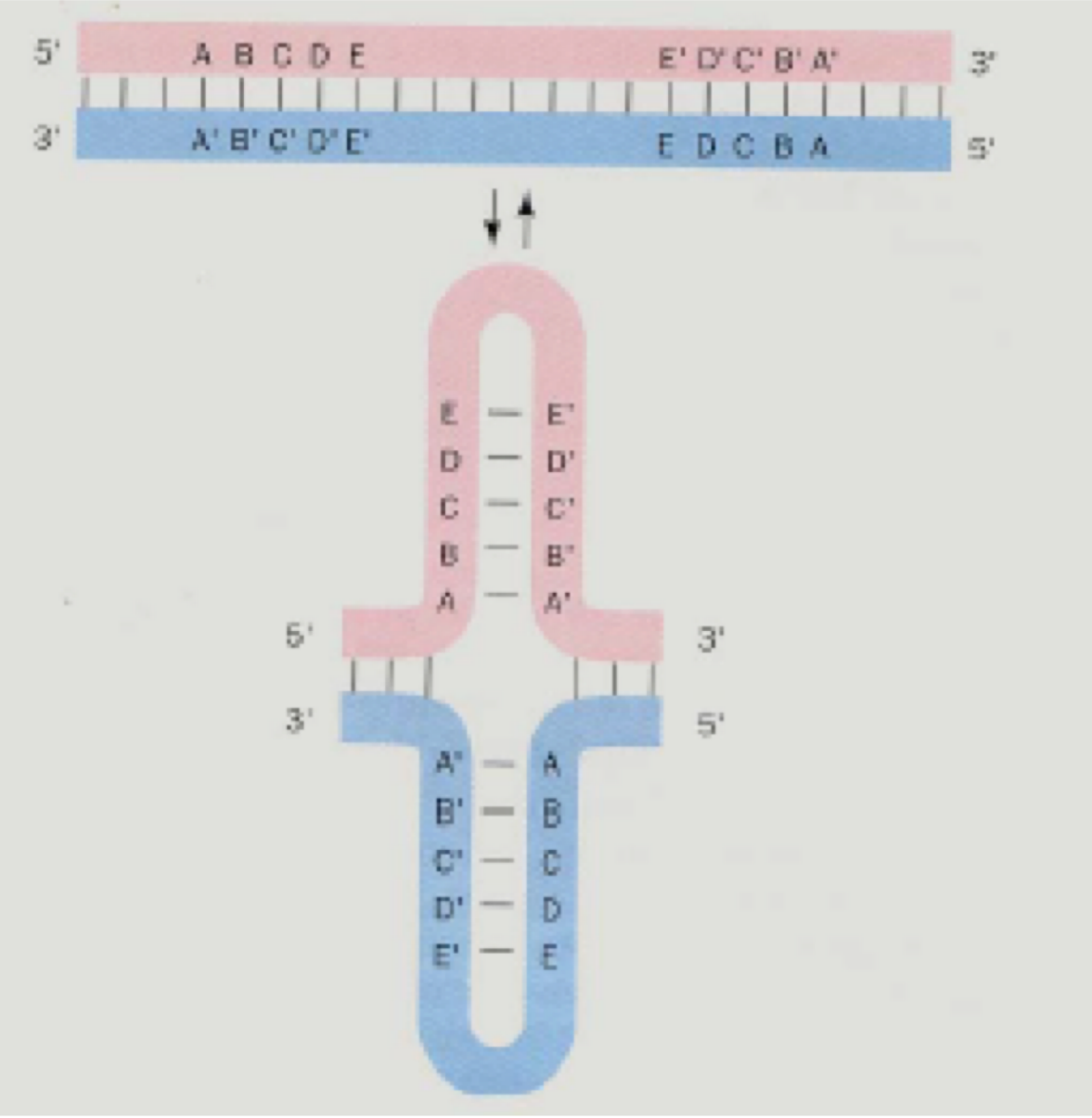
Hairpin is a common secondary/tertiary structure in RNA. It requires complementarity between part of the strand.

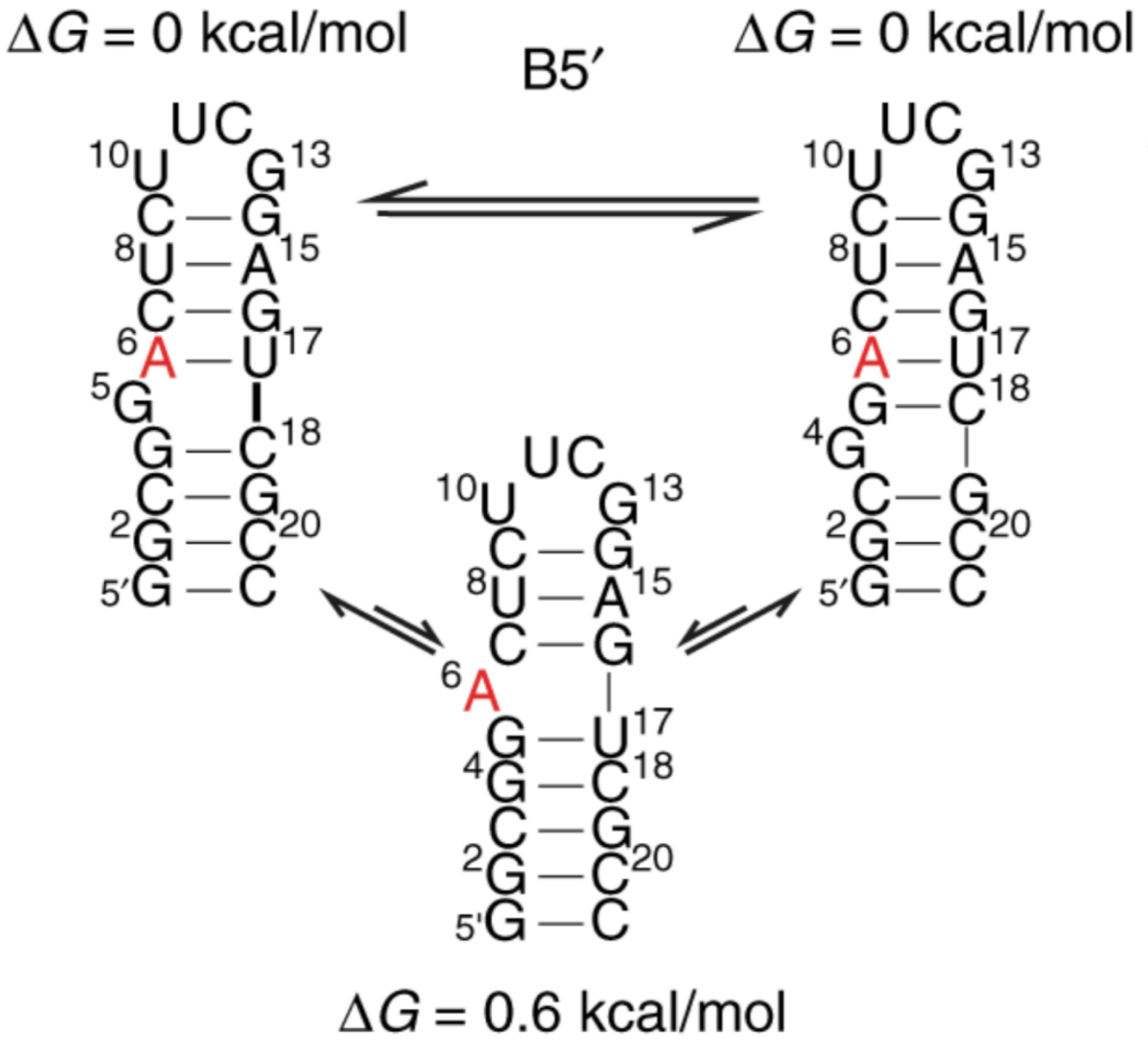


Hairpins

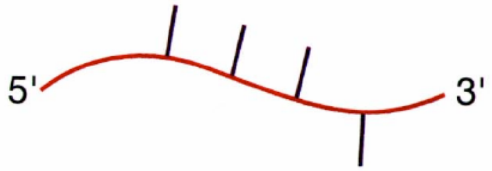


Hairpins

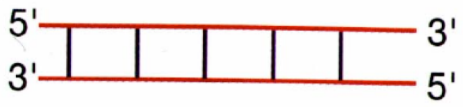



ΔG = 0 kcal/mol
B5′
ΔG = 0 kcal/mol
ΔG = 0.6 kcal/mol

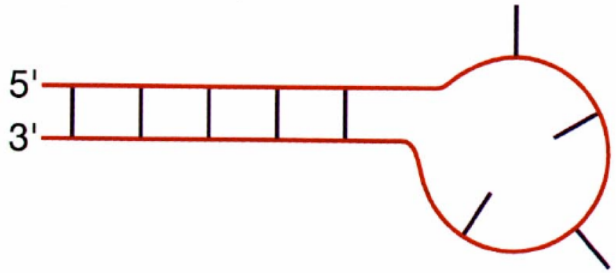
A. Single-stranded RNA



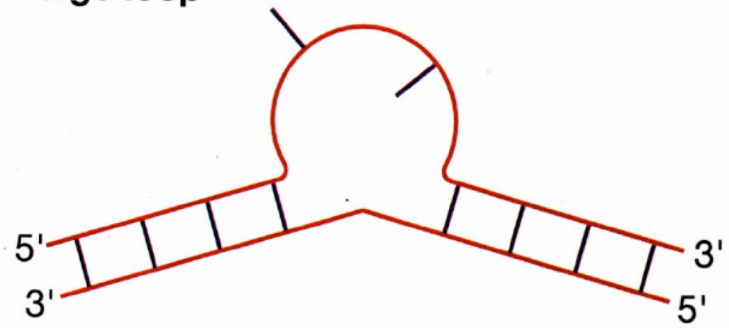
B. Double-stranded RNA helix of stacked base pairs



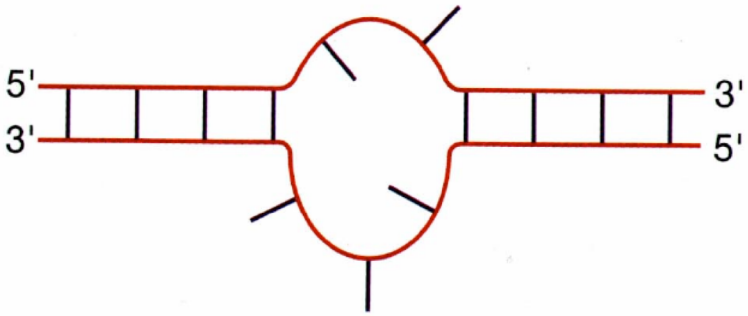
C. Stem and loop or hairpin loop.



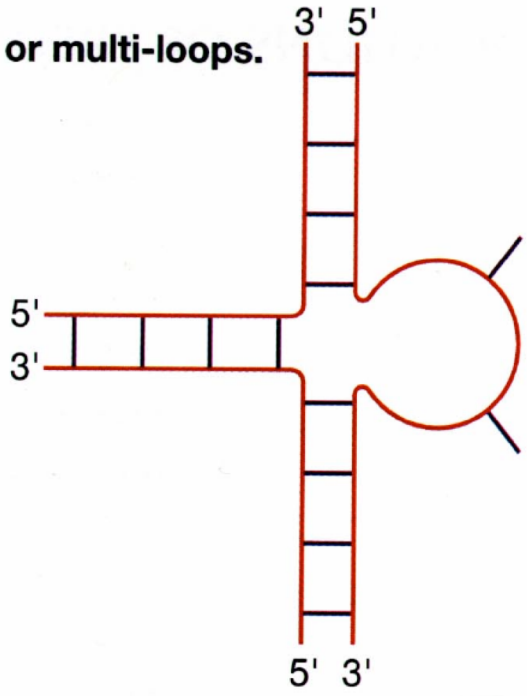
D. Bulge loop



E. Interior loop



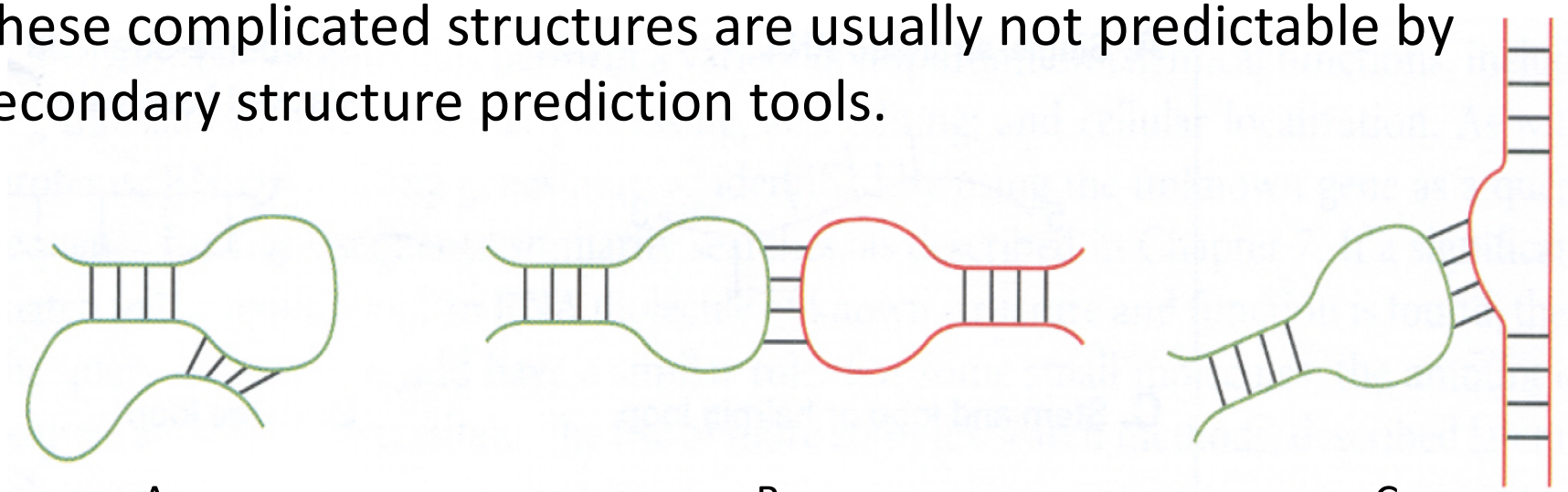
F. Junctions or multi-loops.



RNA tertiary structures

In addition to secondary structural interactions in RNA, there are also tertiary interactions.
pseudoknots (A)
kissing hairpins (B)
hairpin-bulge (C)

These complicated structures are usually not predictable by secondary structure prediction tools.



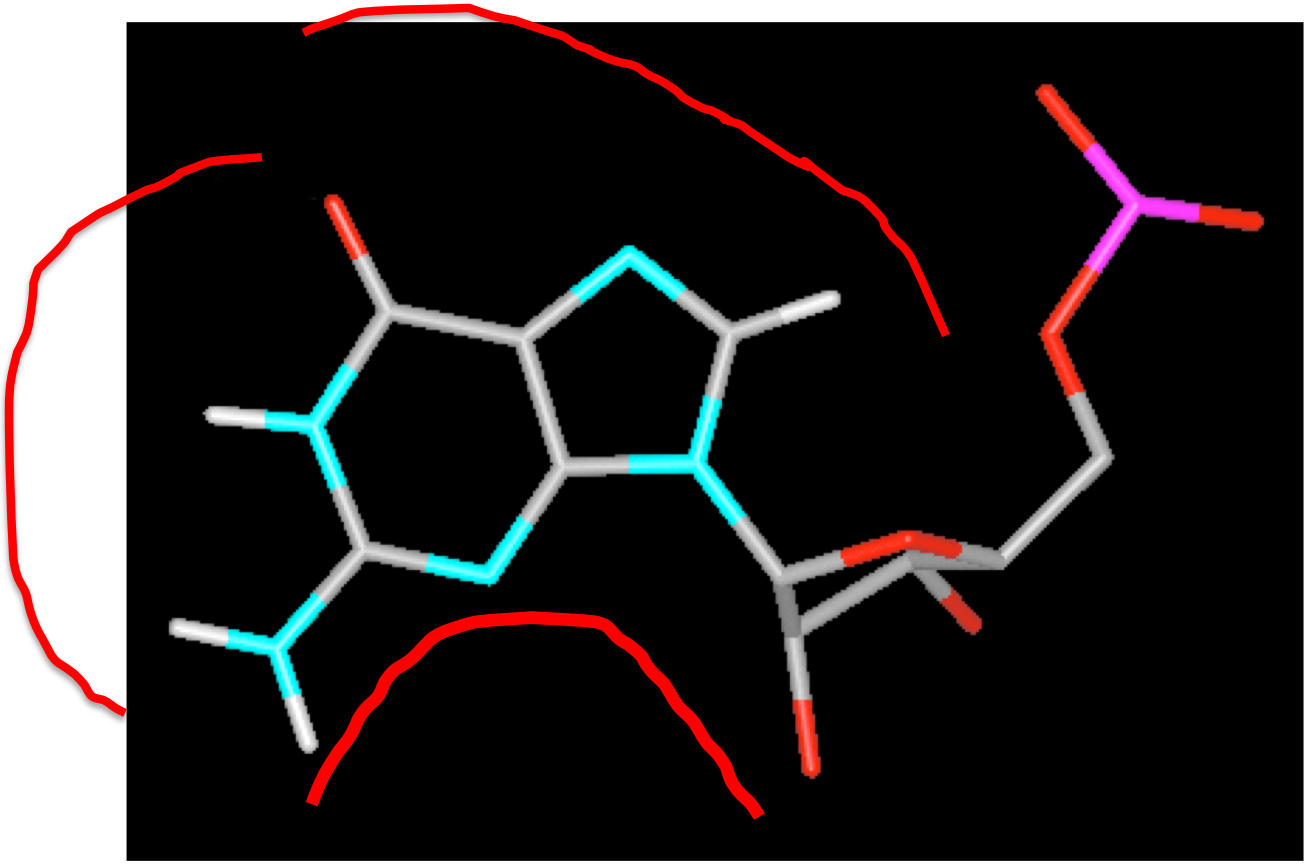
RNA base pairing

- Watson-Crick base pairs
  - Form double stranded helices
  - Define the 2D structure (Main building block)
  - Dependence on monovalent ions
- Non-Watson-Crick base pairs
  - Form RNA motifs
  - Responsible for RNA-RNA recognition & 3D fold
  - Dependence on Divalent ions (Mg2+)

Three Interacting Edges

Purins

Hoogsteen Edge

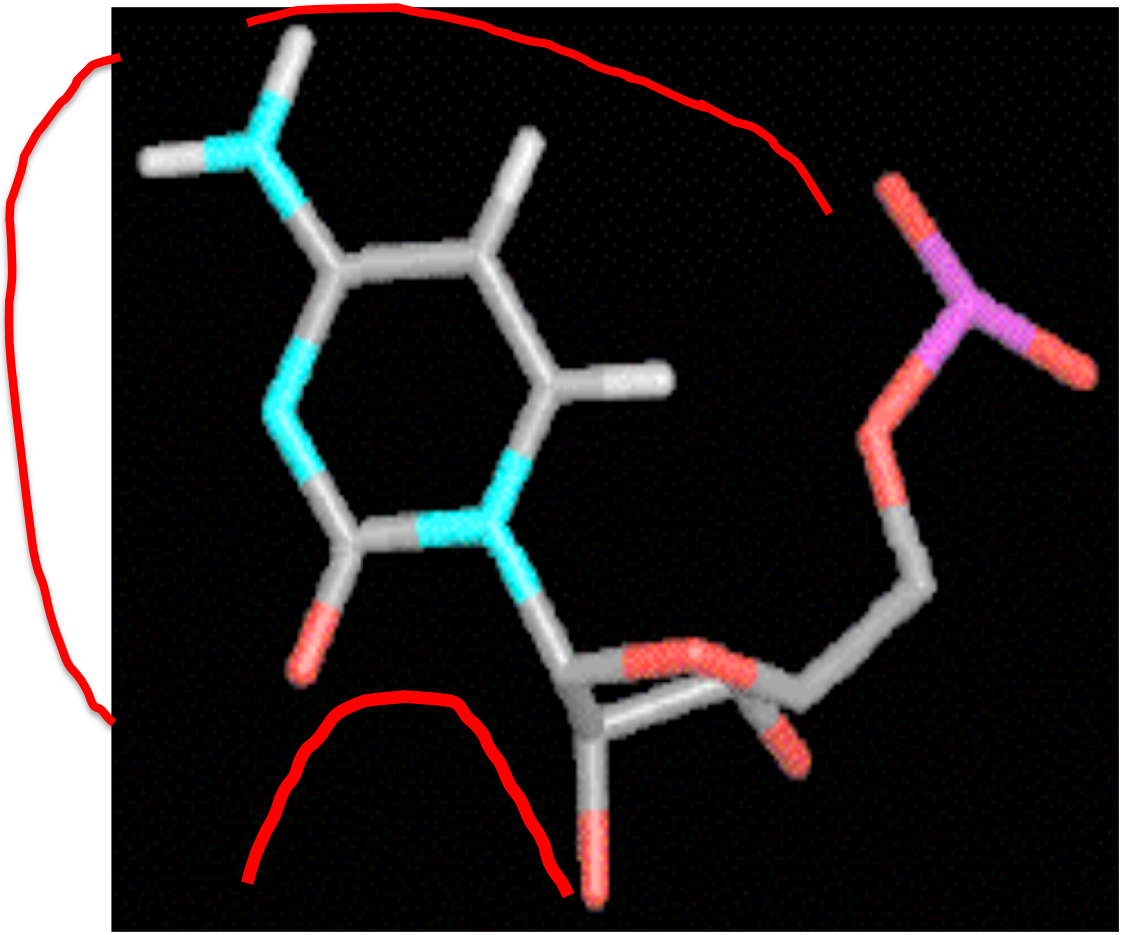


Watson-Crick Edge

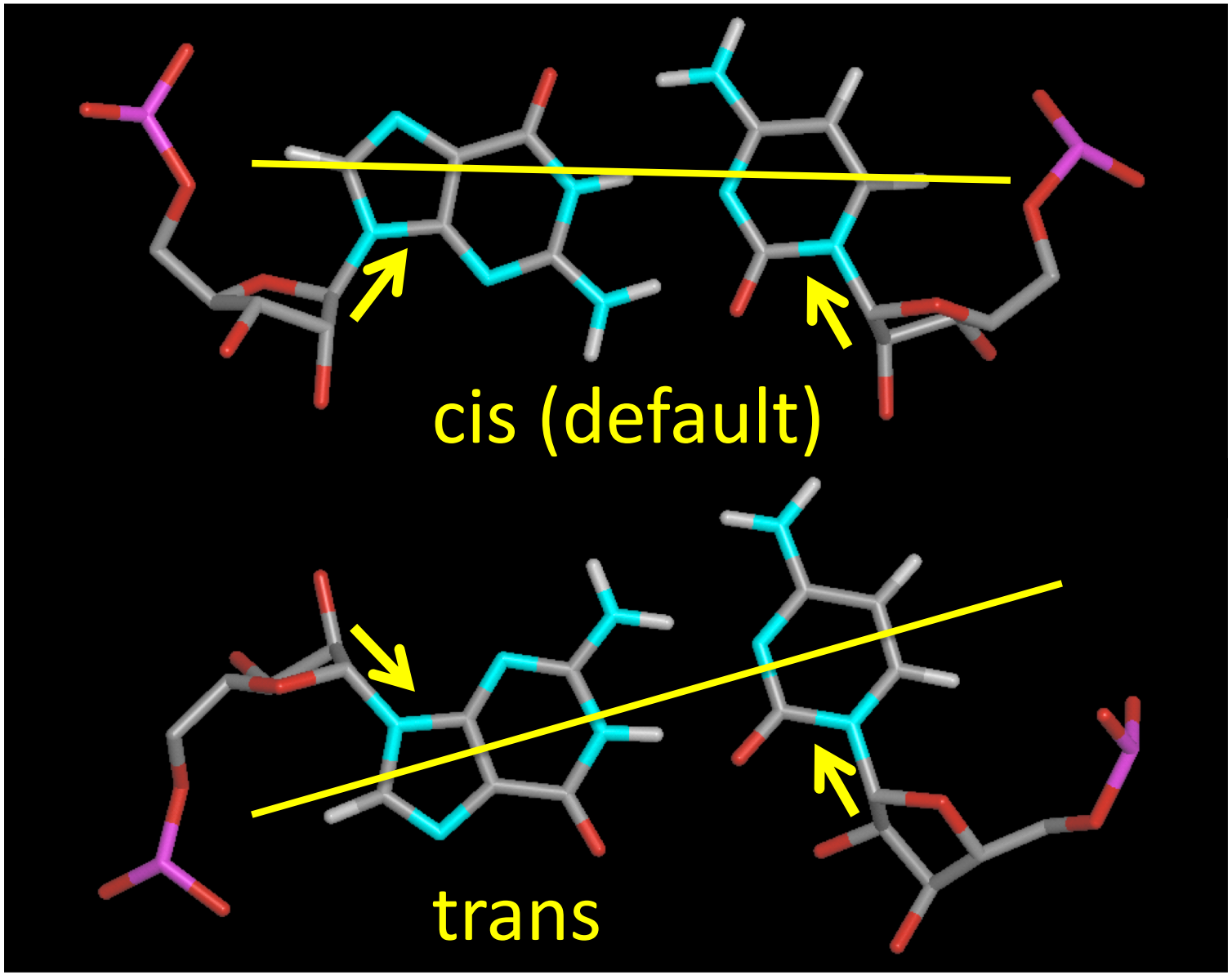
Sugar Edge

Three Interacting Edges

Pyrimidins



Glycosidic bond orientation



Edge-to-Edge Pairing Types

Watson-Crick
Hoogsteen
Sugar-edge

X

Watson-Crick
Hoogsteen
Sugar-edge

X

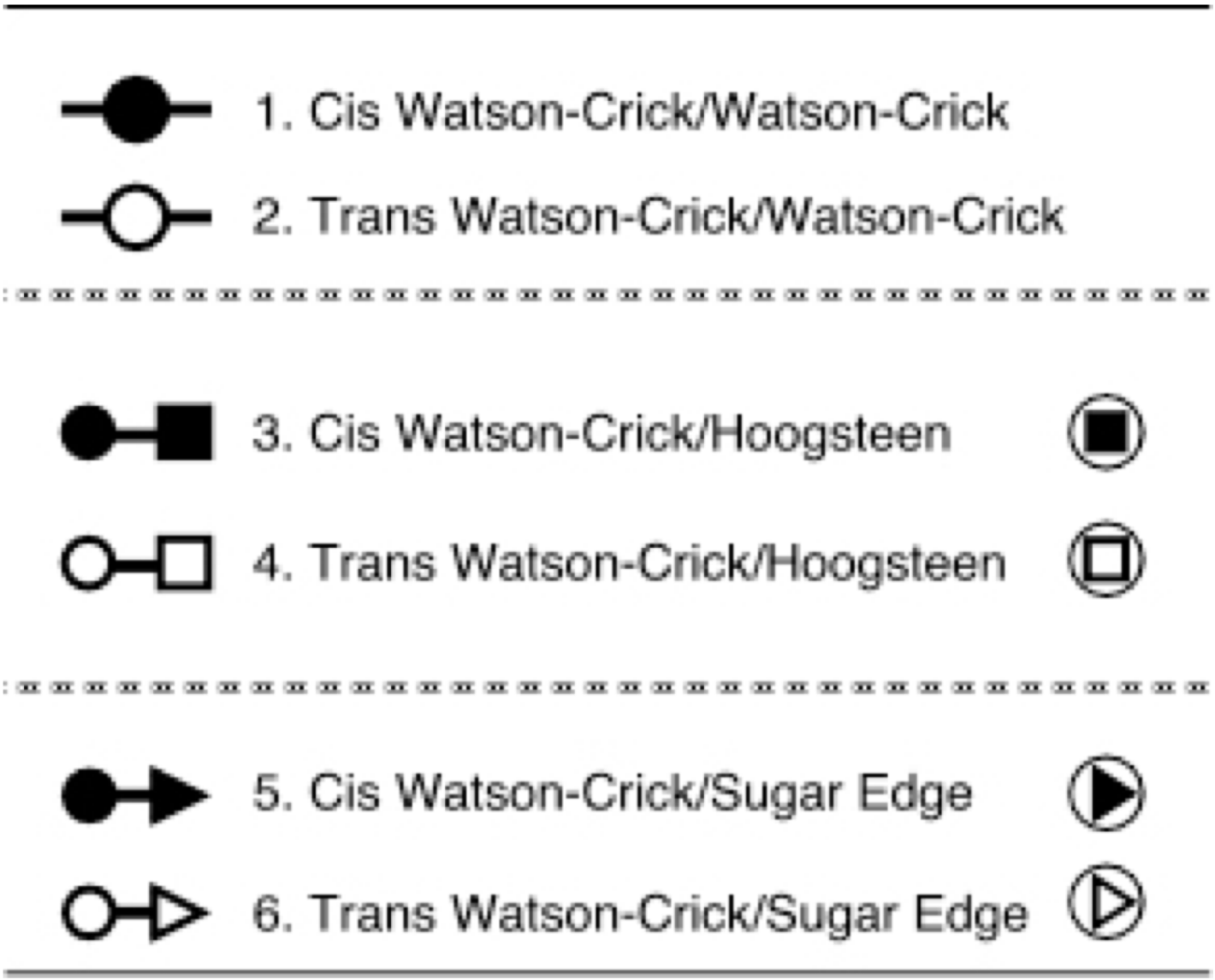
cis
trans

= 12 types

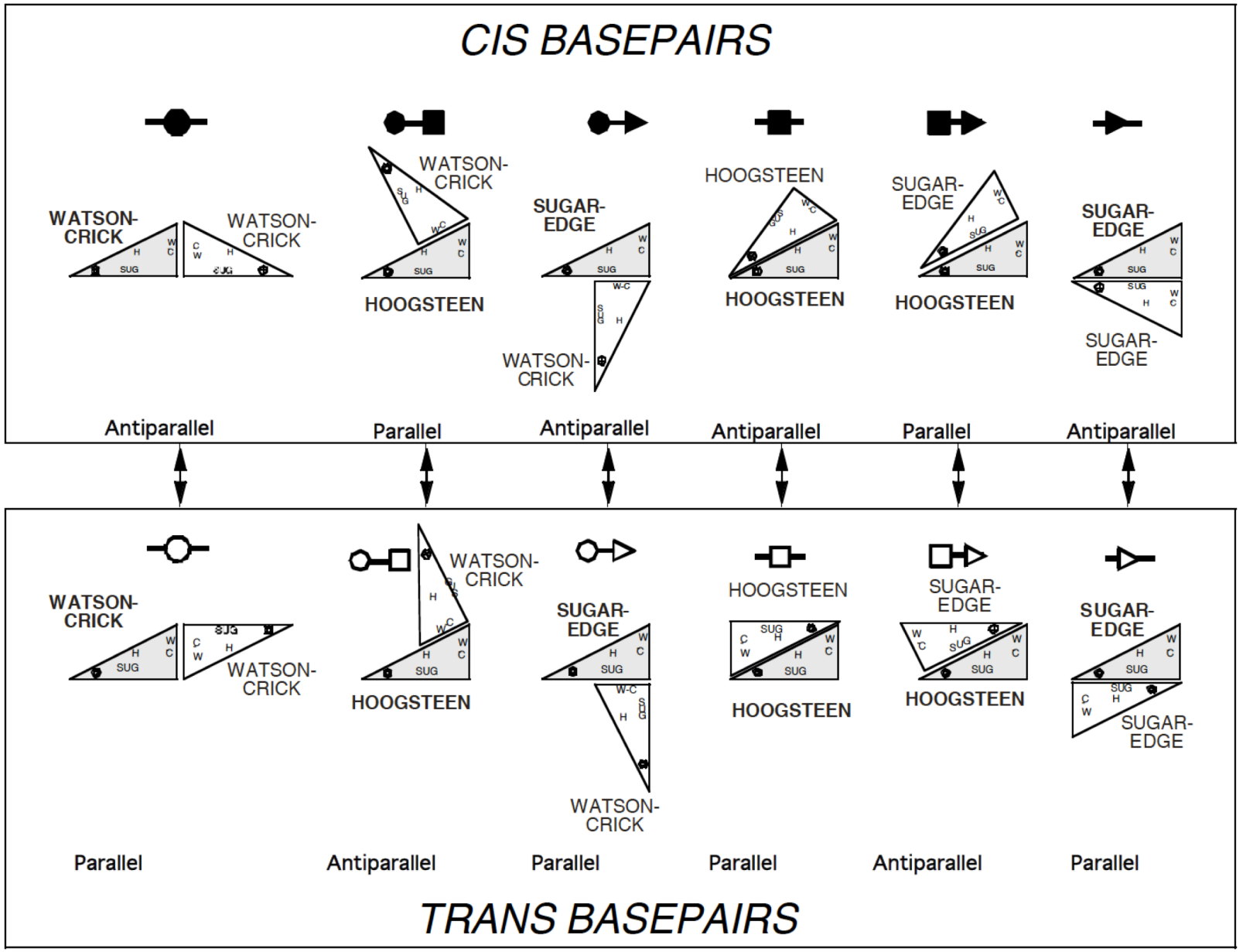
Edge-to-Edge Pairing Types

No.	Glycosidic bond orientation	Interacting edges	Local strand orientation
1	*Cis*	Watson–Crick/Watson–Crick	Antiparallel
2	*Trans*	Watson–Crick/Watson–Crick	Parallel
3	*Cis*	Watson–Crick/Hoogsteen	Parallel
4	*Trans*	Watson–Crick/Hoogsteen	Antiparallel
5	*Cis*	Watson–Crick/Sugar Edge	Antiparallel
6	*Trans*	Watson–Crick/Sugar Edge	Parallel
7	*Cis*	Hoogsteen/Hoogsteen	Antiparallel
8	*Trans*	Hoogsteen/Hoogsteen	Parallel
9	*Cis*	Hoogsteen/Sugar Edge	Parallel
10	*Trans*	Hoogsteen/Sugar Edge	Antiparallel
11	*Cis*	Sugar Edge/Sugar Edge	Antiparallel
12	*Trans*	Sugar Edge/Sugar Edge	Parallel

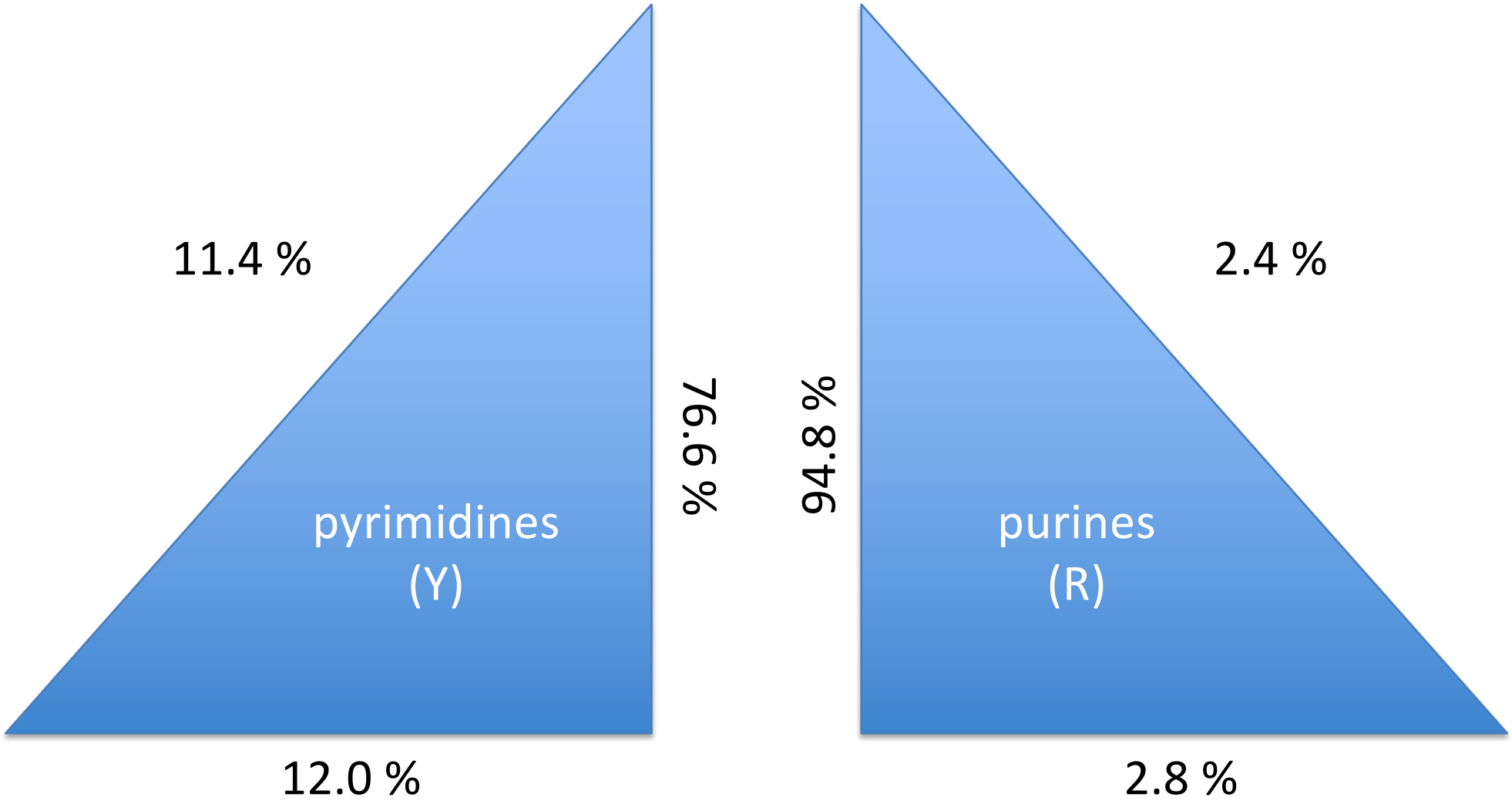
Annotations for Non-Watson-Crick Pairs



Annotations for Non-Watson-Crick Pairs



statistics…



- What is a RNA motif ?
- How can we detect the presence of a motif in a given RNA ?
- How can we compare motifs ?

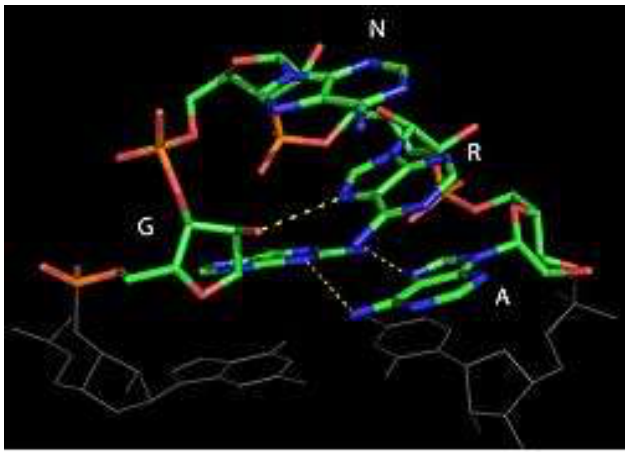
RNA motif

A RNA motif is an ensemble of ordered elements under constraints.

- Sequential motifs :

  Strict : -AUG-

  Fuzzy : -AAUAxAA-



- Structural motifs:

  GNRA, UNCG, CUUG (tetraloops)

  Boxes C/D or H/ACA (snoRNAs)

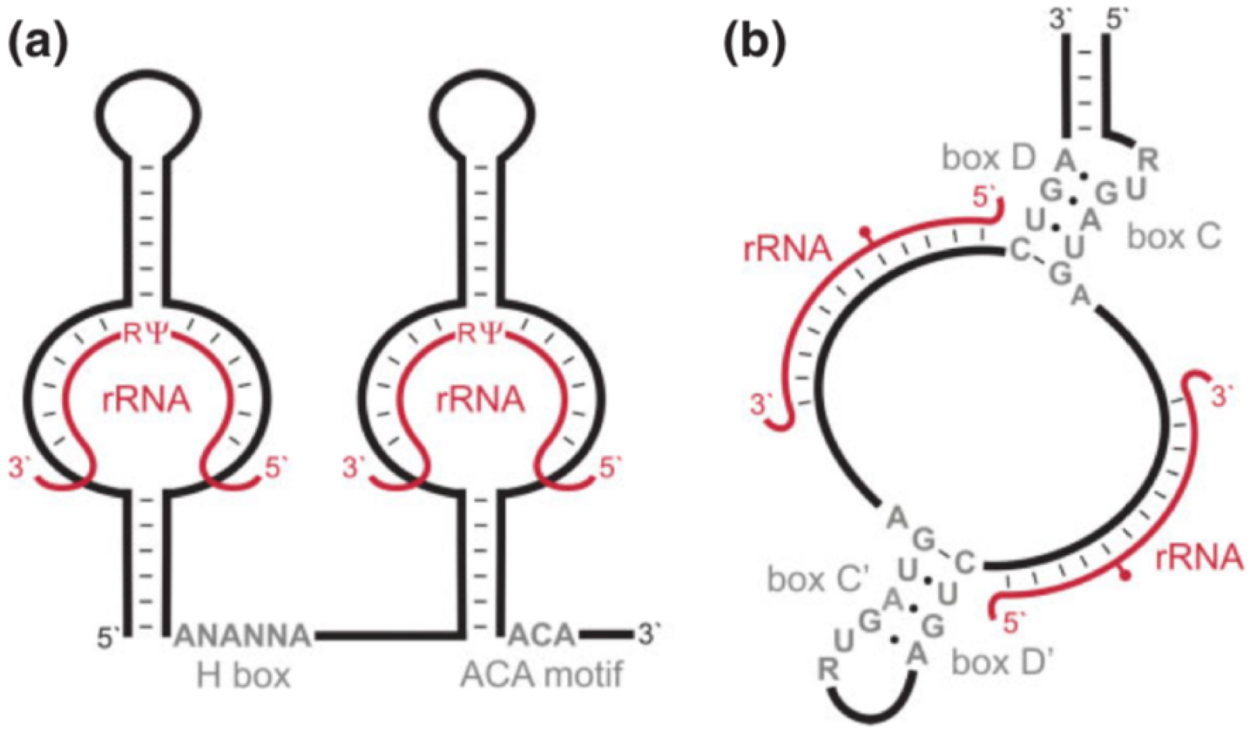


FIGURE 1 | Small nucleolar RNA (snoRNA) structure and function. The schematic structure of H/ACA (a) and box C/D (b) snoRNAs is shown. The conserved box sequences are shown in gray and the rRNA target RNAs in red. The pseudouridylated or methylated residue is indicated by a Ψ (a) or a red circle (b), respectively.

WIREs RNA 2012, 3:397–414. doi: 10.1002/wrna.117

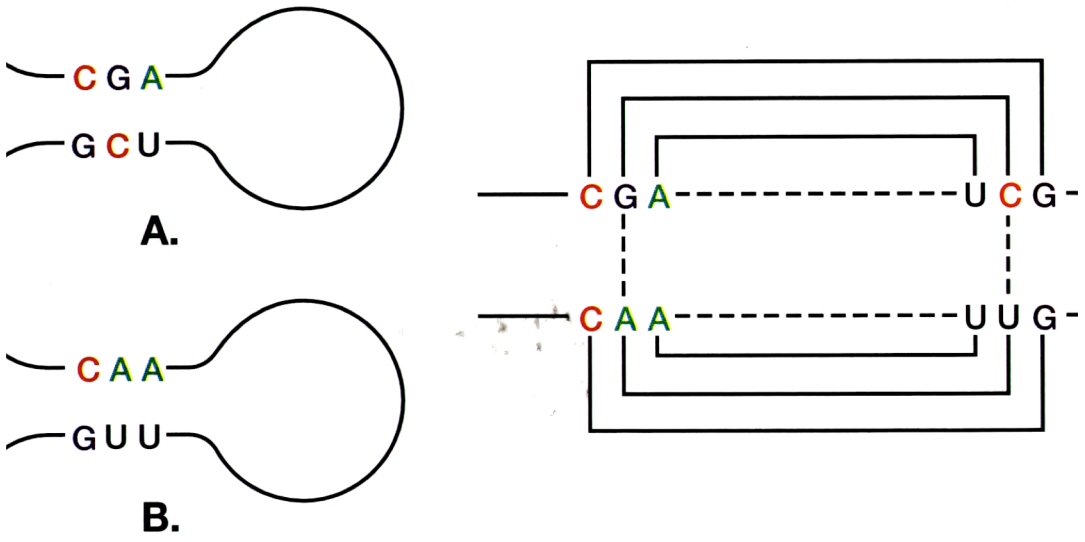
Evolution laws

- Three-dimensional architectures evolve less with time than sequences
- Three-dimensional structures are dictated first by folding rules and secondarily by function
- The phonetic structure of words are more stable than the meaning of words

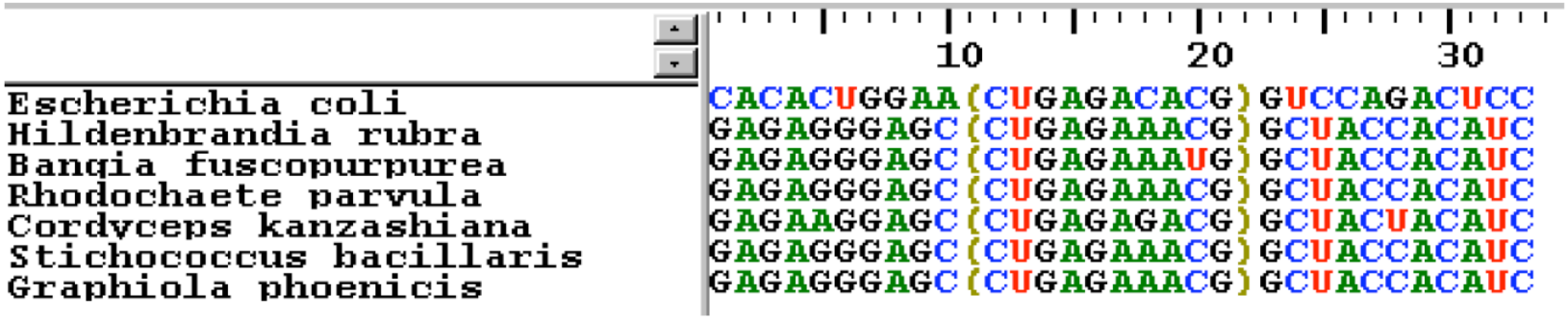
RNA alignments

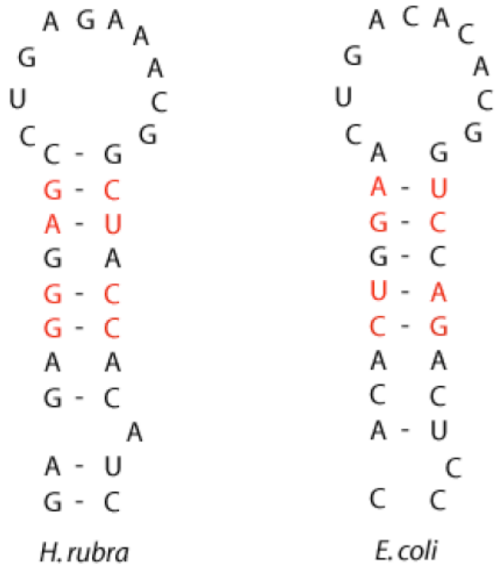
RNA sequences are aligned/compared differently because sequence variation in RNA maintain base-pairing patterns

Alignments of RNA sequences will show covariation at interacting base-pair positions

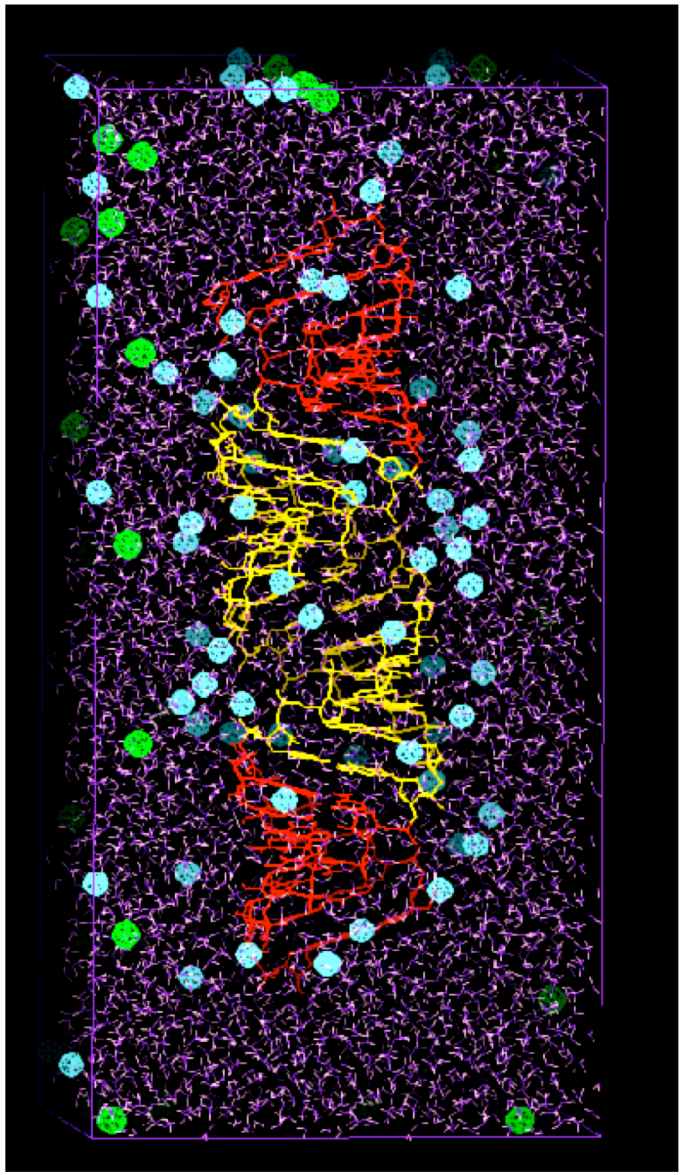


Covariation





… RNA folding procedures…



- Water molecules
- Counter-ions
- Co-ions
- Polyamines, …

RNA base pairing

- Watson-Crick base pairs
  - Form double stranded helices
  - Define the 2D structure (Main building block)
  - Dependence on monovalent ions
- Non-Watson-Crick base pairs
  - Form RNA motifs
  - Responsible for RNA-RNA recognition & 3D fold
  - Dependence on Divalent ions (Mg2+)

RNA/ion interactions

Divalent cations	Monovalent cations	Anions
		Cl^-
Mg^{2+}	Na^+	$SO4^{2-}$
Mn^{2+}	K^+	…
Ca^{2+}	Rb^+	
Sr^{2+}	Cs^+	
…	Tl^+	
	…	

RNA folding procedures

In vitro folding:

Kinetic vs. thermodynamic control

In vivo folding:

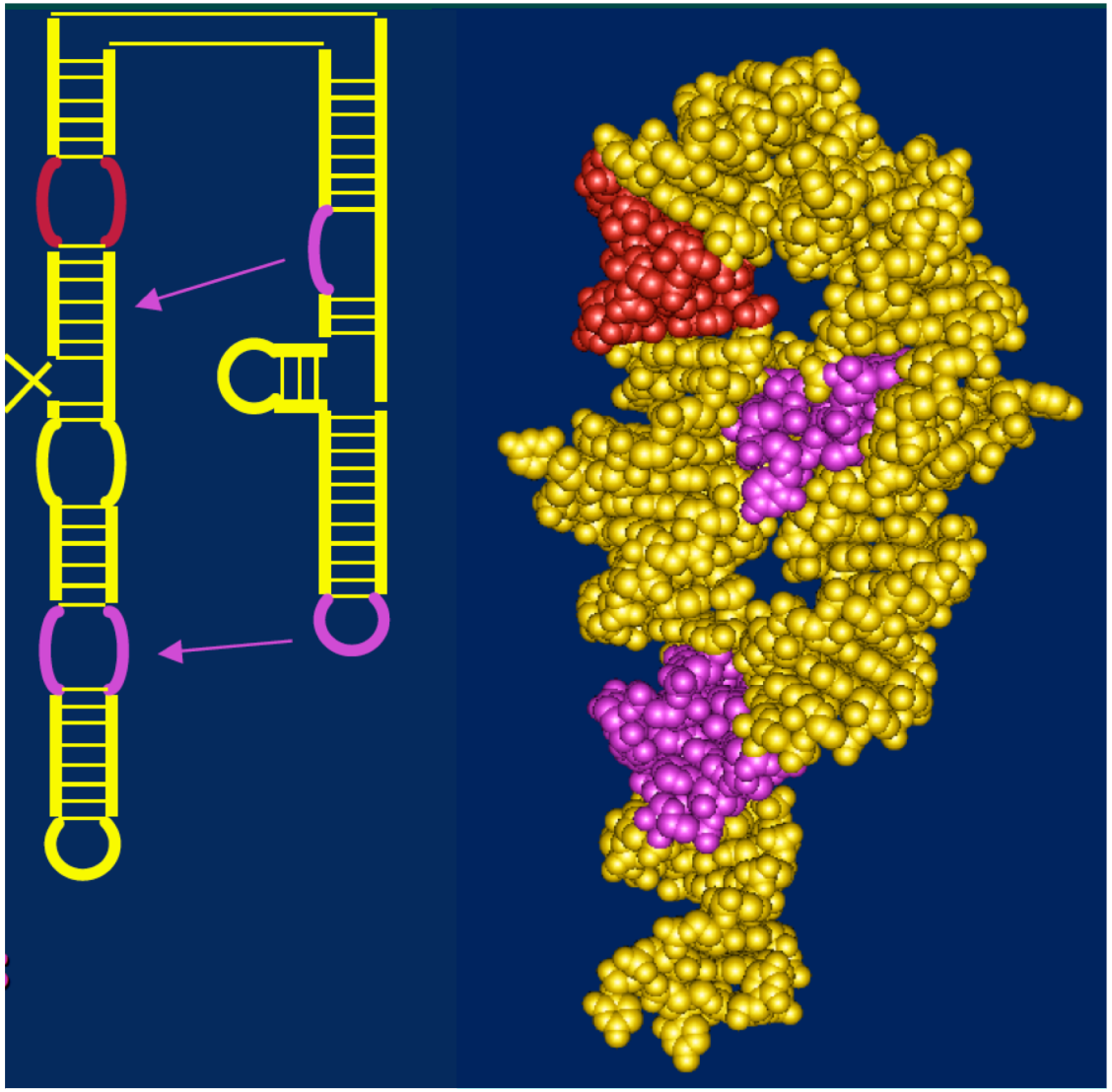
Sequential 5'>3' or co-transcriptional

Modular and hierarchical

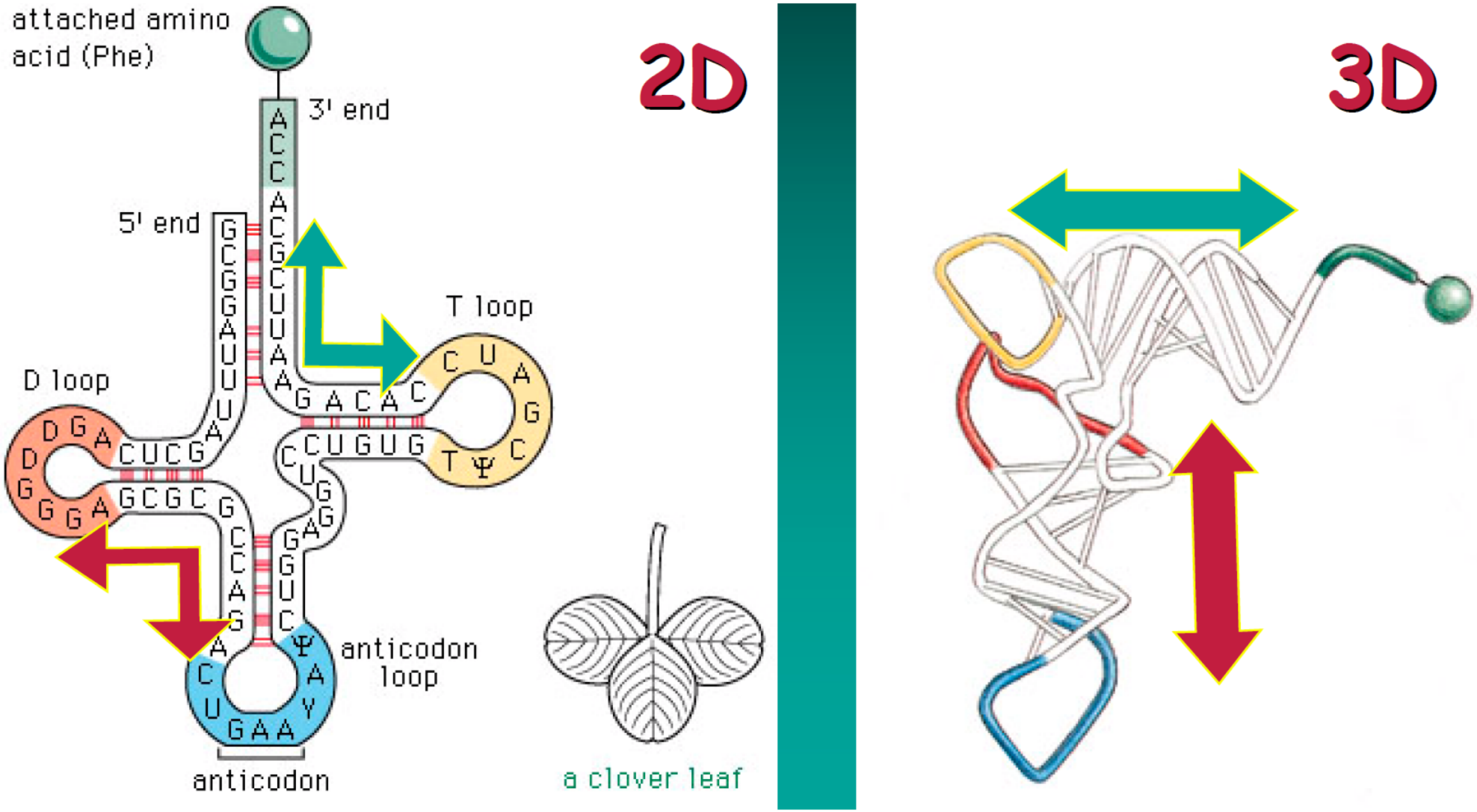
Architectural hierarchy by modular assembly in RNA

- Helices and hairpin loops first form
- Helices build sub-domains by parallel or end-to-end packing
- Local and specific recognition contacts occur cooperatively between preformed sub-domains.

Parallel packing of helices



End-to-end stacking of helices

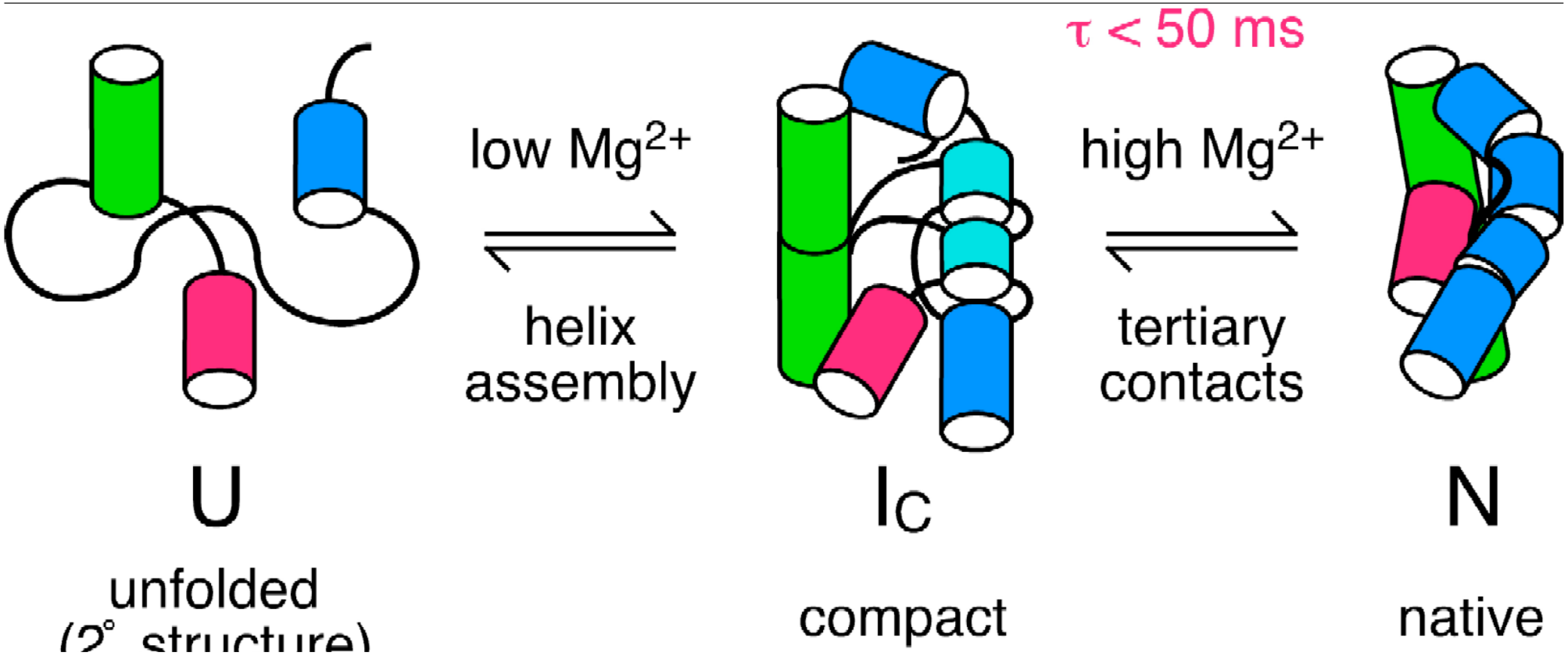


RNA self-assembly & folding

Coupled Architectural & Electrostatic Hierarchies

- Formation of helices that build subdomains by parallel or end-to-end packing & rapid collapse to compact states induced by non-specific ion binding;
- Specific RNA-RNA recognition & cooperative transitions to native state promoted by specific ion binding.

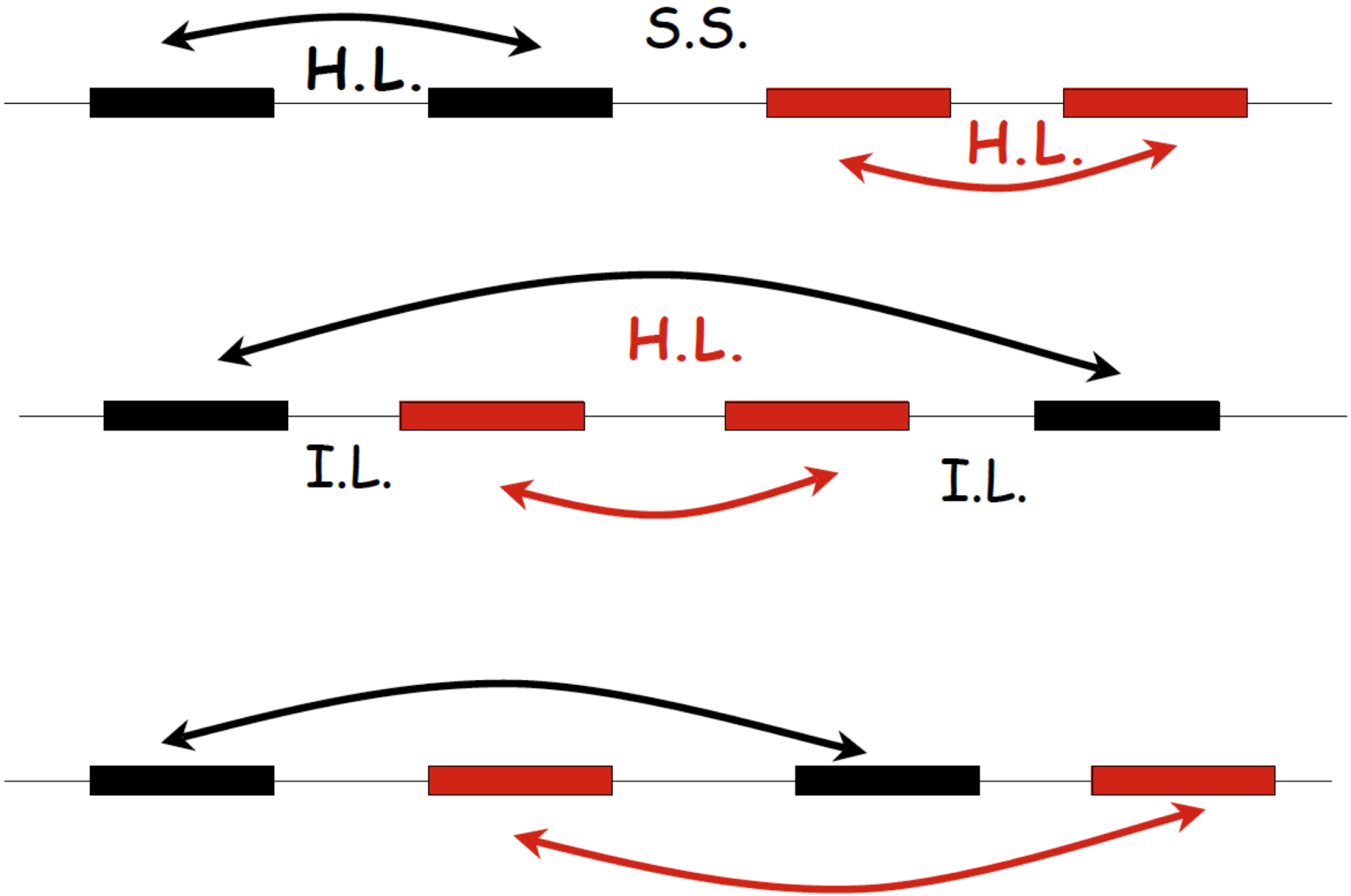
RNA self-assembly & folding



Kinetic values…

- Stacking of single-strands : 1 ms
- Hairpin formation : 10 -100 ms
- Tertiary structure formation : 10 -100 ms
- Native state : 1s - 10 min

Only three ways to pair four segments



Modeling algorithms

- 3D structure : assembly of fragments
- Stress 3D fold rather than sequence (inverse folding)
- Search for a «!consensus!» 3D fold (global architecture)
- 2D Topology (not strongly correlated with sequence) - RNA is right-handed > right-handedness of stacks, of junctions

Basics of RNA structure prediction

Two primary methods of structure prediction

– Covariation analysis/Comparative sequence analysis

- Takes into account conserved patterns of basepairs during evolution (2 or more sequences).
- Pairs will vary at same time during evolution yet maintaining structural integrity
- Manifestation of secondary structure

– Minimum Free-Energy Method

- Using one sequence can determine structure of complementary regions that are energetically stable

Comparative Sequence Analysis

Molecules with similar functions and different nucleotide sequences will form similar structures

- Predicts secondary and tertiary structure from underlying sequence
- Correctly identifies high percentage secondary structure pairings and a smaller number of tertiary interactions
- Primarily a manual method

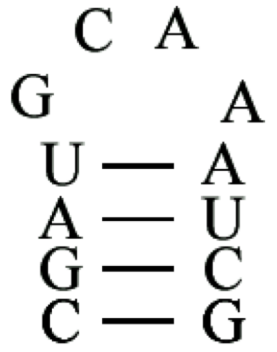
Positional Covariation

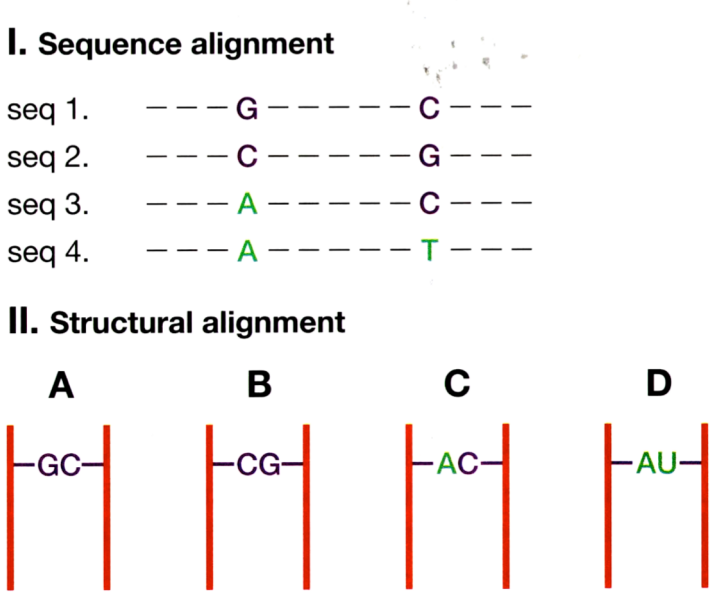
Helix is formed from two sets of sequences that are not identical.

C G A U (G C A A) A U C G

Search for positions that co-vary.

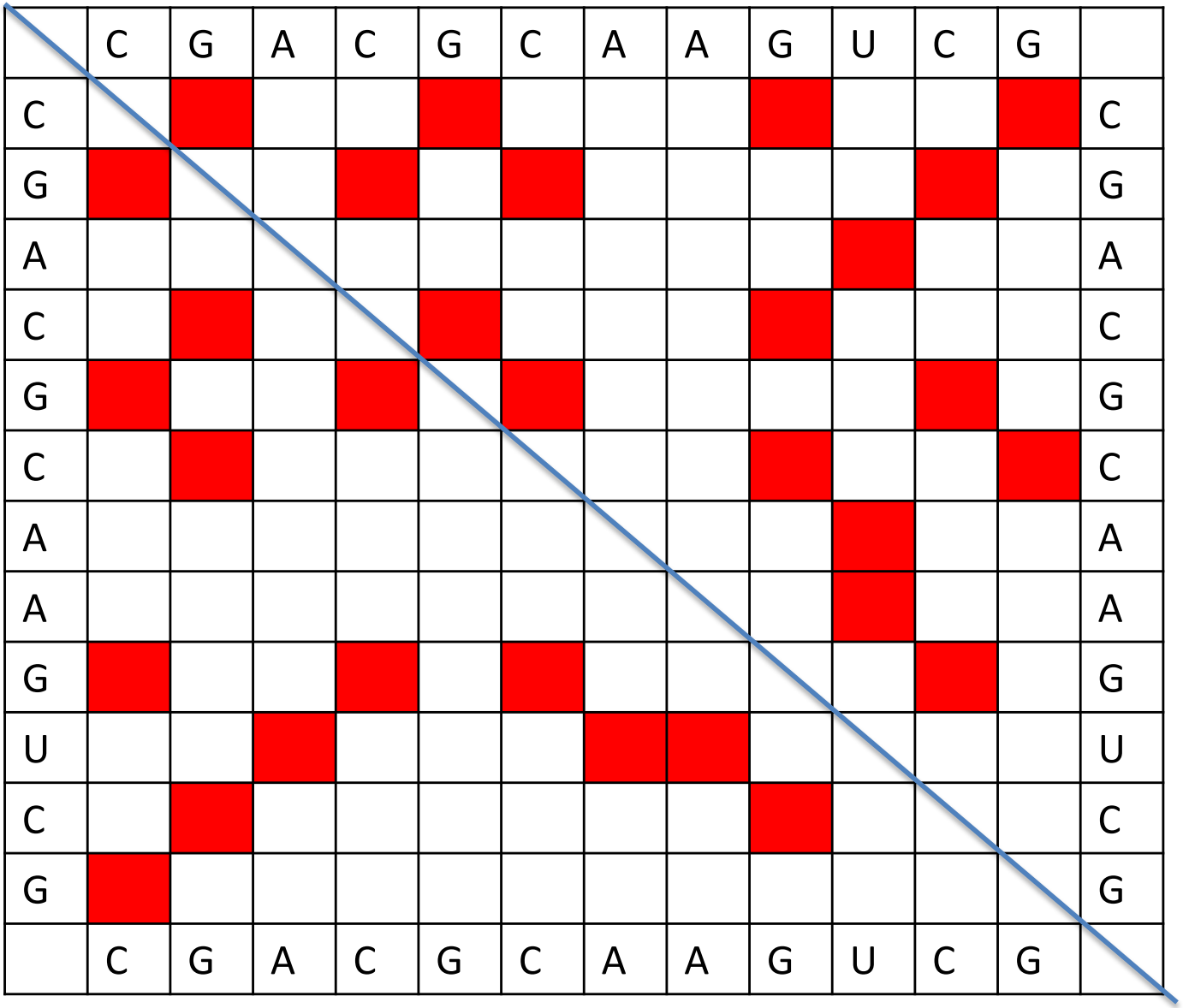
Positions that co-vary with one another are possible pairing partners.





Minimum Free energy method

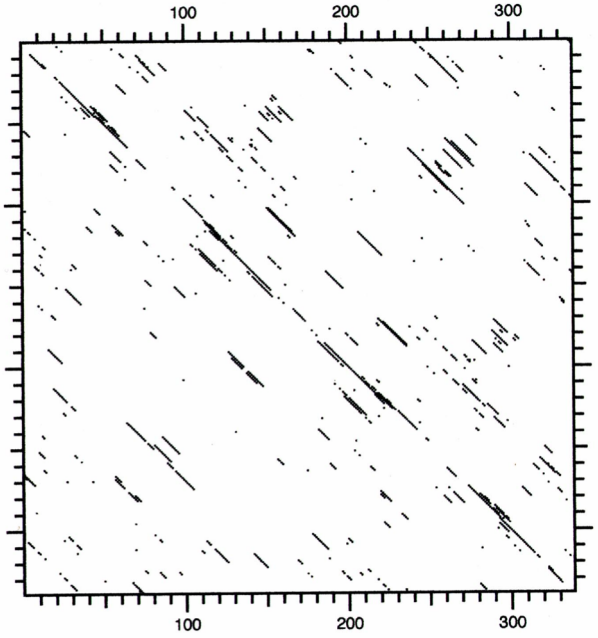
G



Minimum Free energy method

Hypothesis:

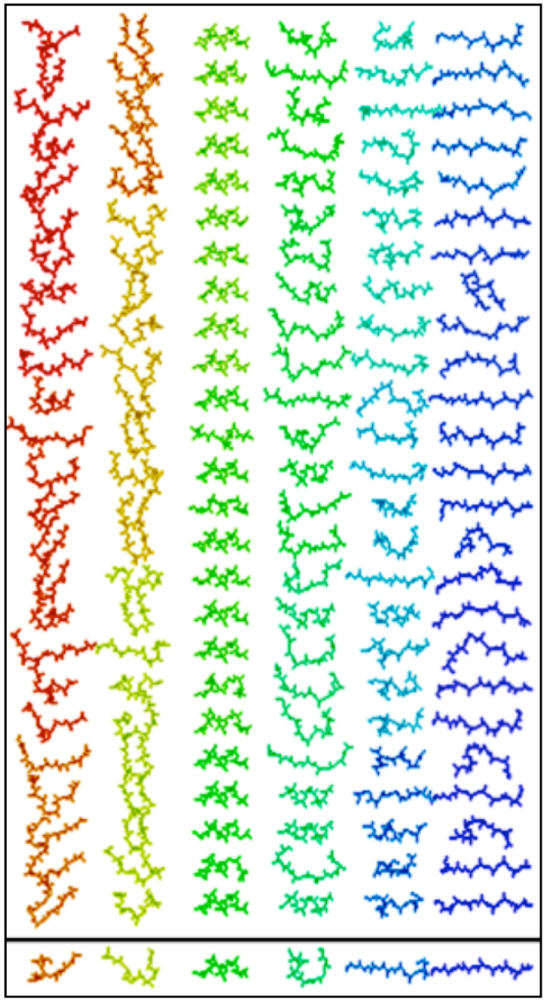
- The native secondary structure is the one with the minimum free energy
- Searching for structures with stable energies
- First a dot matrix analysis is carried out to highlight complementary regions (diagonal indicates succession of complementary nucleotides)
- The energy is then calculated for each predicted structure by summing negative base stacking energies

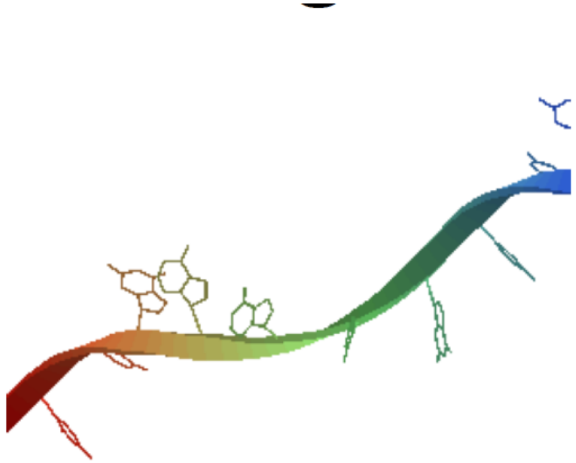


Minimum Free energy method

- Assumption: 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 of the base pair free energies.

De novo modeling





Fragment **A**ssembly of **RNA** (**FARNA**)

Single sequence secondary structure prediction

CentroidHomfold Secondary structure prediction by using homologous sequence information

CyloFold Secondary structure prediction method based on placement of helices allowing complex pseudoknots.

GTFold Fast and scalable multicore code for predicting RNA secondary structure.

IPknot Fast and accurate prediction of RNA secondary structures with pseudoknots using integer programming.

KineFold Folding kinetics of RNA sequences including pseudoknots by including an implementation of the partition function for knots.

Mfold MFE (Minimum Free Energy) RNA structure prediction algorithm.

RNA123 Secondary structure prediction via thermodynamic-based folding algorithms and novel structure-based sequence alignment specific for RNA.

RNAstructure A program to predict lowest free energy structures and base pair probabilities for RNA or DNA sequences. Programs are also available to predict Maximum Expected Accuracy structures and these can include pseudoknots. Structure prediction can be constrained using experimental data, including SHAPE, enzymatic cleavage, and chemical modification accessibility.

Sfold Statistical sampling of all possible structures. The sampling is weighted by partition function probabilities.

...

RNA homology search software

ERPIN "Easy RNA Profile IdentificatioN" is an RNA motif search program reads a sequence alignement and secondary structure, and automatically infers a statistical "secondary structure profile" (SSP). An original Dynamic Programming algorithm then matches this SSP onto any target database, finding solutions and their associated scores.

Infernal "INFERence of RNA ALignment" is for searching DNA sequence databases for RNA structure and sequence similarities. It is an implementation of a special case of profile stochastic context-free grammars called covariance models (CMs).

GraphClust Fast RNA structural clustering method to identify common (local) RNA secondary structures. Predicted structural clusters are presented as alignment. Due to the linear time complexity for clustering it is possible to analyse large RNA datasets.

PHMMTS "pair hidden Markov models on tree structures" is an extension of pair hidden Markov models defined on alignments of trees.

RaveNnA A slow and rigorous or fast and heuristic sequence-based filter for covariance models.

RSEARCH Takes a single RNA sequence with its secondary structure and utilizes a local alignment algorithm to search a database for homologous RNAs.

Structator Ultra fast software for searching for RNA structural motifs employing an innovative index-based bidirectional matching algorithm combined with a new fast fragment chaining strategy.

Benchmarking

Name	Description	Structure[Note 1]	Alignment[Note 2]	Phylogeny
BRalibase I	A comprehensive comparison of comparative RNA structure prediction approaches	yes	no	no
BRalibase II	A benchmark of multiple sequence alignment programs upon structural RNAs	no	yes	no
BRalibase 2.1	A benchmark of multiple sequence alignment programs upon structural RNAs	no	yes	no
BRalibase III	A critical assessment of the performance of homology search methods on noncoding RNA	no	yes	no
CompaRNA	An independent comparison of single-sequence and comparative methods for RNA secondary structure prediction	yes	no	no

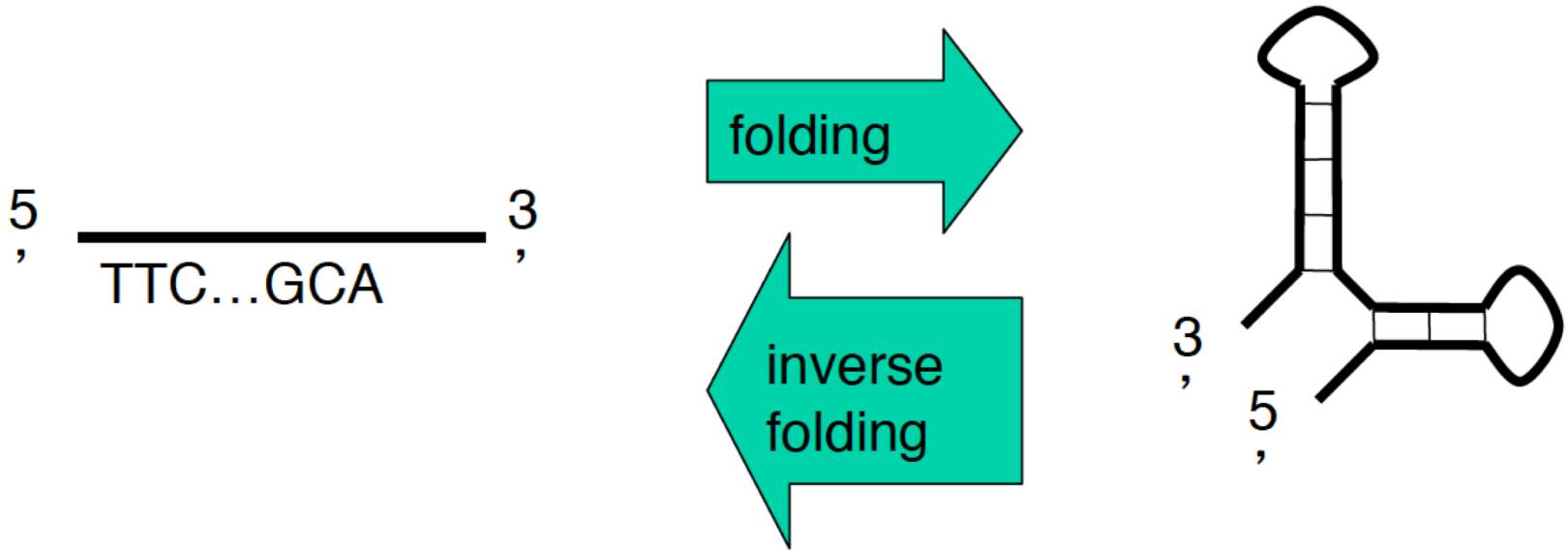
Notes

1. ^ **Structure:** benchmarks structure prediction tools <yes|no>.
2. ^ **Alignment:** benchmarks alignment tools <yes|no>.

Inverse folding

- Another direction in sequence design is designing a sequence that folds into a given secondary structure.
- This problem is called inverse folding, because it is the inverse of the problem of finding the secondary structure of a sequence with the minimum free energy. The inverse folding problem is to find a sequence whose minimum energy structure coincides with the given one

Inverse folding



Main aim: discovery of novel, structured and functional RNAs in transcriptomic data.

Inverse folding

RNAinverse The ViennaRNA package provides RNAinverse, an algorithm for designing sequences with desired structure.

RNAiFold A complete RNA inverse folding approach based on constraint programming and implemented using OR Tools which allows for the specification of a wide range of design constraints.

RNA-SSD/RNA Designer The RNA-SSD (RNA Secondary Structure Designer) approach first assigns bases probabilistically to each position based probabilistic models. Subsequently a stochastic local search is used to optimize this sequence.

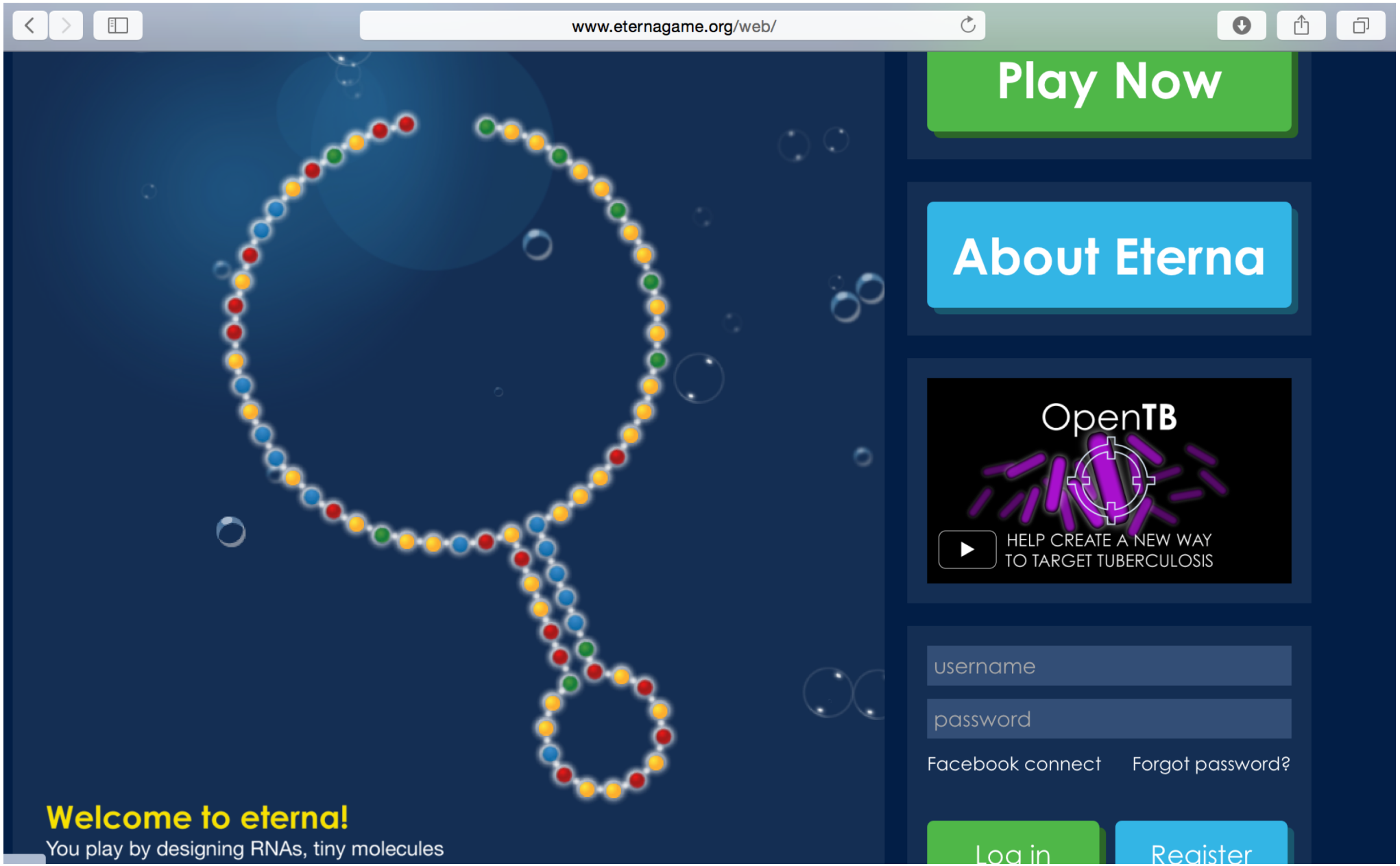
INFO-RNA INFO-RNA uses a dynamic programming approach to generate an energy optimized starting sequence that is subsequently further improved by a stochastic local search that uses an effective neighbor selection method.

RNAexinv RNAexinv is an extension of RNAinverse to generate sequences that not only fold into a desired structure, but they should also exhibit selected attributes such as thermodynamic stability and mutational robustness. This approach does not necessarily outputs a sequence that perfectly fits the input structure, but a shape abstraction, i.e. it keeps the adjacency and nesting of structural elements, but disregards helix lengths and the exact number unpaired positions, of it.

RNA-ensign This approach applies an efficient global sampling algorithm to examine the mutational landscape under structural and thermodynamical constraints.

… and many others

EteRNA- http://www.eternagame.org/web/



Predikujte 3D strukturu RNA:
GCUACGAAGGAAGGAUUGGUAUGUGGUAUAUU
CGUAGC

http://rnacomposer.cs.put.poznan.pl

vyzkousejte vsechny mody predpovedi 2D struktury

– jak se od sebe lisi jednotlive modely?

– ktery z modelu se nejvice blizi experimentalni structure PDB:6E8S?