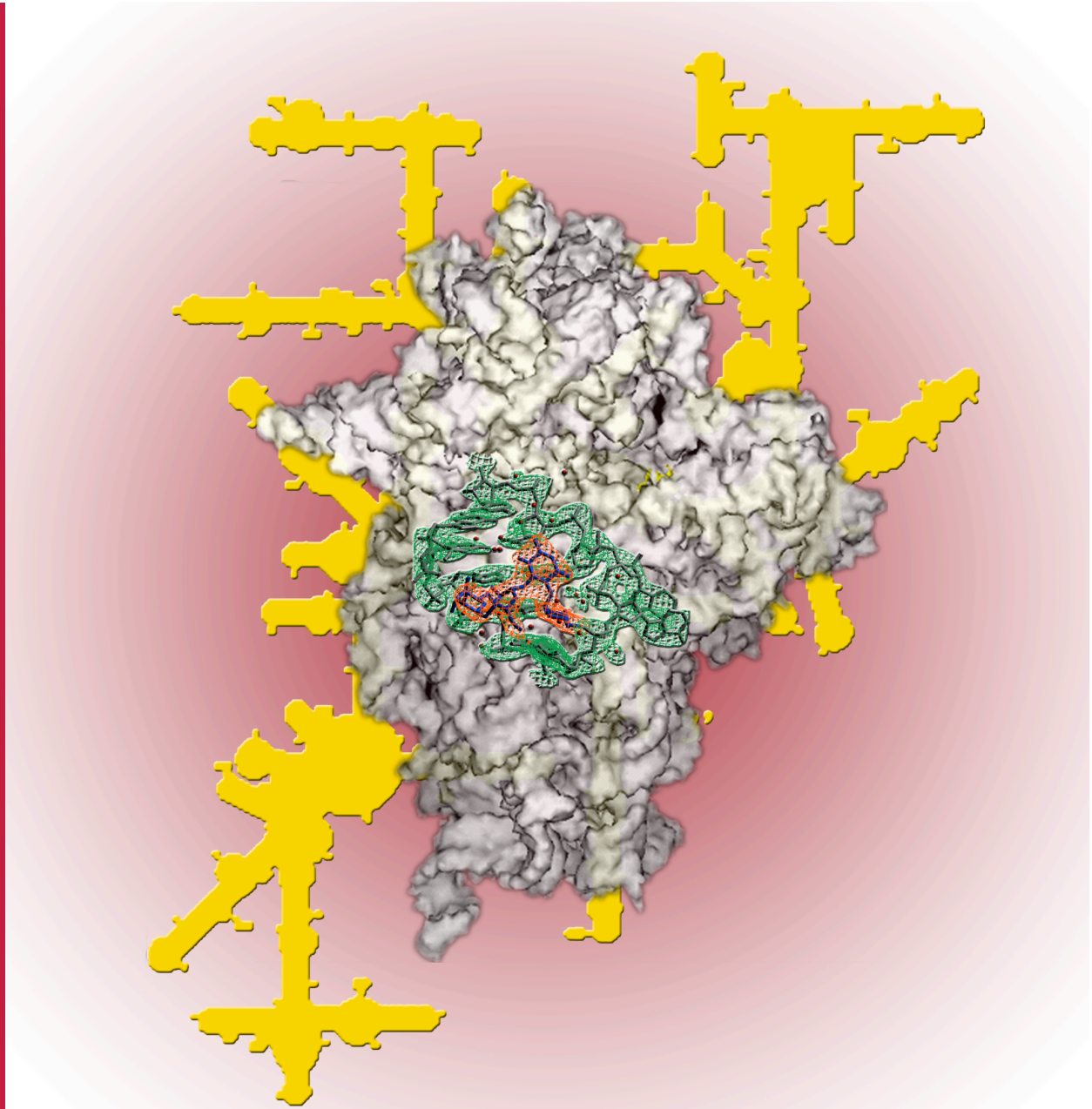
The RNA World IV



E. Westhof

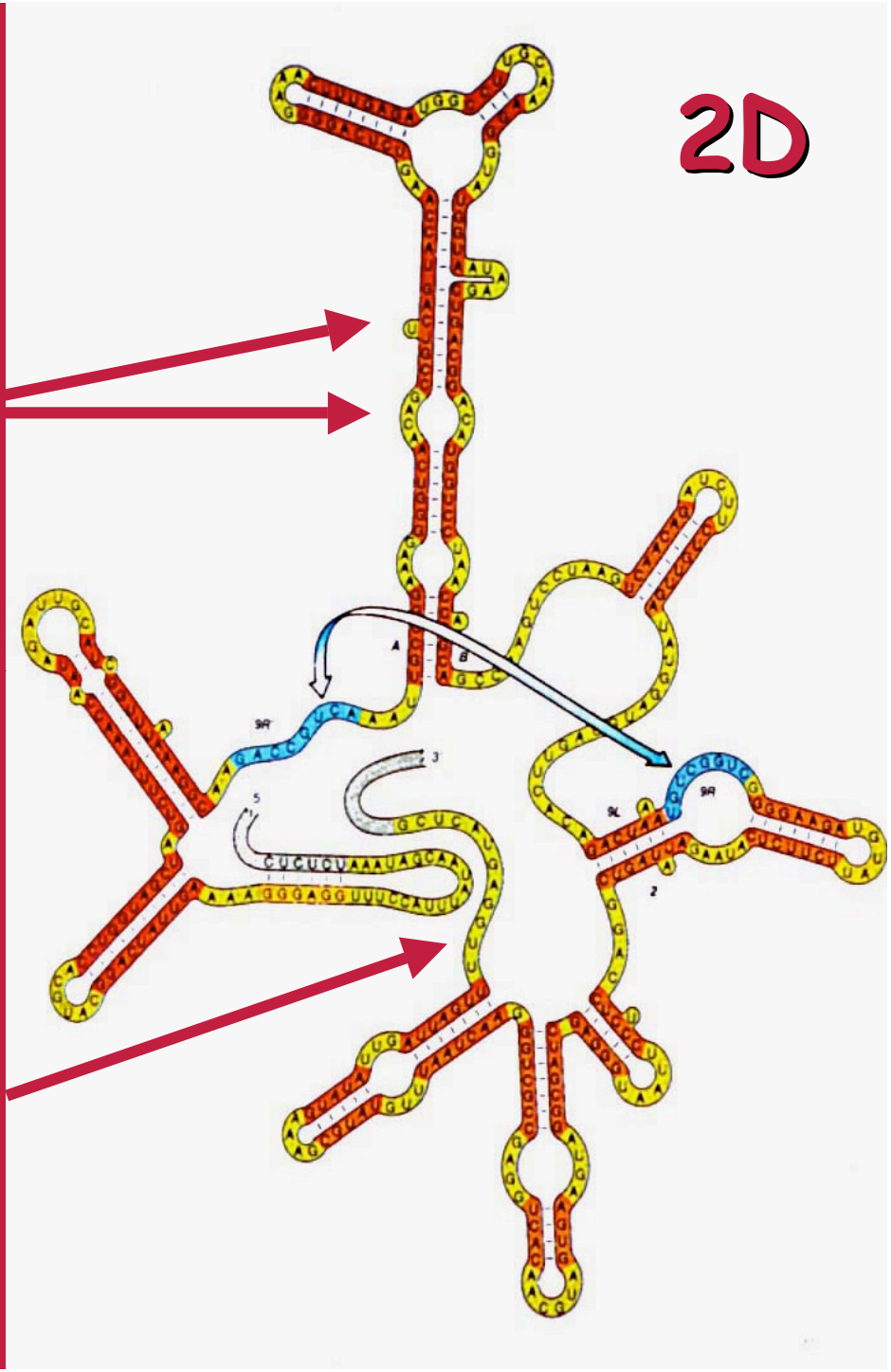
http://www-ibmc.u-strasbg.fr/upr9002/westhof/

Watson-Crick base paired helices

Internal loops (symmetric, Asymmetric, bulge)

Hairpin loops

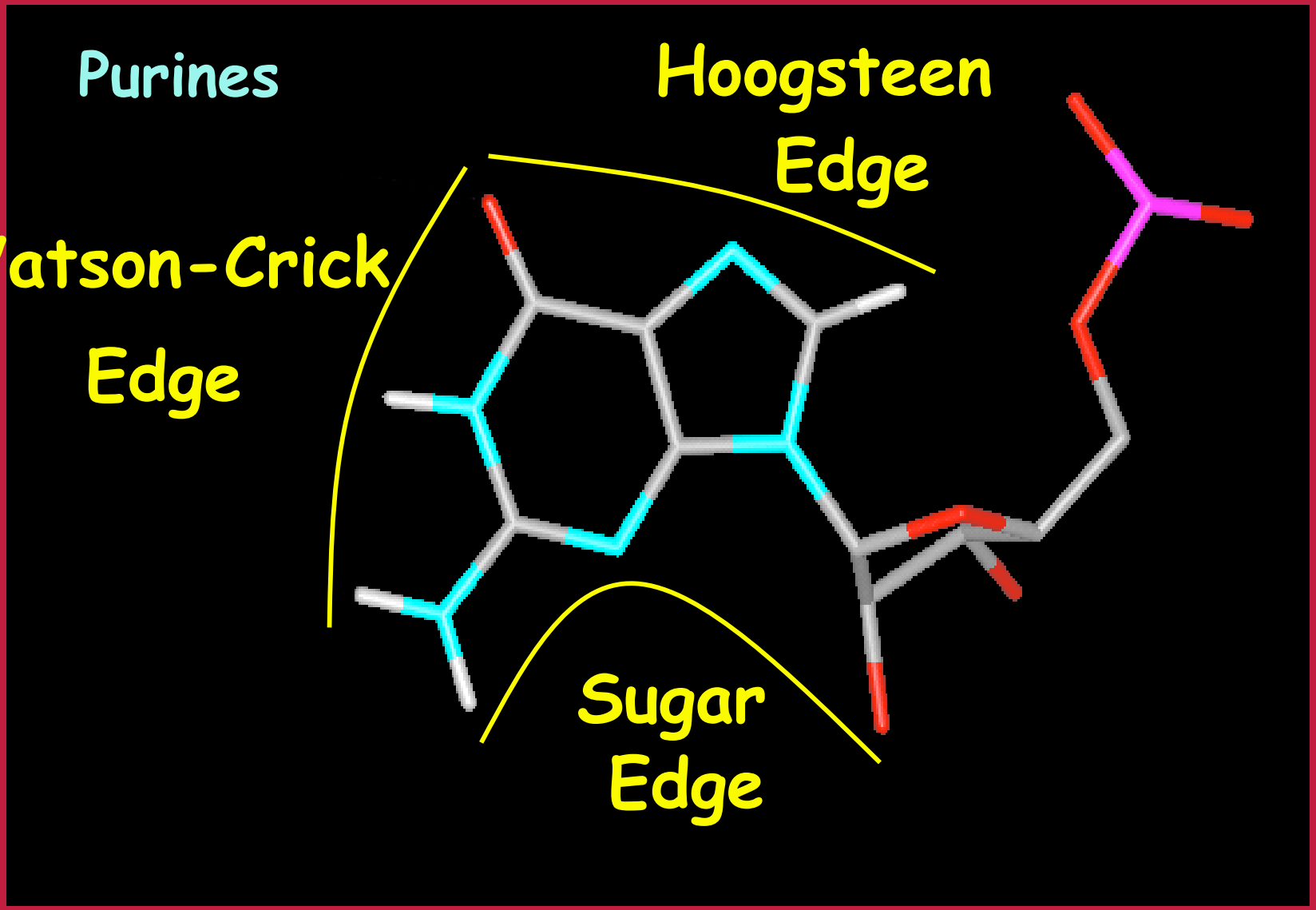
Single-strands junctions



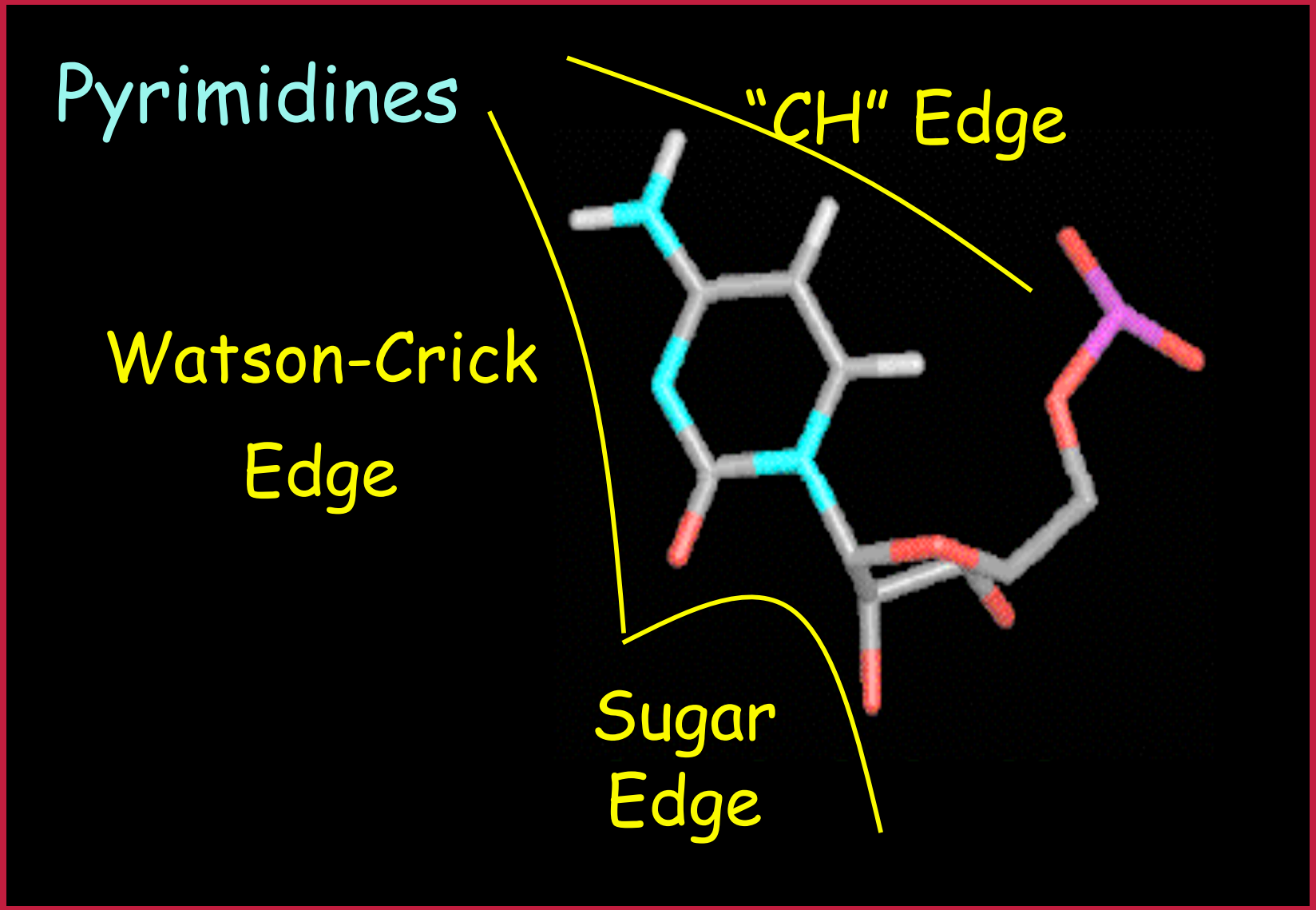
RNA base pairing

- **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 (Mg^{2+})

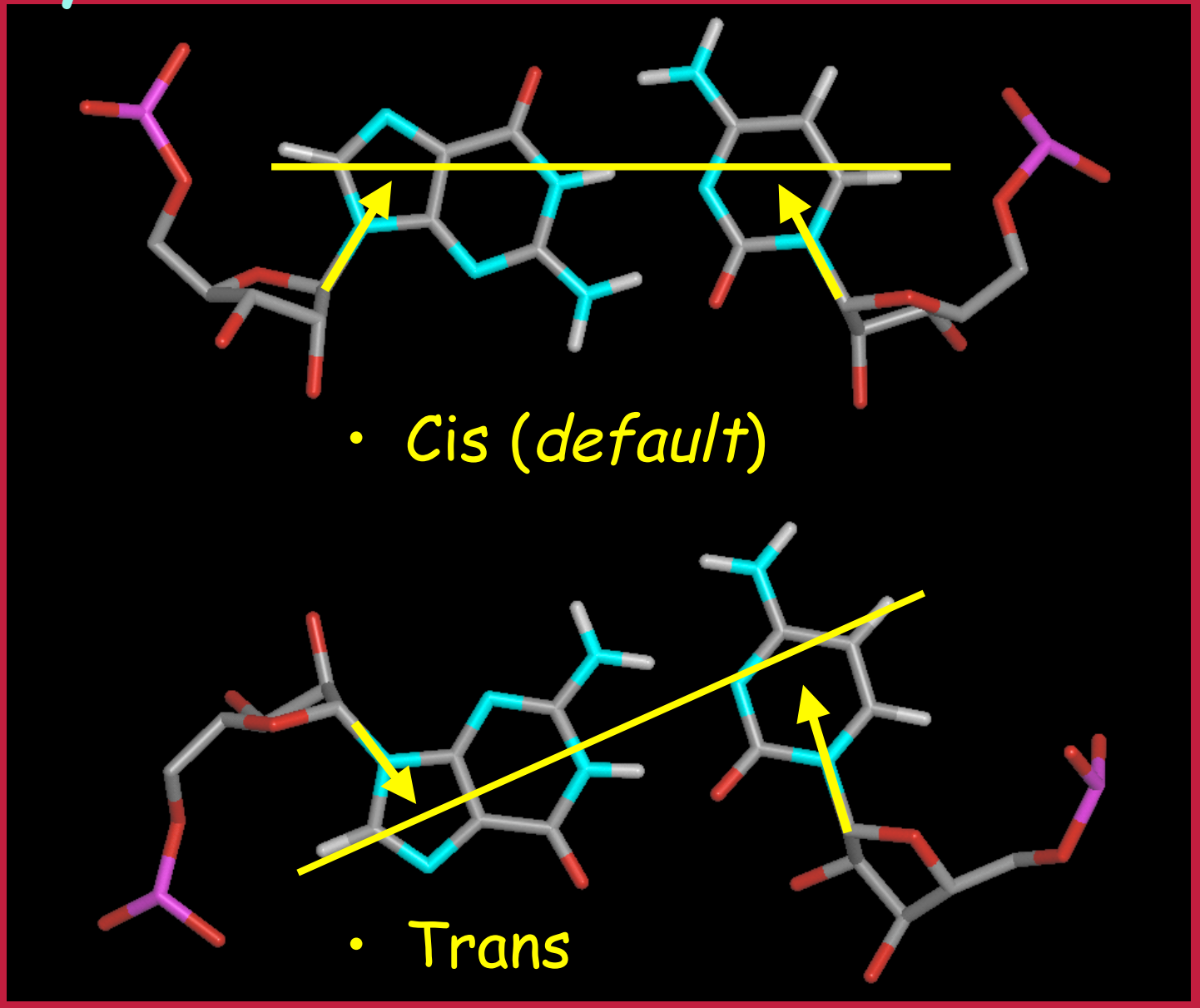
Three Interacting Edges



Interacting Edges



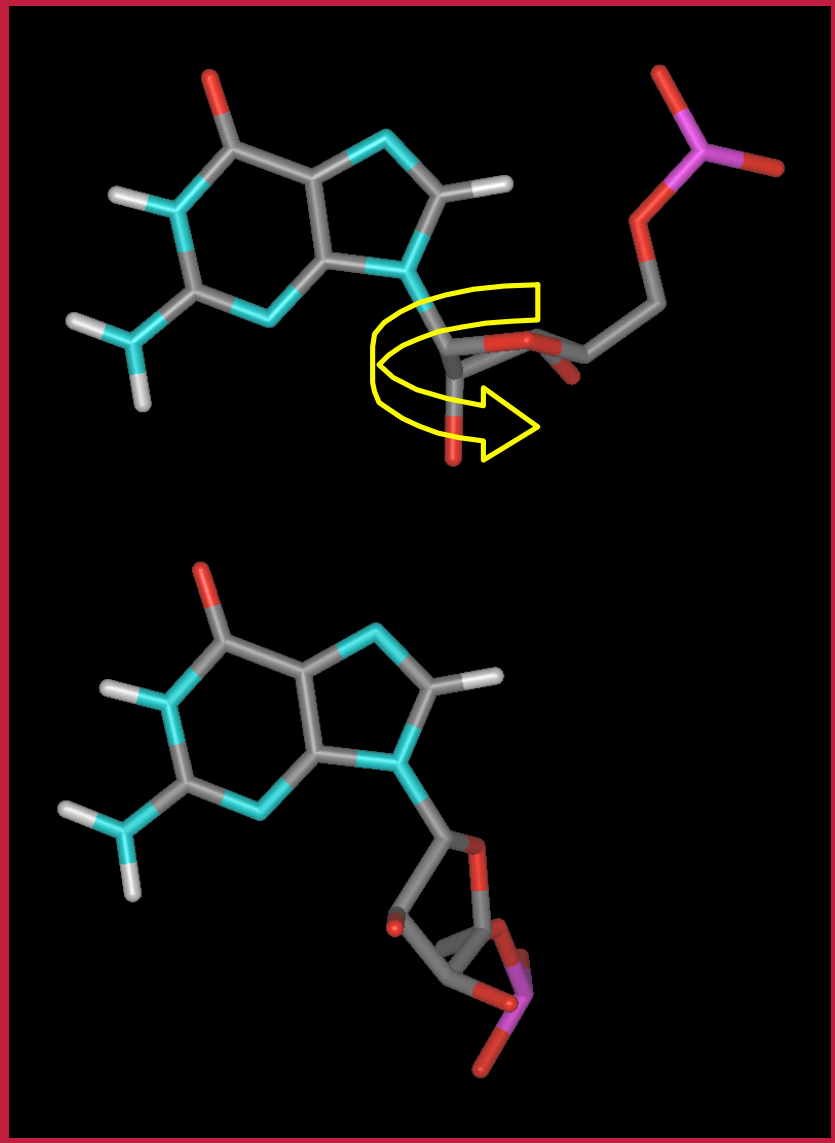
Glycosidic Bond Orientation



- Cis (*default*)
- Trans

Base-Sugar Conformation

- Anti (default)
- Syn (purines only)



Edge-to-Edge Pairing Types

Watson-Crick
Hoogsteen
Sugar-edge

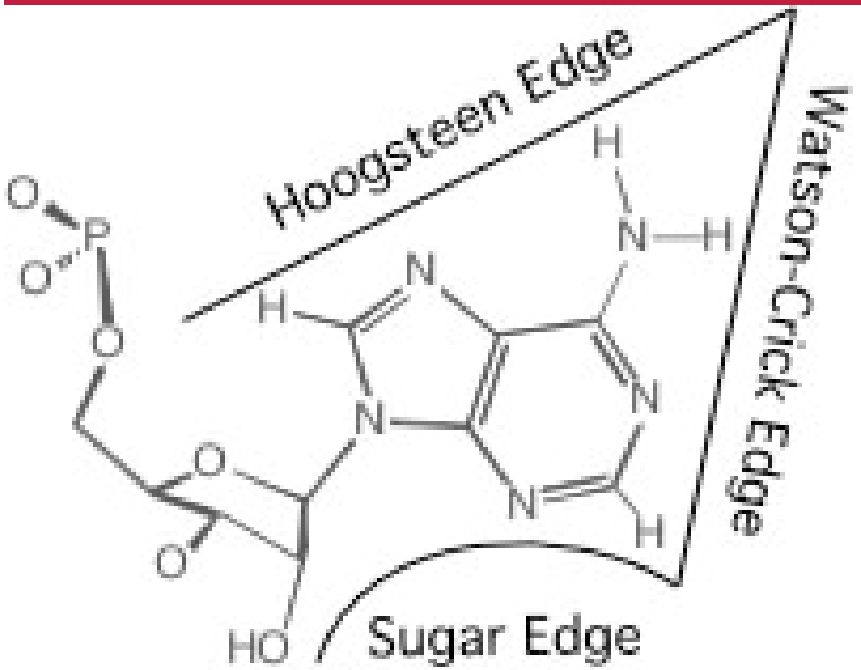
Watson-Crick
Hoogsteen
Sugar-edge

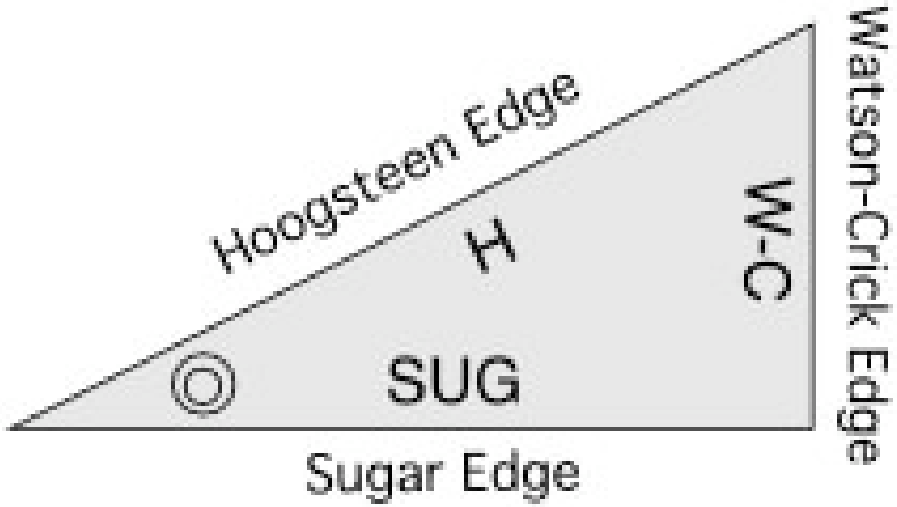
Cis
Trans

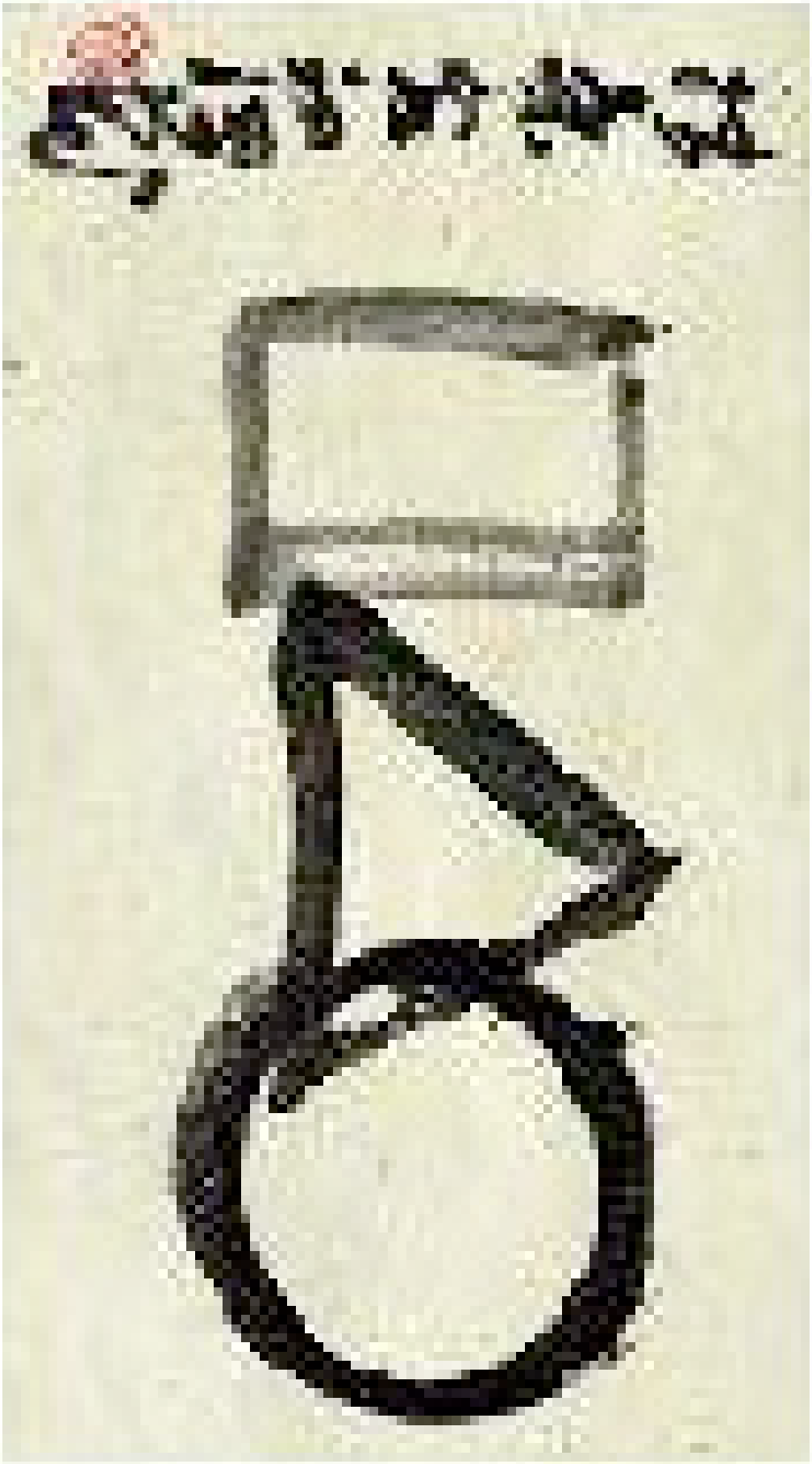
= 12 Basic Types

RNA 7, 499 (2001)

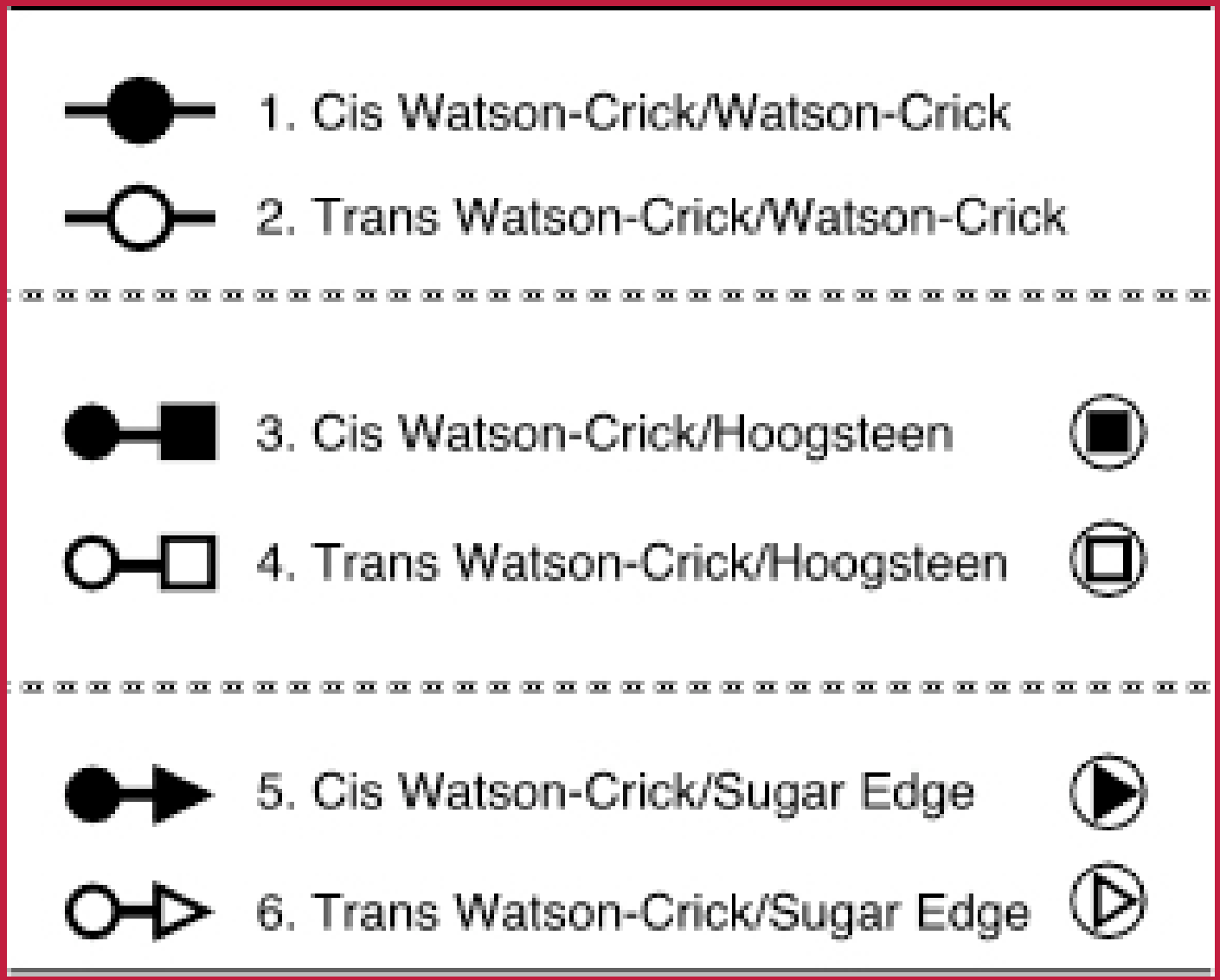
Each base has three edges and each edge carries various H-bonding sites





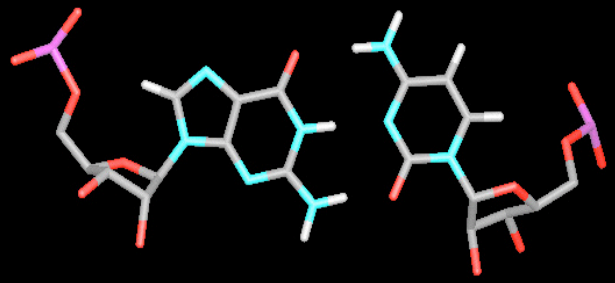


Annotations for Non-Watson-Crick Pairs

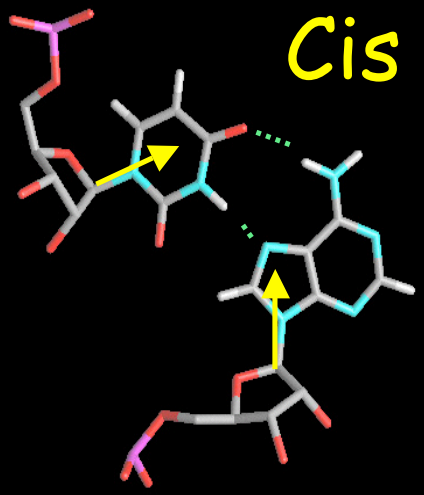


W.C. and Hoogsteen Pairings

W.C.-W.C.
Cis



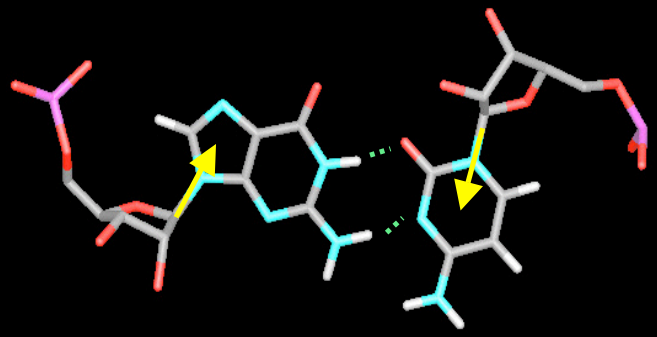
W.C.-Hoog.
Cis



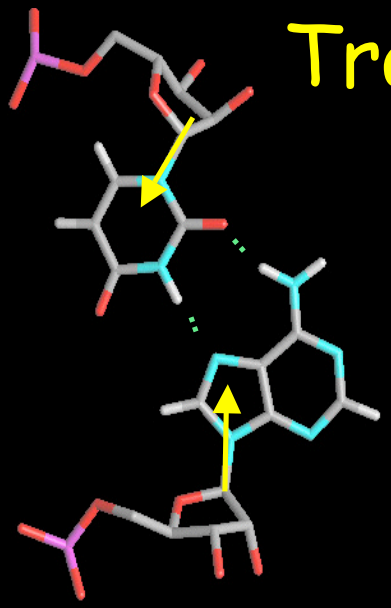
Hoog.-Hoog.
Cis

?

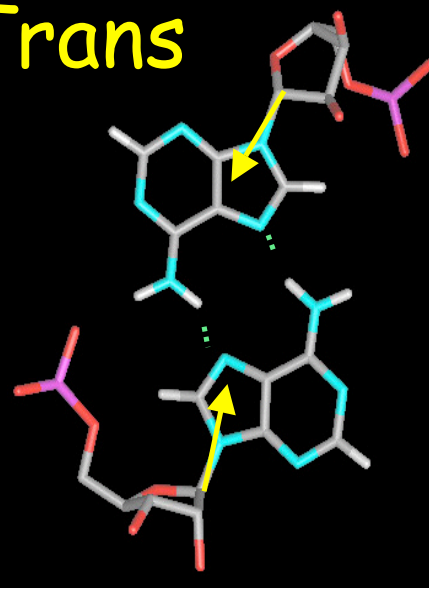
W.C.-W.C.
Trans



Trans

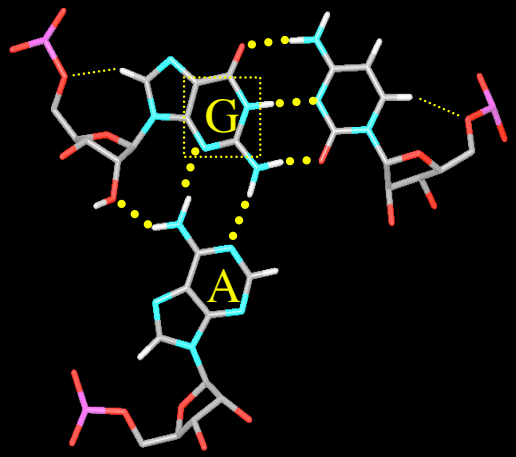


Trans

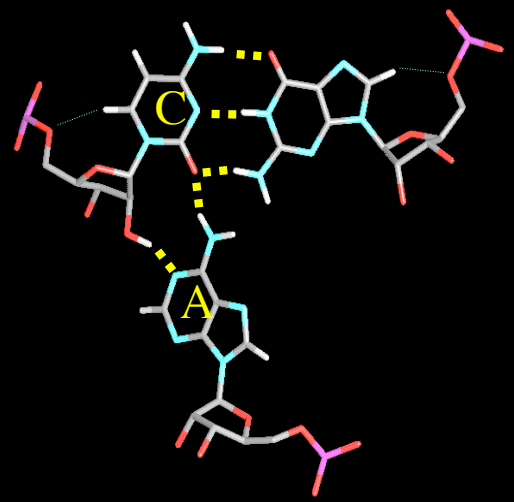


Sugar Edge Pairings

Watson-Crick

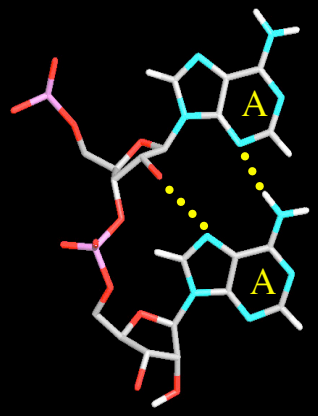


Cis

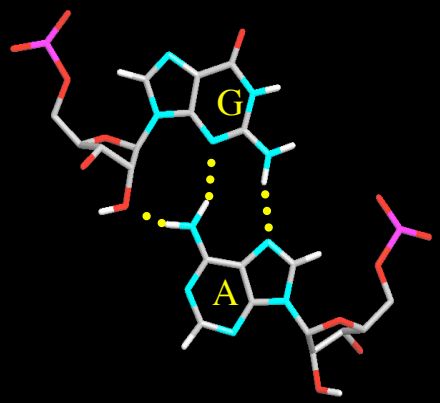


Trans

Hoogsteen

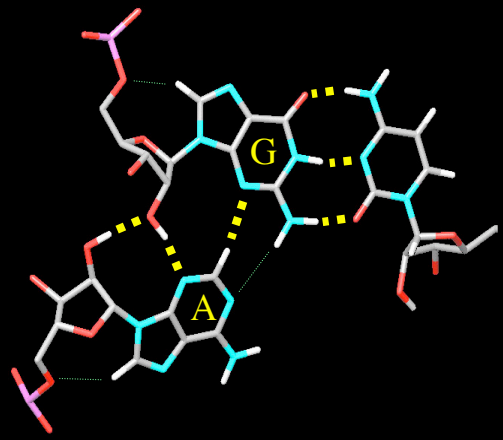


Cis

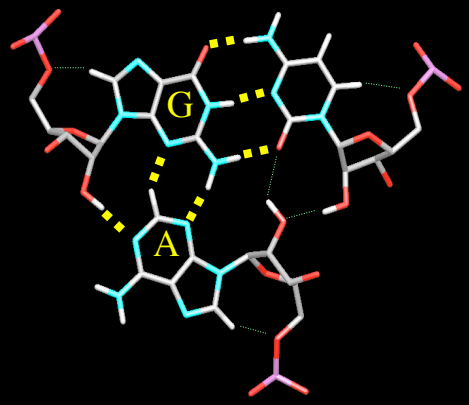


Trans

Sugar Edge



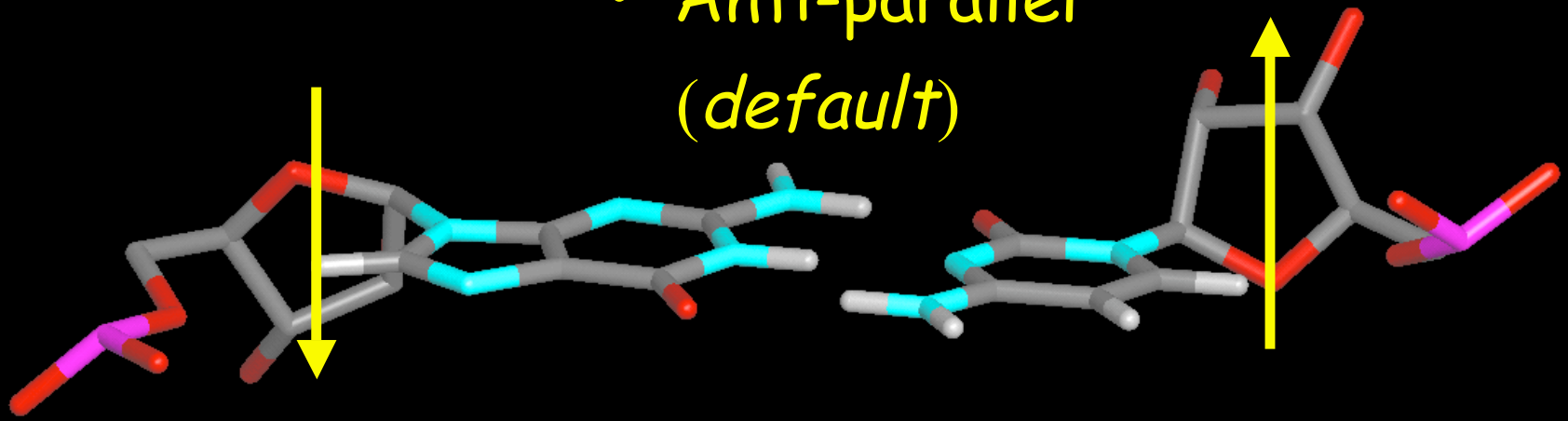
Cis



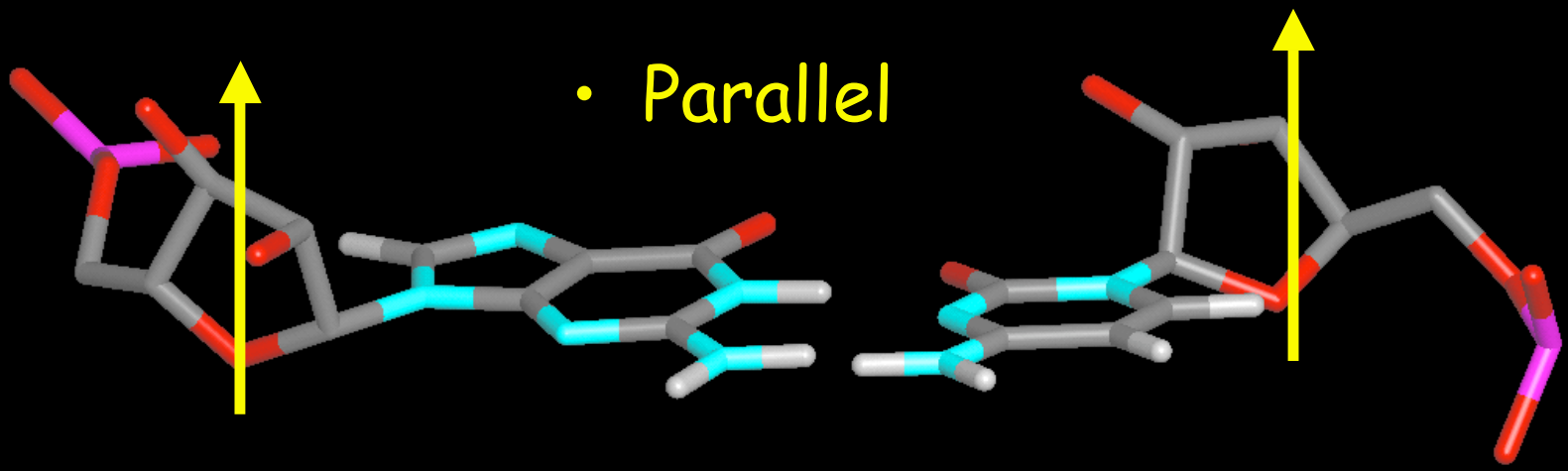
Trans

Local Strand Orientation

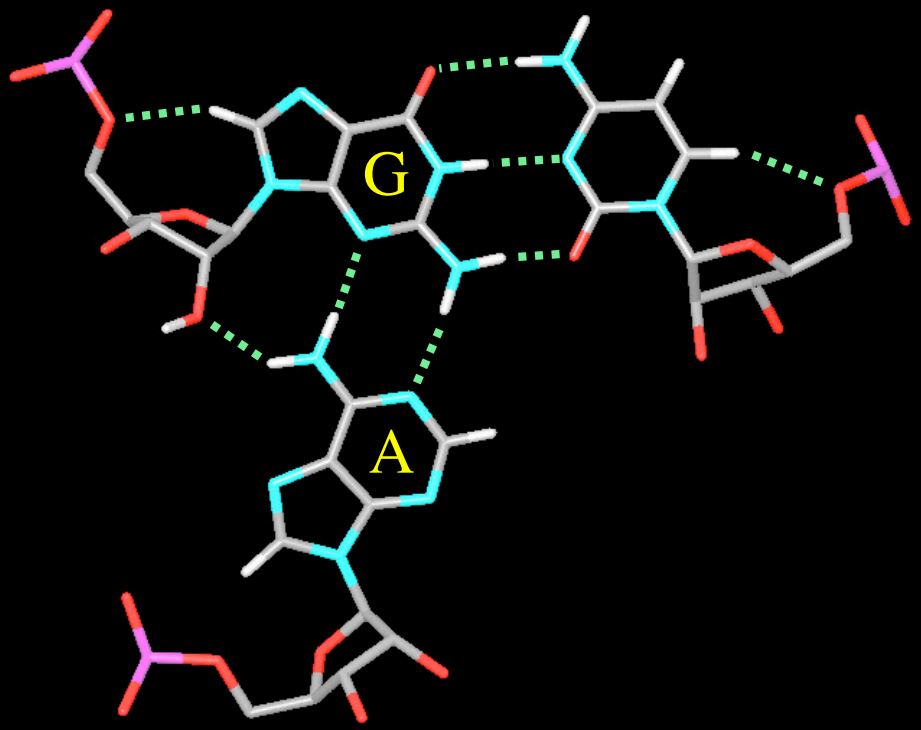
- Anti-parallel (*default*)



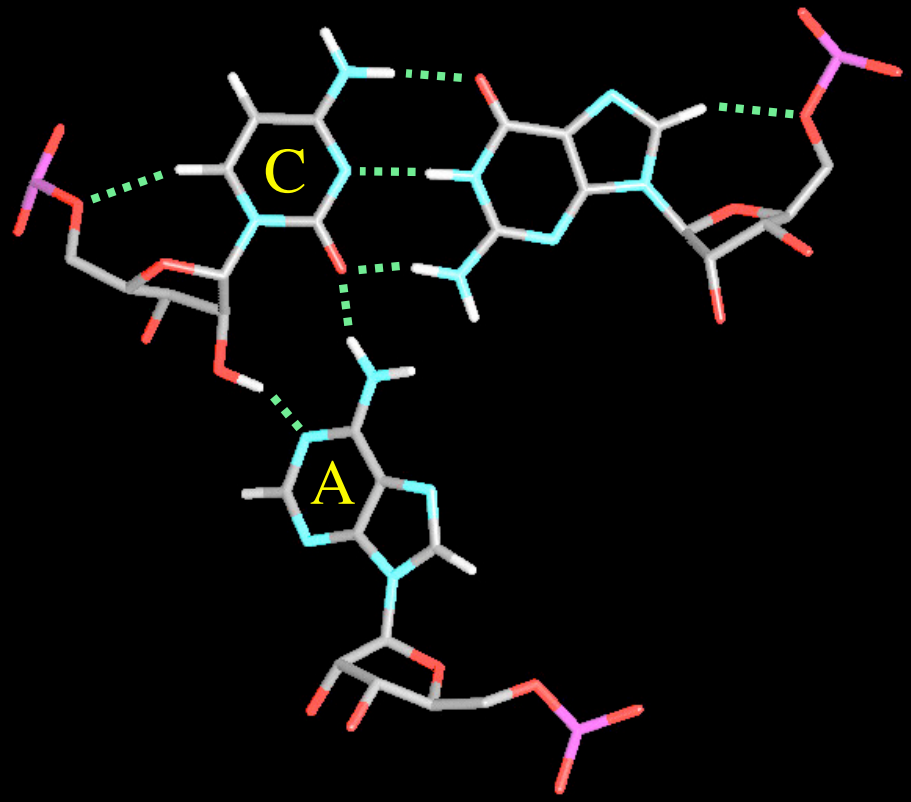
- Parallel



Watson-Crick/Sugar edge Pairs

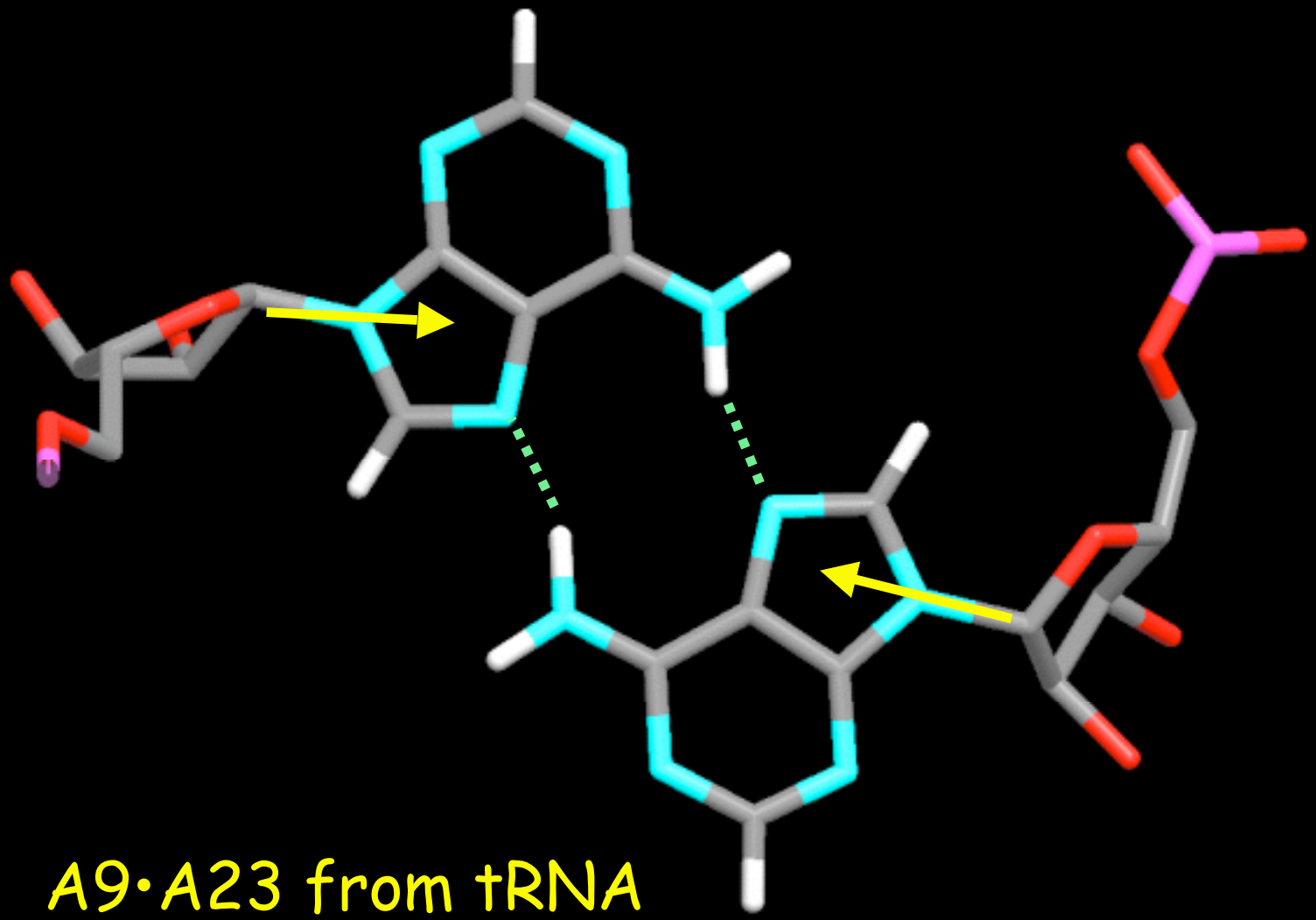


Trans and Parallel



Cis and Anti-Parallel

Trans Hoogsteen/Hoogsteen



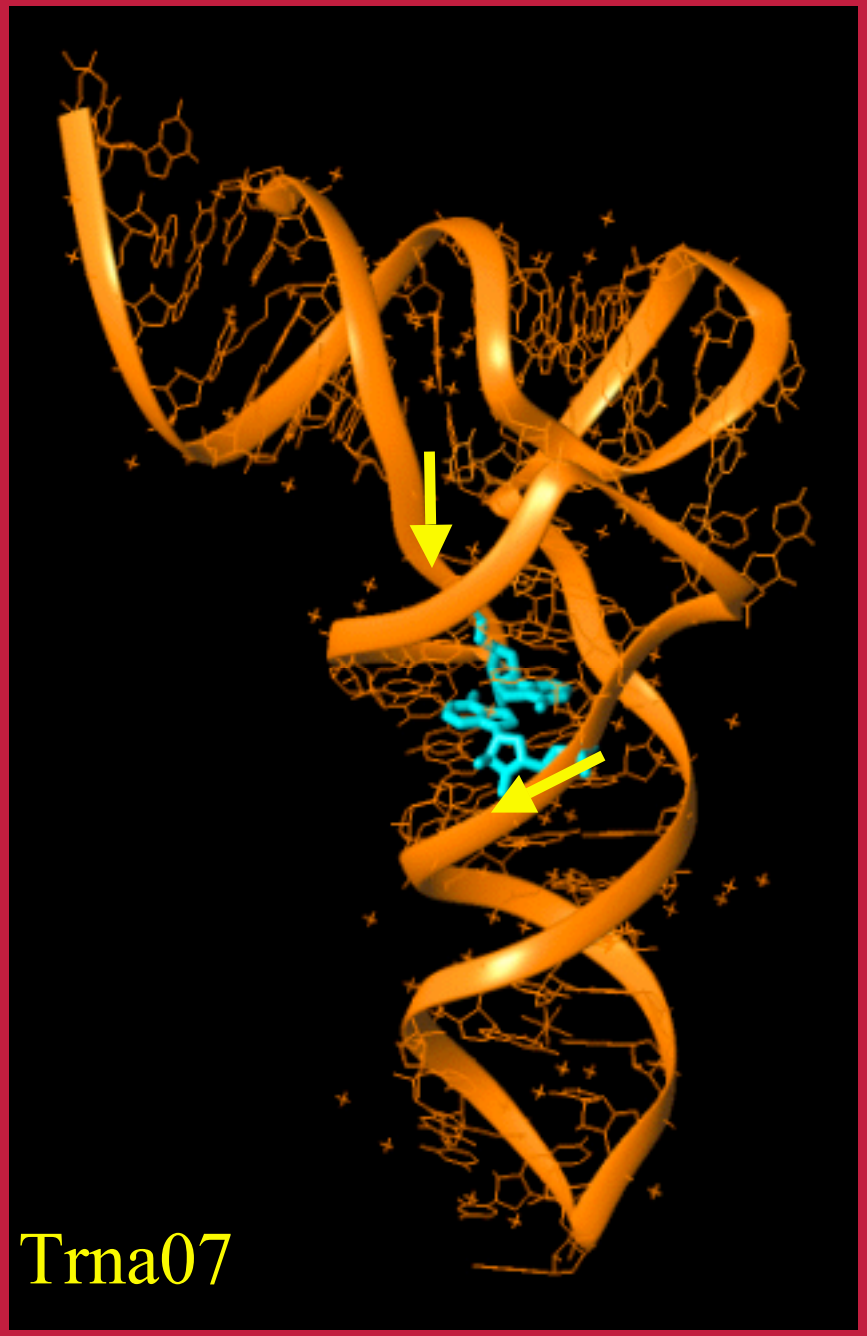
A9·A23 from tRNA

Trans Hoogsteen/Hoogsteen

	Canonical	A9·A23
• Interacting Edges	• W.C./W.C.	• Hoog./Hoog.
• Sugar-Base Conformations	• Anti/Anti	• Anti/Anti
• Glycosidic Bond Orientations	• Cis	• Trans
• Local Strand Orientations	• Anti-Parallel	• Parallel

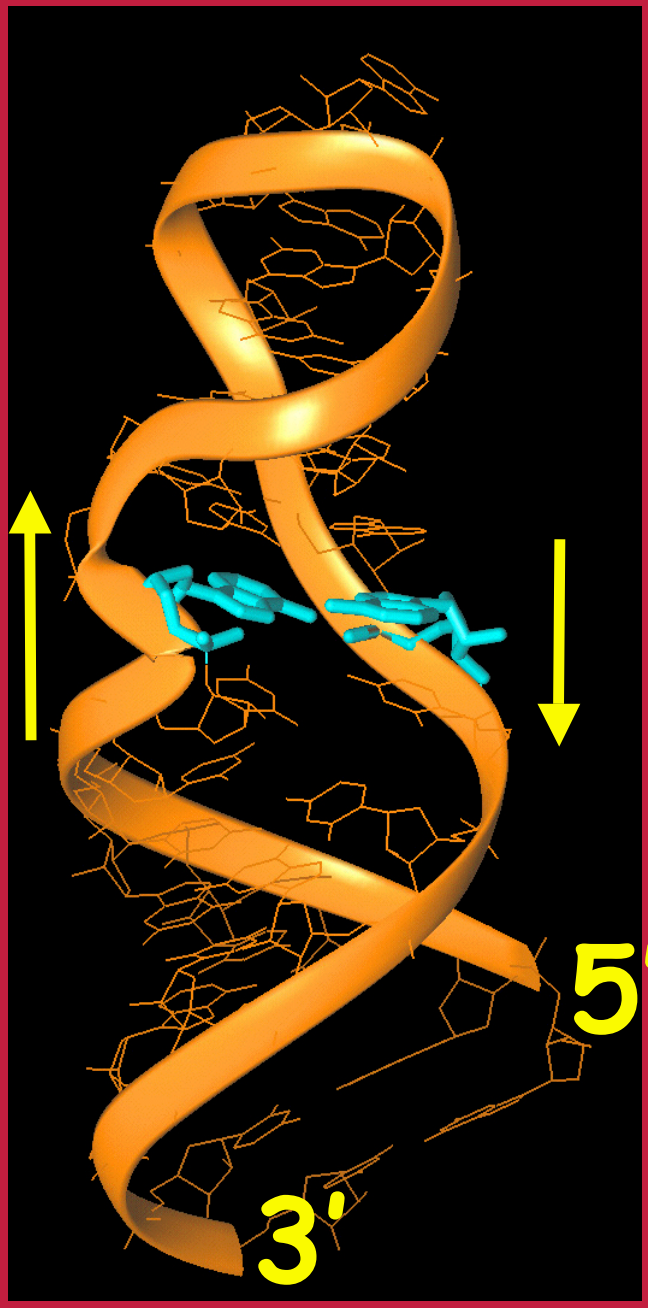
Globally Parallel Strands

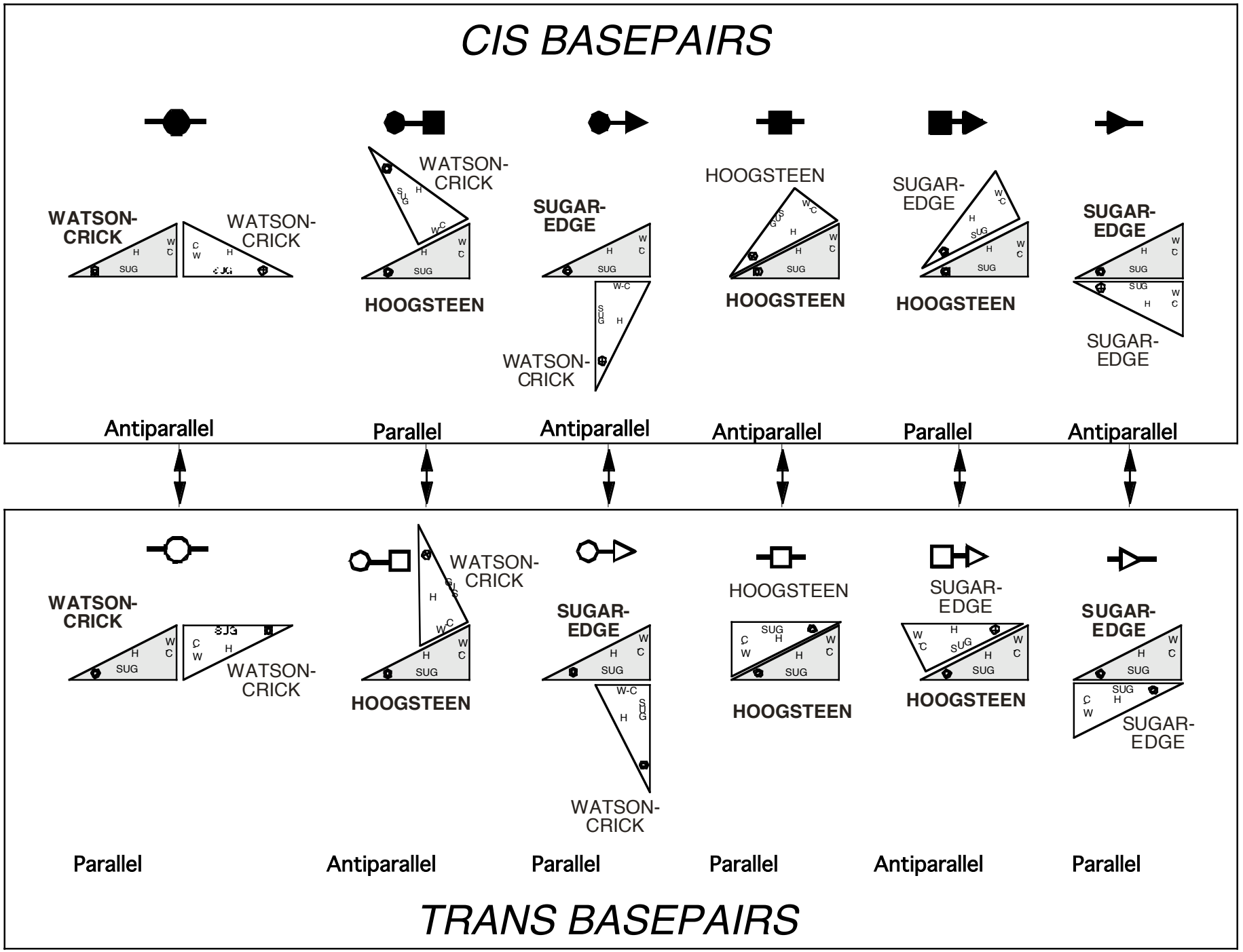
- A9•A23 in tRNA



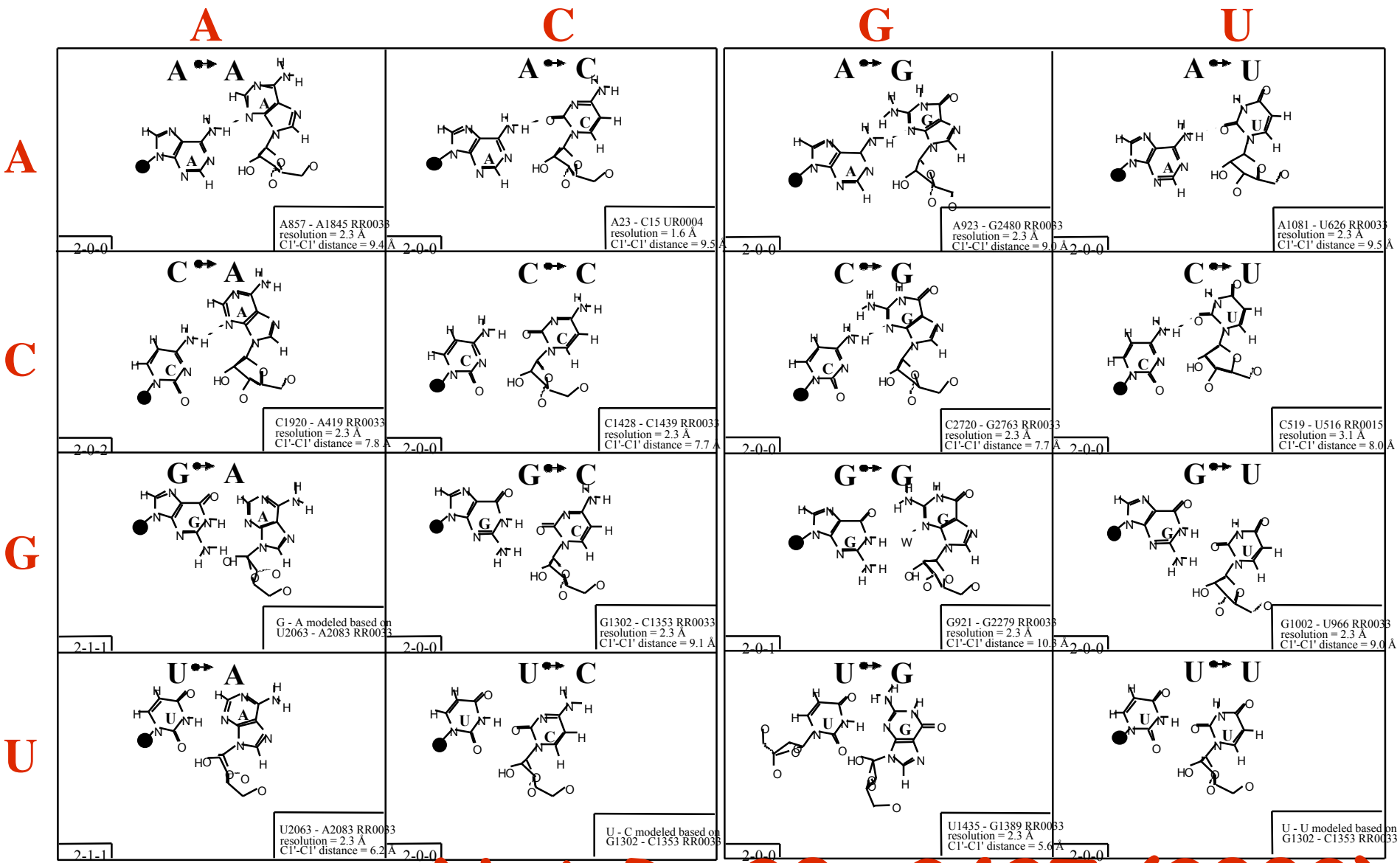
Locally Parallel Strands

- A·A trans Hoogsteen/ Hoogsteen pair in 23S rRNA "sarcin loop"
- Kink occurs next to bulged base
- Strands are globally Anti-parallel

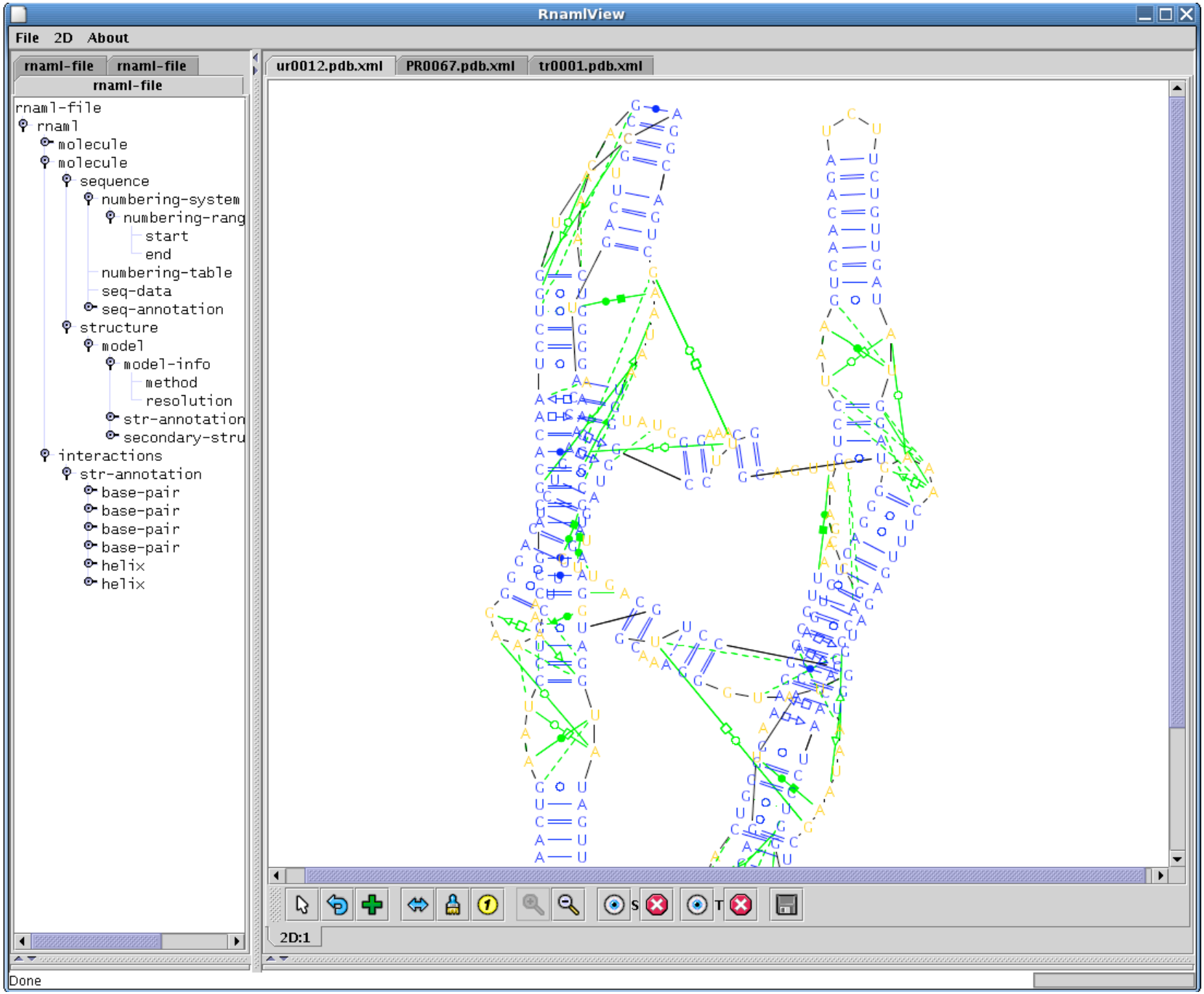



CIS BASEPAIRS
WATSON-CRICK
WATSON-CRICK
WATSON-CRICK
HOOGSTEEN
SUGAR-EDGE
WATSON-CRICK
HOOGSTEEN
HOOGSTEEN
SUGAR-EDGE
HOOGSTEEN
SUGAR-EDGE
SUGAR-EDGE
Antiparallel
Parallel
Antiparallel
Antiparallel
Parallel
Antiparallel
WATSON-CRICK
WATSON-CRICK
WATSON-CRICK
HOOGSTEEN
SUGAR-EDGE
WATSON-CRICK
HOOGSTEEN
HOOGSTEEN
SUGAR-EDGE
HOOGSTEEN
SUGAR-EDGE
SUGAR-EDGE
Parallel
Antiparallel
Parallel
Parallel
Antiparallel
Parallel
TRANS BASEPAIRS

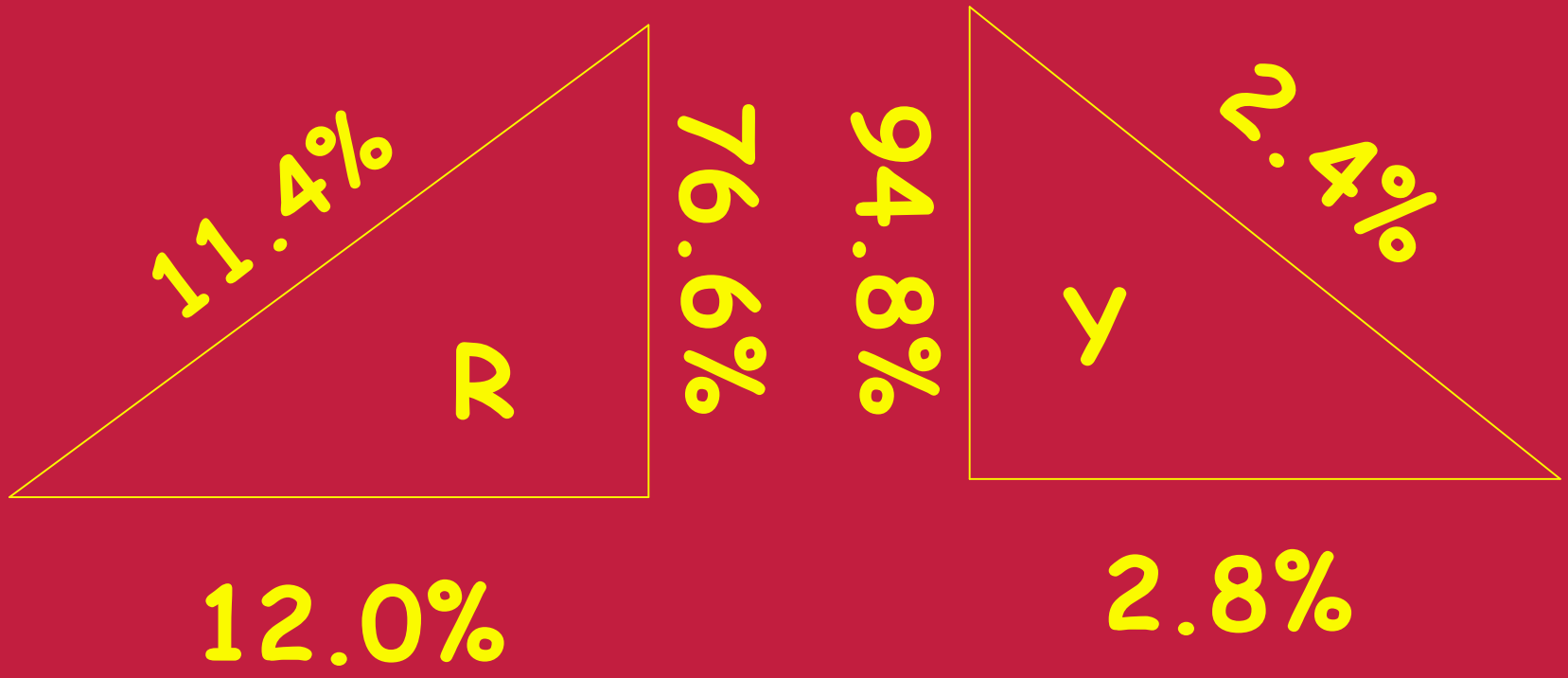
4x4 matrix of cisWatson-Crick/Sugar pairs



N.A.R. 30, 3497 (2002)


RnamlView
File
2D
About
rnaml-file
rnaml-file
rnaml-file
rnaml-file
rnaml
molecule
molecule
sequence
numbering-system
numbering-rang
start
end
numbering-table
seq-data
seq-annotation
structure
model
model-info
method
resolution
str-annotation
secondary-stru
interactions
str-annotation
base-pair
base-pair
base-pair
base-pair
helix
helix
ur0012.pdb.xml
PR0067.pdb.xml
tr0001.pdb.xml
S
T
2D:1
Done

Some statistics …



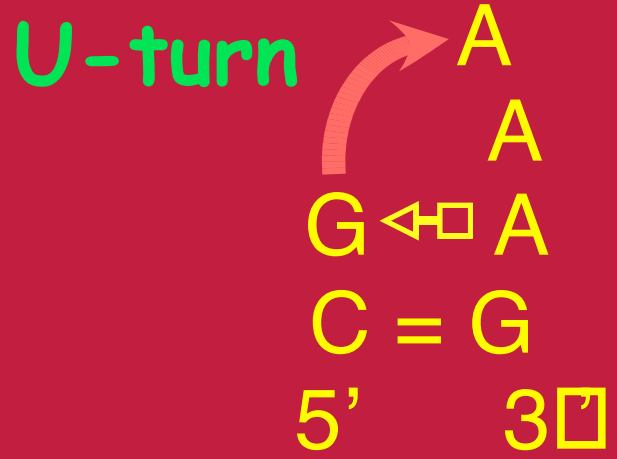
Sugar edge base pairs amongst the most frequent non-Watson-Crick pairs in RNA

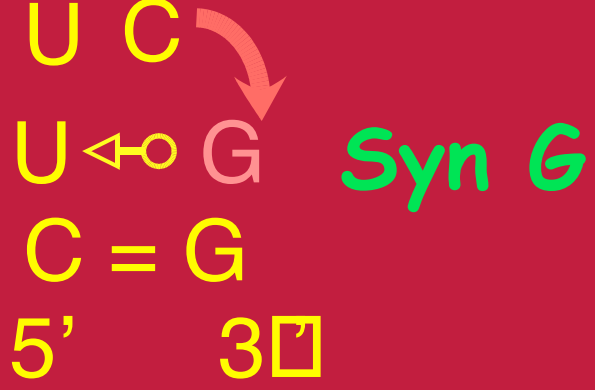
They usually implicate the hydroxyl O2'H

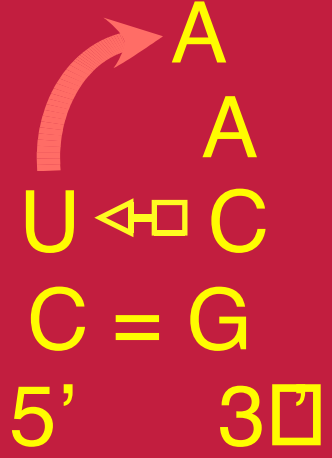
Advantages of the nomenclature

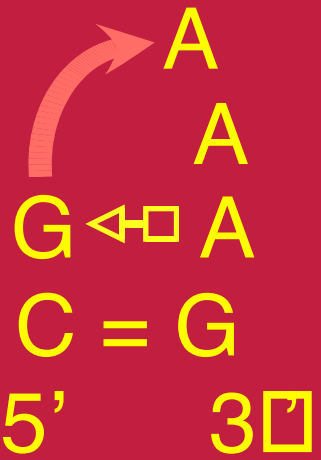
- Based on broad and defined geometrical concepts (edges) and not solely H-bonds
- Dynamical nature of molecules 'frozen' in crystals
- No crystal is perfect (data errors, resolution,…)

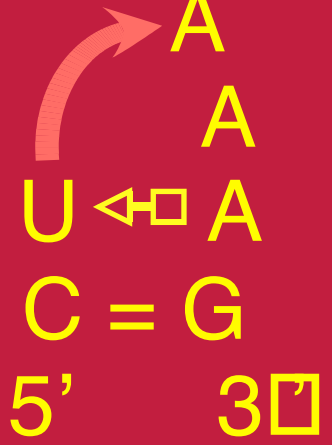
Diversity & Similarity in Tetraloops



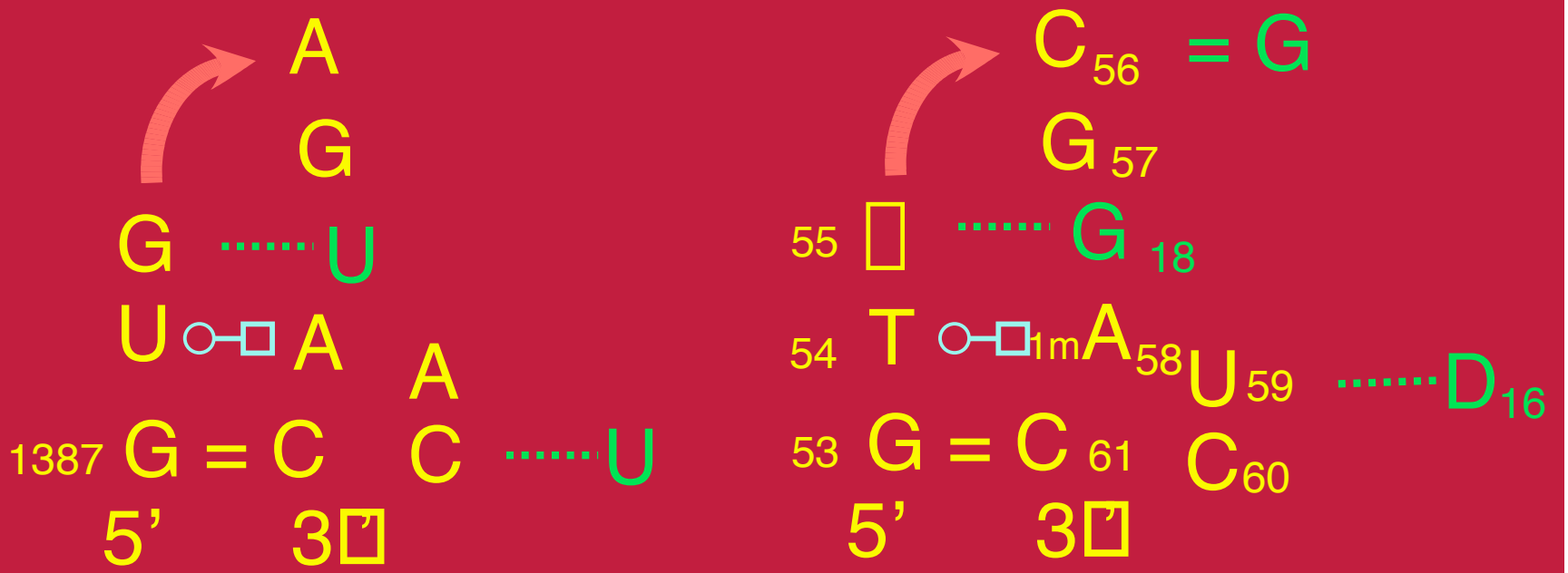


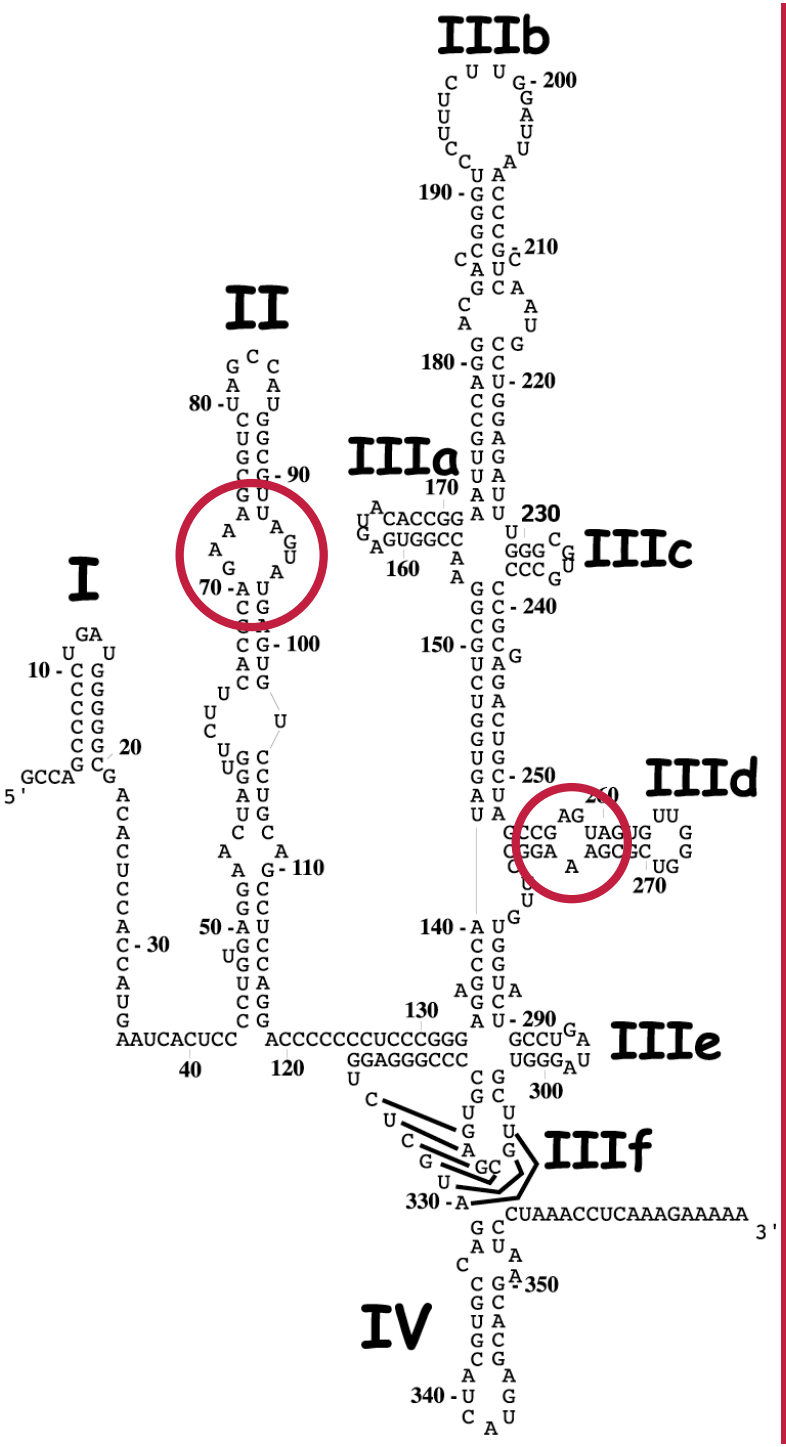






T-loops


23S rRNA
tRNA
A
G
G
U
U
A
A
1387 G = C
C
U
5'
3'
C56 = G
G57
55
G 18
54 T
1mA58
U59
D16
53 G = C 61
C60
5'
3'



How can we detect the presence of a motif in a given RNA ?

How can we compare motifs ?

What is a RNA motif ?

1. The U-turn ?
2. A GoA base pair ?
3. The sarcin loop ?
4. -AAUAAA- ?
5. All of those ?

Why do we bother about classification ?

What are our aims ?

1) Derive accurate & meaningful alignments;

2) Extract from a set of 'aligned' sequences, information about 3D contacts and architecture;

3) Derive rules for RNA evolution;

4) Search genomes for non-coding RNAs.

A MOTIF is an ensemble of ordered elements under constraints.

Sequential motifs :

- Strict : -AUG-
- Fuzzy : -AAUAAA-

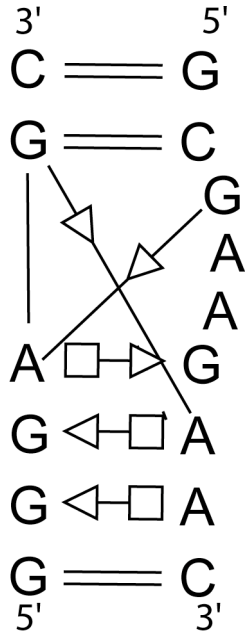
Structural motifs:

- -GNRA-
- Boxes C/D or H/ACA

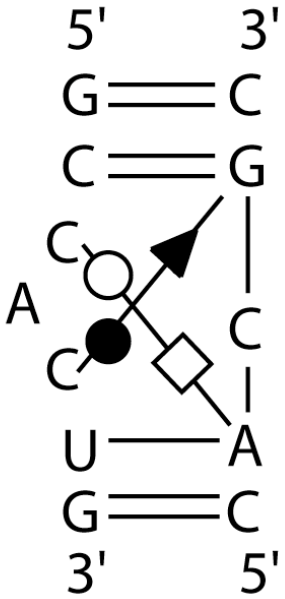
Consequences of this definition :

- Must fit within helices
- Should be recurrent
- Should they fold autonomously ?

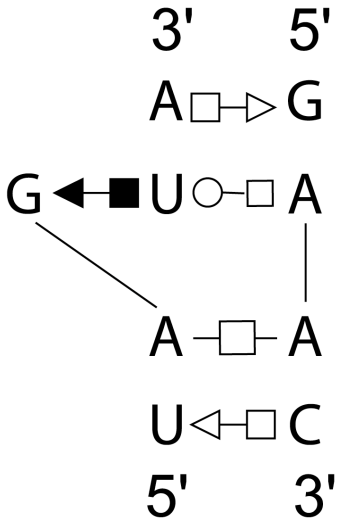
The K-turn



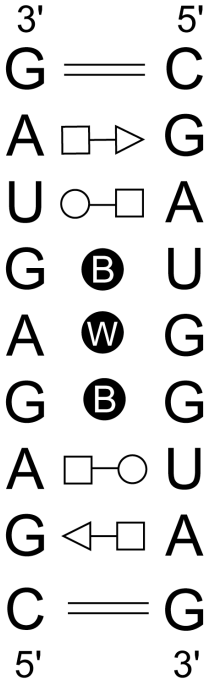
The C-motif

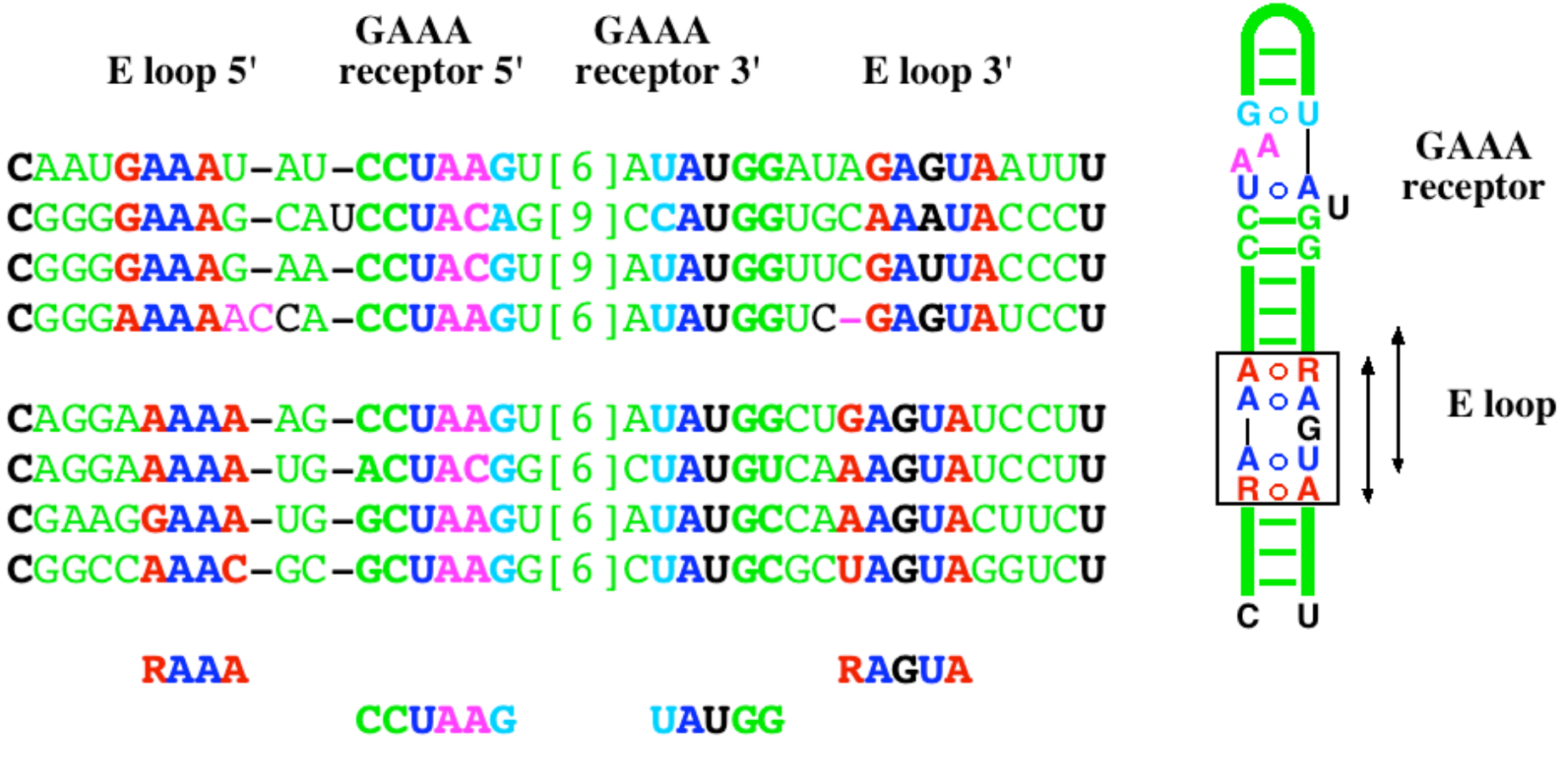


The sarcin motif



The loop-E motif




E loop 5'
GAAA receptor 5'
GAAA receptor 3'
E loop 3'
CAAUGAAAU–AU–CCUAAGU[6]AUAUGGAUAGAGUAAUUU
CGGGGAAAG–CAUCCUACAG[9]CCAUGGUGCAAAUACCCU
CGGGGAAAG–AA–CCUACGU[9]AUAUGGUUCGAUUACCCU
CGGGAAAAACCA–CCUAAGU[6]AUAUGGUC–GAGUAUCCU
CAGGAAAAA–AG–CCUAAGU[6]AUAUGGCUGAGUAUCCUU
CAGGAAAAA–UG–ACUACGG[6]CUAUGUCAAAGUAUCCUU
CGAAGGAAA–UG–GCUAAGU[6]AUAUGCCAAAGUACUUCU
CGGCCAAAC–GC–GCUAAGG[6]CUAUGCGCUAGUAGGUCU
RAAA
RAGUA
CCUAAG
UAUGG
GAAA receptor
E loop

Would such motifs evolve ?

- Chemistry and geometry of base pairs constrain the evolution of motifs

  (> structural convergence)

- Such motifs are spread across the Tree of Life and are used for diverse functions

- Such motifs cannot be used to infer homology

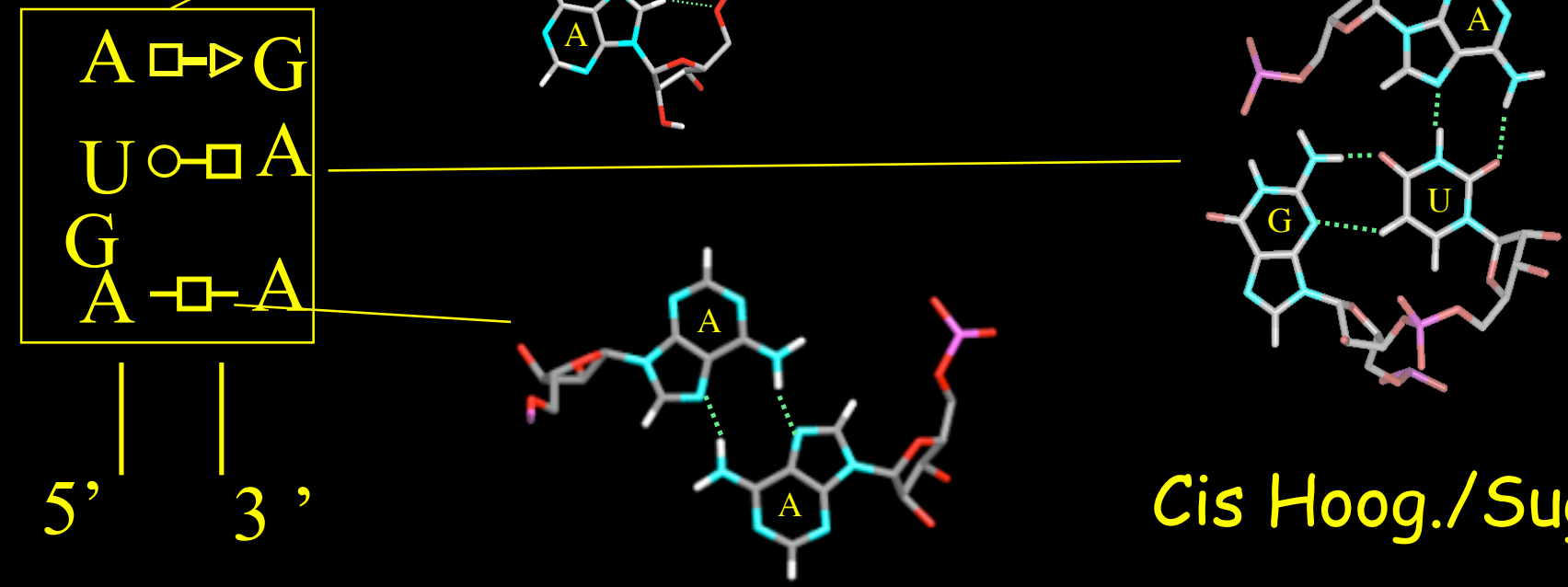
S-Motif Base-pairing

Trans Hoog./Sug.

A•G

Trans W.C./Hoog.

U•A

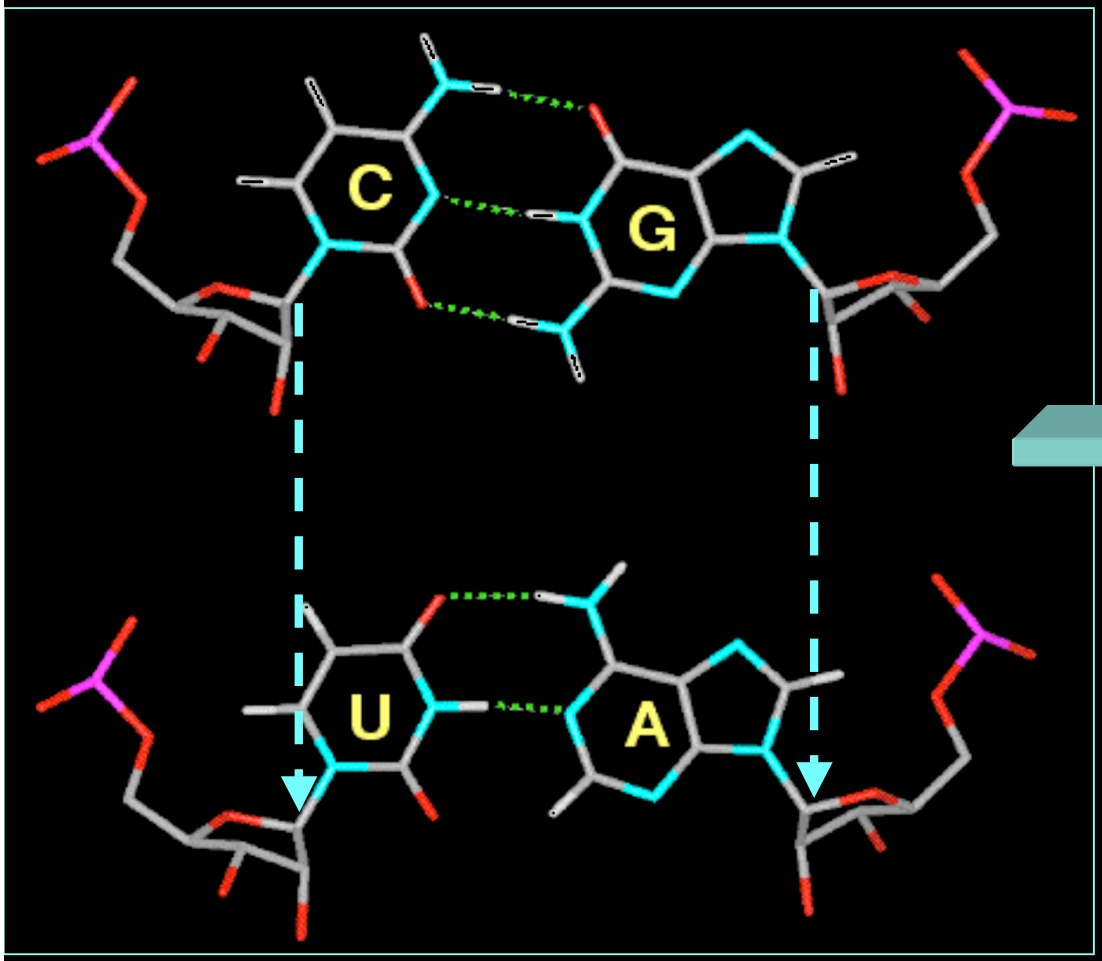


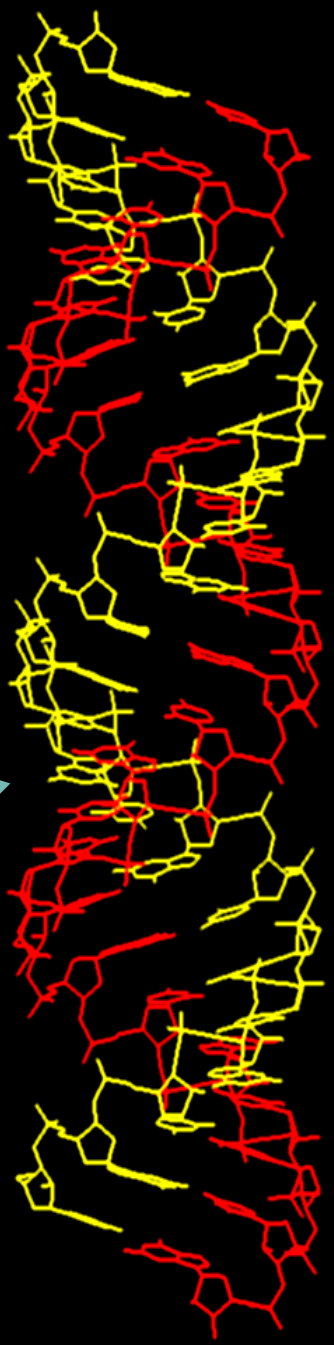
Trans Hoog./Hoog. A•A

Cis Hoog./Sug.

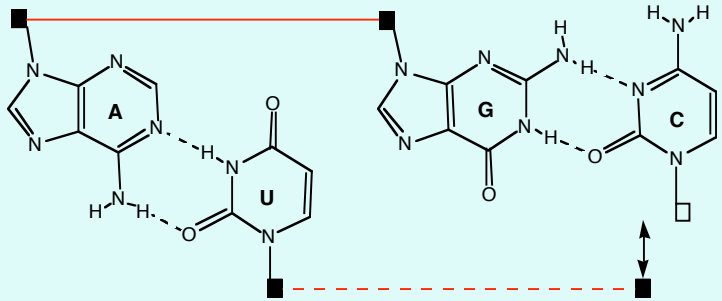
U•G

Watson-Crick pairs are isosteric

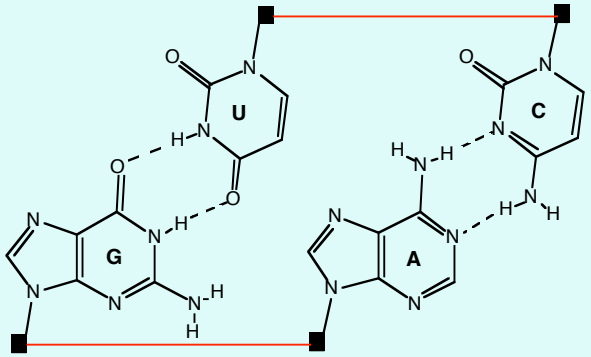




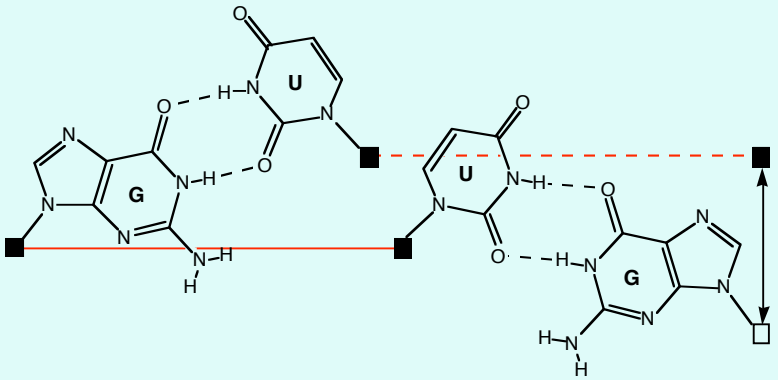
Non-isostérie des appariements Watson-Crick trans pour les bases complémentaires.

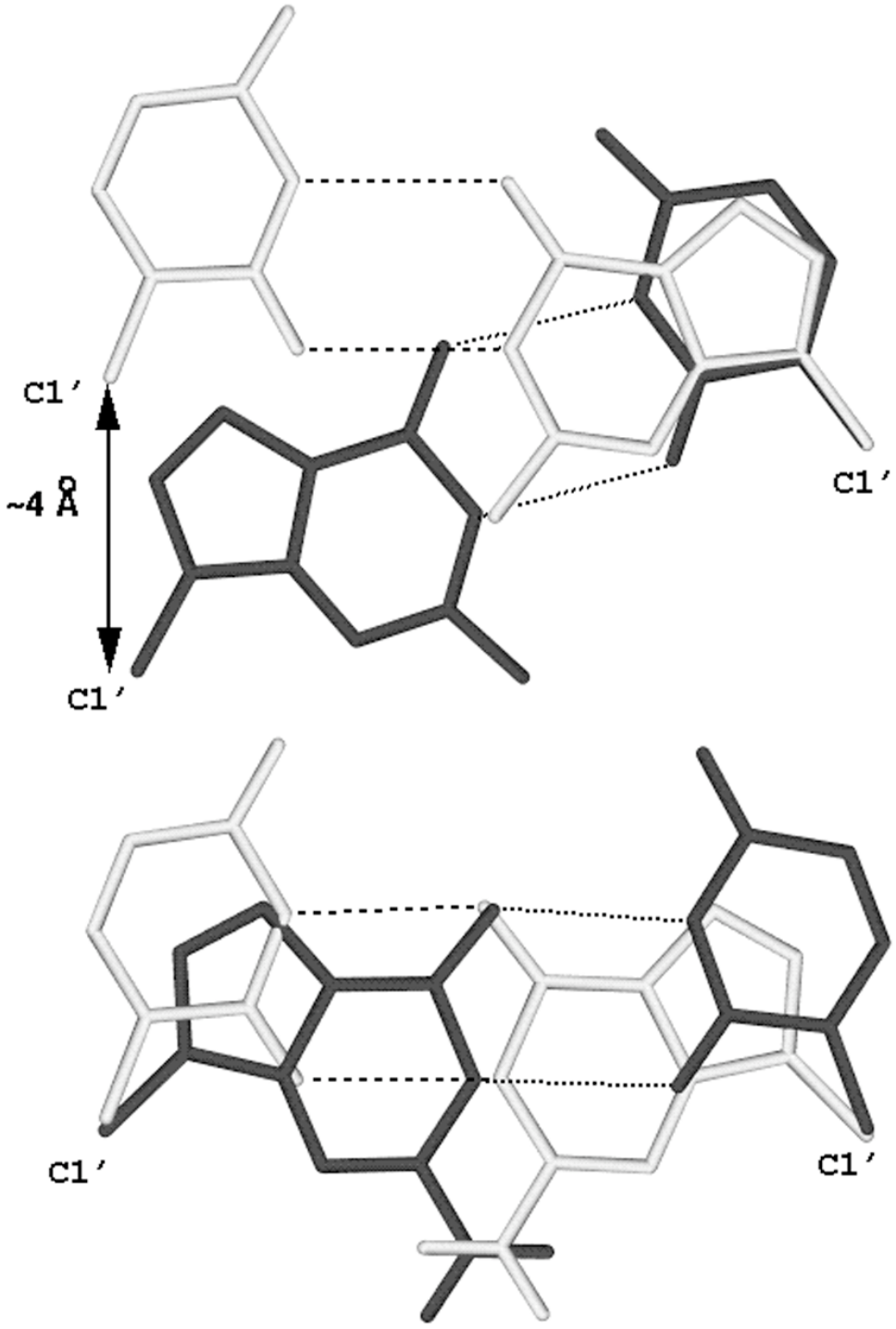


Isostérie des appariements Watson-Crick trans G-U et A-C

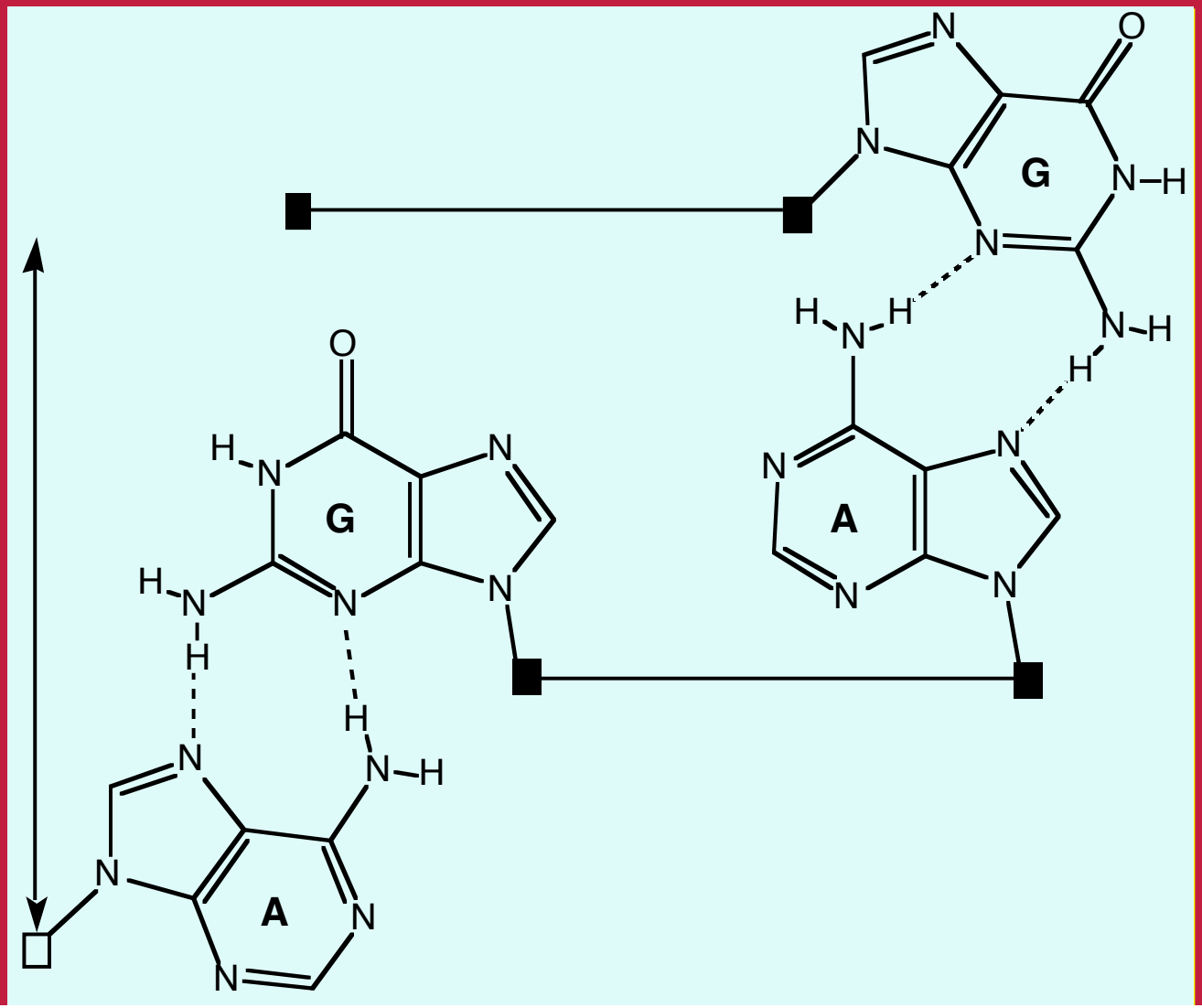


Non-isostérie des appariements Watson-Crick cis G-U et U-G




C1′
~4 Å
C1′
C1′
C1′
C1′

Non-isostericity of Hoogsteen/Sugar base pairs (or sheared AoG pairs)



Subsets of isosteric pairs

- Each of the 12 geometric families comprises 4x4 combinations (16x12=192)
- Only some are chemically allowed
- Those are organized in subsets of isosteric pairs
- The isosteric pairs are revealed by correlating 3D structures with sequence comparisons and modelling

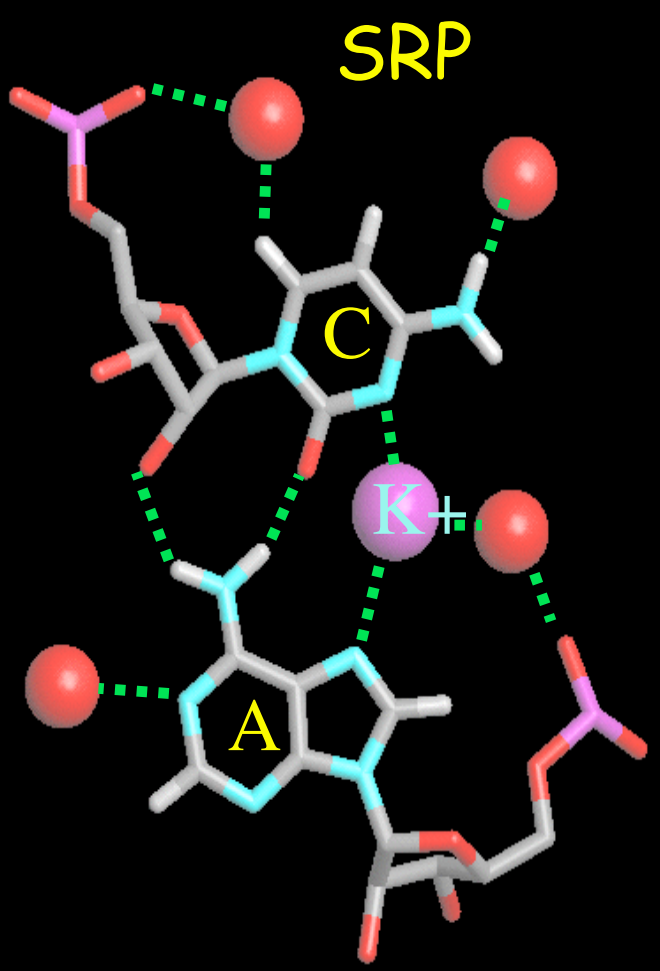
Isostericity Matrix for cis W.C./W.C.

Watson-Crick

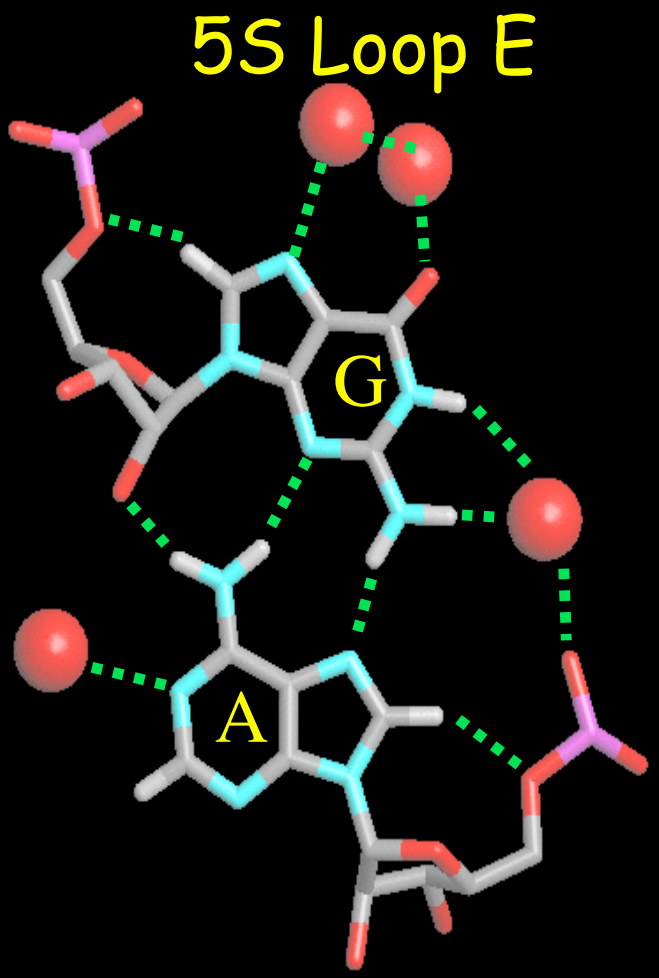
Watson-Crick

	A	C	G	U
A	I5	i3	I2	I1
C	I3	I5	I1	I4
G	I2	I1		i3
U	I1	I4	I3	I5

Trans Hoogsteen/Sugar



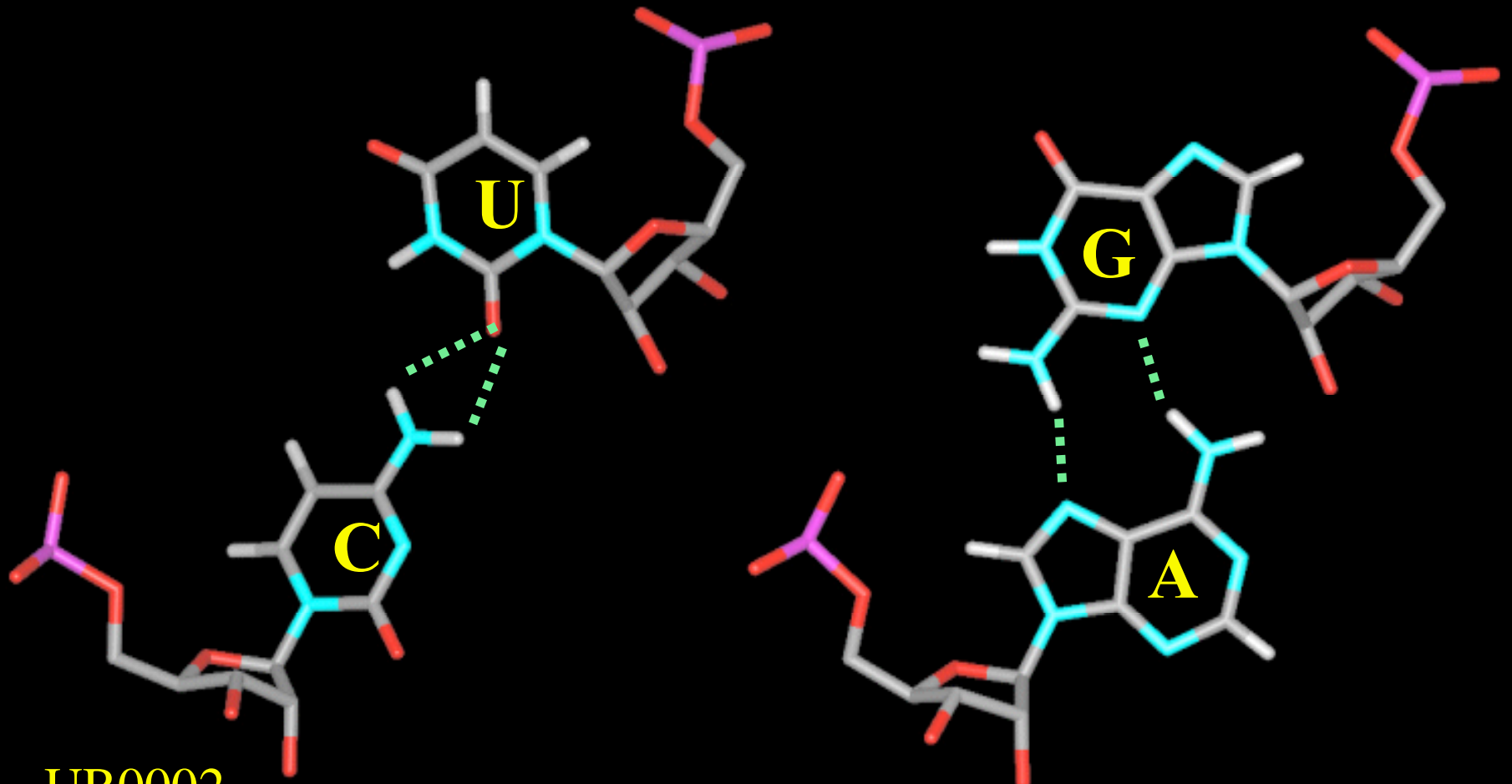
NDB file PR0021
Batey et al. (2000)



NDB file URL064
Correll et al. (1997)

C·U Isosteric to "Sheared" A·G

(*trans* Hoogsteen/Sugar edge)



UR0002

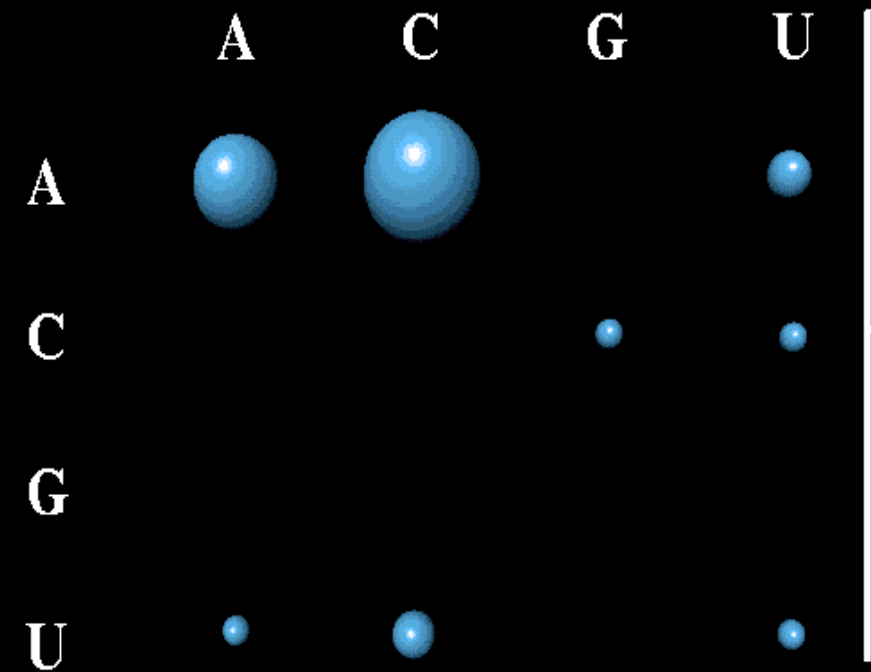
Isostericity Matrix
trans Hoogsteen/Sugar Edge

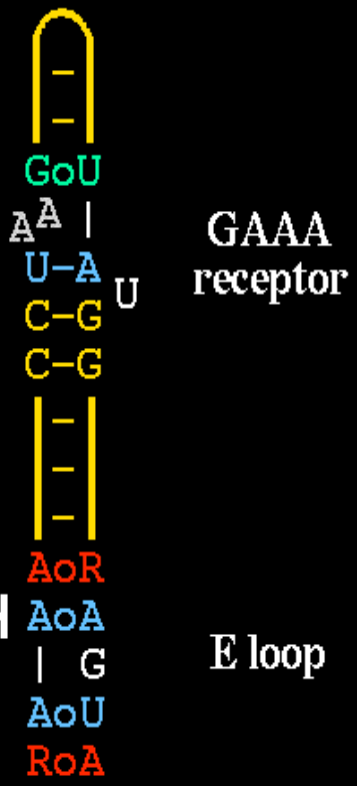
Hoogsteen \ Sugar Edge	A	C	G	U
A	I1	I1	I1	I1
C	I1	I1		I1
G			I2	
U	I2		I2	

Sequence signature of RNA folds

E loop 5'	GAAA receptor 5'		GAAA receptor 3'	E loop 3'
CAAUGAAAU	-AU-	CCUAAGU	[6] AUAUGGAUAGAGUAAUUU	
CGGGGAAAG	-CAU-	CCUACAG	[9] CCAUGGUGCAAAUACCCU	
CGGGGAAAG	-AA-	CCUACGU	[9] AUAUGGUUCGAUUACCCU	
AGGGAAAAACCA	-CCUAAGU		[6] AUAUGGUC-GAGUAUCCU	

RAAA CCUAAG UAUGG RAGUA

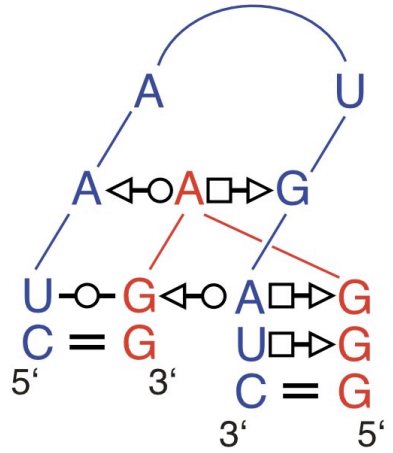


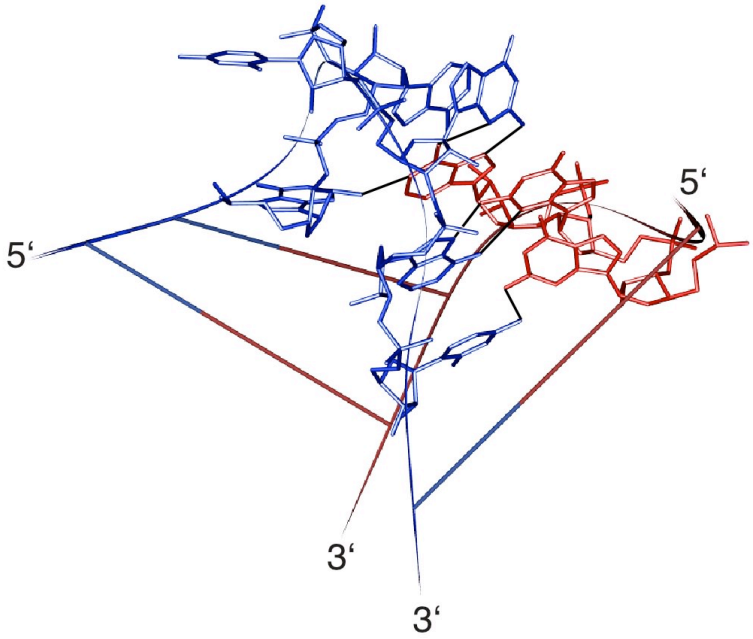


GAAA receptor

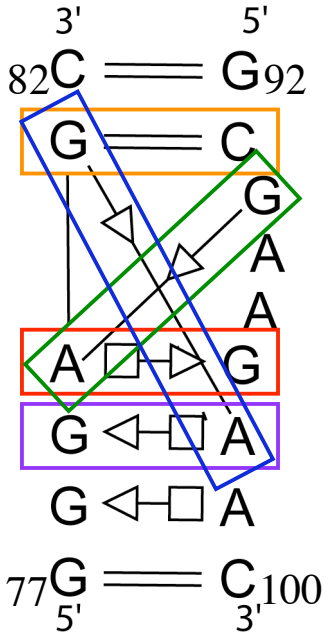
E loop

K-Turn





The K-turn motif



Watson-Crick

cis	A	C	G	U
A	I4	i2	I3	I1
C	I2	I6	I1	I5
G	I3	I1		i2
U	I1	I5	I2	I6

Sugar-edge

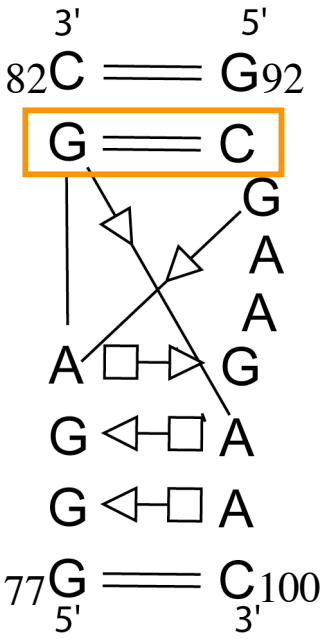
trans	A	C	G	U
A	I1		(I2)	
C	I1		I2	
G	I1		I2	
U	I1		I2	

Sugar-edge

trans	A	C	G	U
A	I1	I1	I1	I1
C	I1	I1		I1
G			I2	
U	I2		I2	

Cis Watson-Crick/Watson-crick

The K-turn motif



Watson-Crick

cis	A	C	G	U
A	I4	i2	I3	I1
C	I2	I6	I1	I5
G	I3	I1		i2
U	I1	I5	I2	I6

Watson-Crick

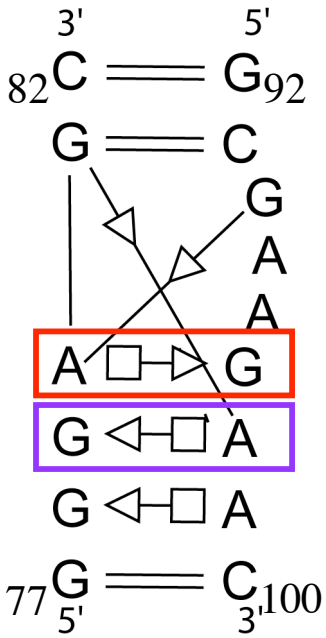
G85

C97

trans	A	C	G	U	-
A			0 (7)	7 (0)	
C	0 (6)		801 (786)	0 (7)	0 (3)
G					
U	6 (0)		11 (17)		
-			0 (3)		3 (0)

Trans Hoogsteen/Sugar-edge

The K-turn motif



Sugar-edge

Hoogsteen

trans	A	C	G	U
A	I1	I1	I1	I1
C	I1	I1		I1
G			I2	
U	I2		I2	

G97

A80

trans	A	C	G	U	-
A	4 (4)		675 (673)	145 (145)	0 (2)
C					
G					
U					
-			0 (2)		2 (0)

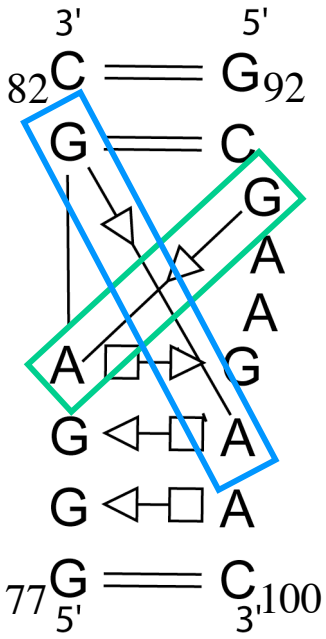
G79

A98

trans	A	C	G	U	-
A	126 (125)	0 (1)	583 (580)	113 (112)	0 (3)
C					
G					
U			1 (1)		
-	0 (1)	1 (0)	0 (3)	0 (1)	3 (0)

Cis Sugar-edge/Sugar-edge

The K-turn motif



Sugar-edge

Sugar-edge / trans	A	C	G	U
A	I1		(I2)	
C	I1		I2	
G	I1		I2	
U	I1		I2	

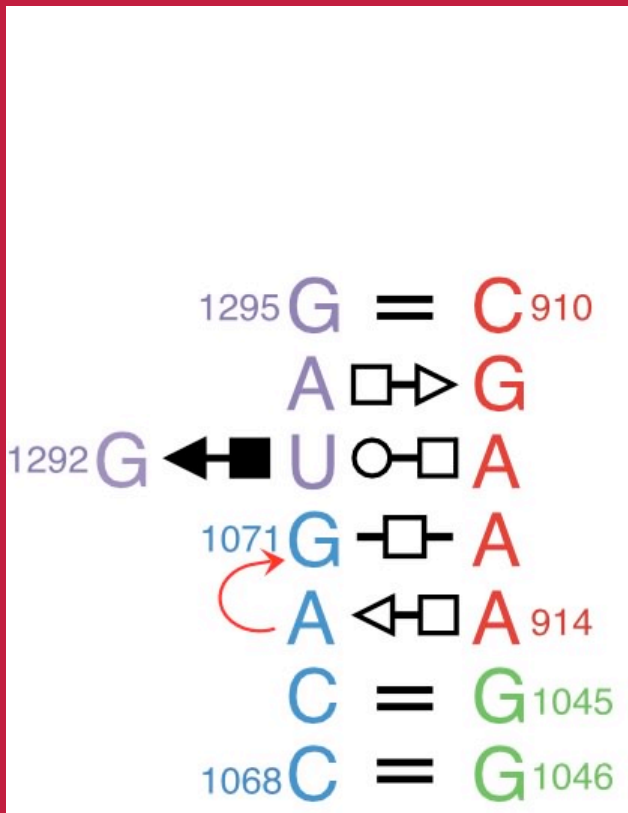
A80

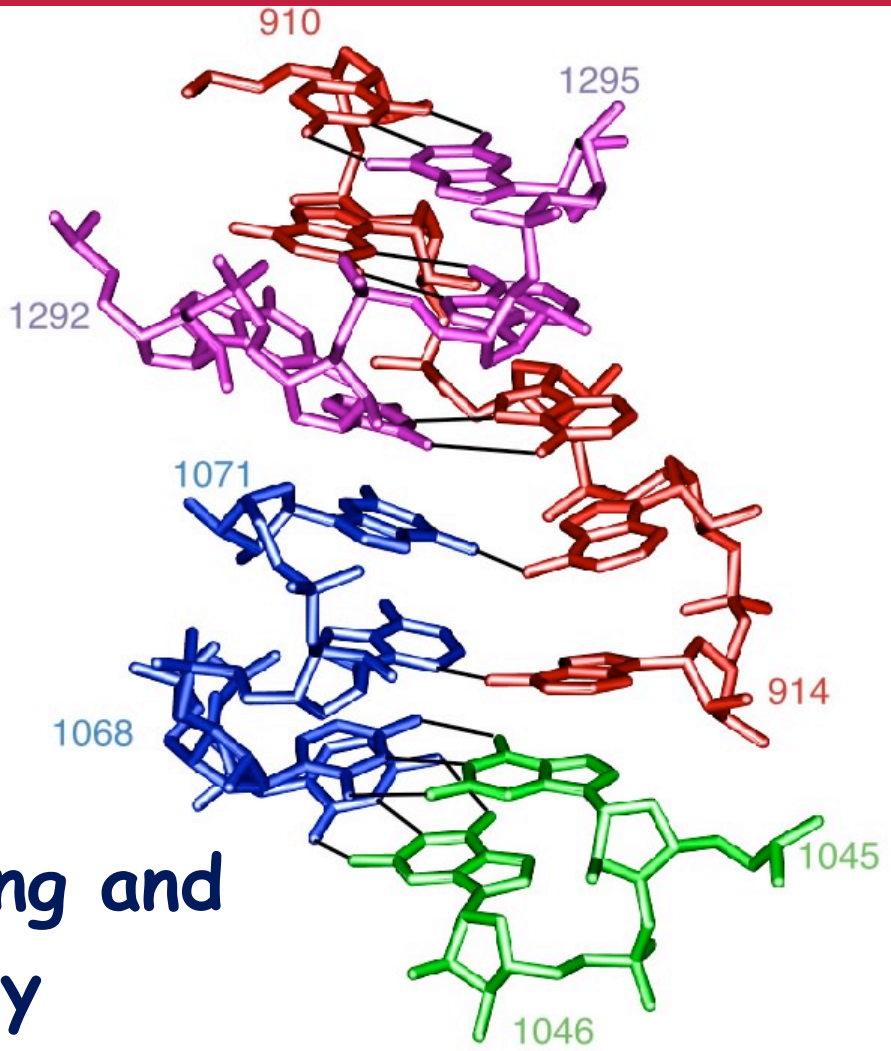
G94 / trans	A	C	G	U	-
A	38 (38)				
C	15 (15)				
G	130 (130)				
U	1 (1)				
-	640 (641)				3 (2)

A98

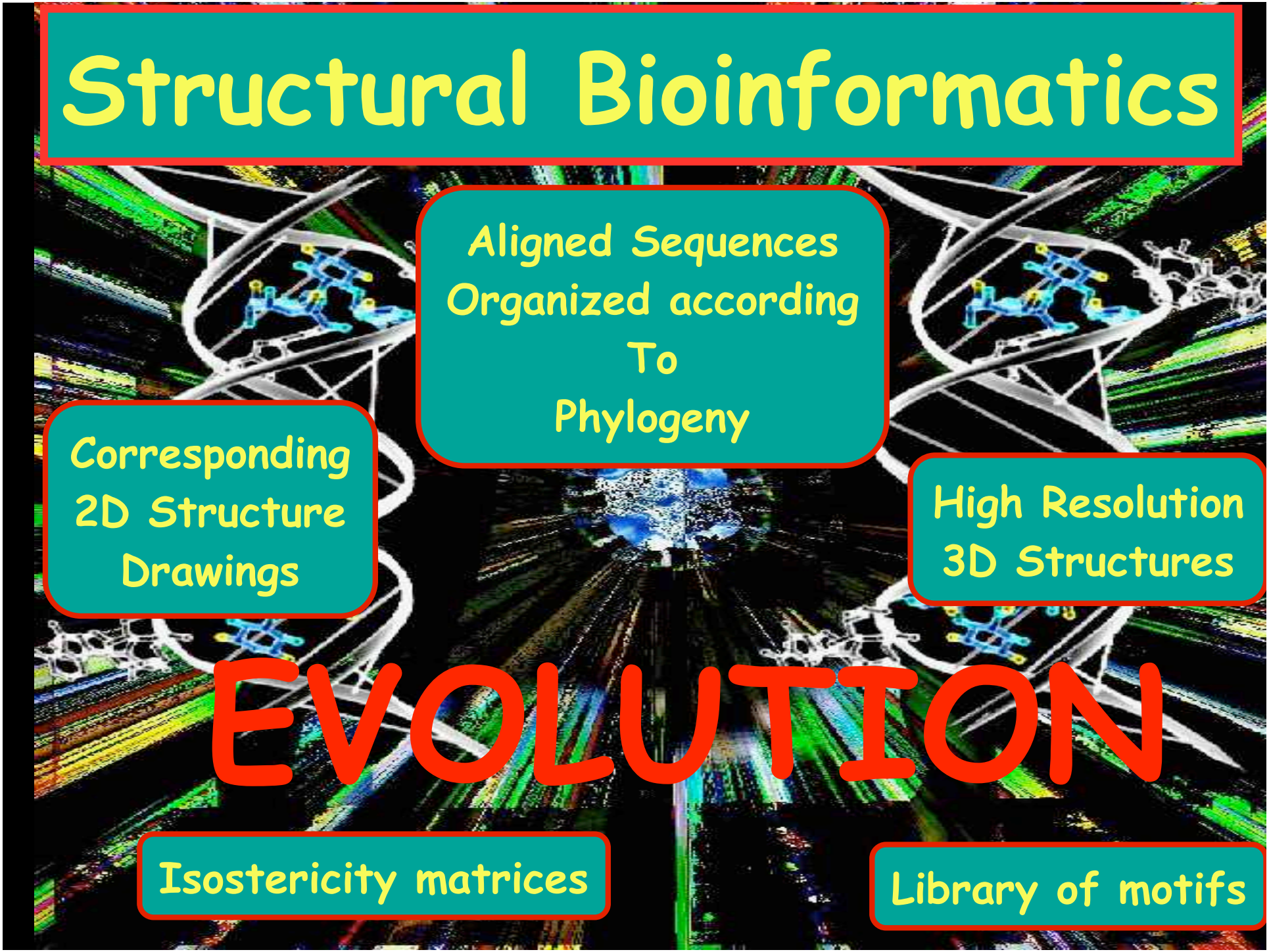
G81 / trans	A	C	G	U	-
A	6 (6)				
C					
G	810 (807)			1 (1)	1 (4)
U	7 (7)				
-	0 (3)				3 (0)

Composite motifs : association of several strands

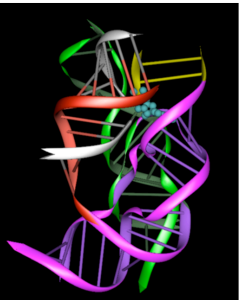




Importance of stacking and Geometry for stability


Structural Bioinformatics
Aligned Sequences
Organized according
To
Phylogeny
Corresponding
2D Structure
Drawings
High Resolution
3D Structures
EVOLUTION
Isostericity matrices
Library of motifs

Levinthal's Paradox



L = 100 nucleotides > 4^{100} (10^{60}) sequences

$1{,}8^{400}$ (10^{102}) structures 2D

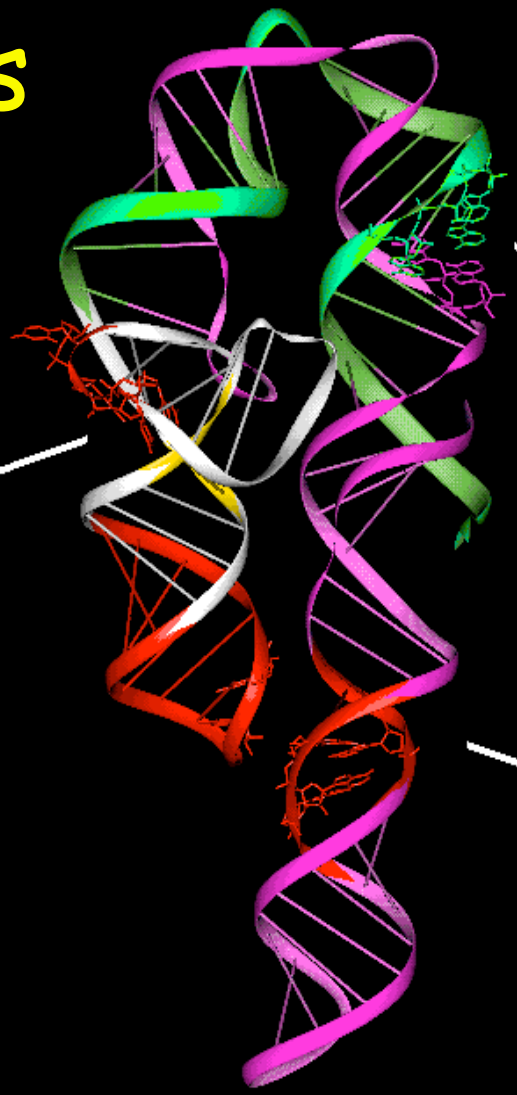
7 torsion angles / nucleotide

Torsion angles restricted & in the preferred conformers all the short range interactions are favourable

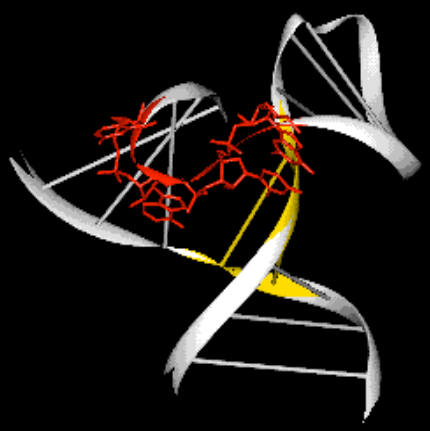
RNA Tectonics

- **Modular and Hierarchical assembly of RNA molecules using motifs**

Each structure can be parsed in key motifs



Nucleation fold

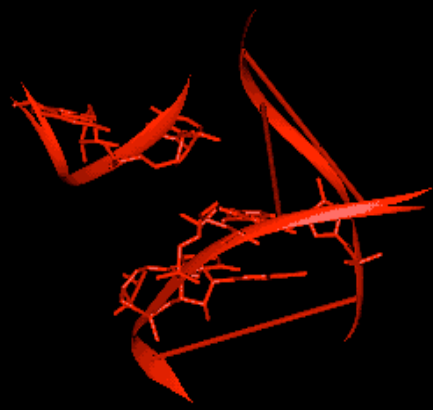


P5abc 3-way junction

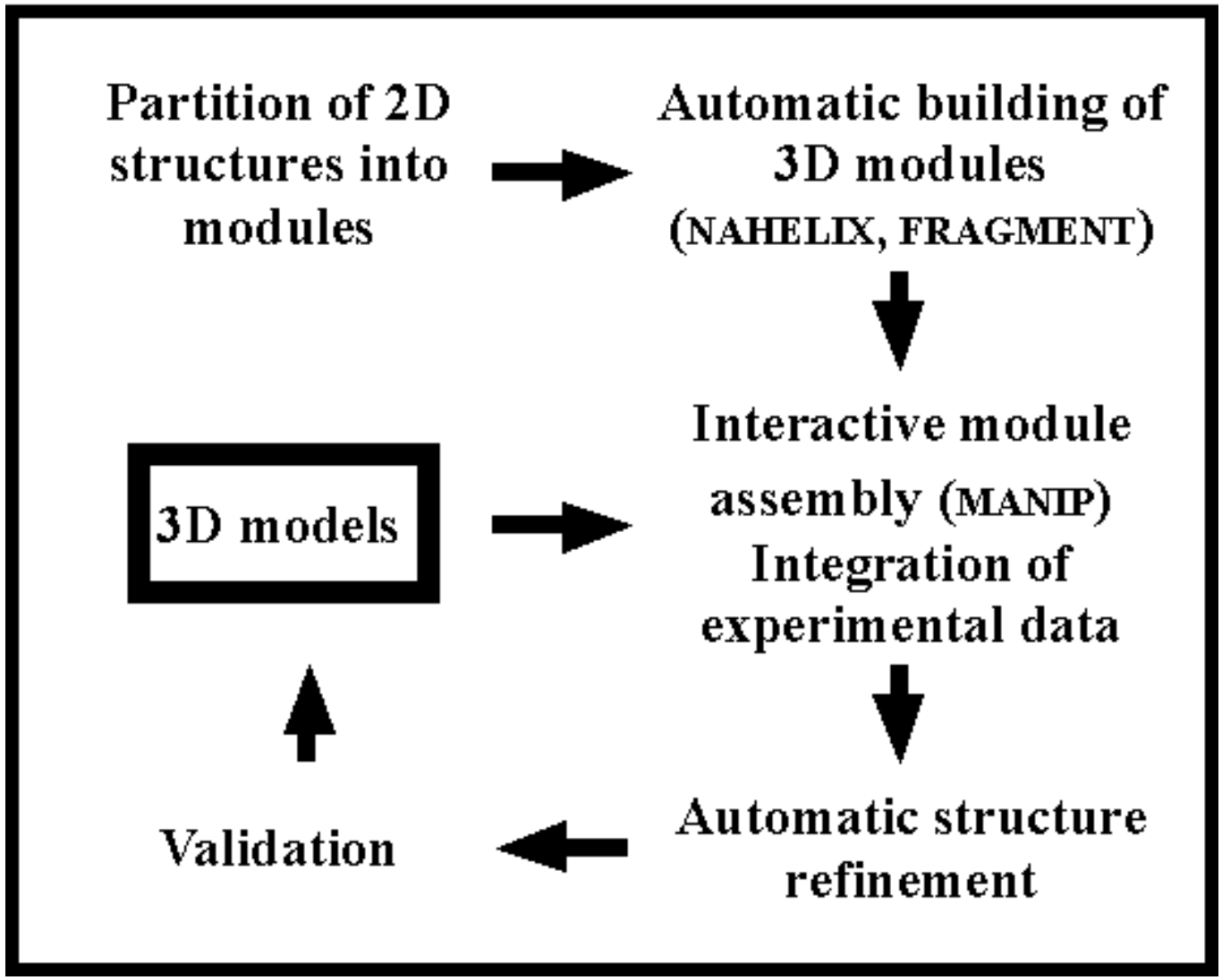
GoU recognition fold

J4/5

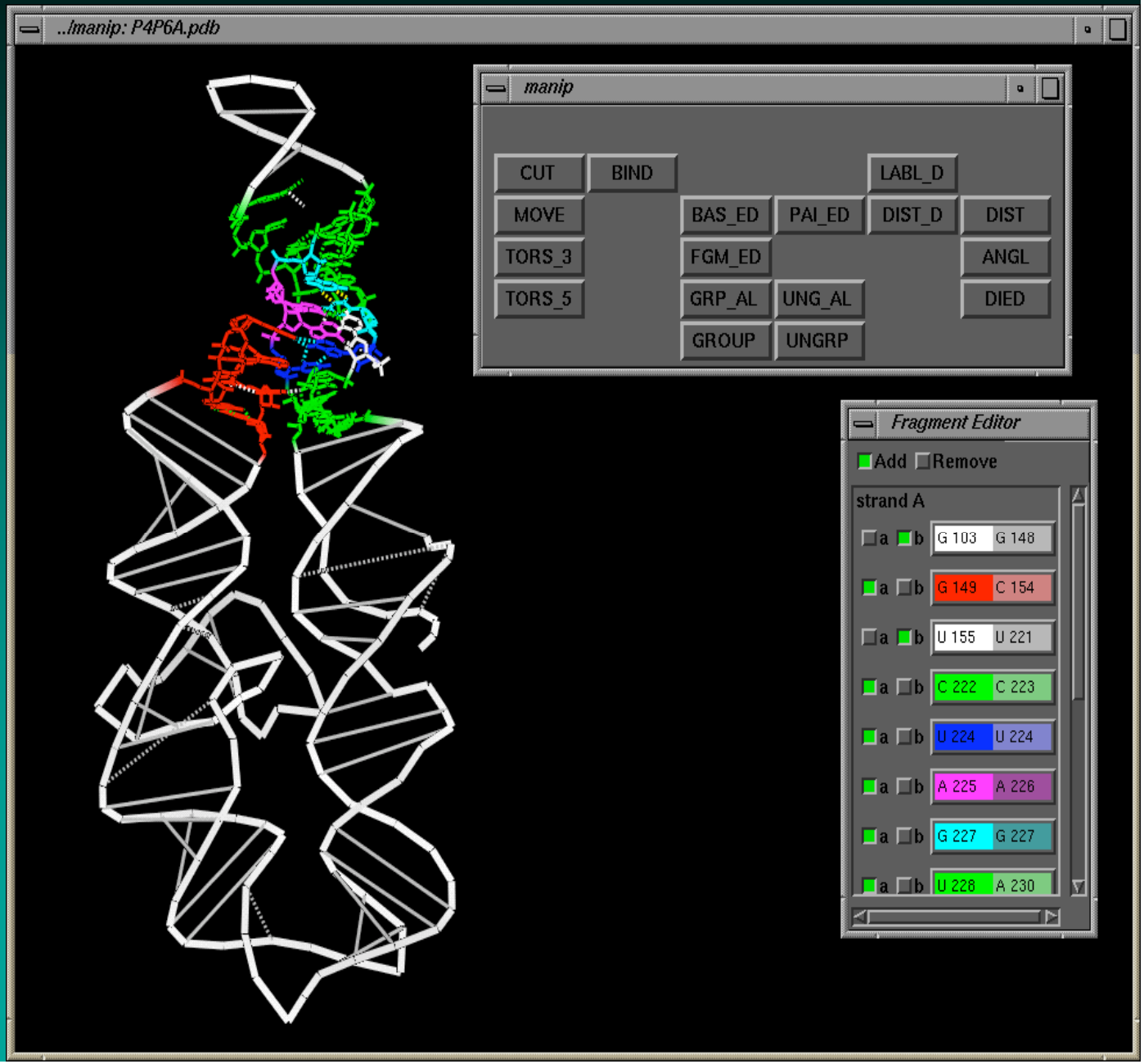
RNA self-assembly motif



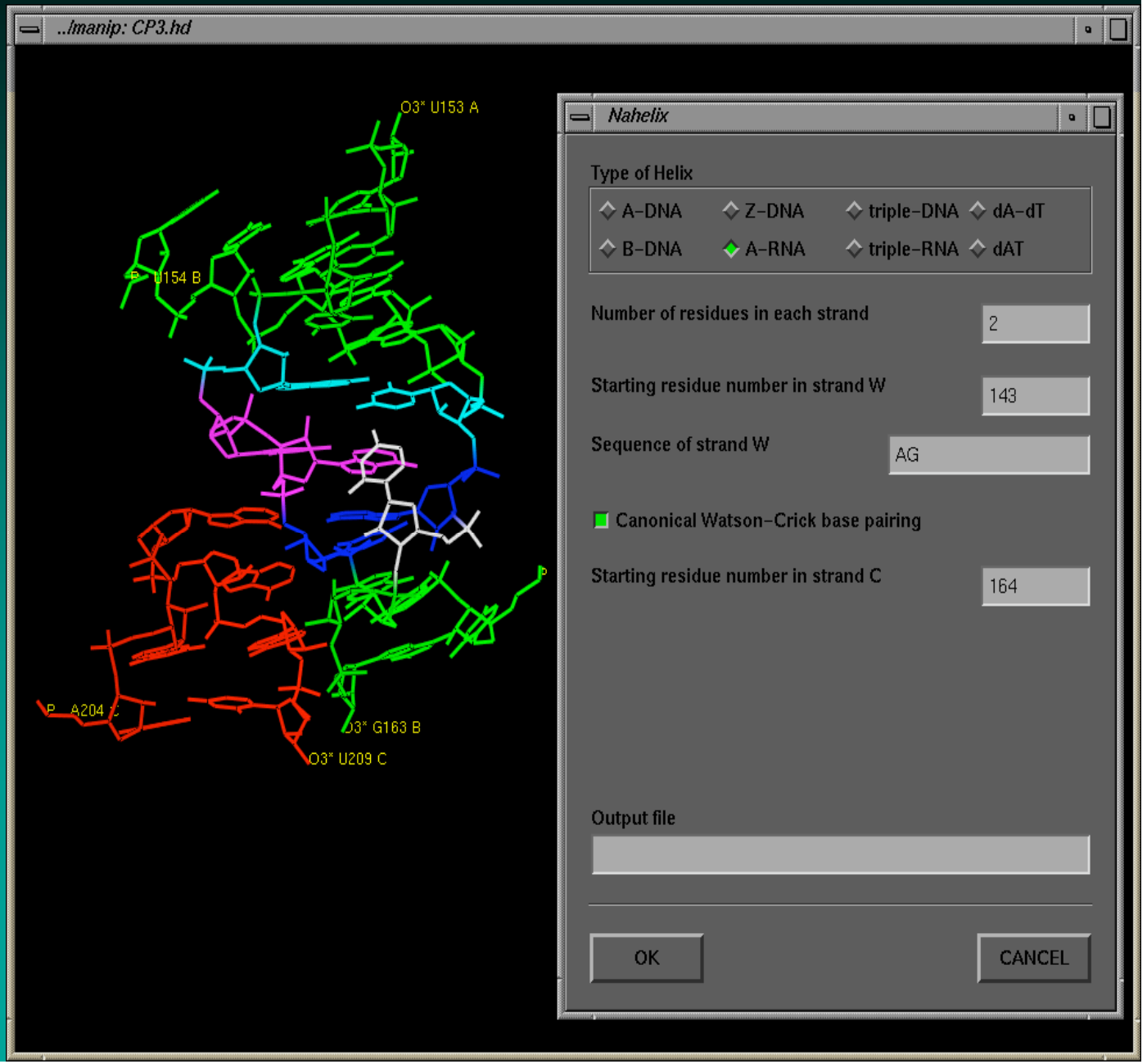
GAAA ··· 11-nt receptor


Partition of 2D structures into modules
Automatic building of 3D modules (NAHELIX, FRAGMENT)
Interactive module assembly (MANIP) Integration of experimental data
Automatic structure refinement
Validation
3D models

Modelling tool : MANIP



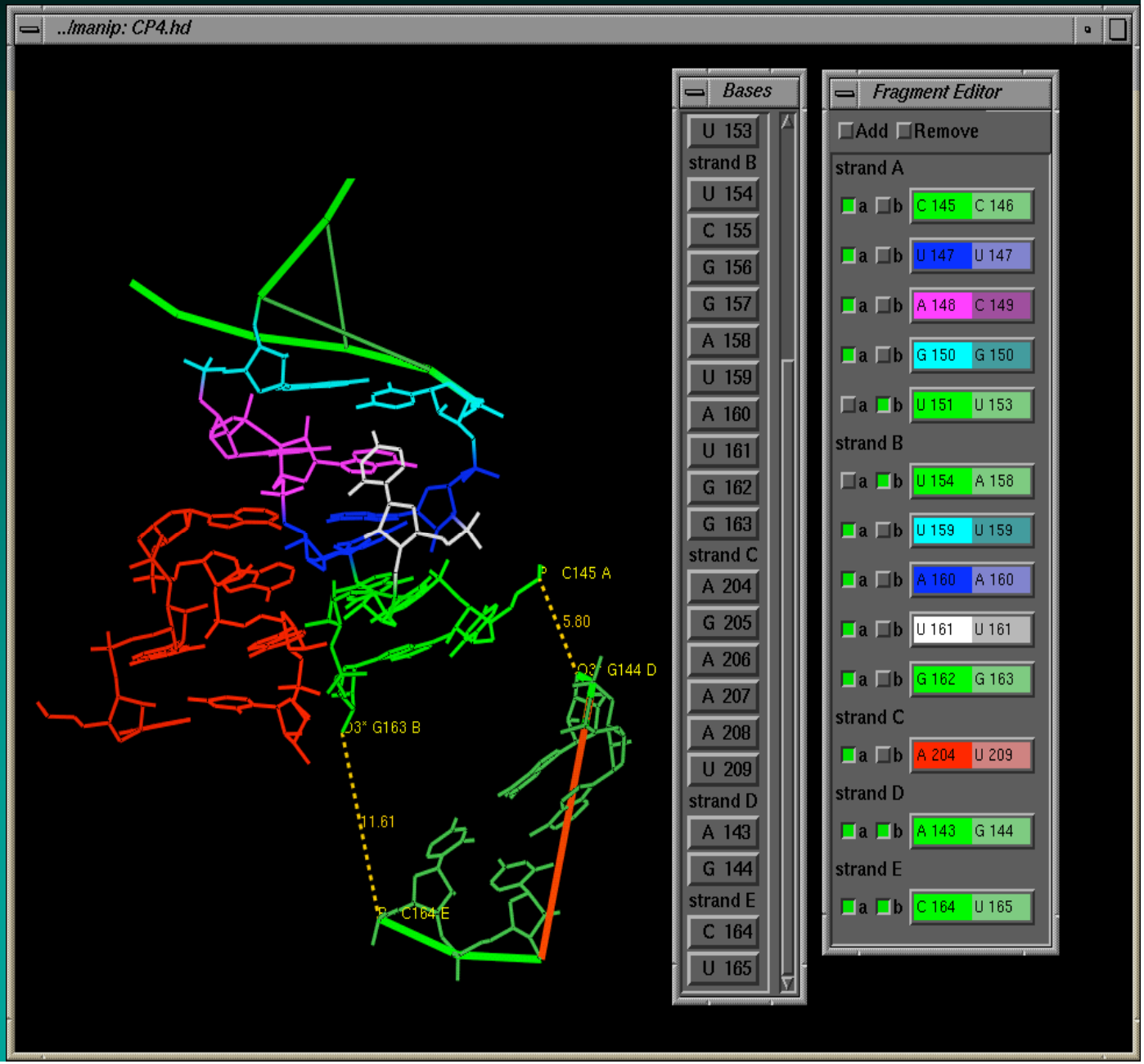
NAHELIX


../manip: CP3.hd
O3* U153 A
P U154 B
P A204 C
O3* G163 B
O3* U209 C
Nahelix
Type of Helix
A-DNA
Z-DNA
triple-DNA
dA-dT
B-DNA
A-RNA
triple-RNA
dAT
Number of residues in each strand
2
Starting residue number in strand W
143
Sequence of strand W
AG
Canonical Watson-Crick base pairing
Starting residue number in strand C
164
Output file
OK
CANCEL

FRAGMENT

NUCLIN

NUCLSQ



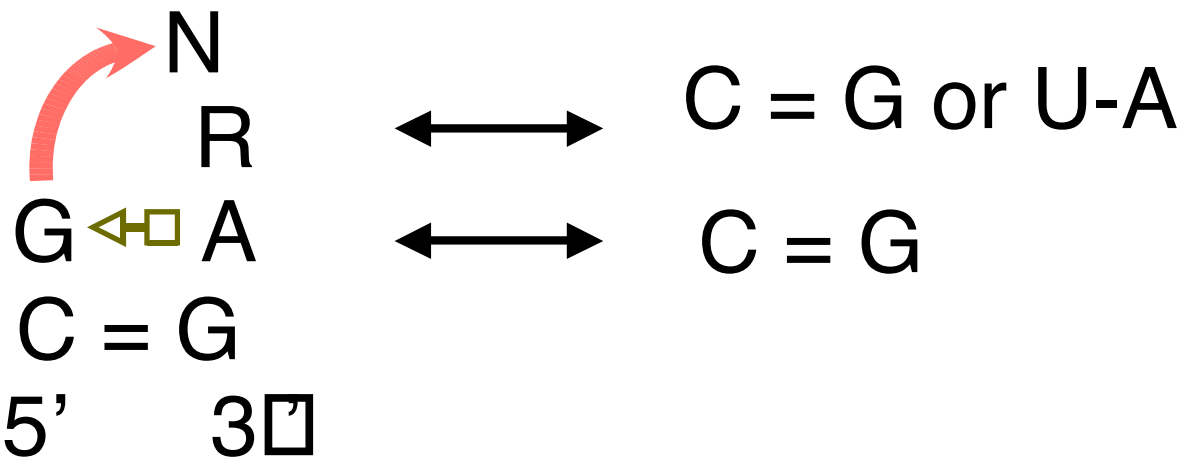
Properties of the modules:

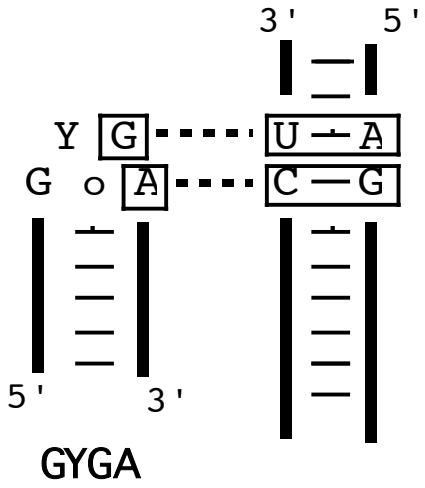
- interfaces should be easily identified;
- they should interact via defined protocols;
- they should be easily modified and should evolve independently.

Properties of the Protocols:

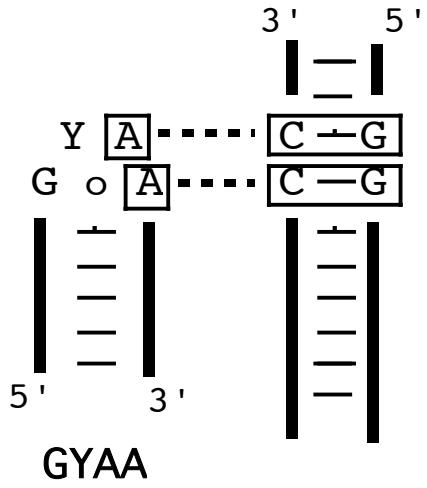
- Rules of association between modules;
- Should be robust & should promote evolution.

GNRA - helix/receptors

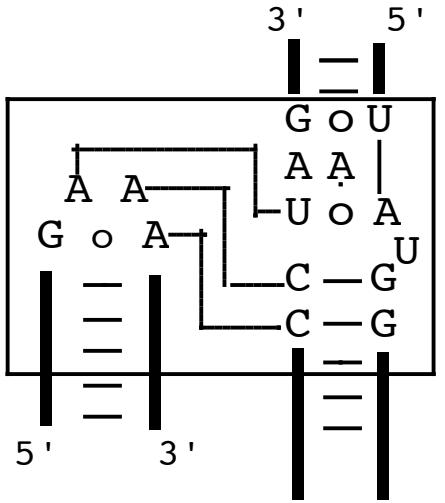




GYGA



GYAA



11nt-receptor motif

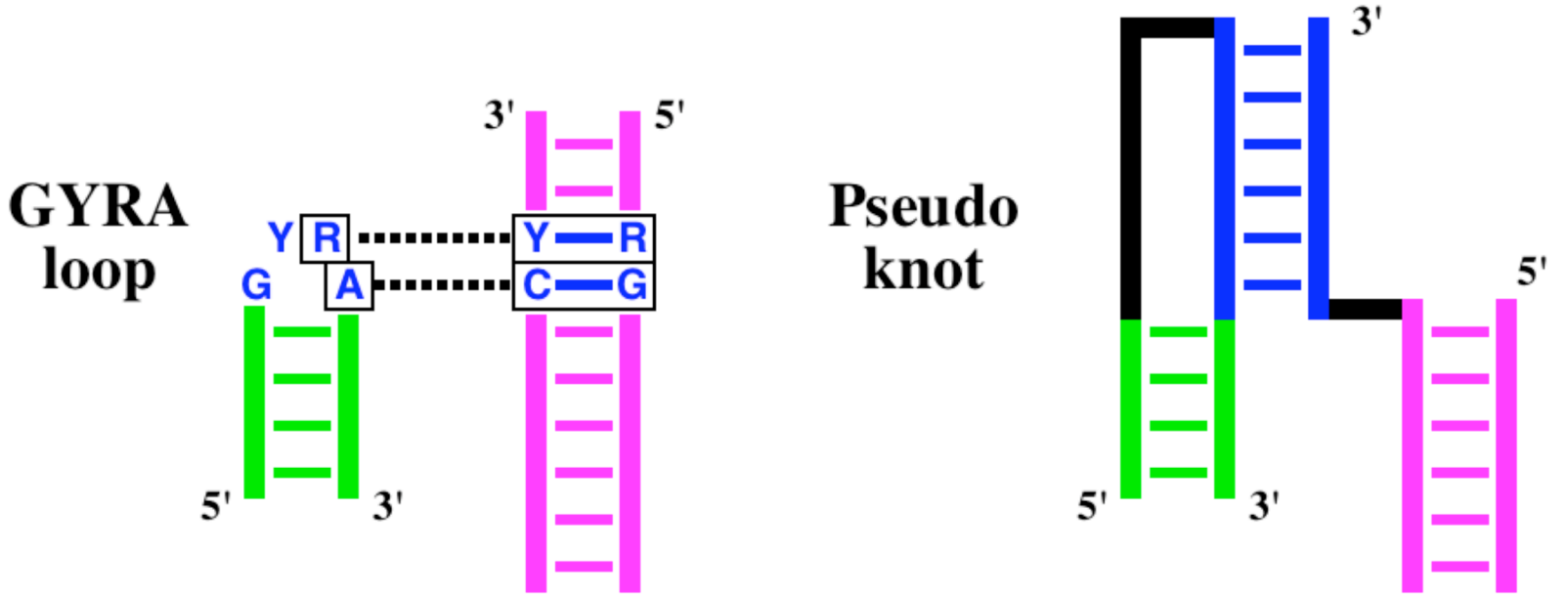
Swap between GNRA/helix and loop-loop motifs

GYRA loop

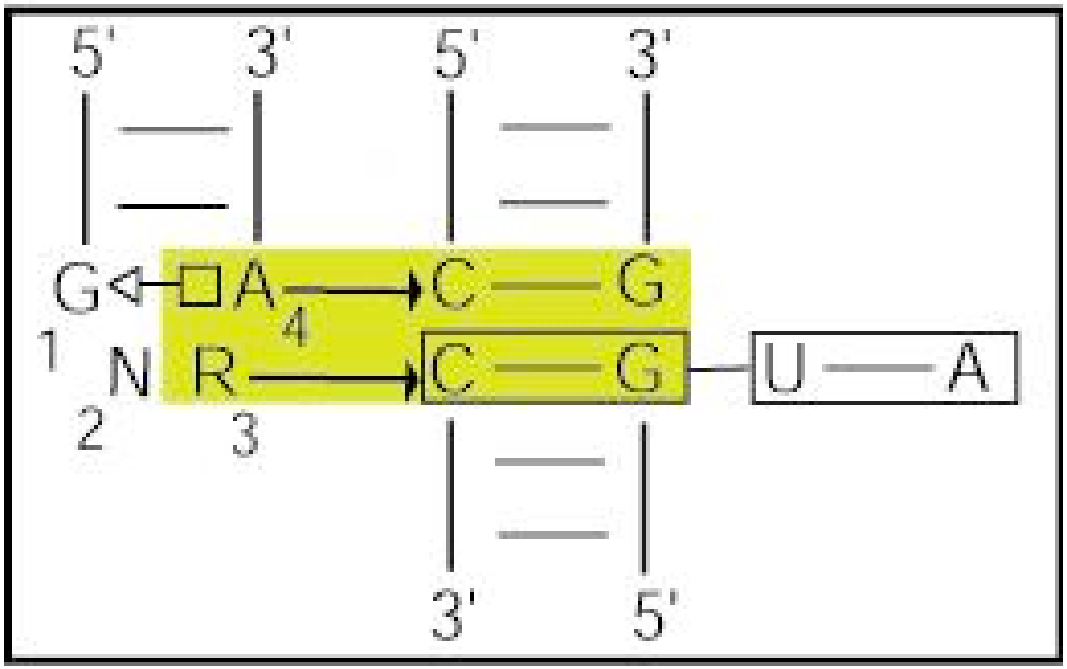
AGAGUCGAUU[ ]GGGC----GUGAGCCU[ ]GAUCGACUCU----
CAGGAUAGGG[ ]CGGG----GCAACCCG[ ]CCCUAUCCUG----

Pseudoknot

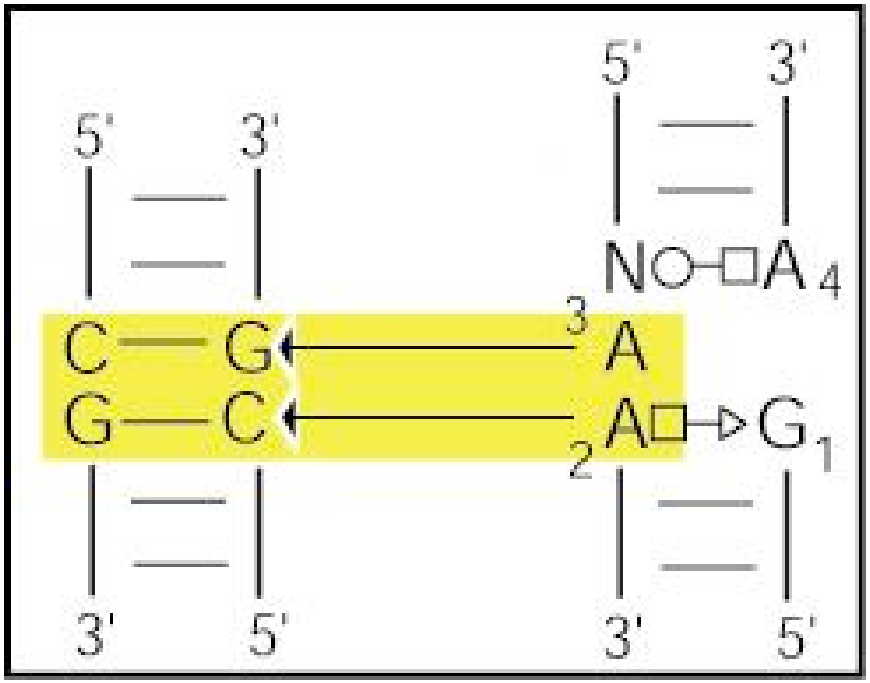
----GCUGUC[ ]GAUG-UGAUAGCCAUA[ ]GACAGCAUGCUAUC
---CGCUGUC[ ]GACGCUAAUAGACGUC[ ]GGCAGCGUUCUAUU

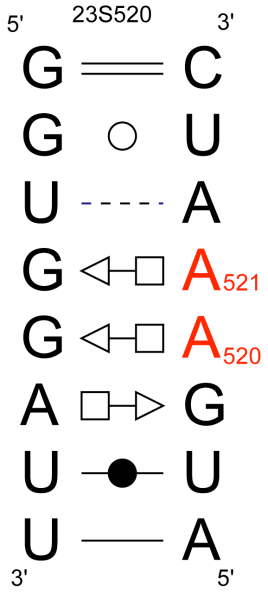


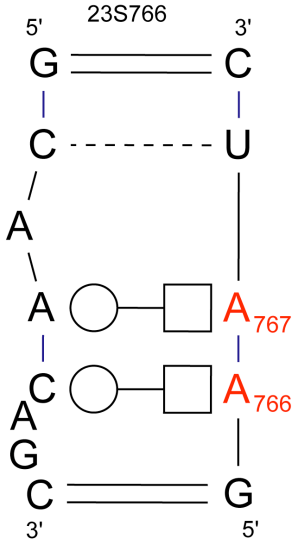
GNRA

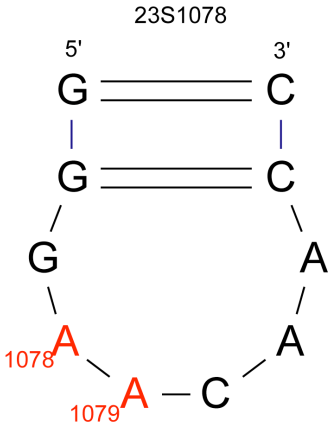


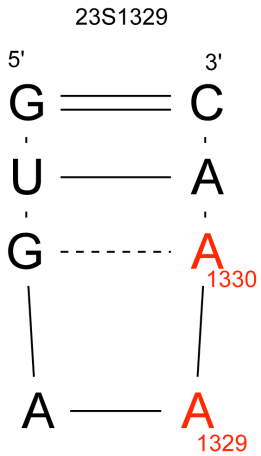
A-rich internal loop

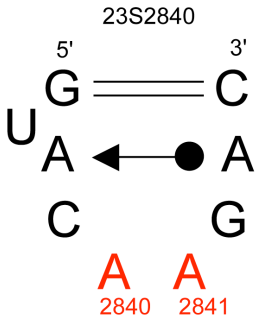




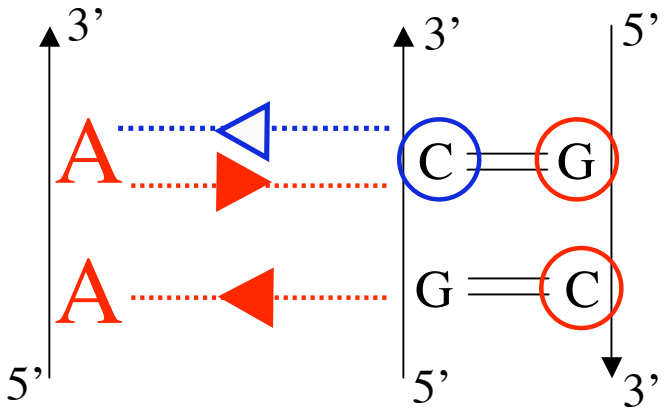




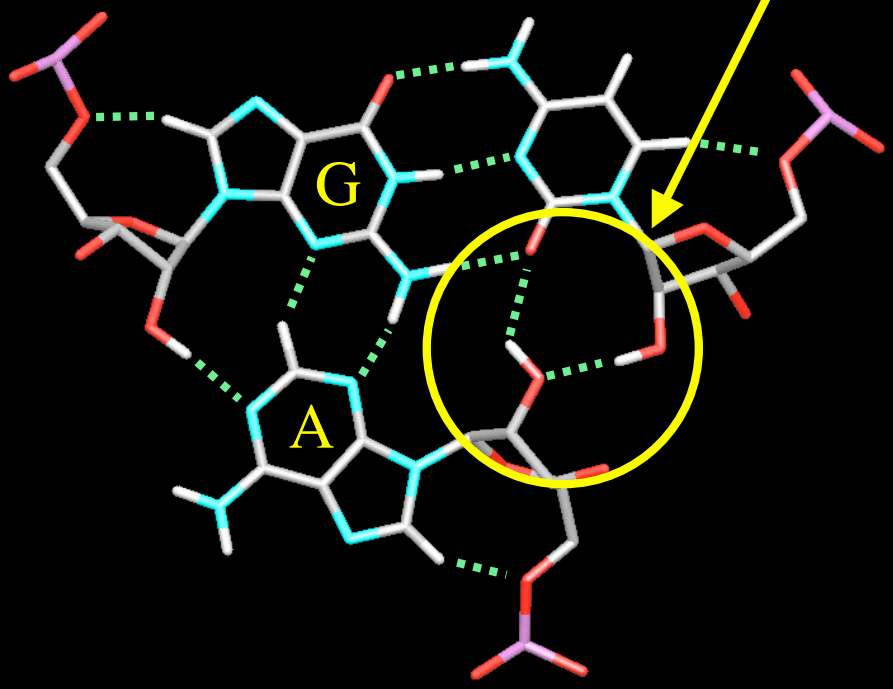


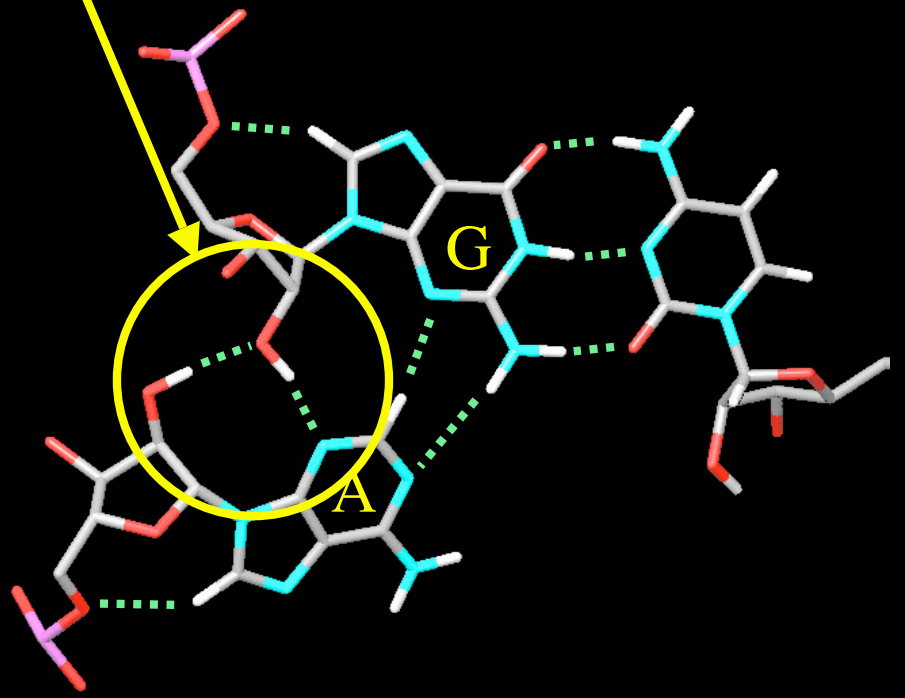


Always the same base pairs with two adenines presented by various motifs



Sugar Edge/ Sugar Edge Ribose zipper

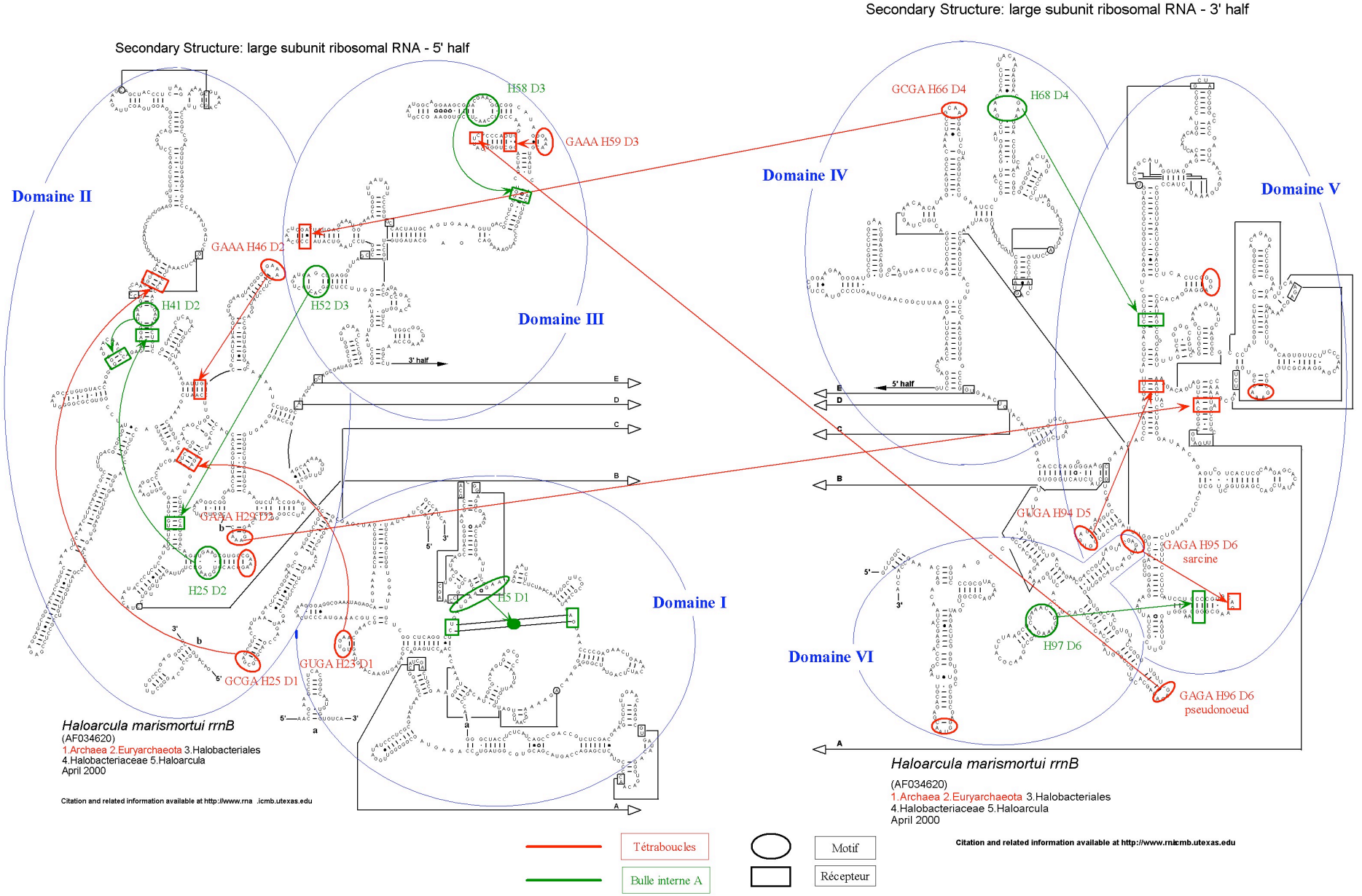




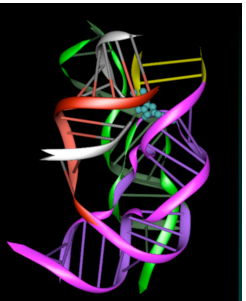
Trans and Parallel

Cis and Anti-Parallel

Variety of A-minor motifs in 50S particle



Laws of evolution

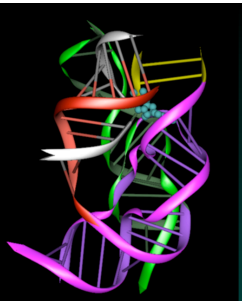


- Three-dimensional architectures evolve less with time than sequences

- Three-dimensional structures are dictated first by folding rules and secondarily by function

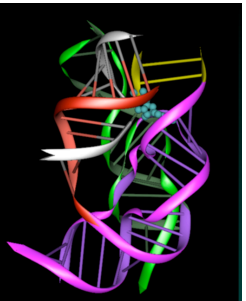
The phonetic structure of words are more stable than the meaning of words

Modelling algorithm



- 3D structure : assembly of fragments
- Stress 3D fold rather than sequence (inverse folding)
- Search for a «consensus» 3D fold (global architecture)

Modelling Constraints



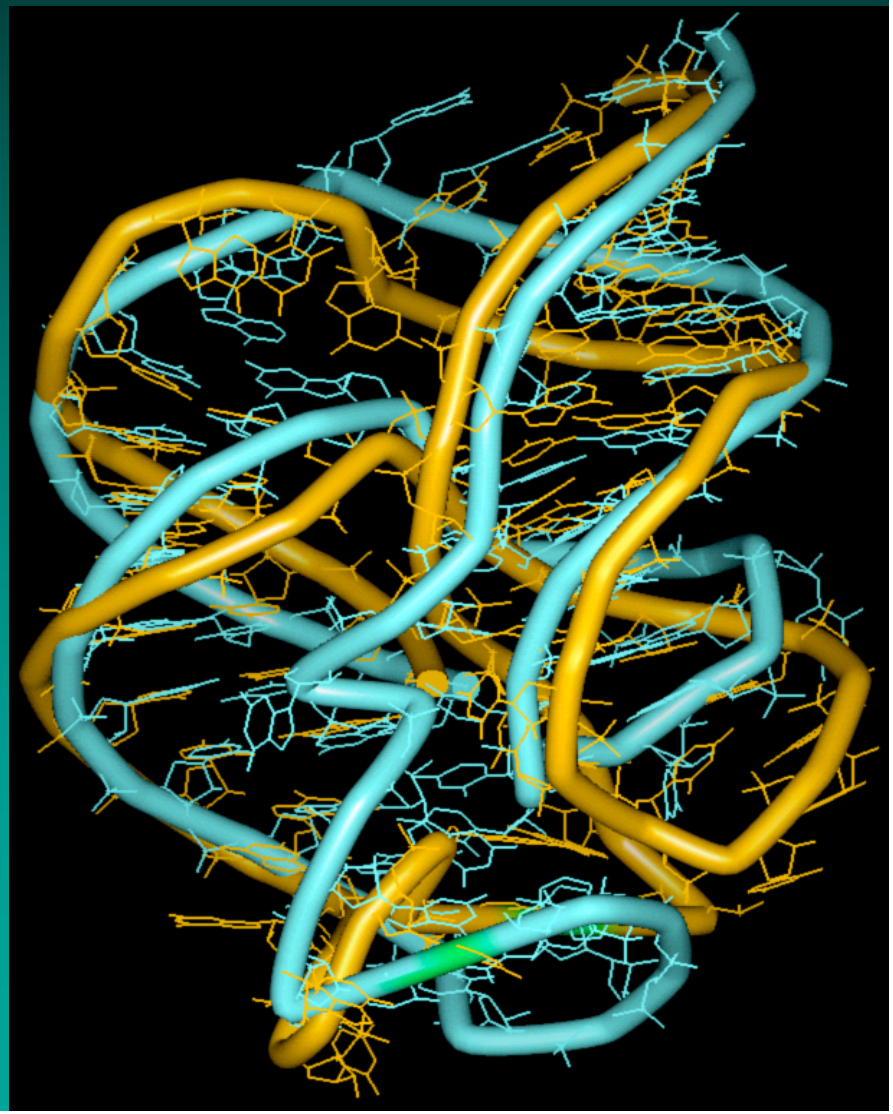
- 2D Topology (not strongly correlated with sequence) :
  - RNA is right-handed > right-handedness of stacks, of junctions
  - Type of junction : 3-way or 4-way…
  - Positions of single-strands w/r to the junction

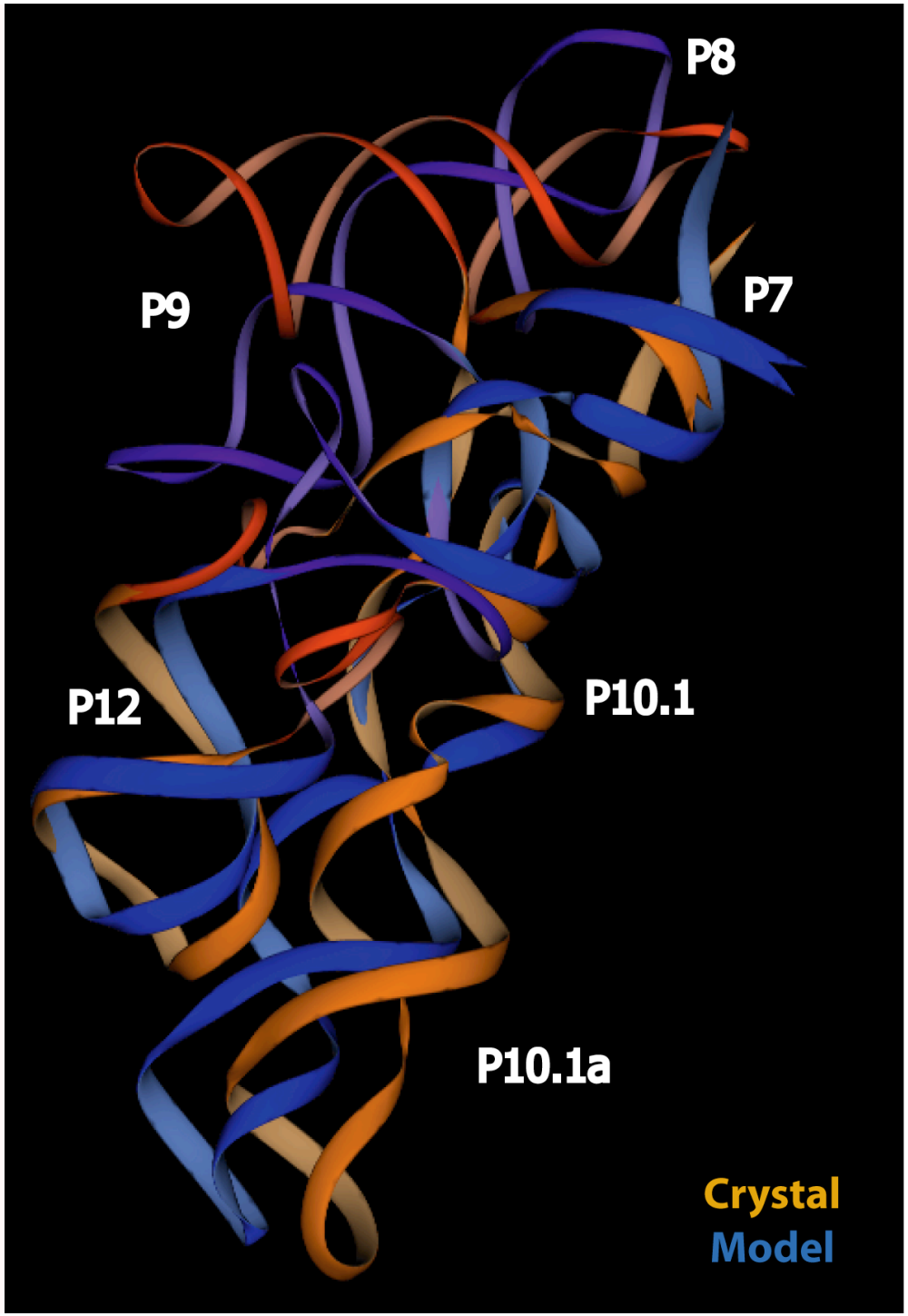
HDV ribozyme : X-ray vs. Model (r.m.s. 8.0 Å)

Ferré d'Amaré et al.
Nature 395, 567 (1998)

Tanner et al.
Current Biol. 4, 488 (1994)

Two invariant G=C base pairs were not modelled





***rmsd* X-ray structure/model**

overall:	**11.2 Å**
minus P8-P9:	**8.6 Å**
minus P8-P9 L11/12	**5.5 Å**

***nrmsd* X-ray structure/model**

Carugo & Pongor (2001) *Prot Sci*

overall:	**4.8 Å**
minus P8-P9:	**3.9 Å**
minus P8-P9 L11/12	**2.7 Å**

2003
1998

Rms (> 1500 atoms)

9.06 Å

Normalized rms
100 atoms)

3.85 Å