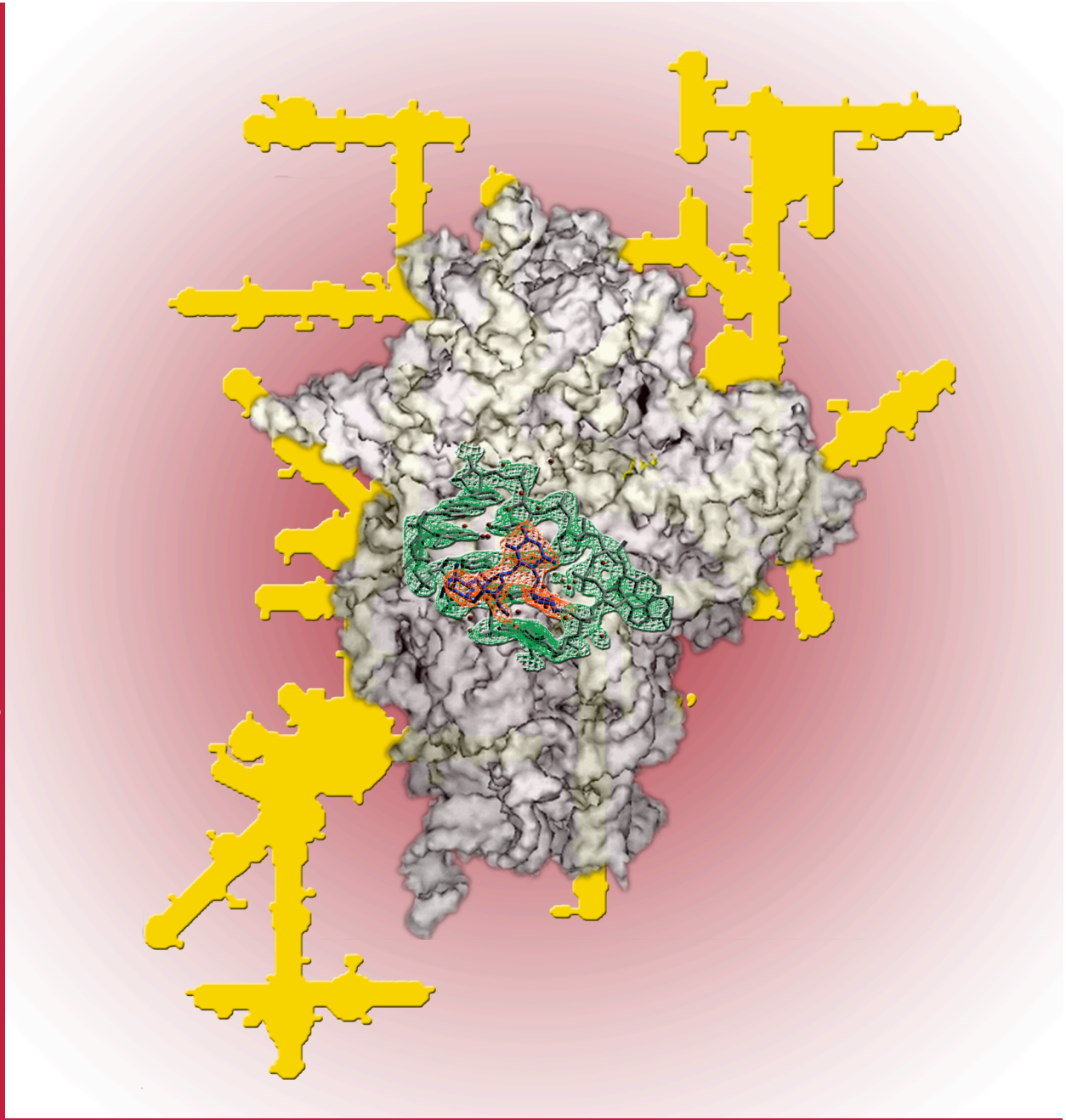
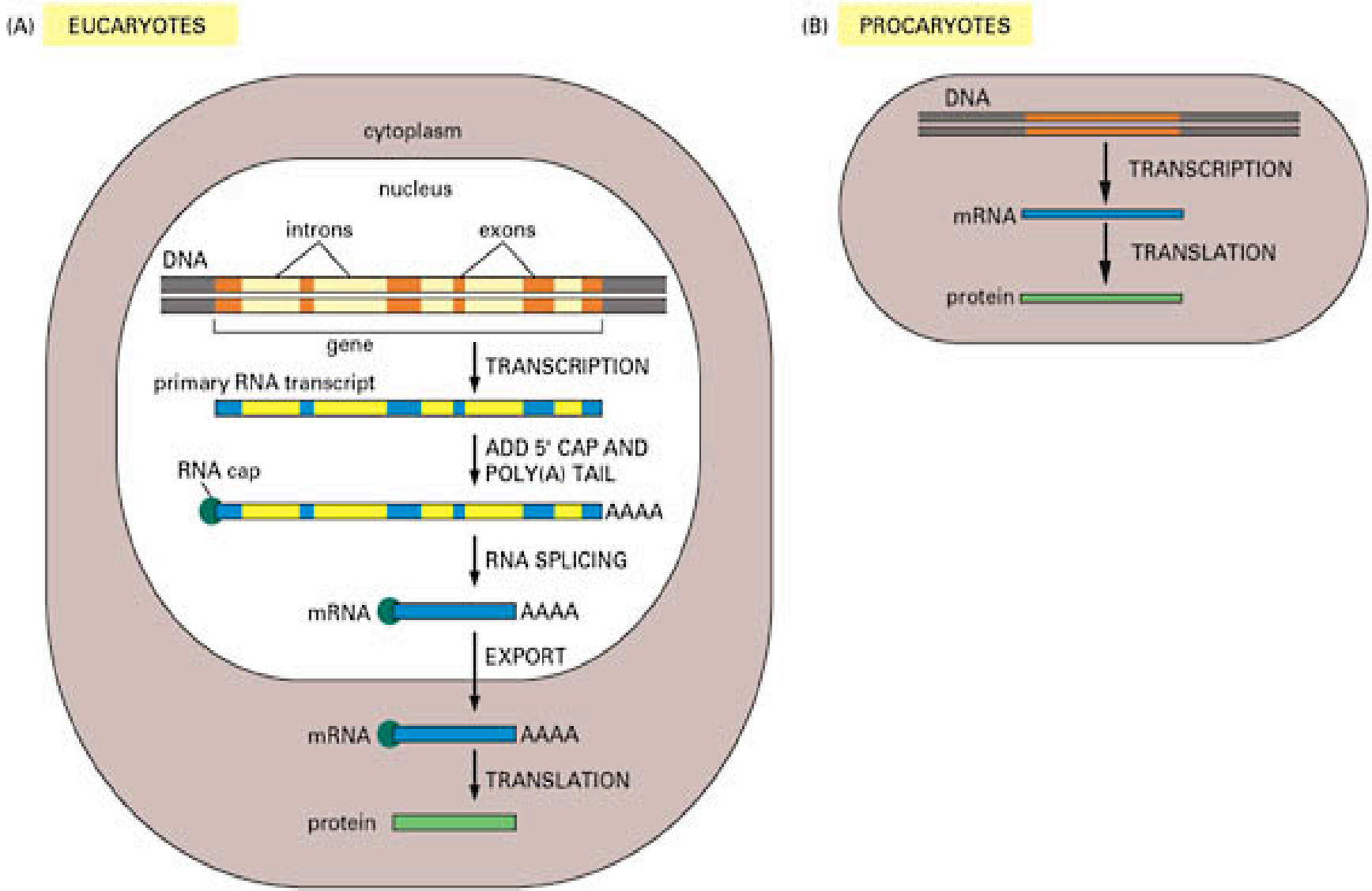
The RNA World II



E. Westhof

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

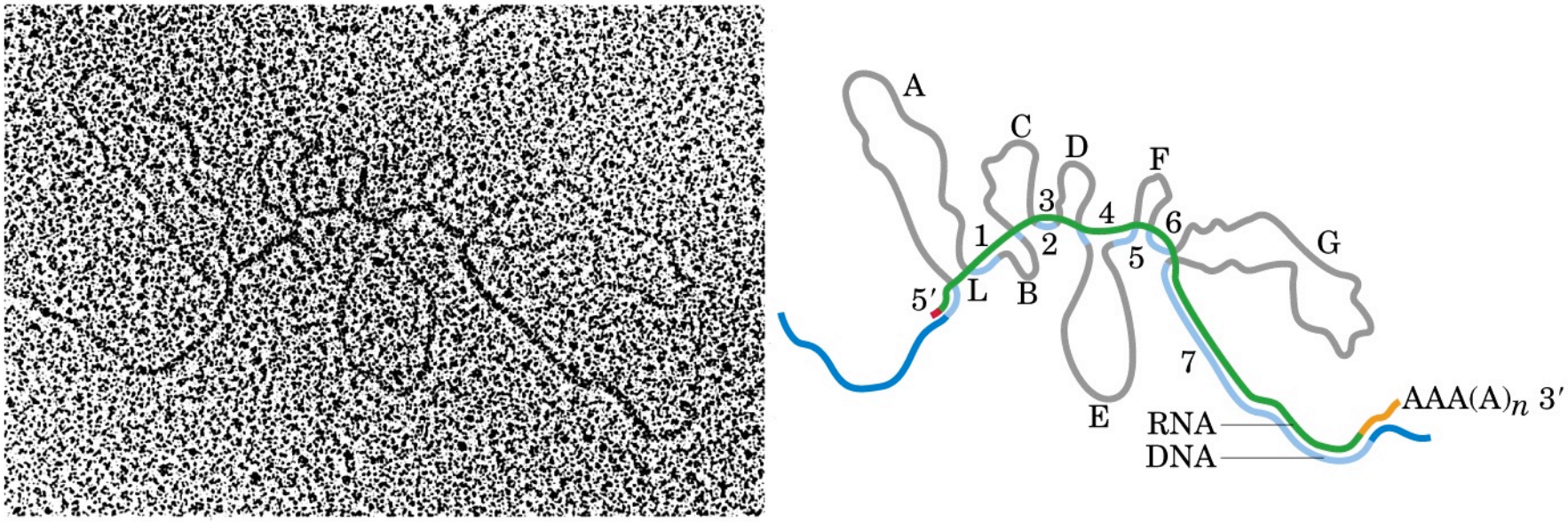
Eukaryotes -*vs*- Prokaryotes



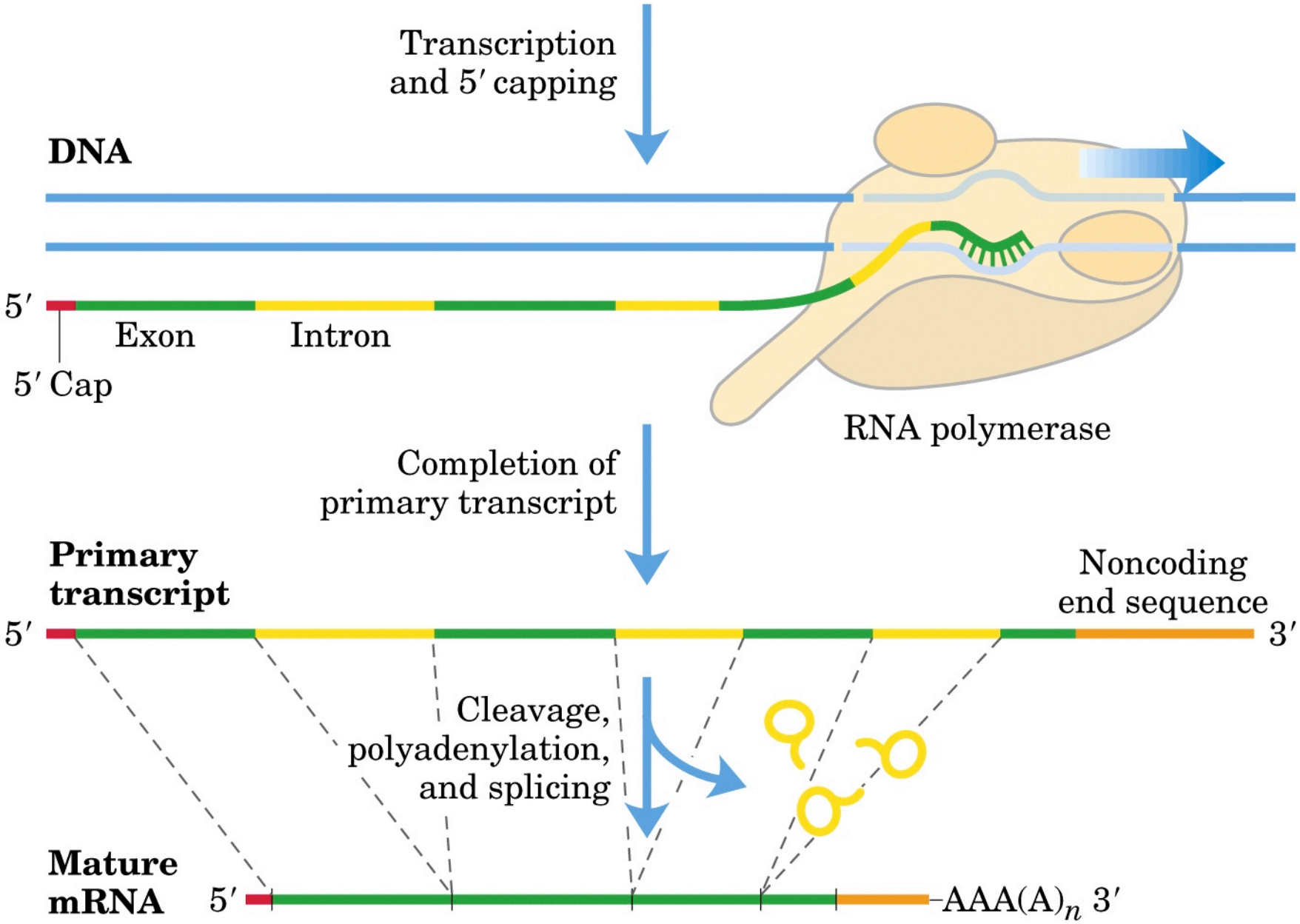
Local RNA structures in untranslated regions (UTR) have known roles in regulation of gene expression :

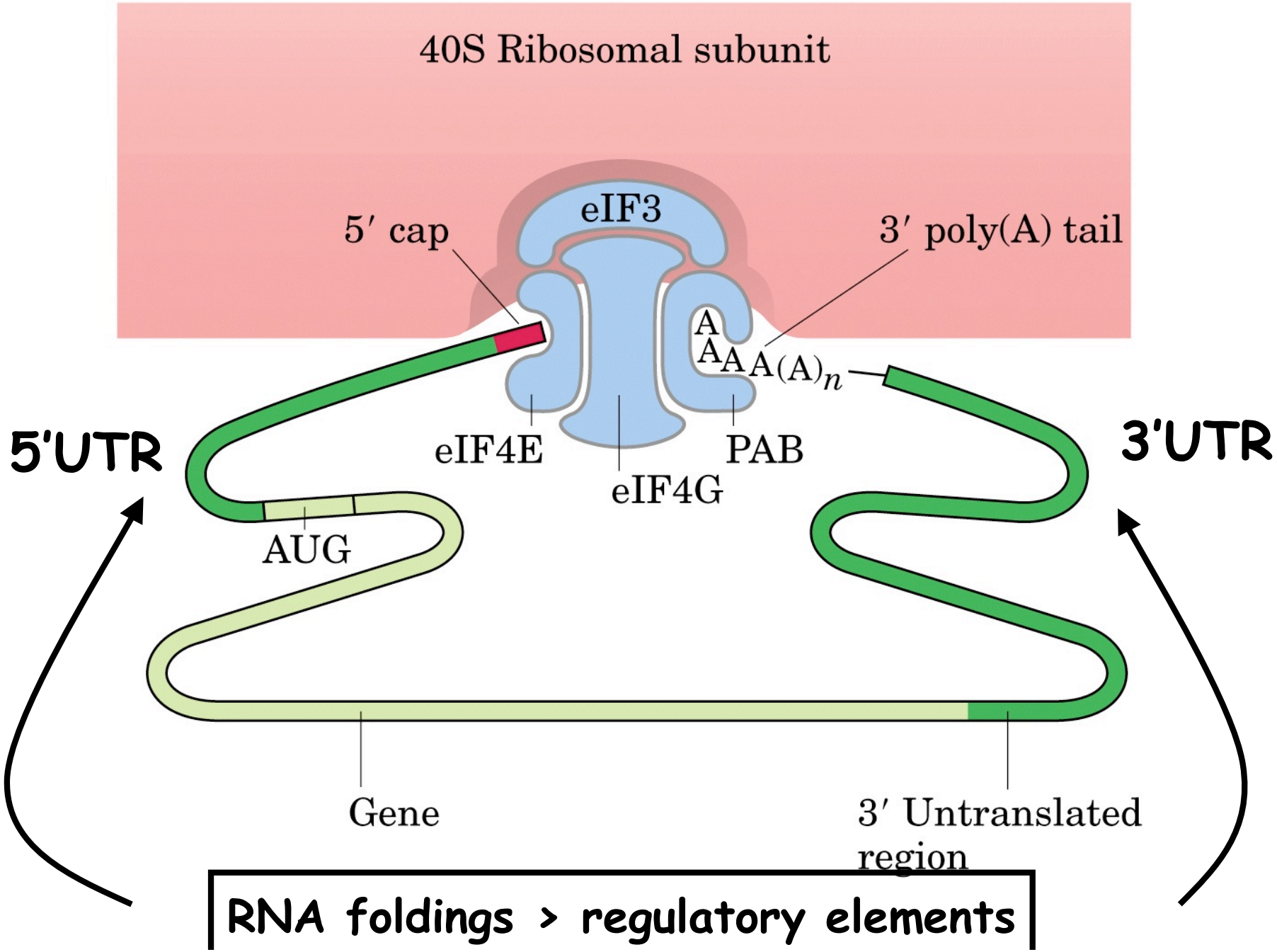
- mRNA stabilization
  - 5' UTR elements in bacteria reduce mRNA degradation
  - 3' UTR elements in eukaryotes control mRNA degradation
- mRNA translation
  - Control and Rate of translation
  - IRES (viruses)
- mRNA localization
  - Transport - development
- mRNA processing
  - Splicing of introns (alternative)

In the coding regions, redundancy of the genetic code leaves (some) room for RNA sec. structure on top



(a) (b)


Transcription
and 5′ capping
DNA
5′
Exon
Intron
5′ Cap
RNA polymerase
Completion of
primary transcript
Primary
transcript
Noncoding
end sequence
5′
3′
Cleavage,
polyadenylation,
and splicing
Mature
mRNA
5′
AAA(A)n 3′



RNA foldings > regulatory elements

Classes of functional RNAs

I. Sequence is the function

   - coding segments

   - non-coding where Watson-Crick recognition is the key element (siRNA, …)

II. Tertiary fold is the function

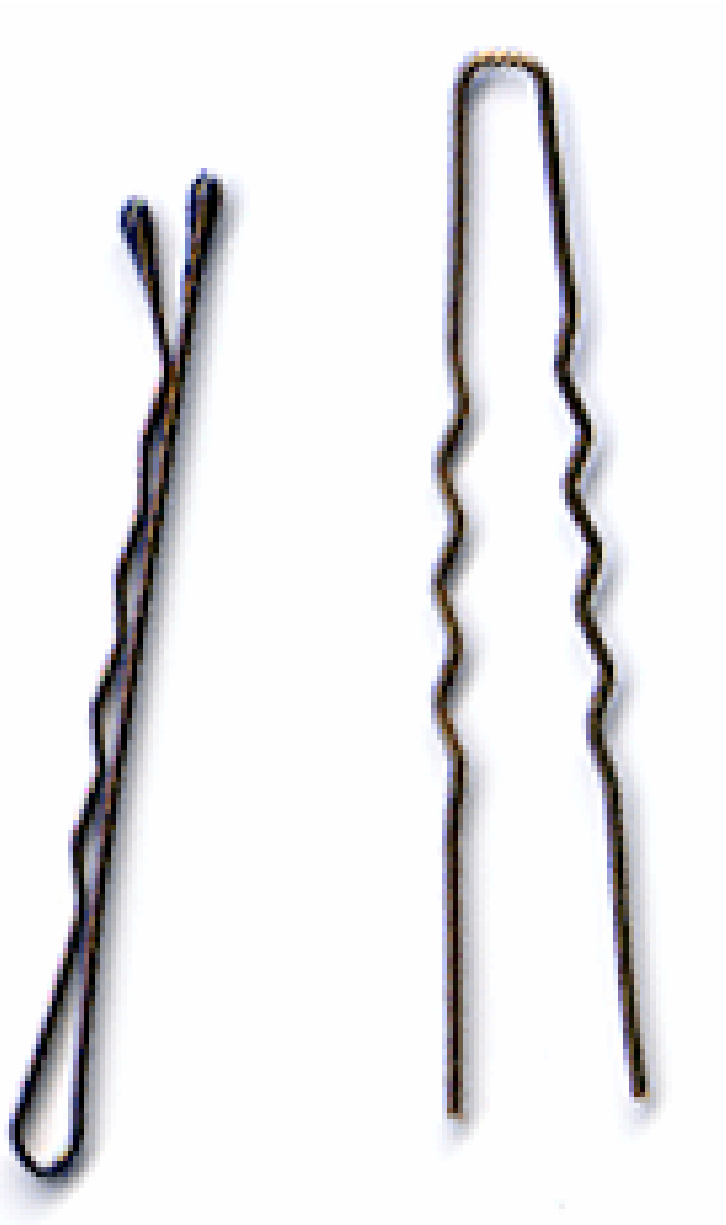
   - rRNAs, tRNAs, …

   - catalytic RNAs (RNaseP, HDV,…)

III. Most present both characters

   - telomerase RNA, snoRNAs,…

Hairpins



Terminators

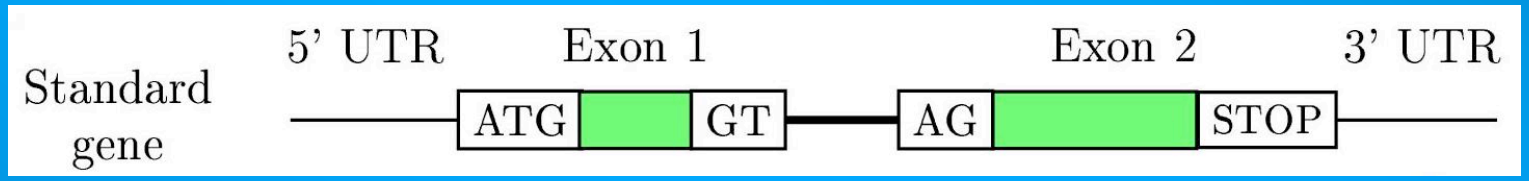
- Stem/loop
  - structural only
- 3’-U tail

Rho-independent

- C-rich
- G-poor
- “loose” consensus

Rho-dependent

Gene structure

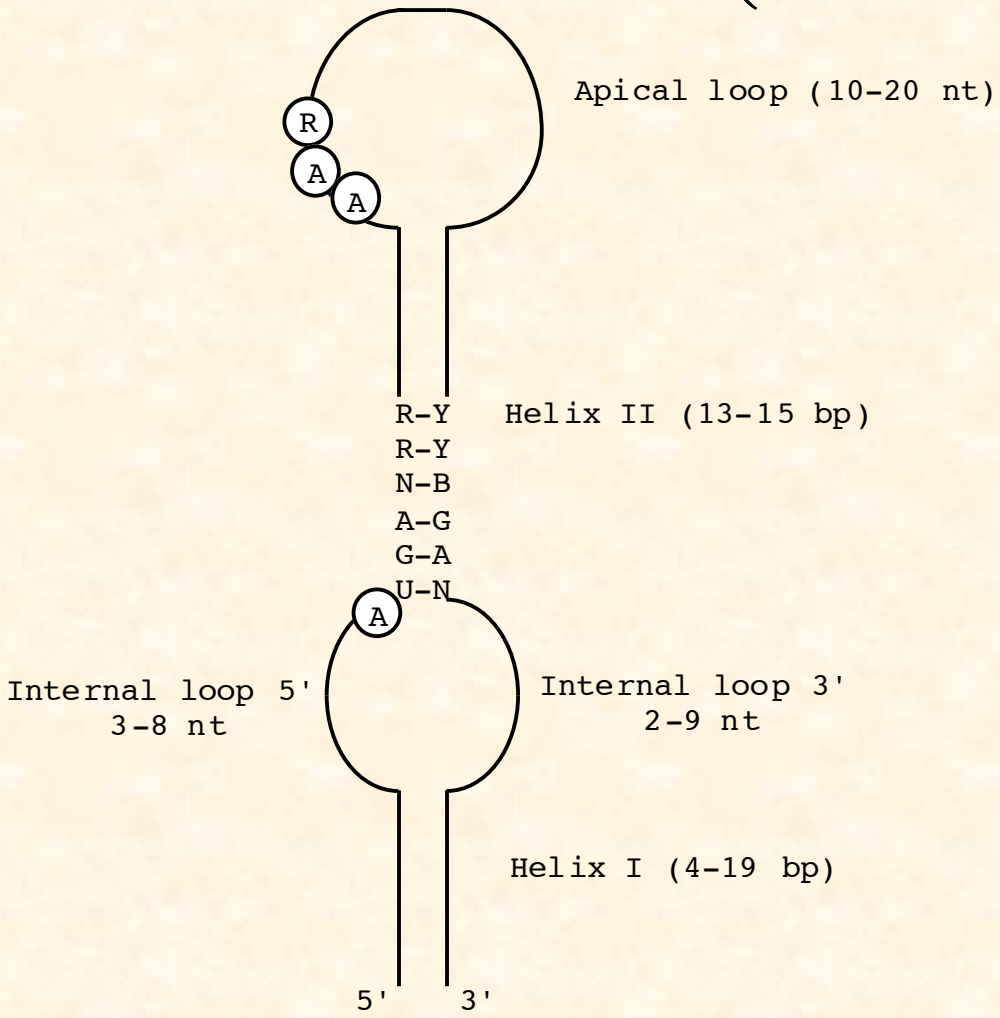


1. Start sites
2. Splice sites
3. Stop sites
4. Signals in exons, introns and UTRs

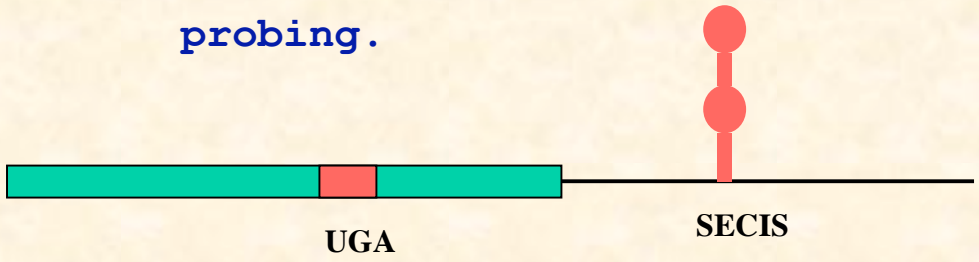
Selenoproteins

1. Selenocysteine (Sec) is the 21st aa
2. UGA (STOP) is the codon for Sec
3. There is a $tRNA^{sec}$ for the UGA codon
4. Recoding:
    1. RNA structure: the SECIS element
    2. SECIS binding protein (SBP2)
    3. Specific elongation factor (eEFsec)

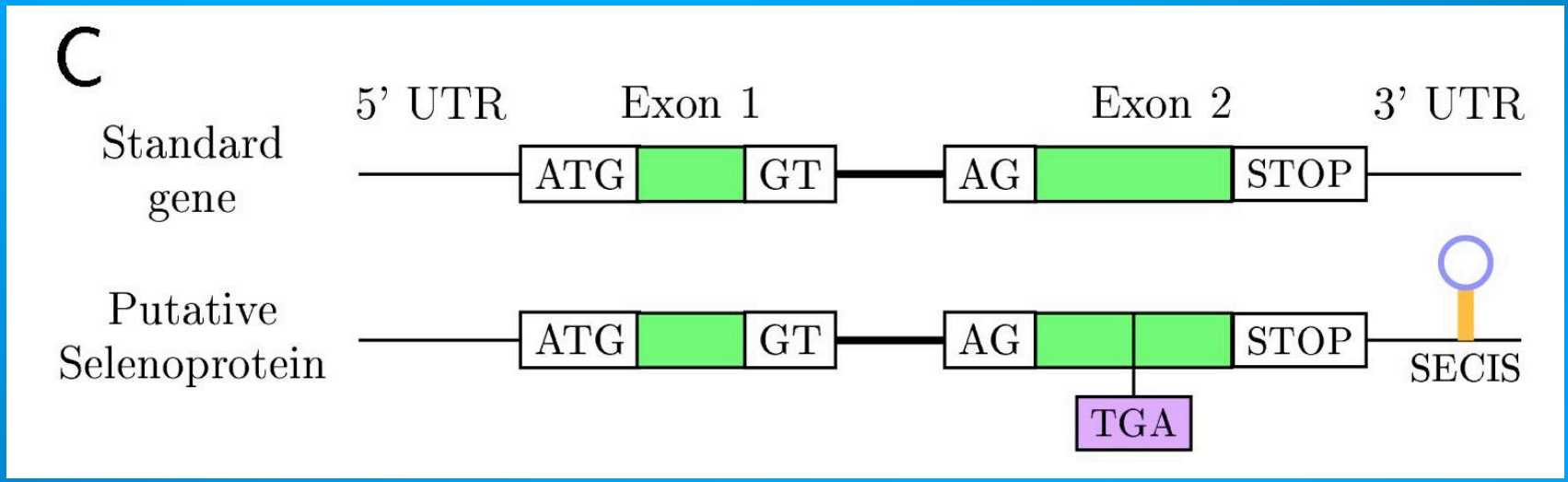
SElenoCysteine Insertion Sequence (SECIS)

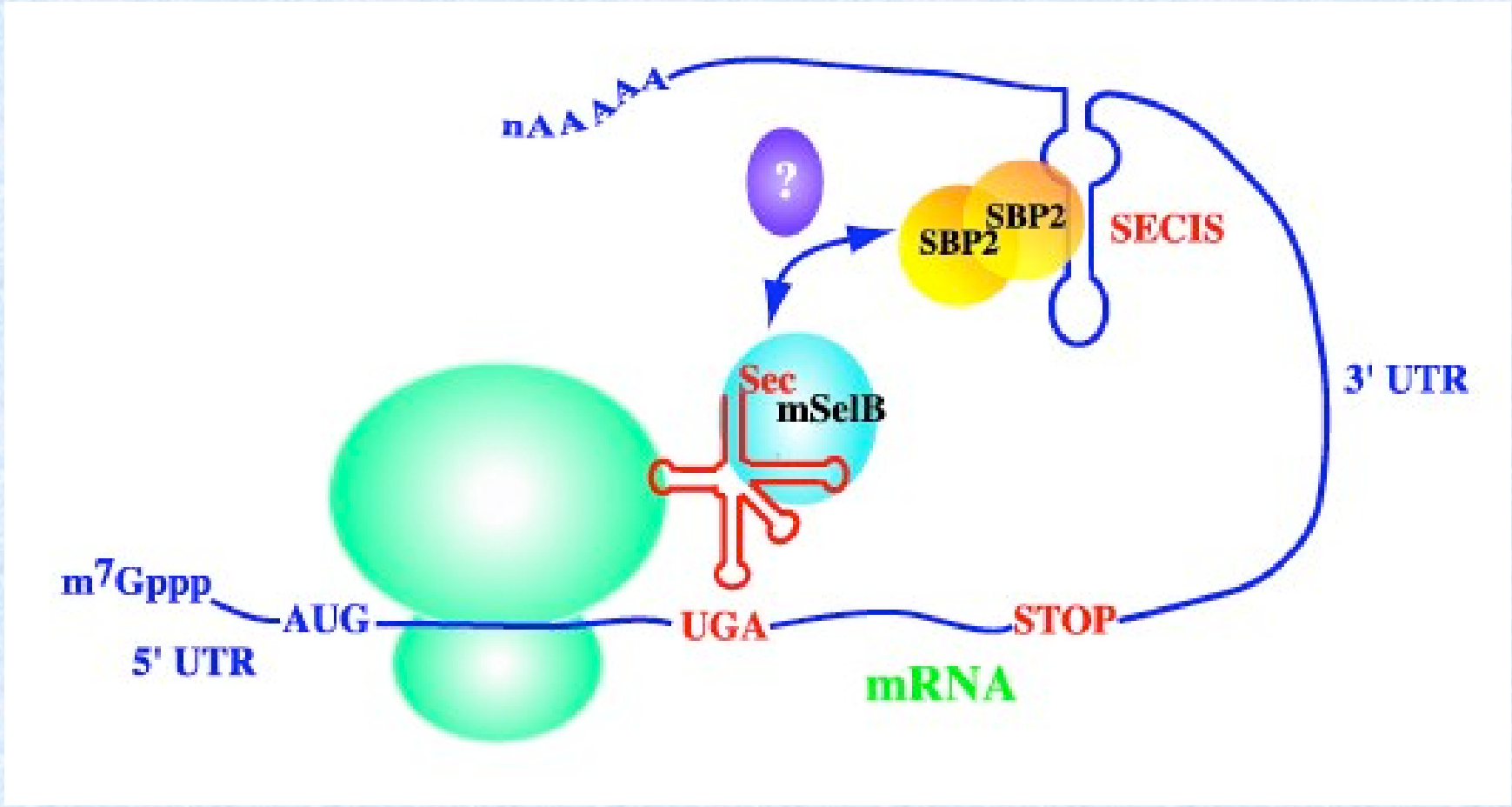


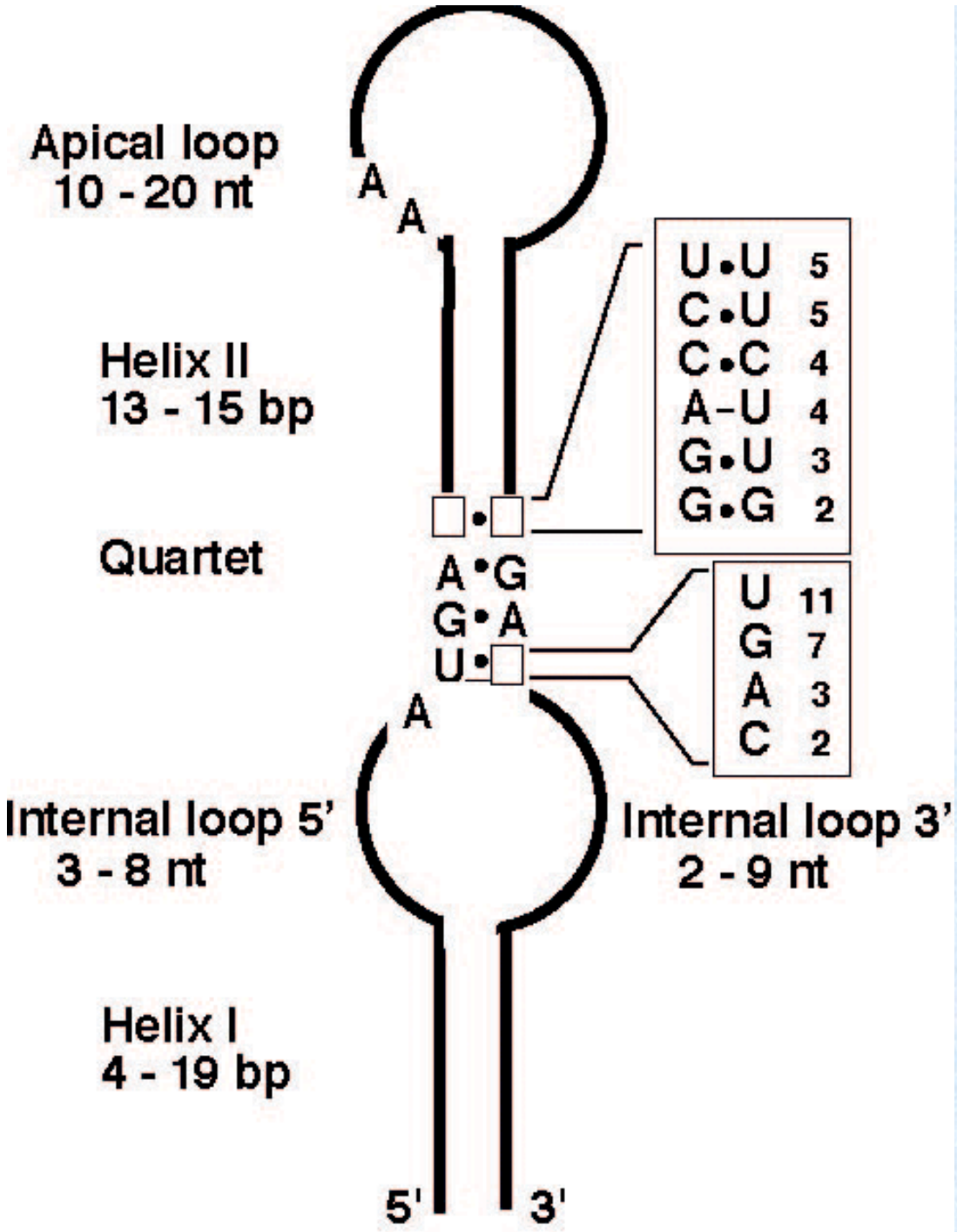
Specific incorporation of selenocysteine in selenoproteins is directed by UGA codons residing within the coding sequence of the corresponding mRNAs. Translation of UGA, usually a termination codon, as selenocysteine requires a conserved stem-loop structure called "SElenoCysteine Insertion Sequence" (SECIS) lying in the 3'UTR region of selenoprotein mRNAs. The consensus structure of SECIS element determined by comparative analysis of several selenoprotein mRNAs as well as on both RNase and chemical probing.



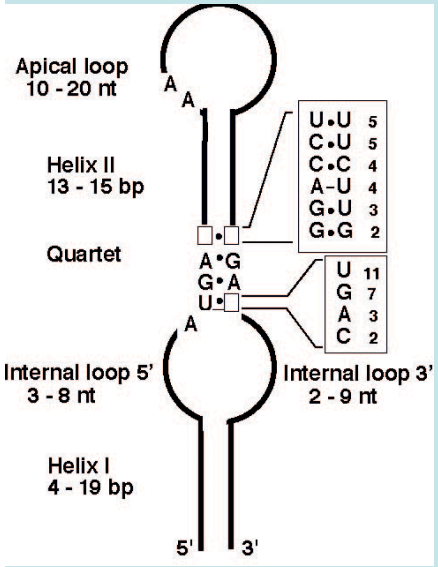
Selenoproteins




nAAAAA
?
SBP2
SBP2
SECIS
3' UTR
Sec
mSelB
m7Gppp
AUG
UGA
STOP
5' UTR
mRNA



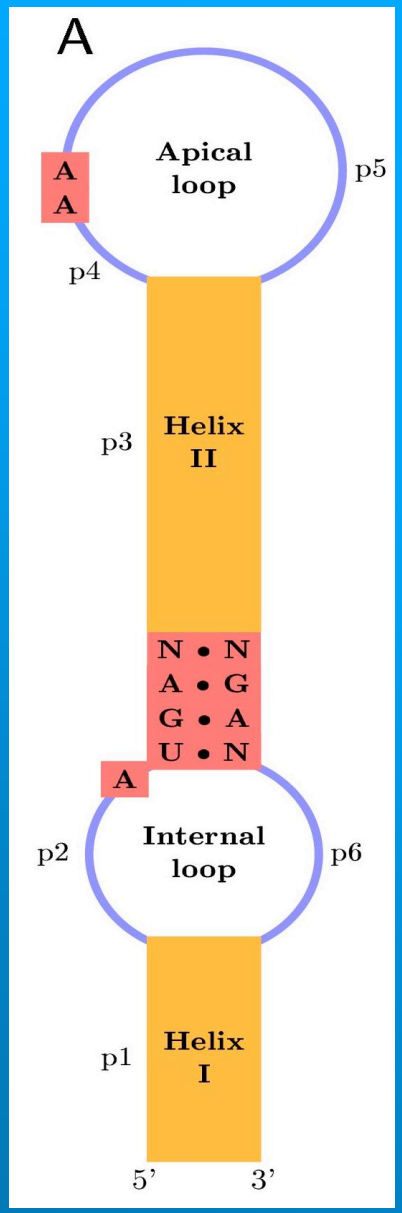
SECIS element Consensus structure



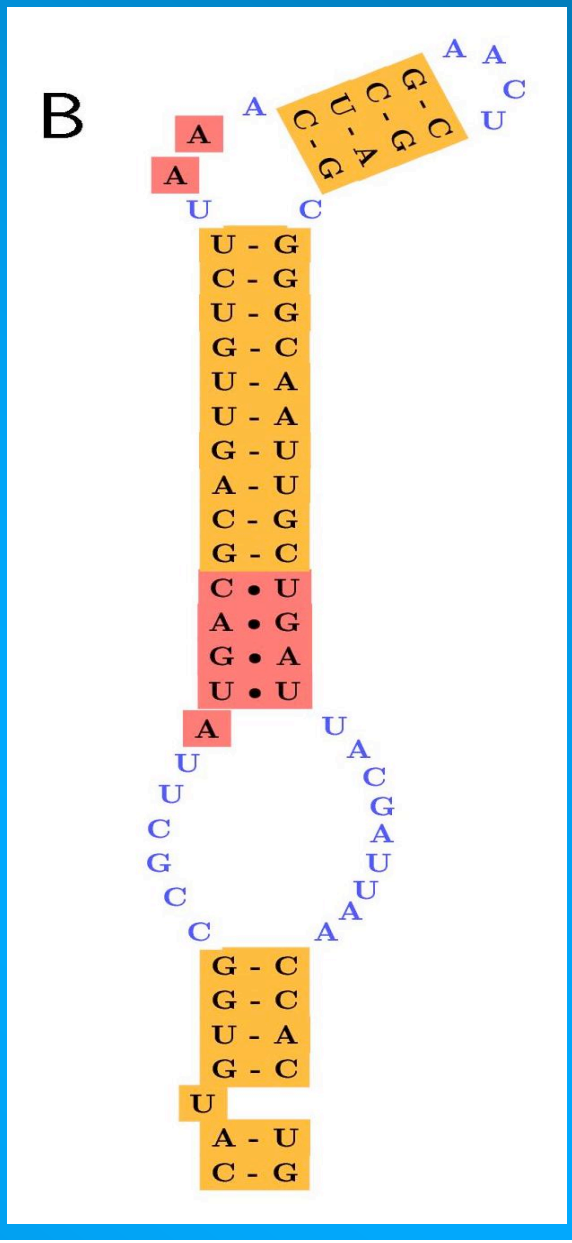
SECIS consensus structure with corresponding alignment

		Helix I	Int.loop 5′		Helix II		Apical loop		Helix II′		Int.loop 3′	Helix I′	
					Quartet					Quartet′			
ovine cell.GPx	699 -	UUUU	---GCCC	A	UGAA	GGUGUUCCCUCU	AA	ACCUACG----------UGGAGGAAUGCC		UGAU	GUCCAGG--	AAAA	- 760
uman cell.GPx	668 -	UUUC	---AUCU	A	UGAG	GGUGUUUCCUCU	AA	ACCUACGA---------GGGAGGAACACC		UGAU	CUUACA---	GAAA	- 728
ouse cell.GPx	911 -	UCUU	-----CC	A	UGAU	GGUGUUUCCUCU	AA	AUUUGCA----------CGGAGAAACACC		UGAU	UUCCAG---	GAAA	- 968
at cell.GPx	1011 -	UUUU	-----CC	A	UGAC	GGUGUUUCCUCU	AA	AUUUACA----------UGGAGAAACACC		UGAU	UUCCAG---	AAAA	- 1068
ovine pl.GPx	1309 -	CAUC	----GCC	A	UGAA	GGAGGGGC-CCG	AA	G-CCCGCGUGGG-----CGG-GCCUCCCU		UGAG	CCUGUCUGA	GGGG	- 1372
uman pl.GPx	1342 -	GUCU	-----UC	A	UGAG	GGAGGGGC-CC-	AA	AGCCCUUGUGGGC-----GGA-CCUCCCC		UGAG	CCUGUCUG-	AGGG	- 1403
ouse pl.GPx	1215 -	UGUC	----UCU	A	UGAA	GGAGGGGC-CCG	AA	G-CCCUUGUGGG-----CGG-GCCUCCCC		UGAG	CCCGUCUGU	GGUG	- 1278
at pl. GPx	1199 -	UGUC	----UCC	A	UGAA	GGAGGGGC-CCG	AA	G-CCCUUGUGGG-----CGG-GCCUCCCC		UGAG	CCCGUCUGU	GGUG	- 1262
uman GPx - GI	831 -	UCAC	----AGA	A	UGAU	GGCACCUU-CCU	AA	ACCCUCA----------UGG-GUGGUGUC		UGAG	AGGC-----	GUGA	- 885
abbit GPx	653 -	UUUC	---AUCC	A	UGAG	GGCGUUCCCCCG	AA	AACAAA----------UGGAGGAACGCC		UGAU	GUCCGG---	GAAA	- 711
ig heart PHGPx	2712 -	GCAC	-----CC	A	UGAC	AGU-CUGC-CUA	AA	AACCAGCC--CUGGU--GGG-GCAG-ACU		CGAG	AACCUGGC-	GUGC	- 2773
at PHGPX	720 -	GCAC	-----UC	A	UGAC	GGU-CUGC-CUG	AA	AA-CAGCCCGCUGGU--GGG-GCAG-UCC		CGAG	GACCUGGC-	GUGC	- 782
.mansoni GPx	539 -	UUCU	-CGCUAU	A	UGAC	GAUGGCAAUCUC	AA	AUGUUCA---------UUGGUUGCC-AUU		UGAU	GAAAUC---	AGUU	- 600
elenoprotein W	367 -	CCGC	----UUC	A	UGAC	AGGAAGGA-CUG	AA	AUGUCUCAAAGACCUG-UGGUCUUUC-UU		CGAU	GUUCU----	GCGG	- 431
um.sel.P 1st	1445 -	UGCU	----UUA	A	UGAG	AAUAG-AAACGU	AA	ACUAUGACCUAG-----GGGUUUUCUGUU		GGAU	AAUU-----	AGCA	- 1505
at sel.P 1st	1467 -	UUAC	---AUUG	A	UGAG	AACAG-AAACAU	AA	ACUAUGACCUAG-----GGGUUU-CUGUU		GGAU	AGCUC----	GUAA	- 1528
um.sel.P 2nd	1881 -	AUAG	----UCA	A	UGAU	GGUUUAAUAGGU	AA	ACCAAA----------CCCUA-UAAACC		UGAC	CUCCU----	UUAU	- 1937
at sel.P 2nd	1846 -	AUAA	----UCA	A	UGAC	GGUUUAAUAGAG	AA	ACUGAG----------UCCUA-UGAACC		UGAA	CUCCU----	UUAU	- 1901
at DI type 1	1528 -	AUUU	---GUUU	A	UGAU	GGUC-ACAGUGU	AA	AGUUCACAC--------AGCUGUGACU--		UGAU	UUUUAA---	AAAU	- 1586
um DI type1	1732 -	AUUU	---GUUU	A	UGAU	GGCC-ACAGCCU	AA	AGUACACAC--------GGCUGUGACU--		UGAU	CAAAAGA--	AAAU	- 1792
at DI type3	1684 -	CUGC	-----UG	A	UGAC	GAACC-GCCUCU	AA	CUGGGCUUGACCAC---GGGUCGGCUC--		UGAA	UUGCA----	GCAG	- 1741
at DI type3	1602 -	CCCC	ACUGCUG	A	UGAC	GAACUAU-CUCU	AA	CUGGUCUUGACCAC---GAGCUAGUUC--		UGAA	UUGCA----	GGGG	- 1667
enop. DI type3	1300 -	UGUU	--UGCAA	A	UGAC	GACCGAUU-UUG	AA	AUGGUCUCACGGCCAA-AAACUCGUGUC-		CGAC	AUC------	AACC	- 1364

RNA SECIS



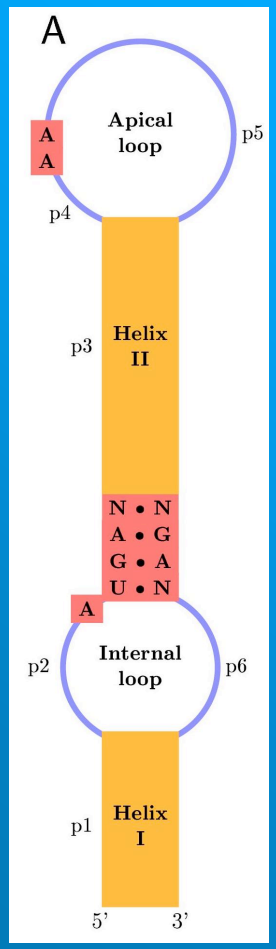
SECIS consensus



dSelK

Search for RNA structures

SECIS



PatScan pattern

B

Pairing rules — Helix I and II: **r1={at,ta,gc,cg,tg,gt}**

	Helix I	Internal loop	Quartet	Helix II
5'	p1=5...15	p2=1...7	A TGA N	p3=9...12

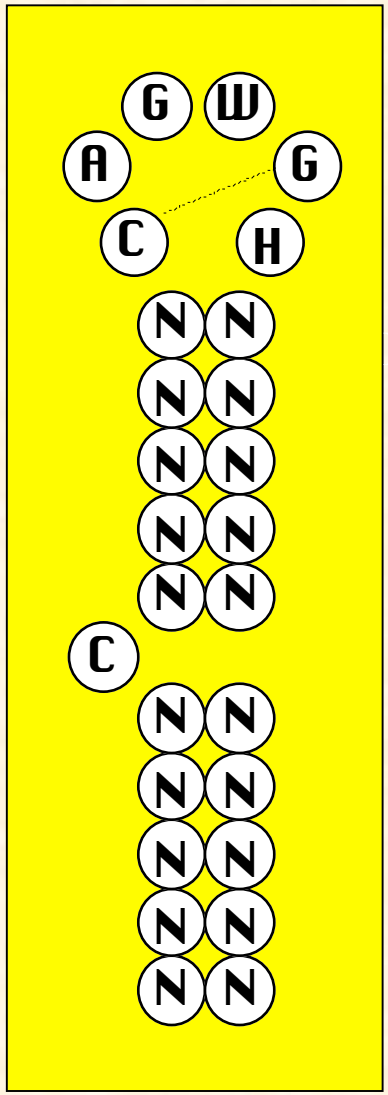
Apical loop: p4=0...2 | AA | p5=6...17

	Helix II	Quartet	Internal loop	Helix I
3'	r1~p3[2,1,1]	N GAN	p6=3...9	r1~p1[2,1,1]

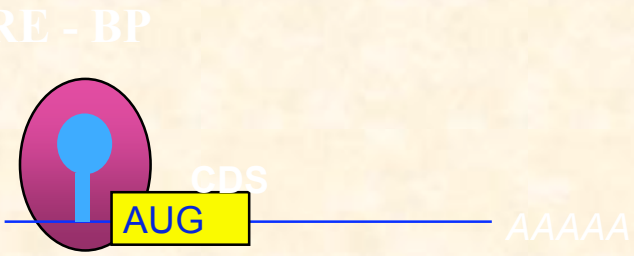
- 1- Matches 5' chain p3
- 2- Follows r1 pairing rules
- 3- 2 mismatches allowed
- 4- 2 bulges allowed

- 1- Matches 5' chain p1
- 2- Follows r1 pairing rules
- 3- 2 mismatches allowed
- 4- 2 bulges allowed

Iron Responsive Element (IRE)



Post-transcriptional regulation of cellular iron homeostasis regulated by UTR regions

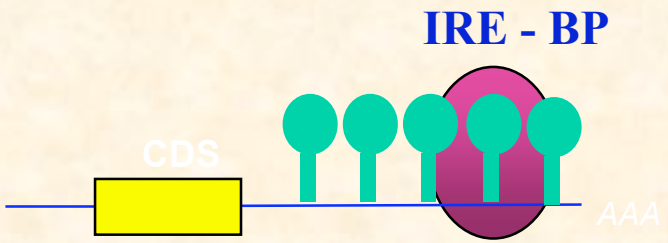


ferritin mRNA

(iron sequestration)

IRE occupied by IRE-BP inhibiting translation initiation

Fe low ➡ IRE-BP on



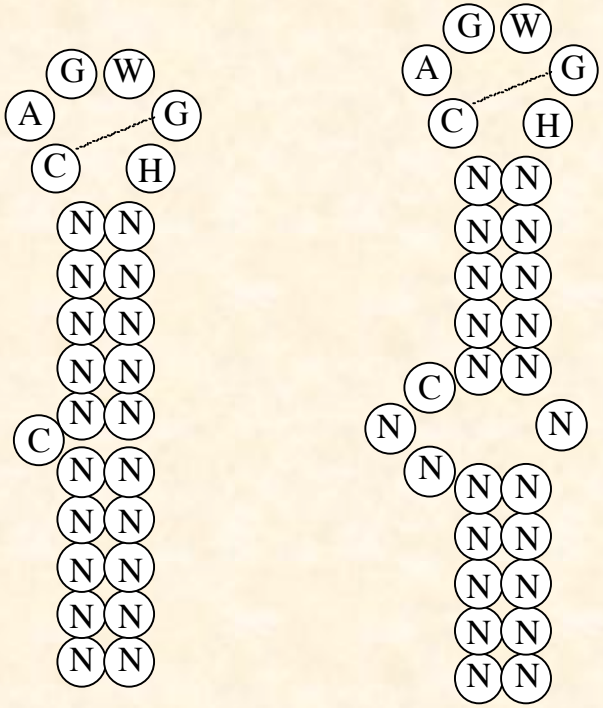
transferrin receptor mRNA

(mediates iron uptake in eukaryotic cells)

one or more IREs occupied by IRE-BP protecting mRNA from mRNA degradation

Fe high ➡ IRE-BP off

Iron Responsive Element



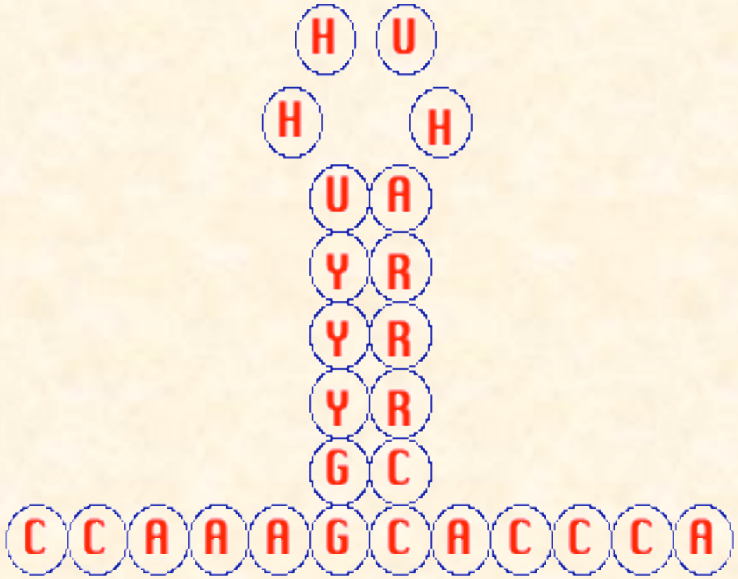
(b)

```
r1={au,ua,gc,cg,gu,ug}
(p1=2...8 c p2=5...5 cagwgh r1~p2 r1~p1 |
p3=2...8 nnc p4=5...5 cagwgh r1~p4 n r1~p3)
```

(c)

```
5HSA001988 :[13,35] :GTT C GTCCT CAGTGC AGGGC AAC
5HSA013930 :[34,56] :CTG C TTCAG CAGTGC TTGGA CGG
5HSA003829 :[8,30]  :TTG C TTCAA CAGTGT TTGGA CGG
5HSA003858 :[35,57] :CTG C TTCAA CAGTGC TTGGA CGG
...
```

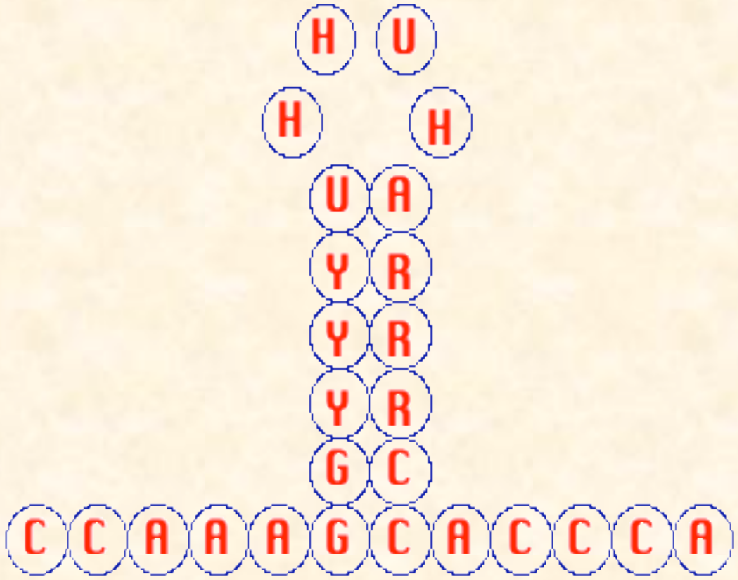
Histone 3’UTR mRNA element



Metazoan histone 3'-UTR mRNAs, lacking a polyA tail, contain a highly conserved stem-loop structure with a six base stem and a four base loop. This stem-loop structure plays a different role in the nucleus and in the cytoplasm. In the nucleus, it is involved in pre-mRNA processing and nucleocytoplasmic transport, whereas in the cytoplasm it enhances translation efficiency and regulates histone mRNA stability. The trans-acting factor which interacts with the 3'-UTR hairpin structure of histone mRNAs is a 31 kDa stem-loop binding protein in mammals (SLBP) present both in nuclei and polyribosomes. In mammals in addition to SLBP histone mRNA processing requires at least one additional factor: the U7 snRNP, which binds a purine-rich element 10-20 nt downstream of the stem-loop sequence (Histone Downstream Element, HDE).

The histone 3'-UTR hairpin structure is peculiar in that the bases of the stem are conserved unlike most functional hairpin motifs where conserved bases are found in single stranded loop regions only. The sequence of the stem an flanking sequences are critical for binding of the SLBP.

Histone 3'UTR mRNA element



PatSearch pattern:

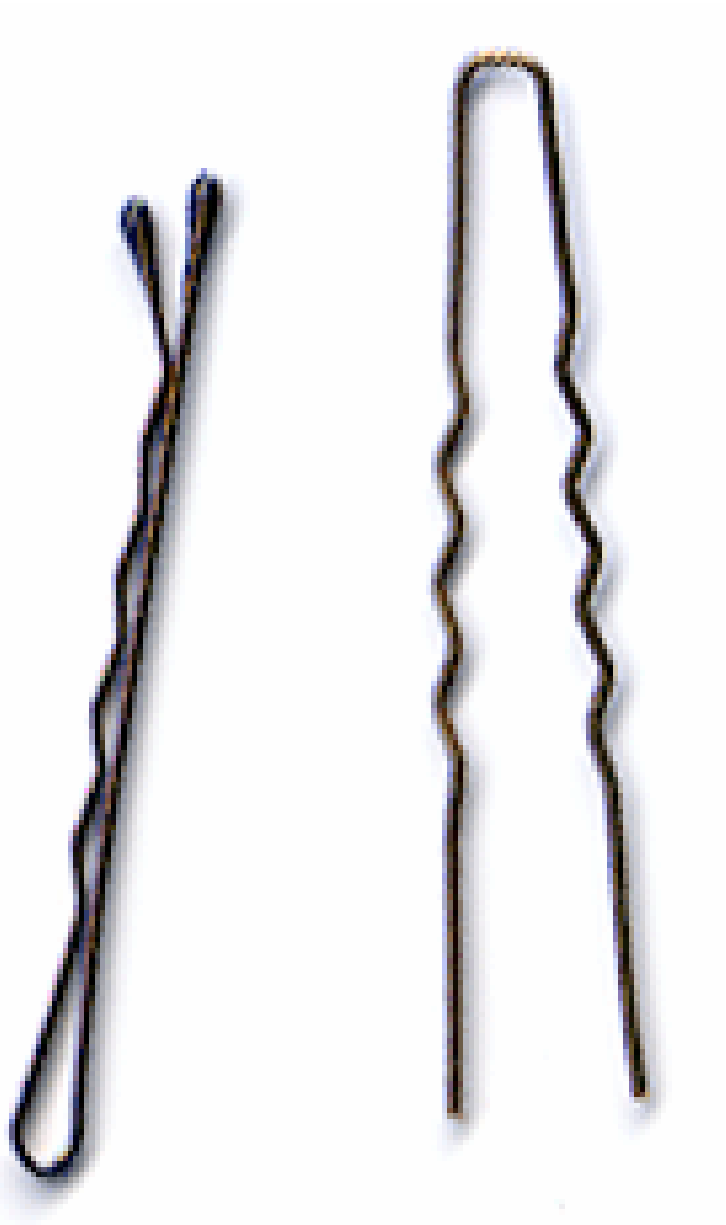
r1={au,ua,gc,cg,gu,ug}

n mmm p1=ggyyy u hhuh a r1~p1

mm 0…3

(m=a/c; y=c/u; h=not g)

Hairpins



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.

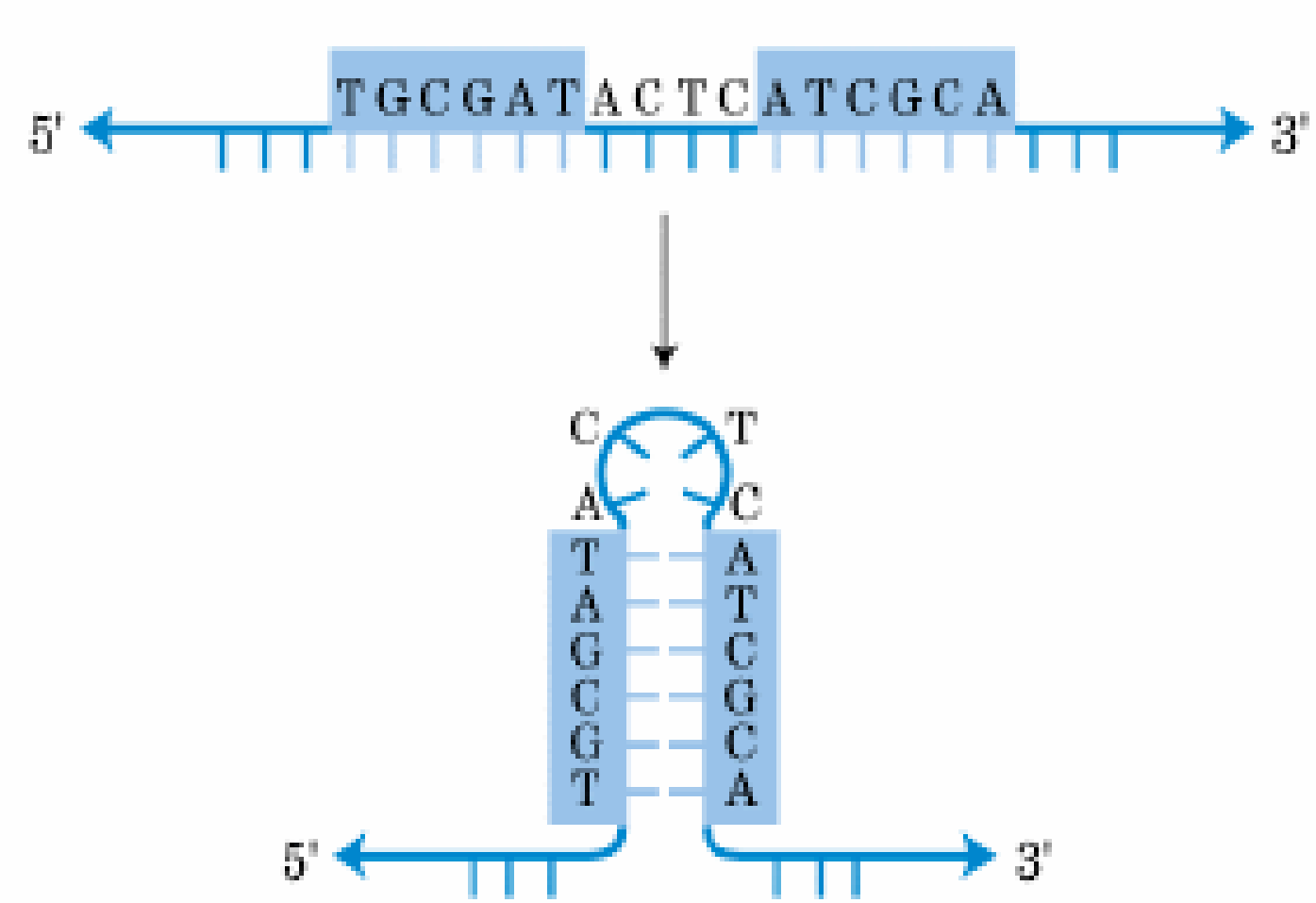
RNA secondary structures

- G-C and A-U form hydrogen bonded base pairs and are said to be complementary.

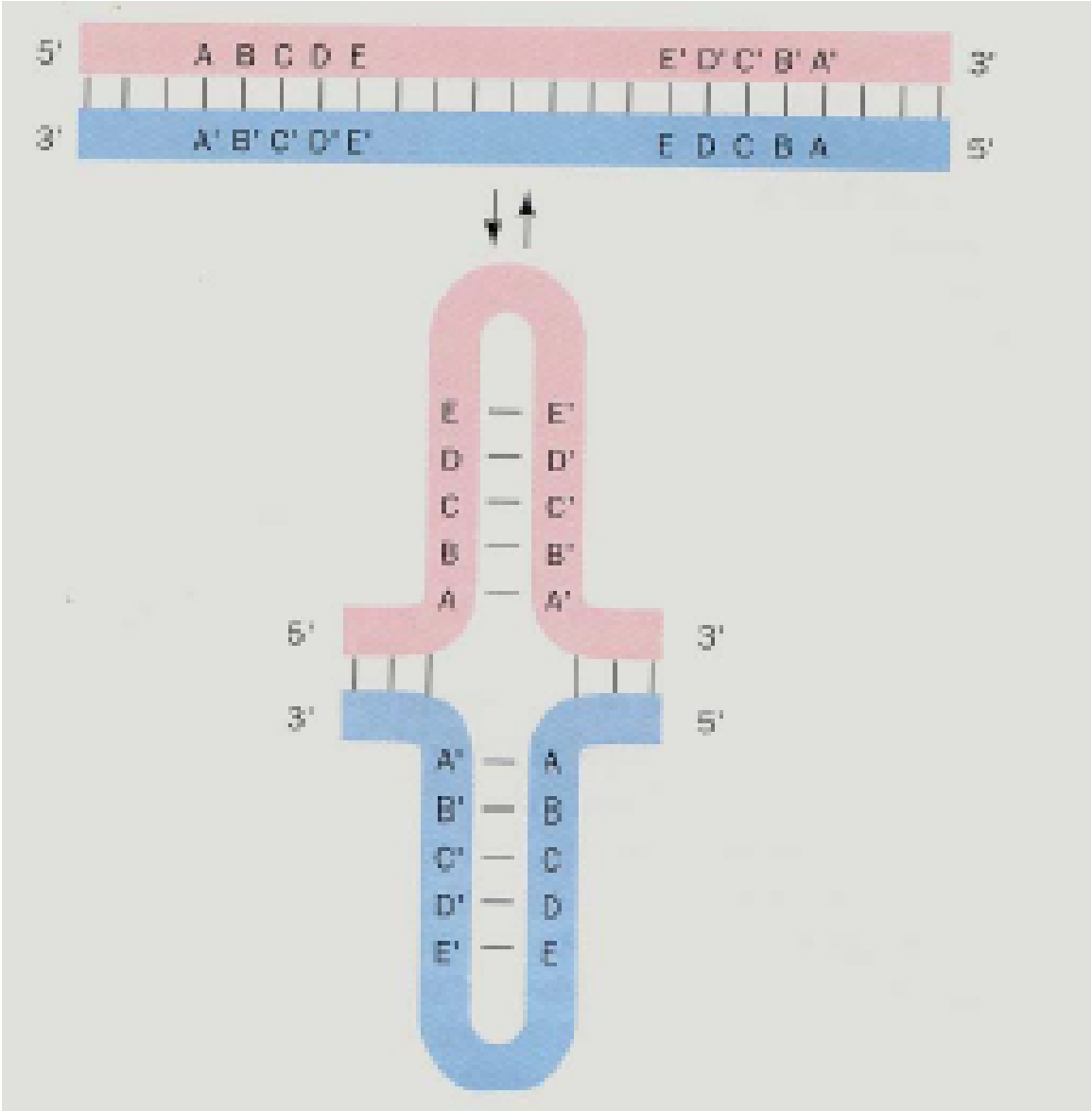
- Base pairs are approximately coplanar and are almost always *stacked* onto other base pairs in an RNA structure. Contiguous base pairs are called *stems*.

- Unlike DNA, RNA is typically produced as a single stranded molecule which then folds intra-molecularly to form a number of short base-paired stems. This base-paired structure is called RNA secondary structure.

Hairpins



Hairpins



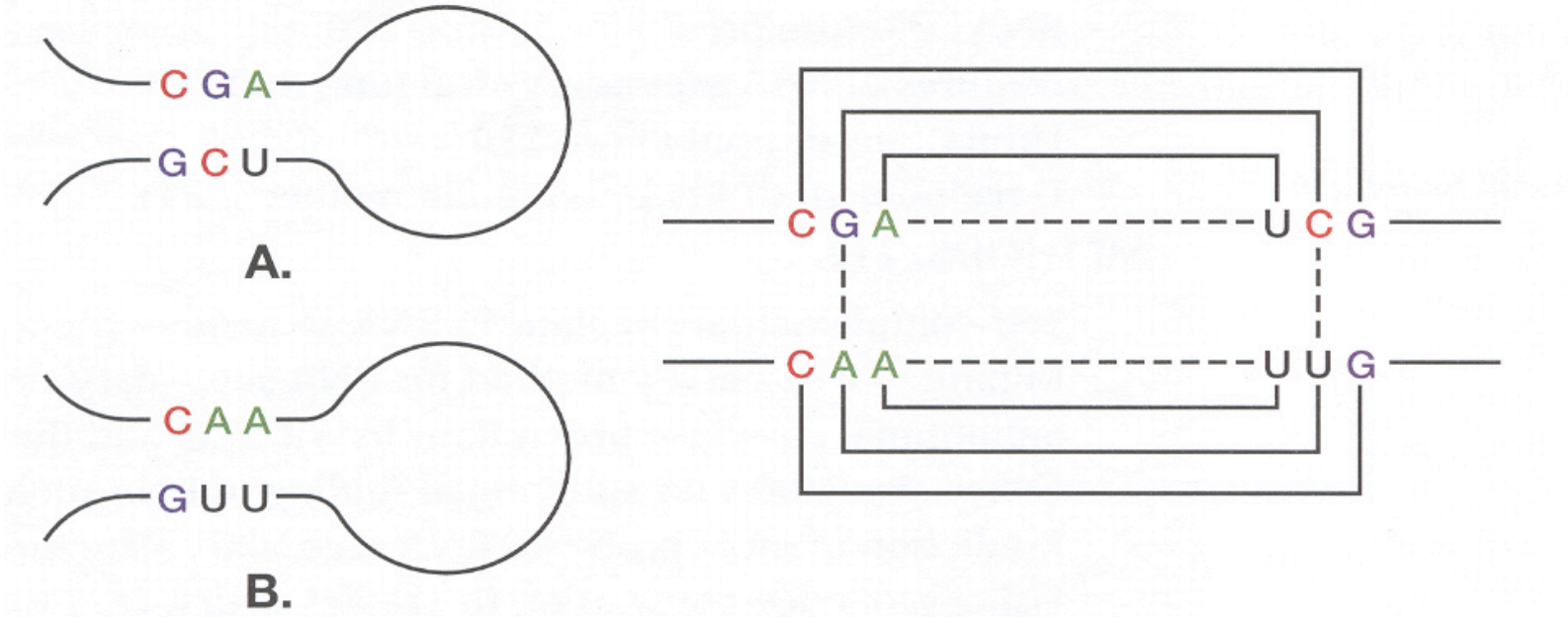
RNA secondary structures

- Single stranded subsequences bounded by base pairs are called *loops*. A loop at the end of a stem is called a *hairpin loop*. Simple substructures consisting of a simple stem and loop are called *stem loops* or *hairpins*.
- Single stranded bases within a stem are called a *bulge* or *bulge loop* if the single stranded bases are on only one side of the stem.
- If single stranded bases interrupt both sides of a stem, they are called an *internal* (*interior*) *loop*.
- There are *multi-branched loops* from which three or more stems radiate.

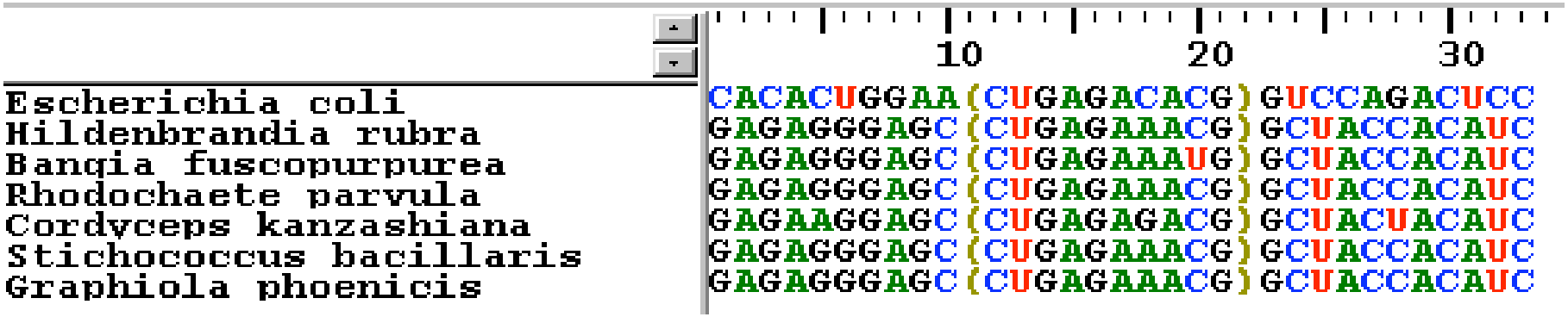
RNA secondary structures

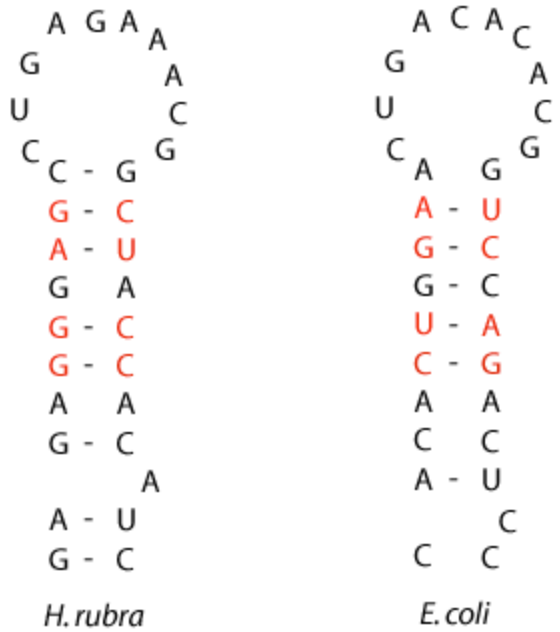
- Sequences variations in RNA sequences maintain base pairing patterns that give rise to double-stranded regions (secondary structures) in molecules.

- Alignments of RNA sequences will show covariation at interacting base-pair positions, see figure below.

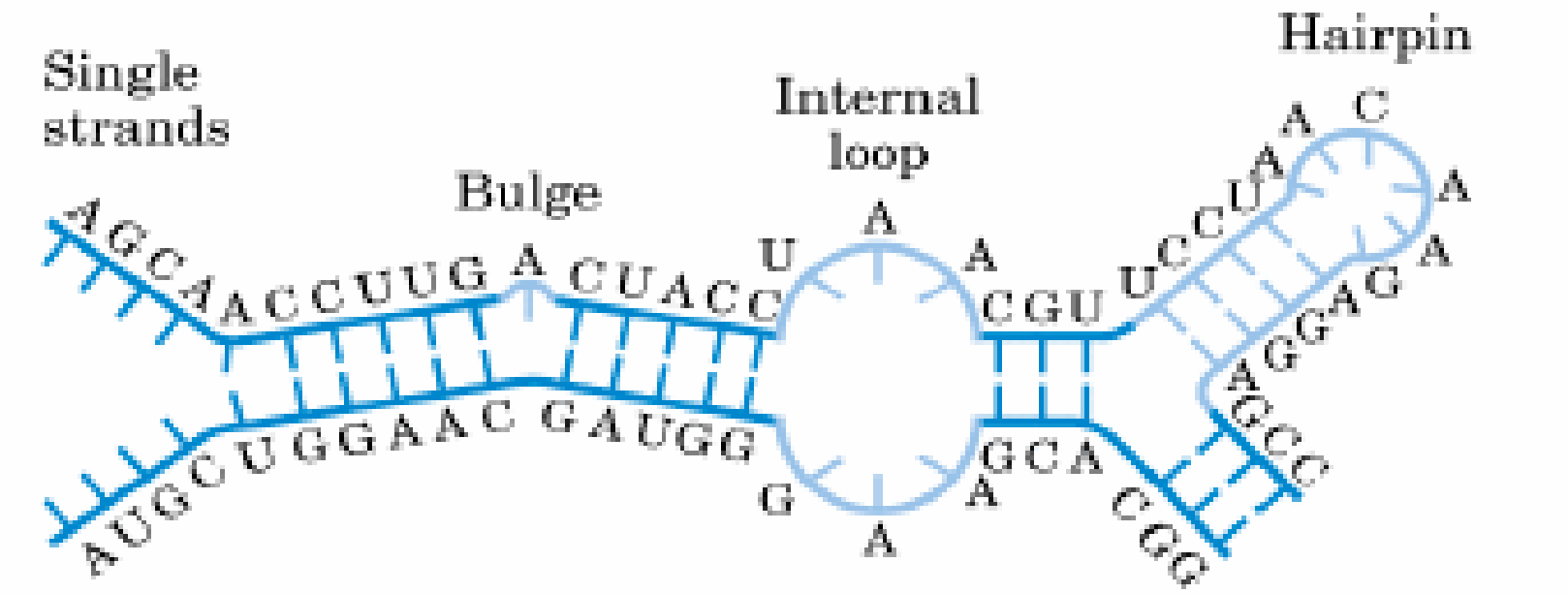


Covariation



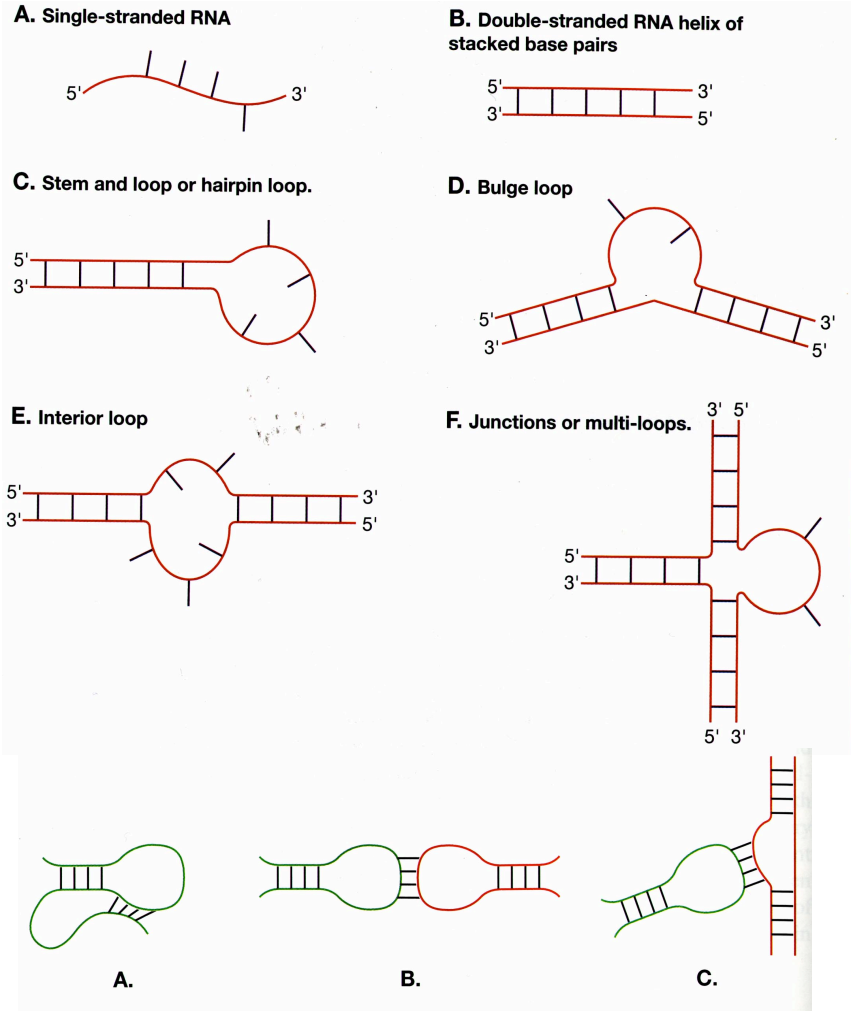


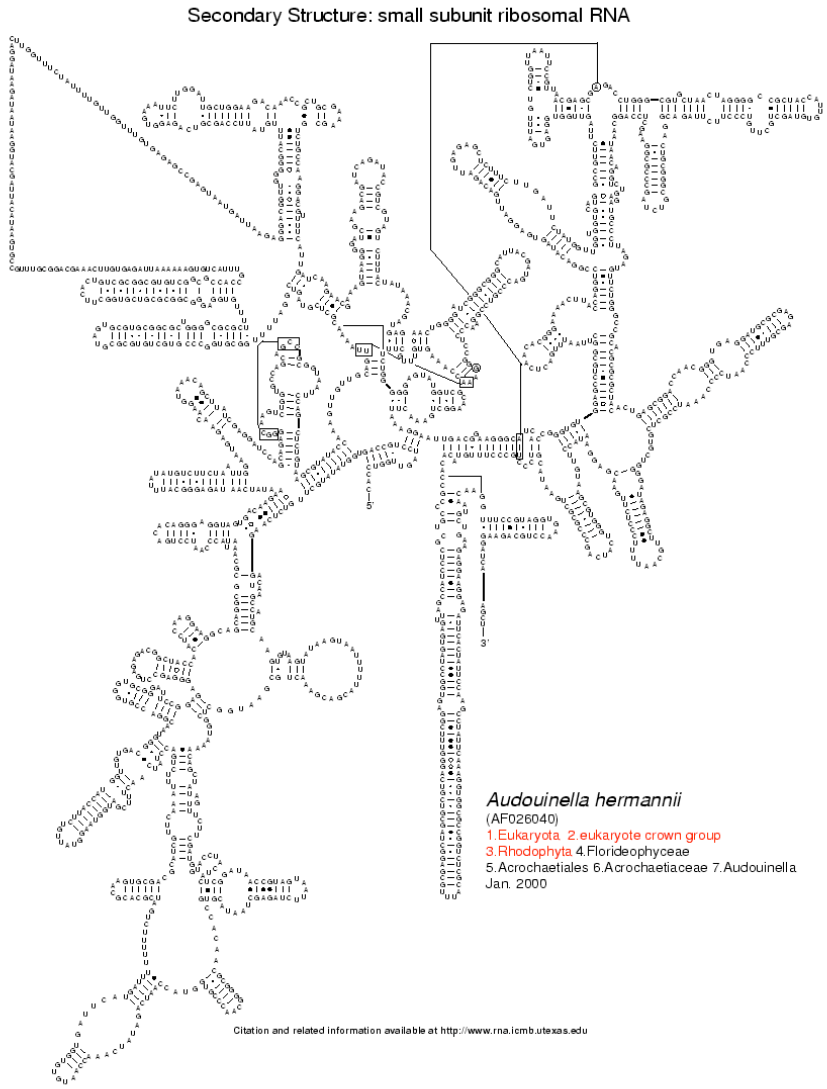
Hairpins



- The bending in Hairpin loops facilitates the binding of some proteins to the DNA
- Short base sequences (example UUCG) which are found at the end of RNA hairpins facilitates the folding of RNA into its precise three dimensional structure

Features of RNA secondary structure

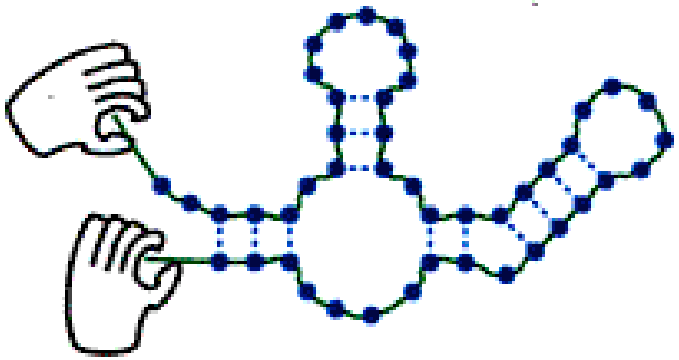




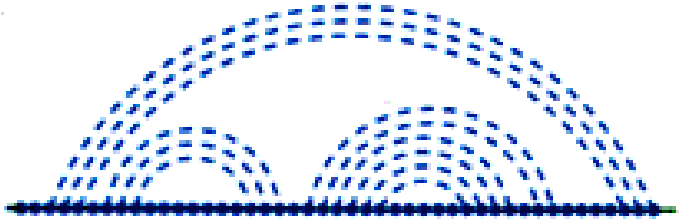
- Grammatically correct string of parentheses

```
..(((.(((.......))).(((((.....)))).)).....)))
AGCTACGGAGCGATCTCCGAGCTTTCGAGAAAGCCTCTATTAGC
```

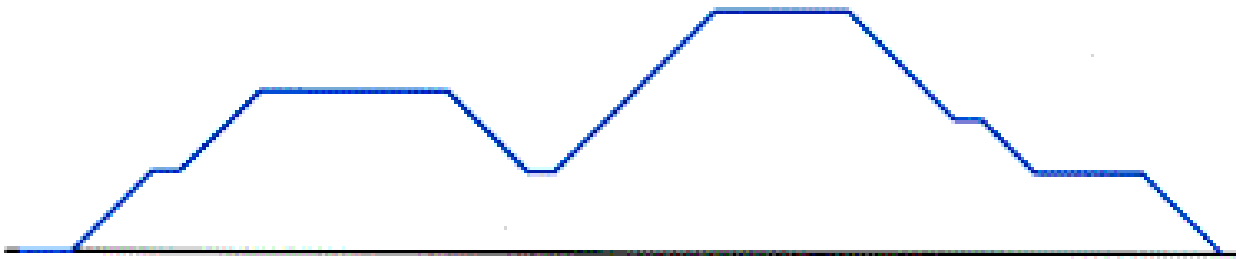
- Planar graph

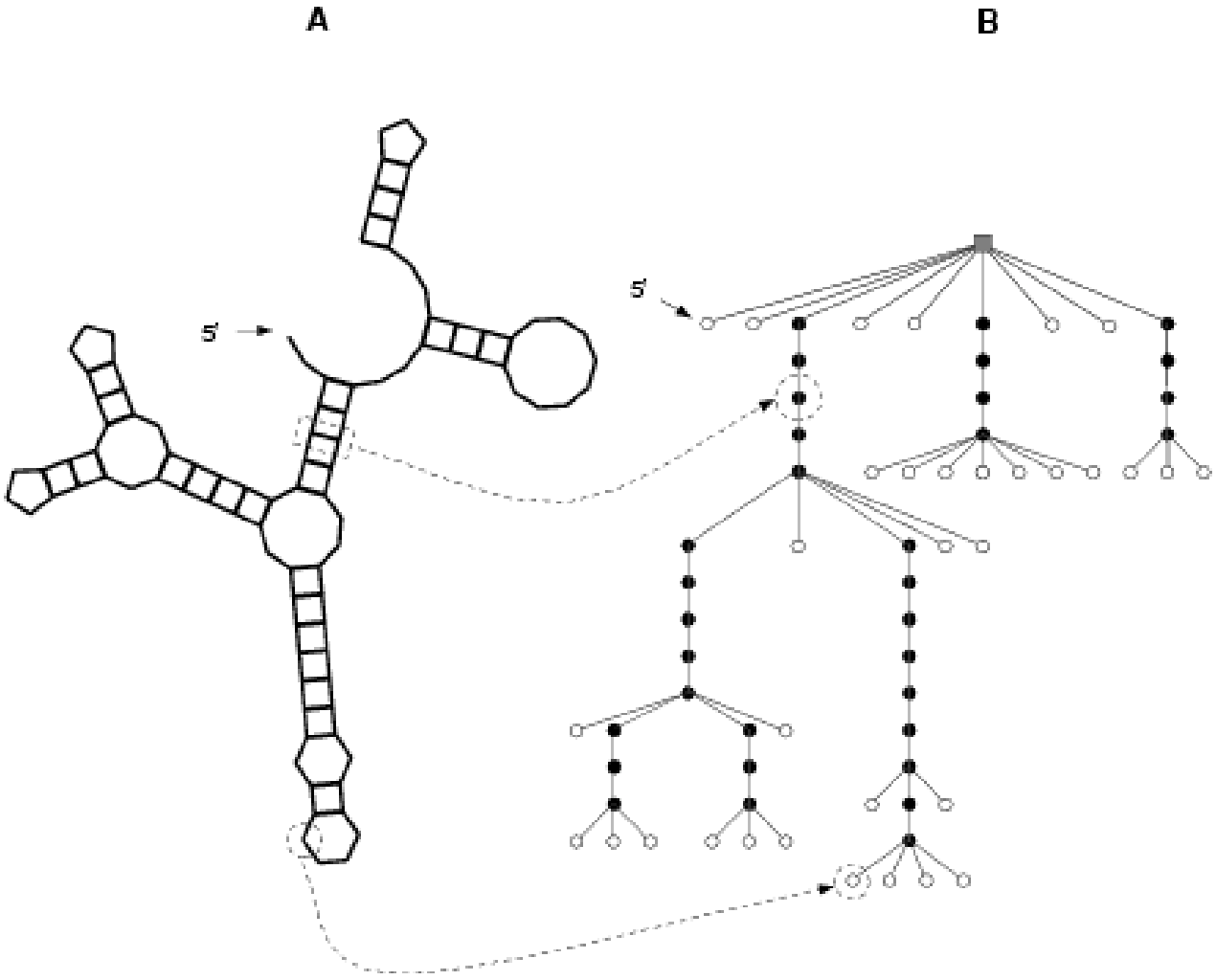


- Arch diagram



- Mountain diagram

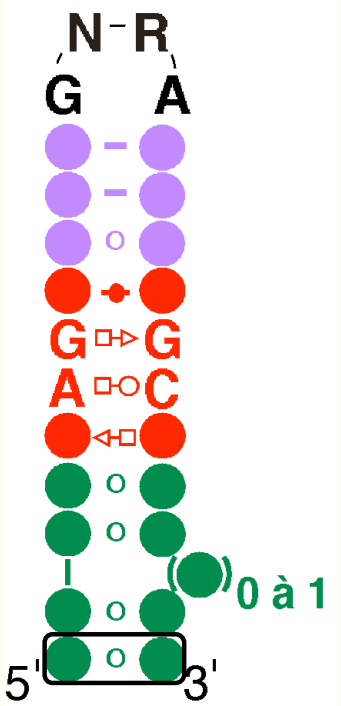




Secondary structure as it's tree representation

```
((((((((((....))))))))))).)
GCUGGAGGGAAGCAAUUCAGCACAG-C
GCUUCAGUGGAGCGAUUCAGCGGAGAU
AGCUCAGUGGAGCAAUCCAGCAGGCGU
CAUGGAGGUUGGAAACAAUGCACAU-G
```

(	brin 5’ d’une hélice
.	simple brin
)	brin 3’ d’une hélice

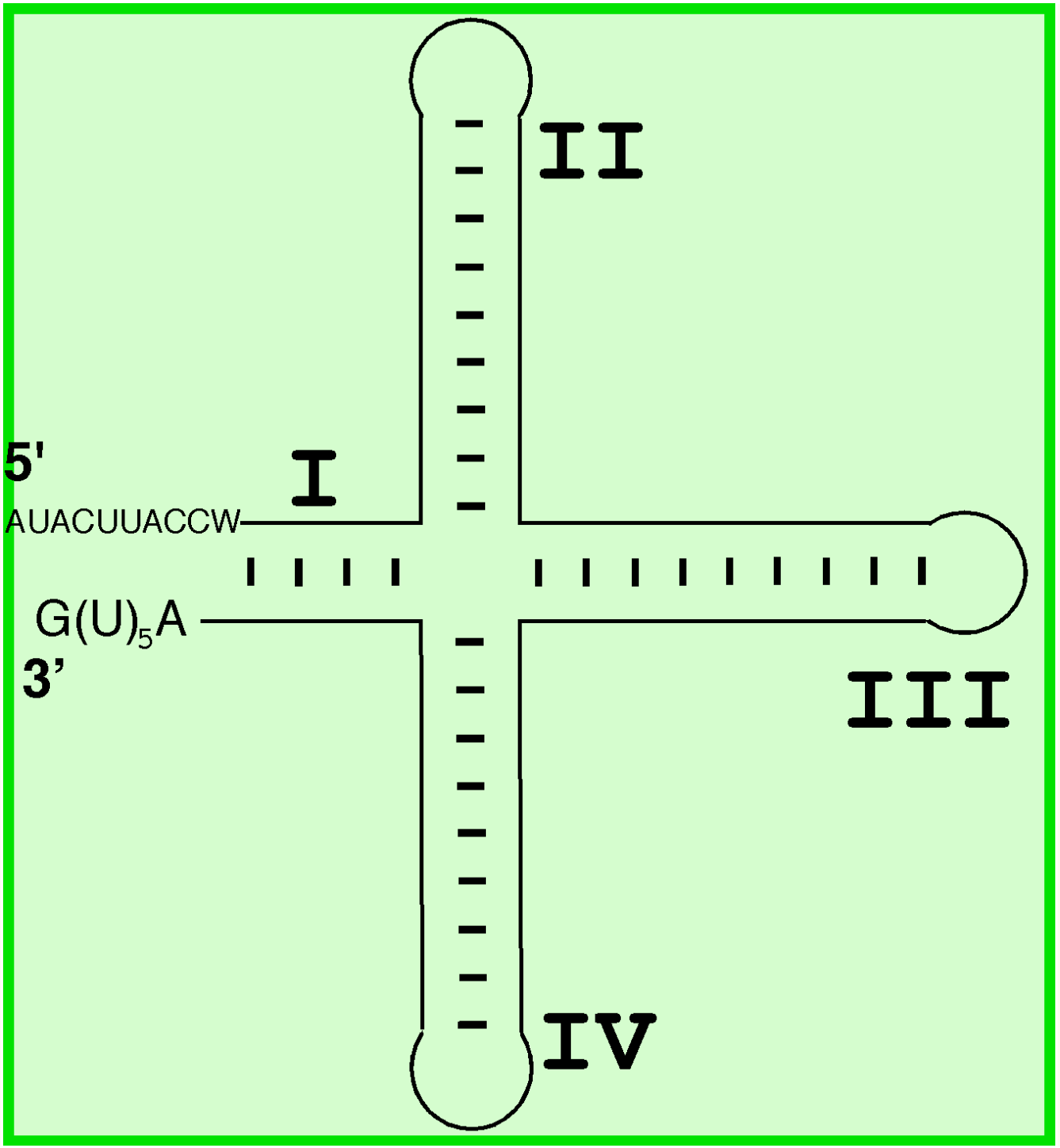


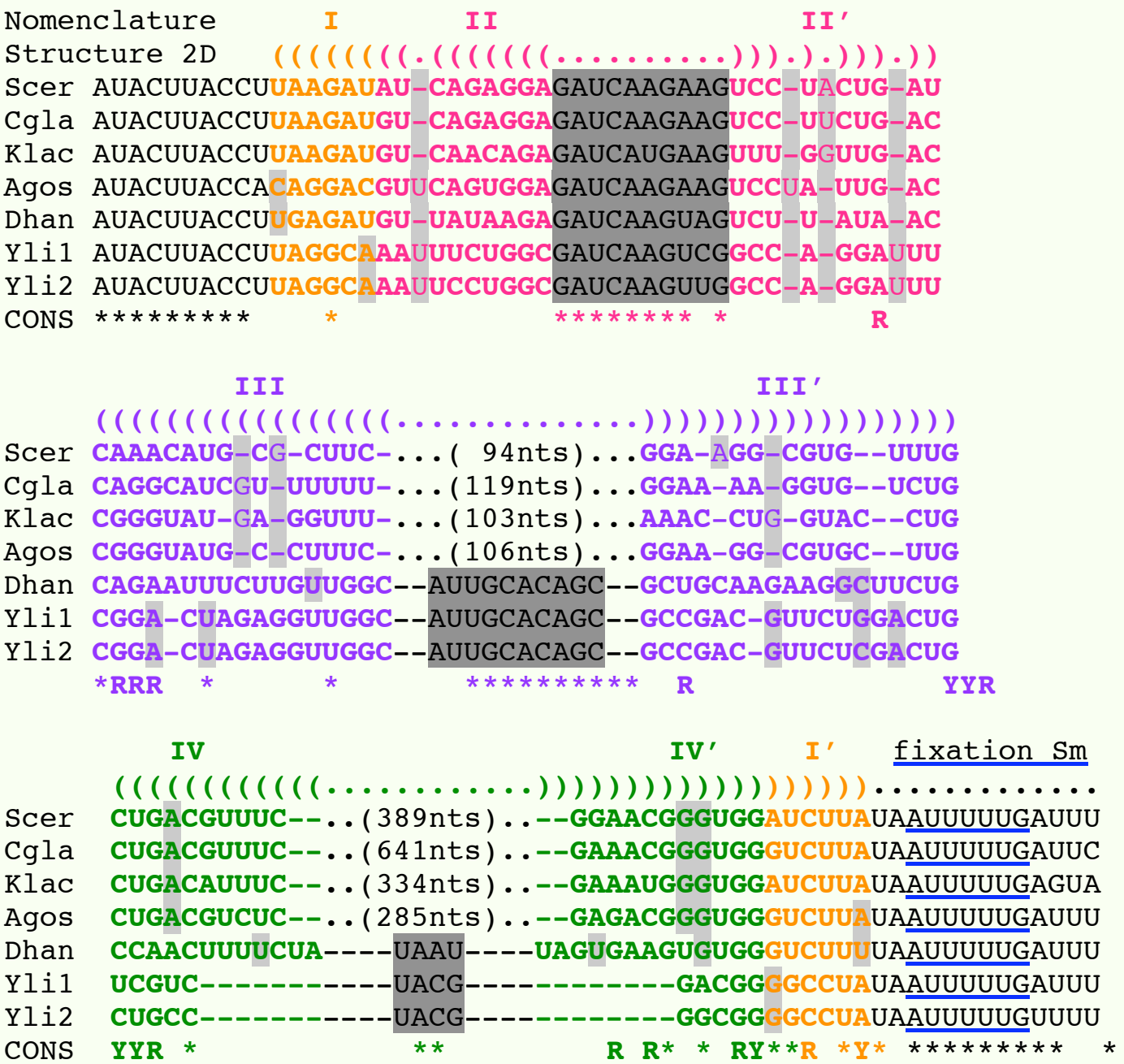
Boucle terminale GNRA

Hélice à 3 pb

Motif structural à 4 pb
non canoniques

Hélice à 4 pb,
avec une bulle interne possible

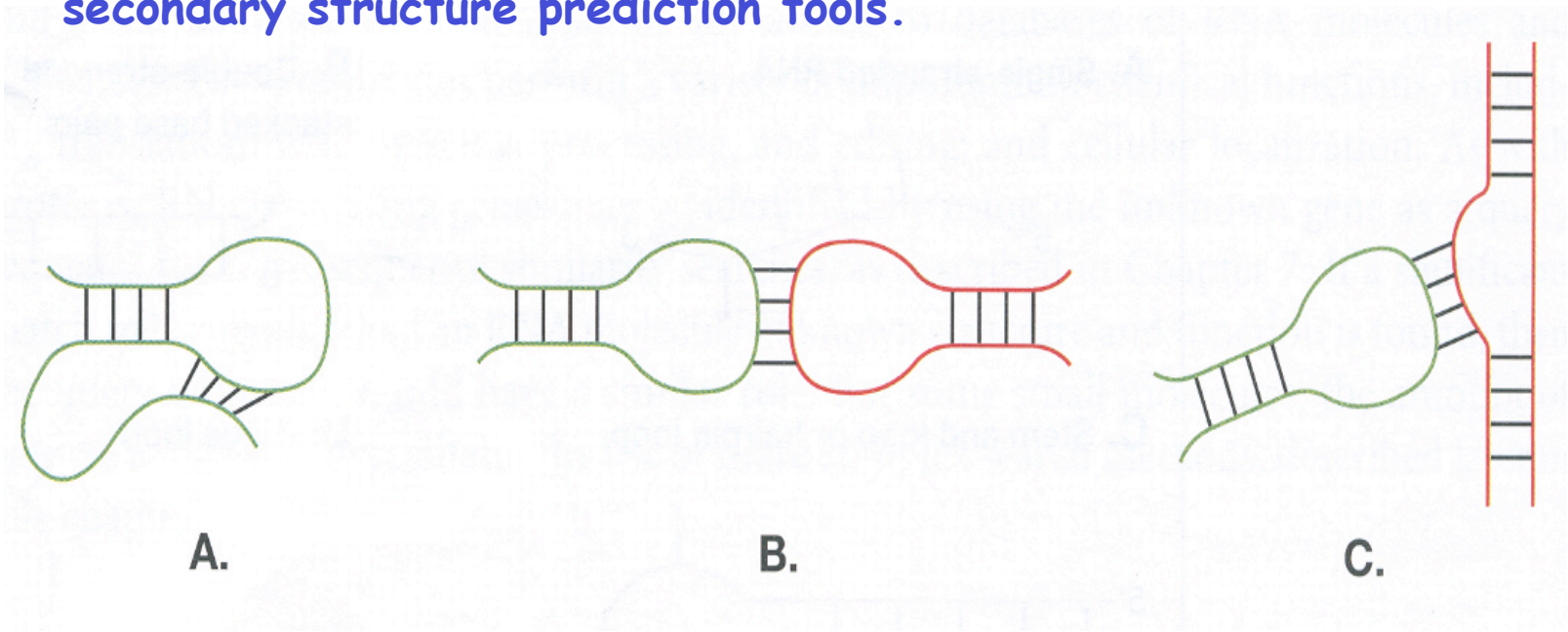

II
5'
I
AUACUUACCW
G(U)5A
3'
III
IV


Nomenclature I II II′
Structure 2D (((((((.(((((((.....))))).).))))))
Scer AUACUUACCUUAAGAUAU-CAGAGGAGAUCAAGAAGUCC-UACUG-AU
Cgla AUACUUACCUUAAGAUGU-CAGAGGAGAUCAAGAAGUCC-UUCUG-AC
Klac AUACUUACCUUAAGAUGU-CAACAGAGAUCAUGAAGUUU-GGUUG-AC
Agos AUACUUACCACAGGACGUUCAGUGGAGAUCAAGAAGUCCUA-UUG-AC
Dhan AUACUUACCUUGAGAUGU-UAUAAGAGAUCAAGUAGUCU-U-AUA-AC
Yli1 AUACUUACCUUAGGCAAAUUUCUGGCGAUCAAGUCGGCC-A-GGAUUU
Yli2 AUACUUACCUUAGGCAAAUUCCUGGCGAUCAAGUUGGCC-A-GGAUUU
CONS ********* * ******** * R
III III′
(((((((((((((((.....)))))))))))))))
Scer CAAACAUG-CG-CUUC-...(94nts)...GGA-AGG-CGUG--UUUG
Cgla CAGGCAUCGU-UUUUU-...(119nts)...GGAA-AA-GGUG--UCUG
Klac CGGGUAU-GA-GGUUU-...(103nts)...AAAC-CUG-GUAC--CUG
Agos CGGGUAUG-C-CUUUC-...(106nts)...GGAA-GG-CGUGC--UUG
Dhan CAGAAUUUCUUGUUGGC--AUUGCACAGC--GCUGCAAGAAGGCUUCUG
Yli1 CGGA-CUAGAGGUUGGC--AUUGCACAGC--GCCGAC-GUUCUGGACUG
Yli2 CGGA-CUAGAGGUUGGC--AUUGCACAGC--GCCGAC-GUUCUCGACUG
*RRR * * ********** R YYR
IV IV′ I′ fixation Sm
((((((((((.....))))))))))..((((((.....
Scer CUGACGUUUC--..(389nts)..--GGAACGGGUGGAUCUUAUAAUUUUUGAUUU
Cgla CUGACGUUUC--..(641nts)..--GAAACGGGUGGGUCUUAUAAUUUUUGAUUC
Klac CUGACAUUUC--..(334nts)..--GAAAUGGGUGGAUCUUAUAAUUUUUGAGUA
Agos CUGACGUCUC--..(285nts)..--GAGACGGGUGGGUCUUAUAAUUUUUGAUUU
Dhan CCAACUUUUCUA----UAAU----UAGUGAAGUGUGGGUCUUUUAAUUUUUGAUUU
Yli1 UCGUC----------UACG----------GACGGGGCCUAUAAUUUUUGAUUU
Yli2 CUGCC----------UACG----------GGCGGGGCCUAUAAUUUUUGUUUU
CONS YYR * ** R R* * RY**R *Y* ********* *

RNA secondary structures

- In addition to secondary structural interactions in RNA, there are also tertiary interactions, illustrated in figure below. These include **A** pseudoknots, **B** kissing hairpins and **C** hairpin-bulge contact.

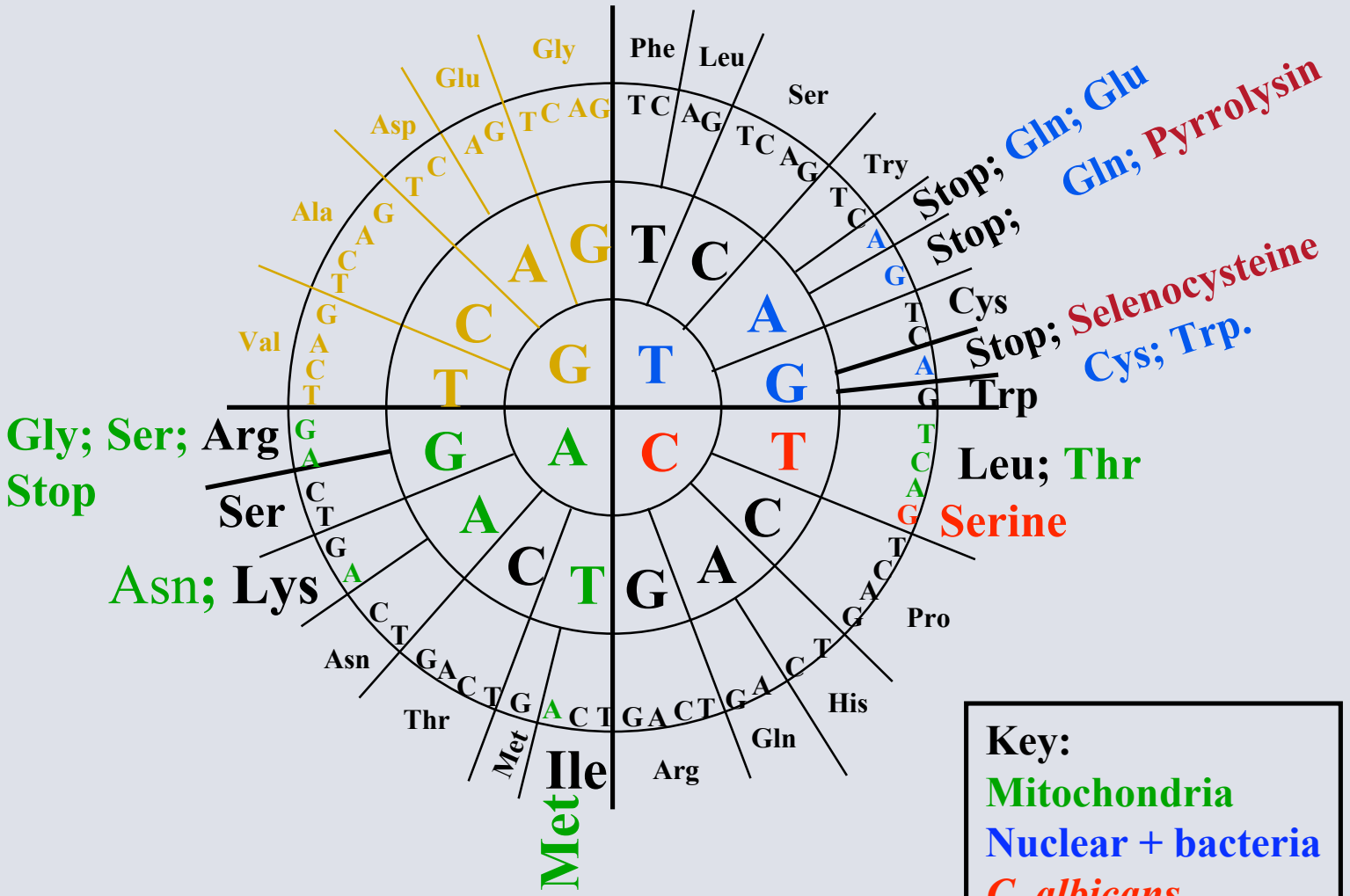
- These complicated structures are usually not predictable by secondary structure prediction tools.



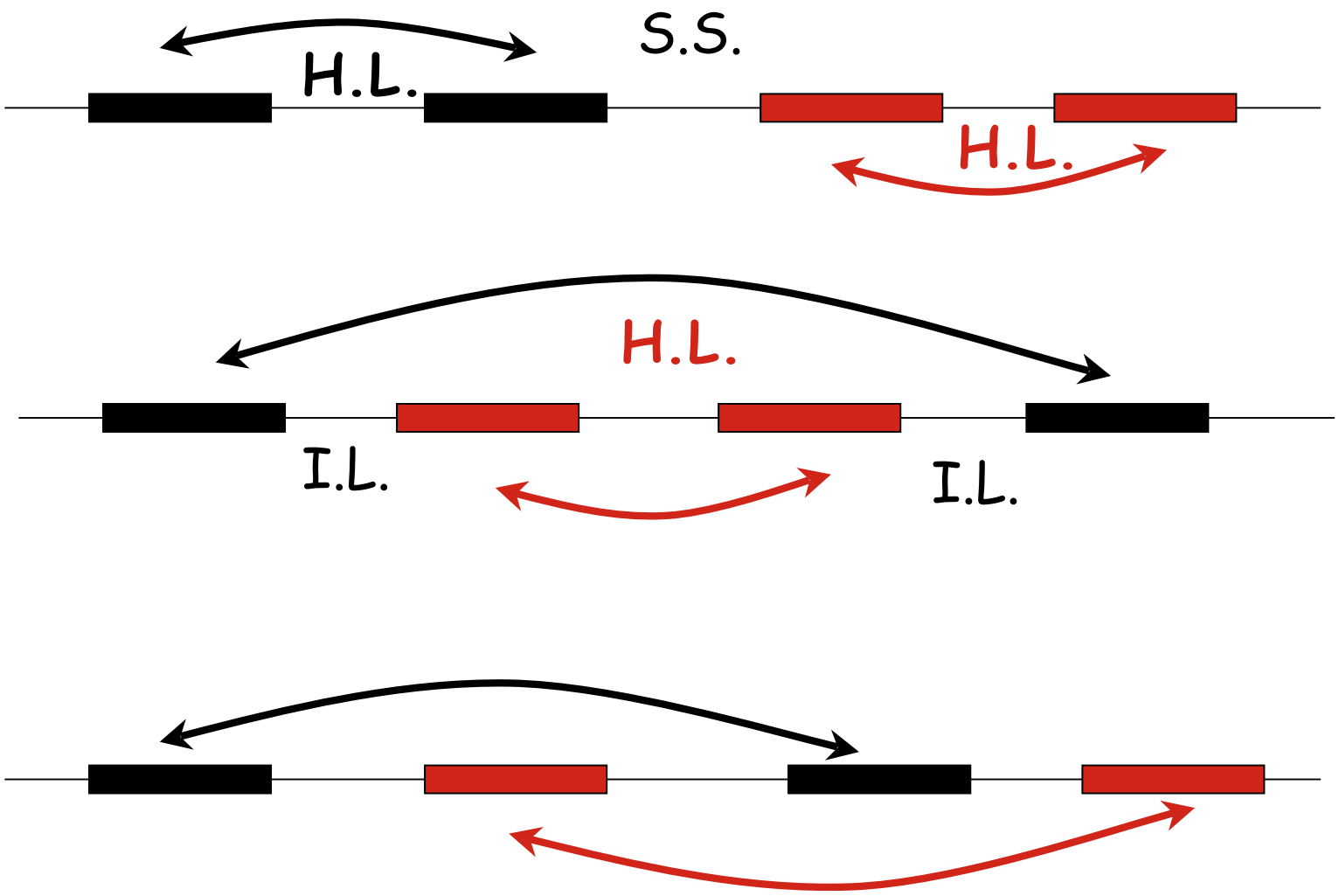
Functional RNA pseudoknots

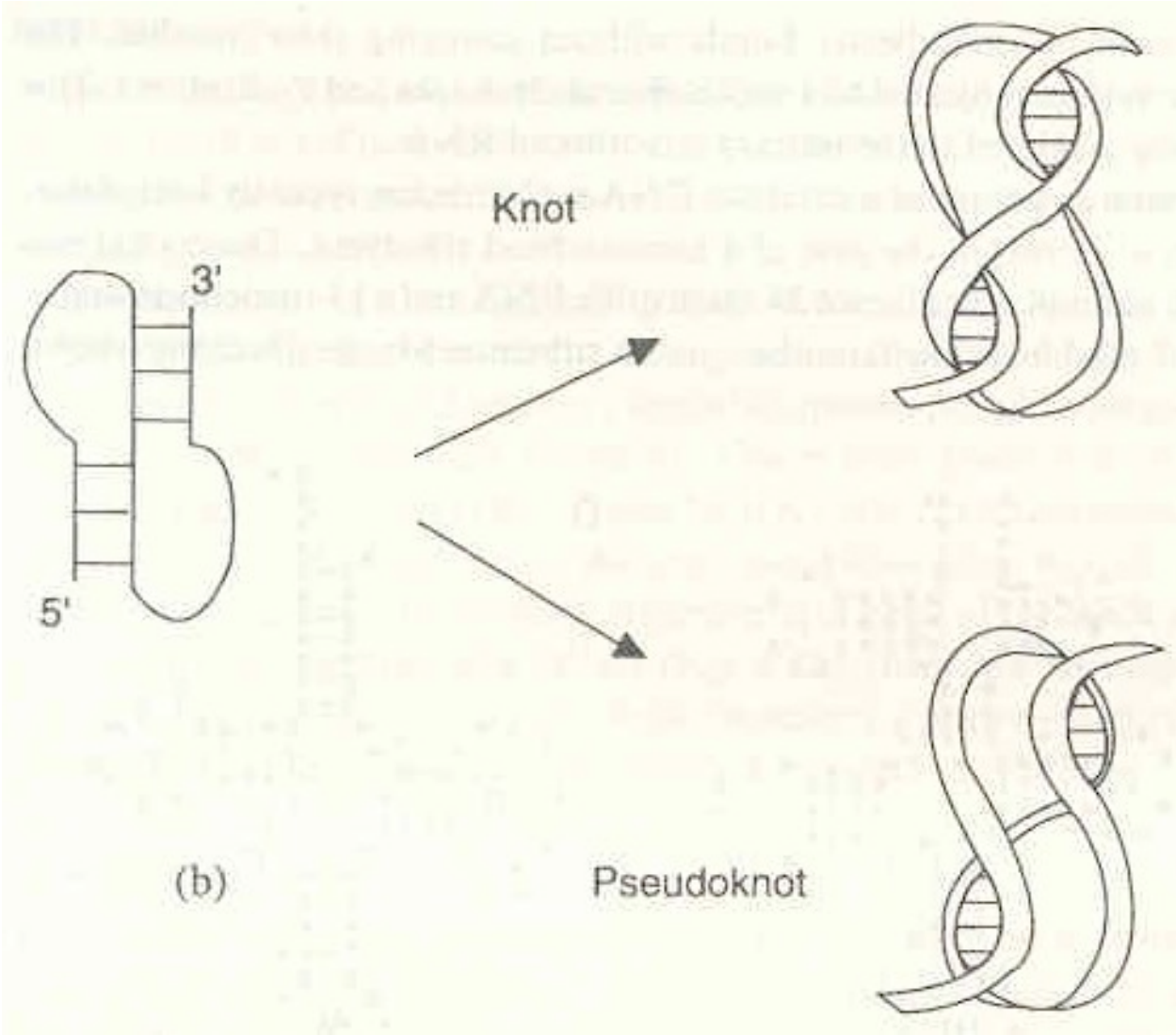
1) Pseudoknots located 5' of a messenger :
    - translational control : S4/ S8
    - frameshift of readingframe : RSV, coronavirus
    - readthrough : retrovirus
2) Pseudoknots located 3' of a messenger :
    - initiation of replication : BMV
    - telomere analog : BMV/TYMV
    - substitution for poly(A) tail : TMV
3) Central Pseudoknots :
    - folding of 16S rRNA
    - control of folding and catalytic activity :
        - group I introns
        - RNase P
        - Hepatitis delta virus

Genetic code & ribosome reprogramming

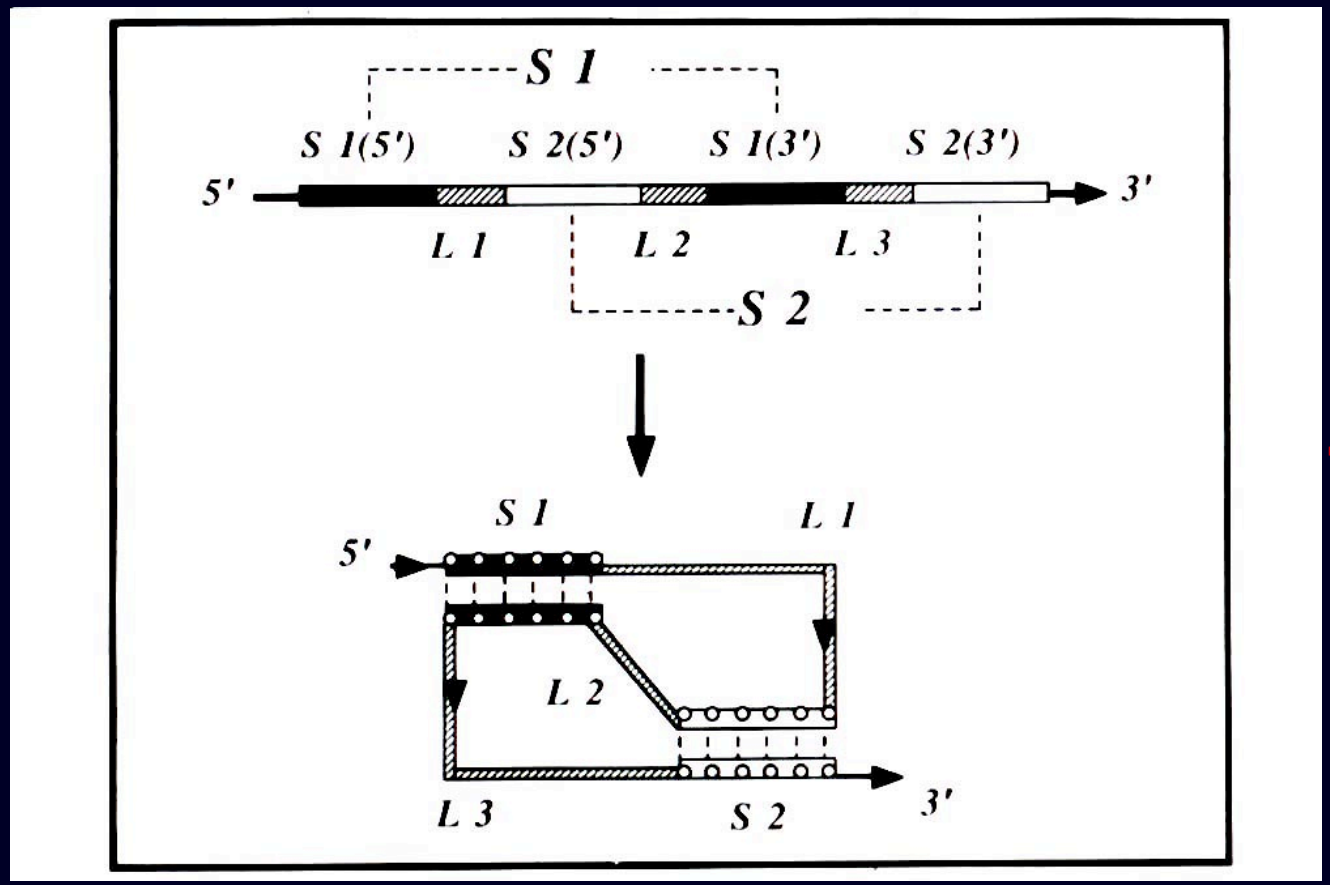


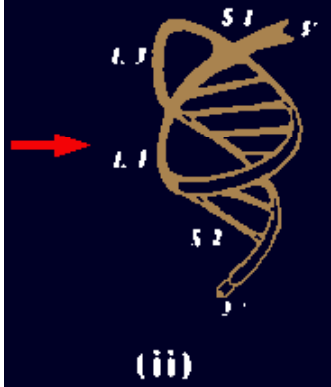
Only three ways to pair four segments

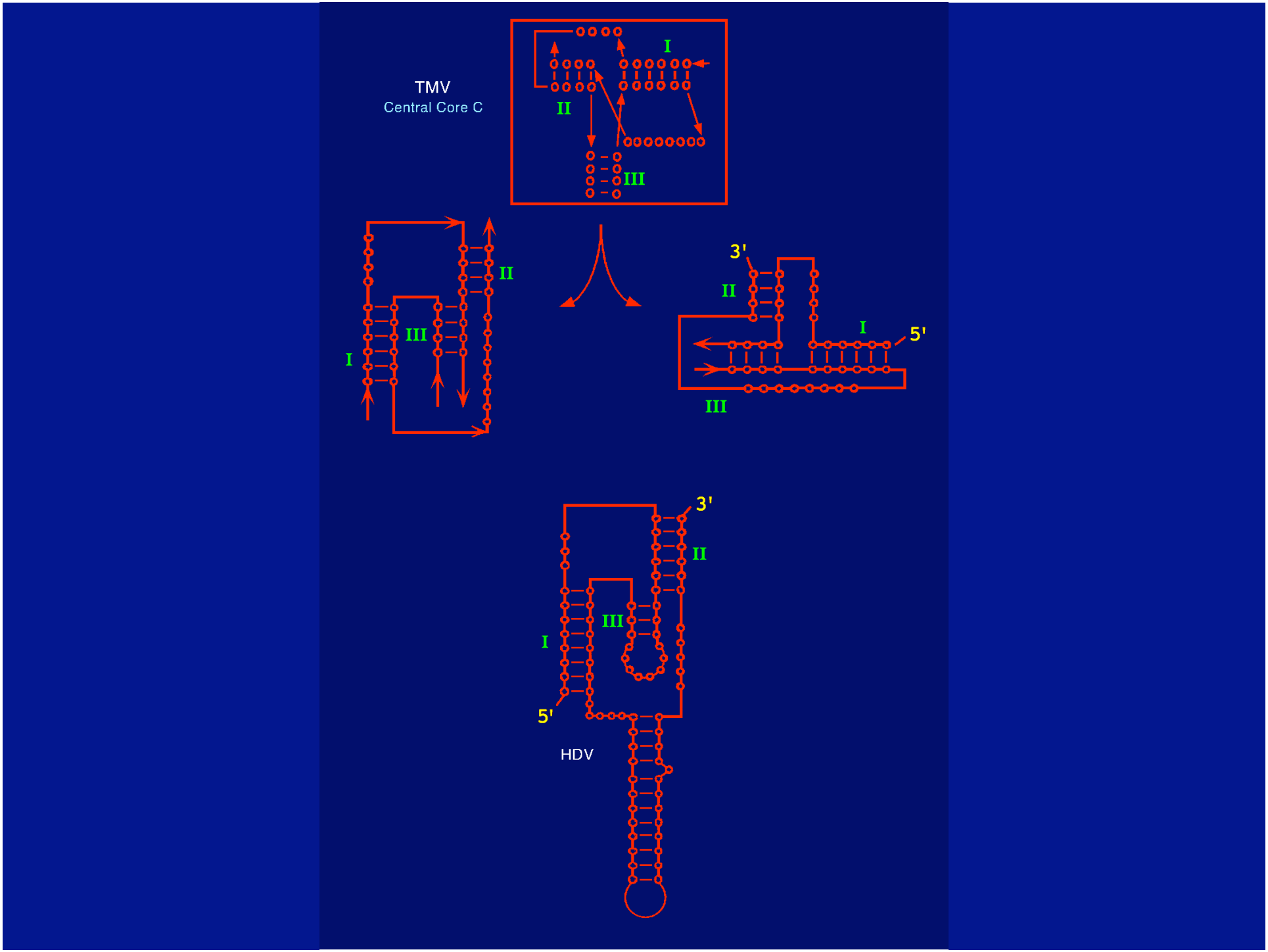


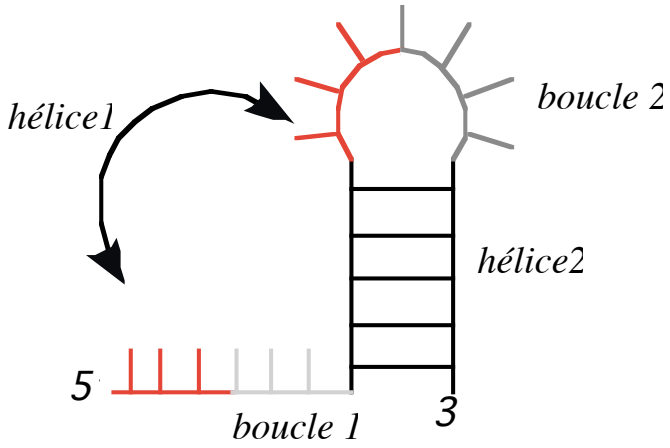

Knot
3'
5'
(b)
Pseudoknot

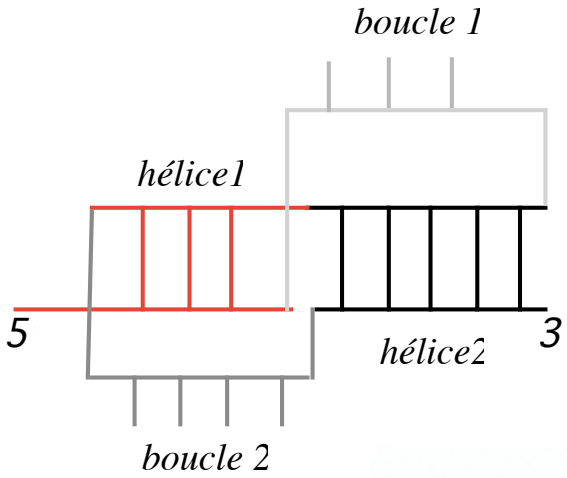
Intramolecular loop/single strand or loop/loop motif = pseudoknot

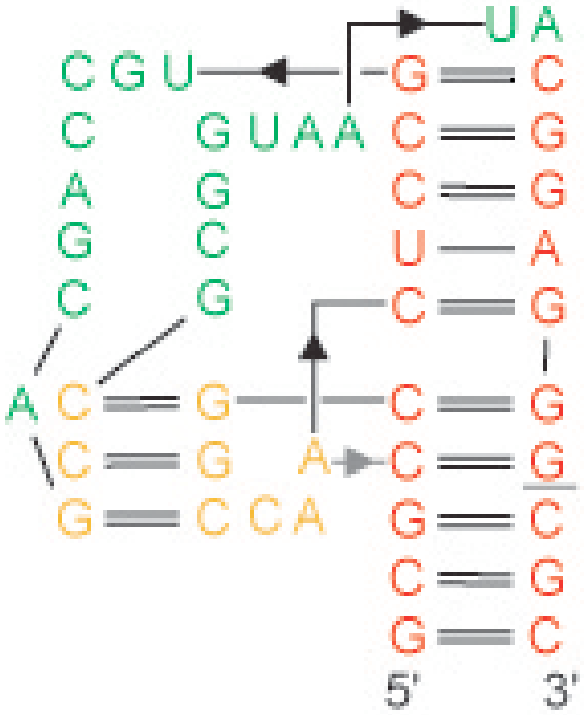


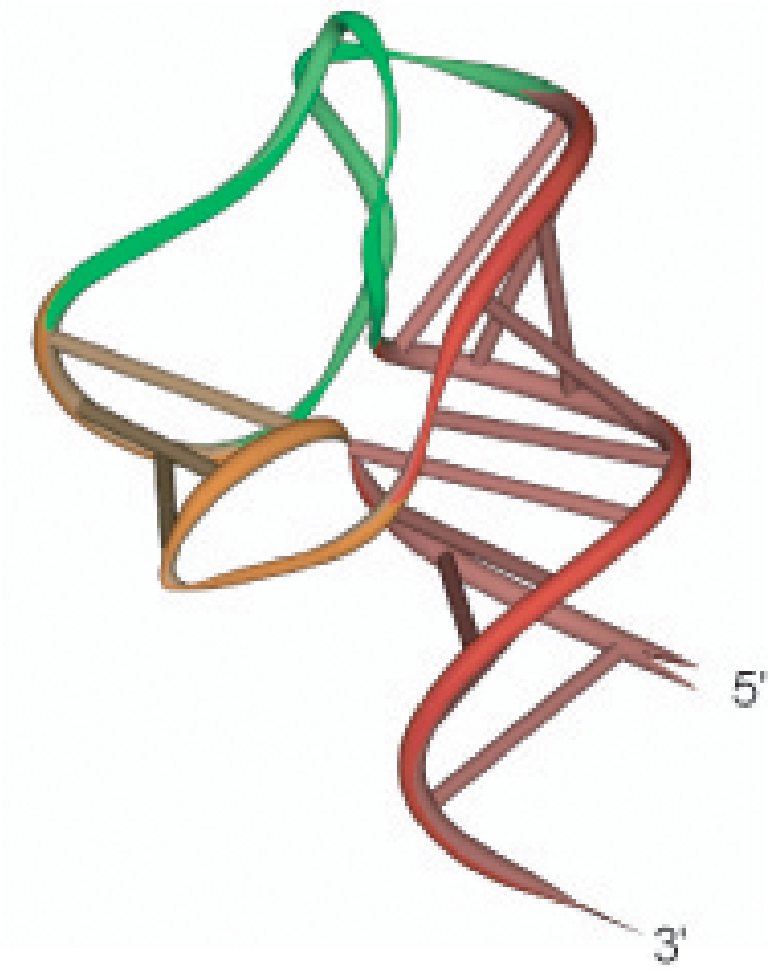



TMV
Central Core C
I
II
III
II
III
I
3'
II
I
5'
III
3'
II
III
I
5'
HDV

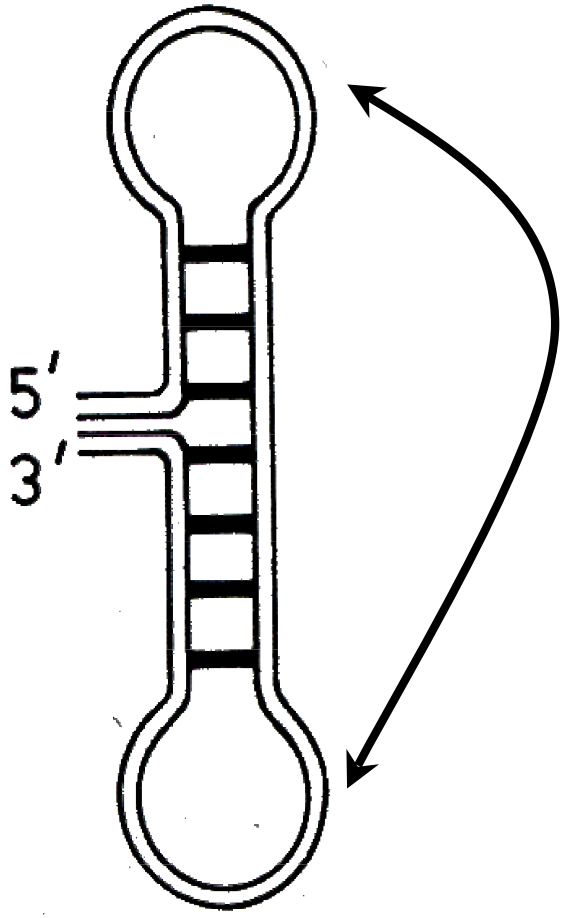

hélice1
boucle 2
hélice2
5
boucle 1
3


boucle 1
hélice1
5
hélice2
3
boucle 2

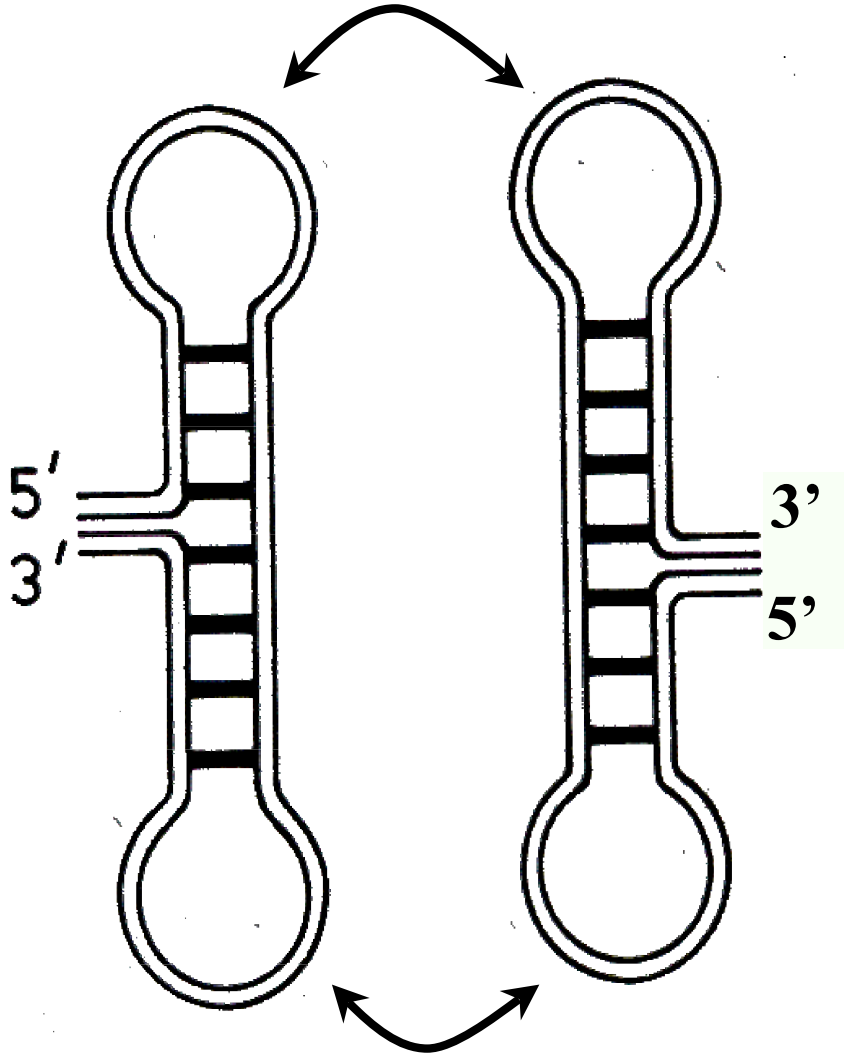

U A
C G U
G C
G U A A
C G
A
G
C
C G
A
C G
C G
G C C A
C G
C G
C G
U A
C G
C G
A
C G
C G
C G
5'
3'

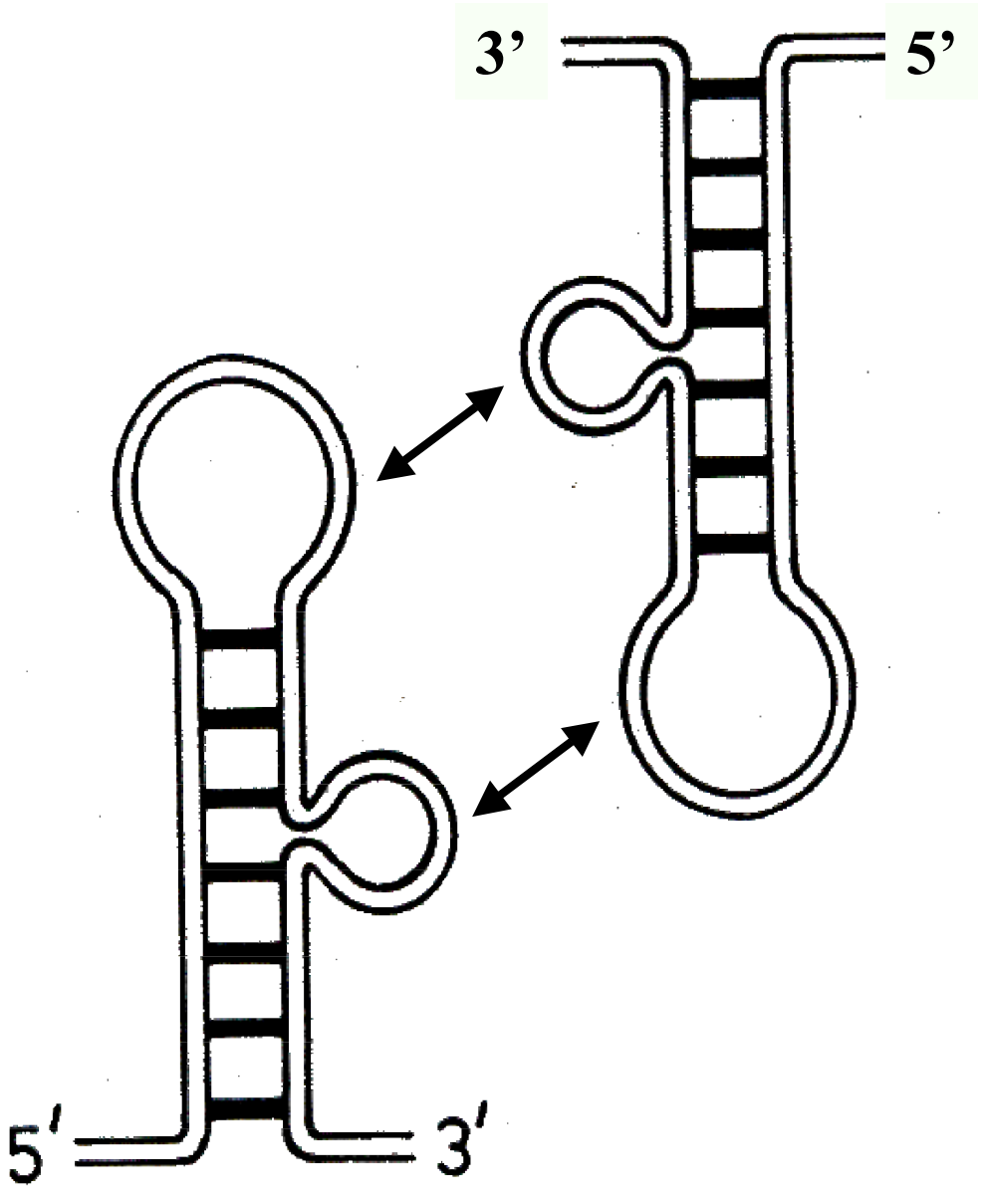

5'
3'

Intra-molecular
loop-loop
or
pseudoknot



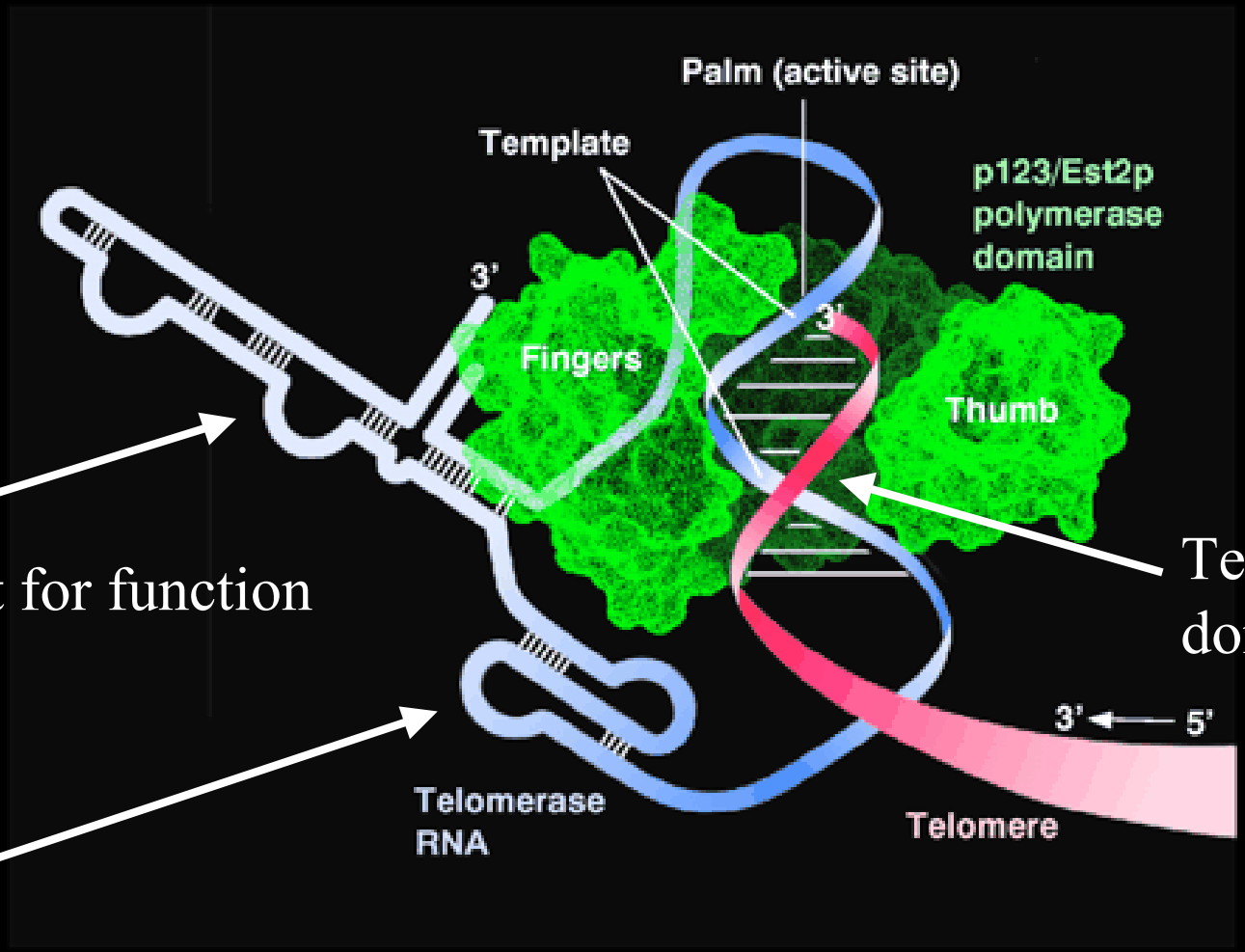
Inter-molecular
loop-loop
or dimer





An example of dimer formation of this type is seen in bicoid mRNA

Telomerase



Important for function

Template domain

pseudoknot

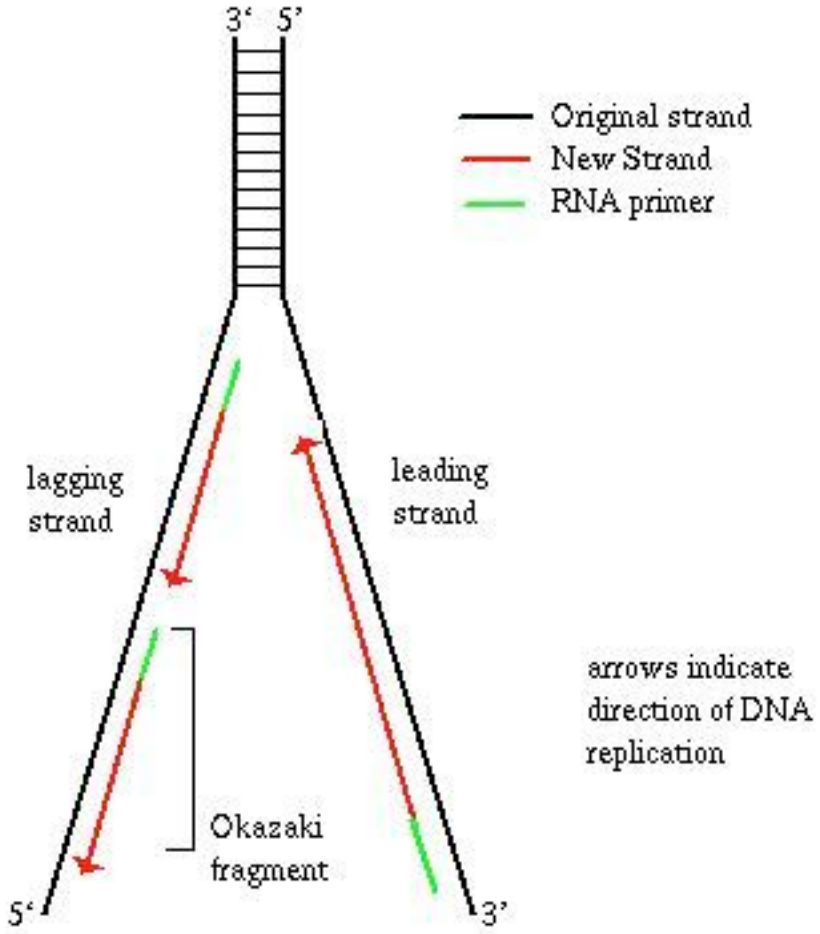
Functions of telomeres

Telomeres are elements at the ends of eukaryotic chromosomes they:

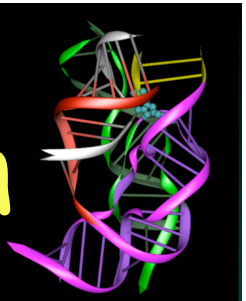
1. Protect chromosomes from end-to-end fusion,
2. Prevent homologous recombination and inappropriate DNA-repair.
3. Have a role in chromosome organisation at meiosis, mitosis and at interphase
4. Reverse the progressive degradation of chromosome termini caused by the incomplete replication of the 5’ end of the DNA double helix in S-phase of the cell cycle.

DNA replication

Loss of 40 to 200 bp per division in human (cited in Wright et al. 1997, see also Counter et al. 1992, Shay et al. 1993



Sequence guides & Recognition



- Recognition of nucleotides :

  Polymerases, Reverse-transcriptases
- Recognition of single-strands :

  Group I and Group II Introns

  Codon-Anticodon Recognition on mRNA
- Recognition of loops :

  Sense-Antisense RNAs

Translation

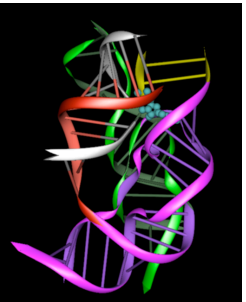
Ribosome Binding Site, Shine-Dalgarno Site

nnGGAGGnnnnnnATG...

typical *E. coli*

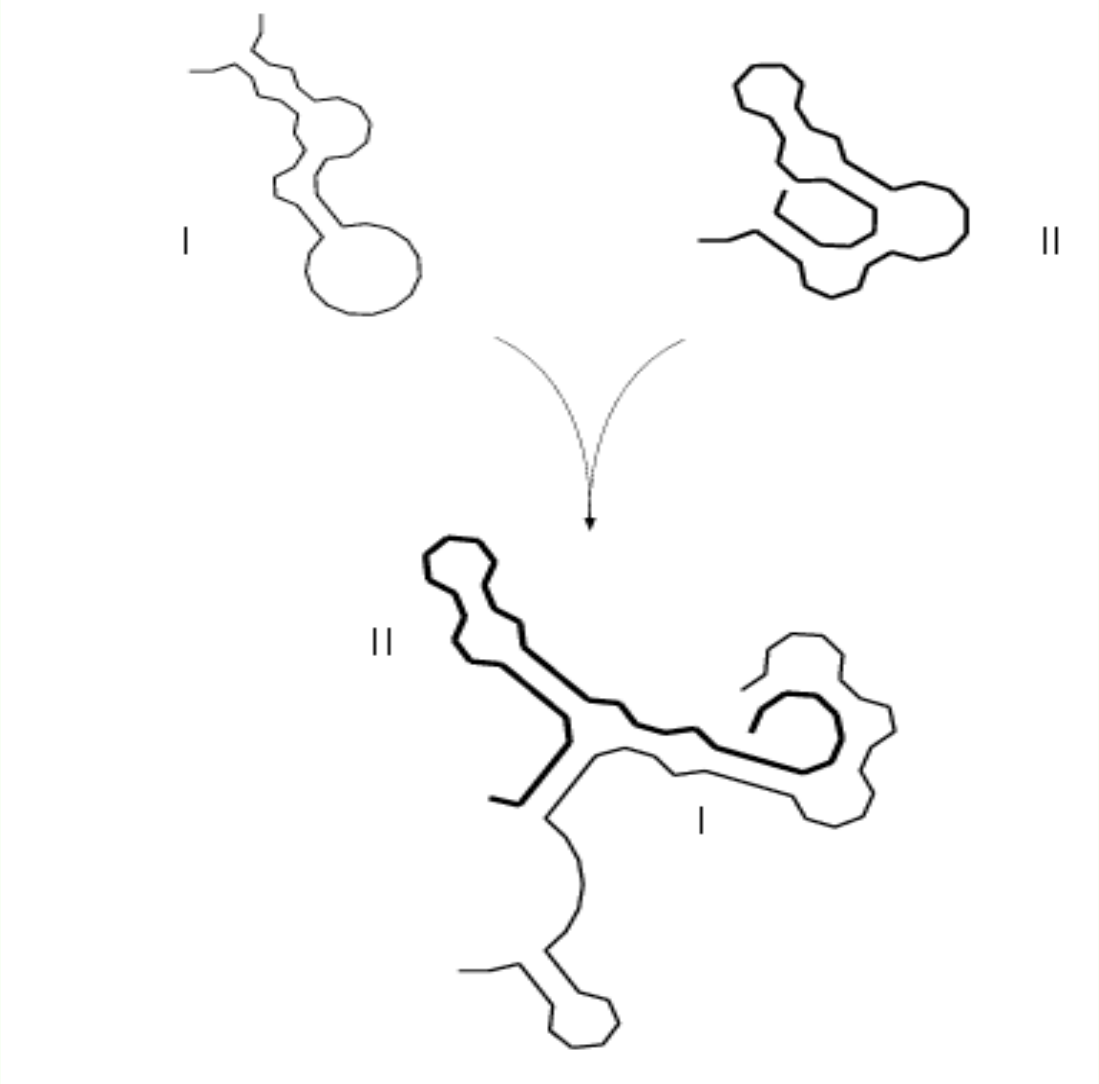
nnaaAGGnnnnnnATG

RNA self-assembly motifs



- Non-specific : Between sugar-phosphate :
  ribose zipper : O2'...O2' or O2'...N3(R)/O2(Y)
- Highly specific : Non-Watson-Crick base pairs
  GNRA tetraloops and receptors :
    GNRA....G-C/A-U base pairs in helices
    GAAA... 11-nt Costa-Michel motif
- Versatile : Watson-Crick base pairings :
  guide sequence, loop...loop, pseudoknots

RNA dimerization or hybridization

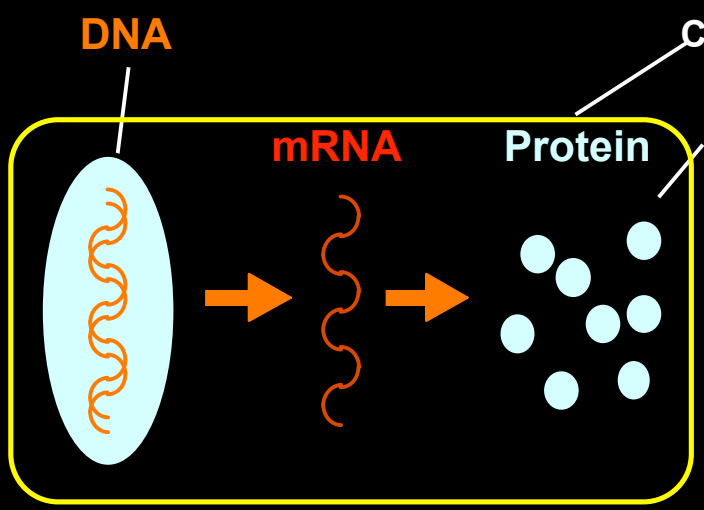


Inhibition by complementary base pairing

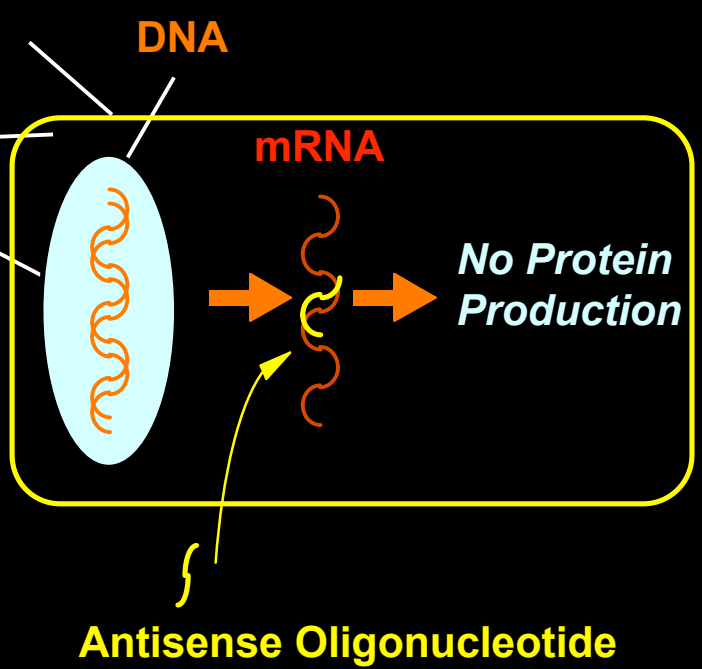
5′…-A-G-G-U-C-A-C-U-U-U-G-C-A-A-C-G-…3′ target mRNA

3′-C-A-G-T-G-A-A-A-C-G-T-T-5′ antisense DNA

Normal Protein Production

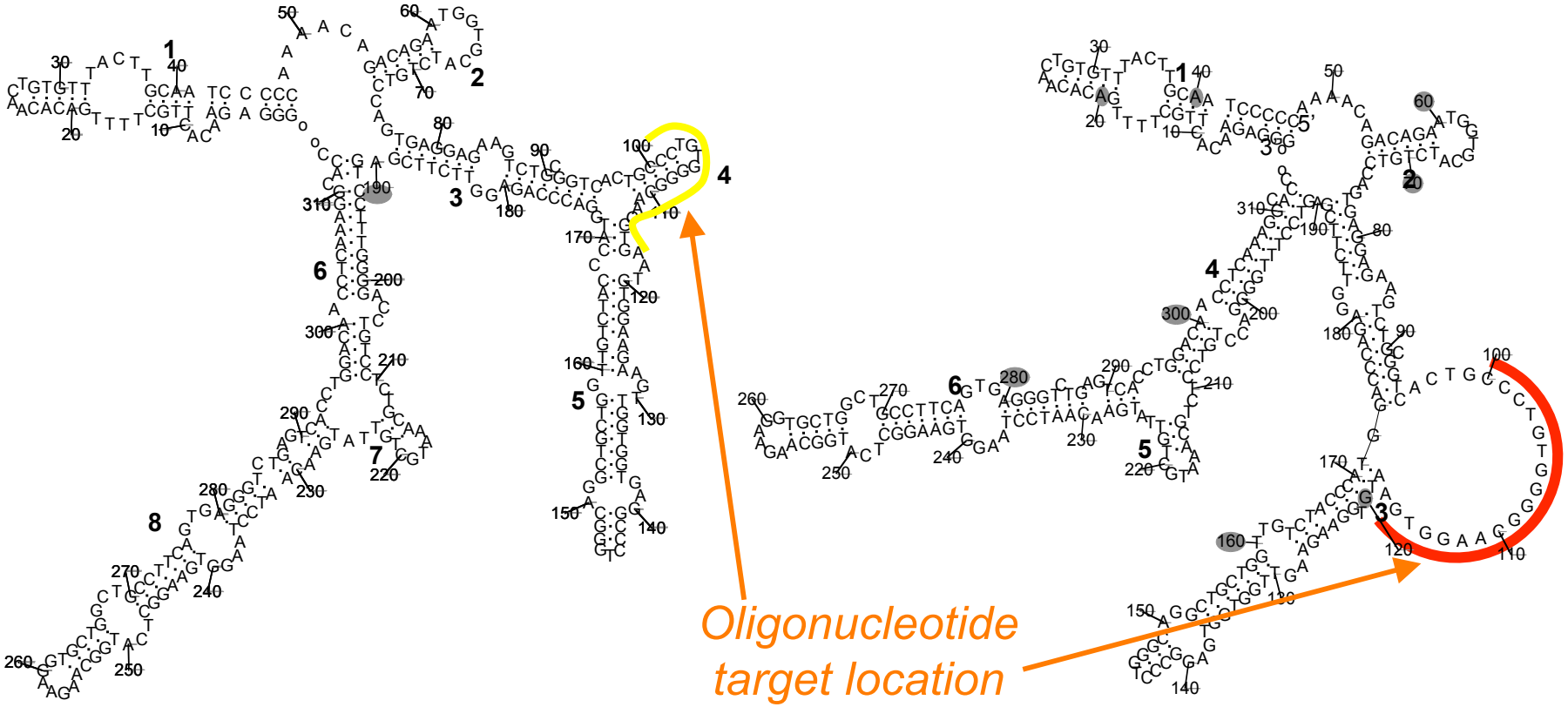


Antisense Inhibition



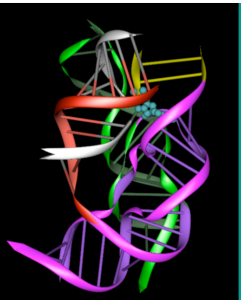
Binding: A Problem of mRNA Structure

mRNA secondary structure predictions (free energy minimization) for rabbit β-globin (RBG)

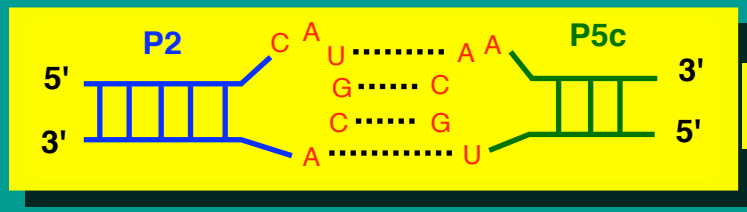


$$\Delta G_{UN} + \Delta G_{RE} = \mathbf{12.5\ kcal/mol}$$

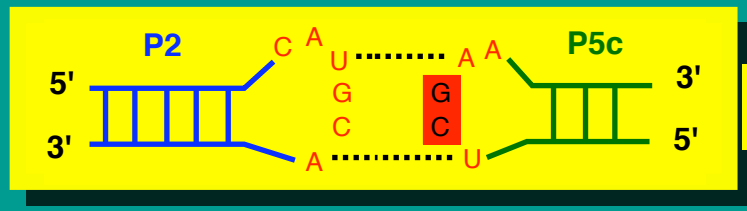
Watson-Crick pairs : loop-loop motifs



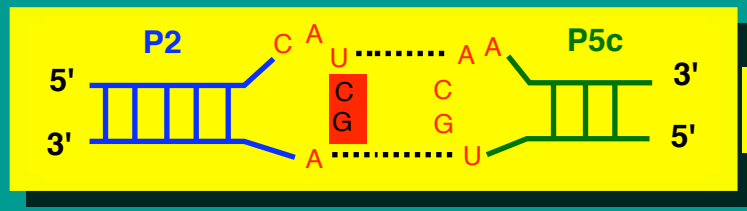
	P2	P14	P2′	P5c	P14′	P5c′
Tt.LSU	AGUUAUCAGGCA	UGCA	CCUGGUAGCU	( )	CCU--	UGCA	AAGG
Aa.LSU	AGGCA-UCGACU	CCCA	UUGA-UGCCU	( )	CUG-U	UGGG	GCAG
Pp.LSU	AGUCA-CAGGUAG	GCA	CCUGGUGGCU	( )	CCU--	UGC	ACAGG
Gg.LSU3	UGGCCGCGCAGUU	GCA	AGCGCGGUCA	( )	CGGAA	UGC	UACCG
Gp.SSU	AGUCGGGU(30)C	GAG	AAACCCGGCU	( )	CCC-A	CUC	CGGGG



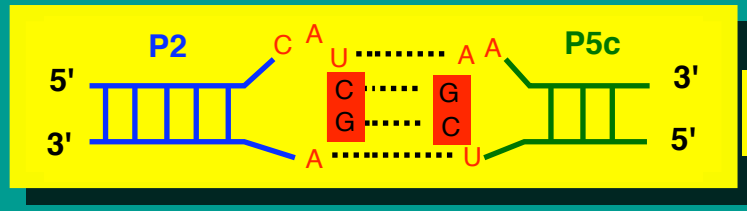
Wild type → Active



Mutant L5c → Inactive



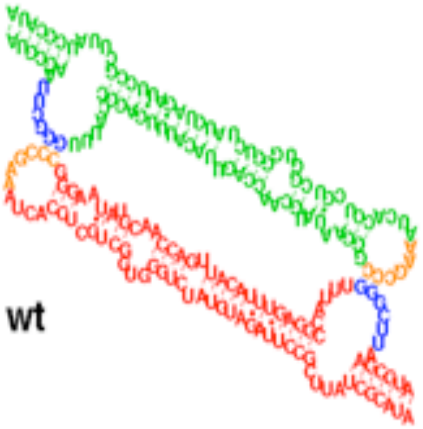
Mutant L2 → Inactive

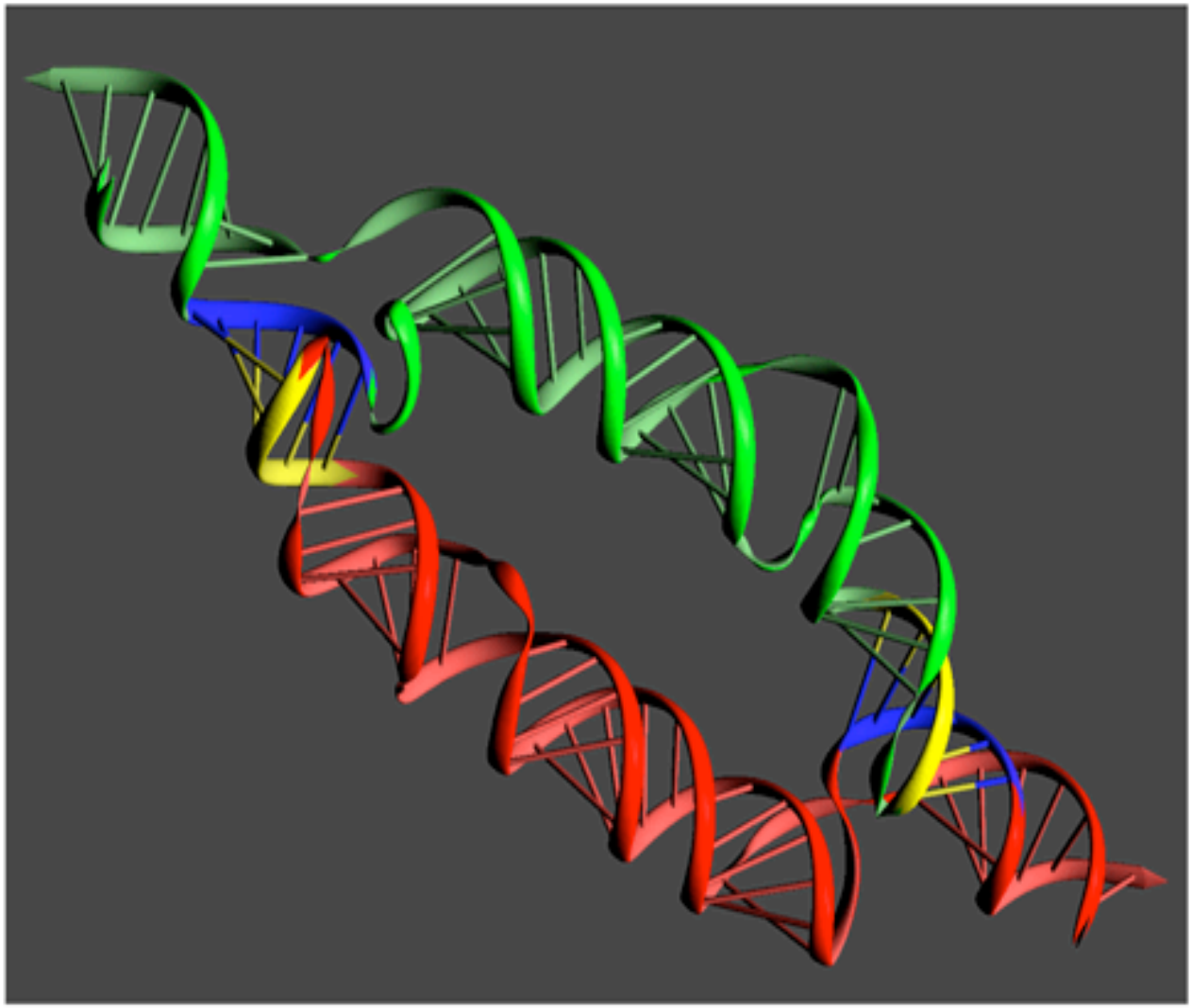


Mutant L5c-L2 → Active

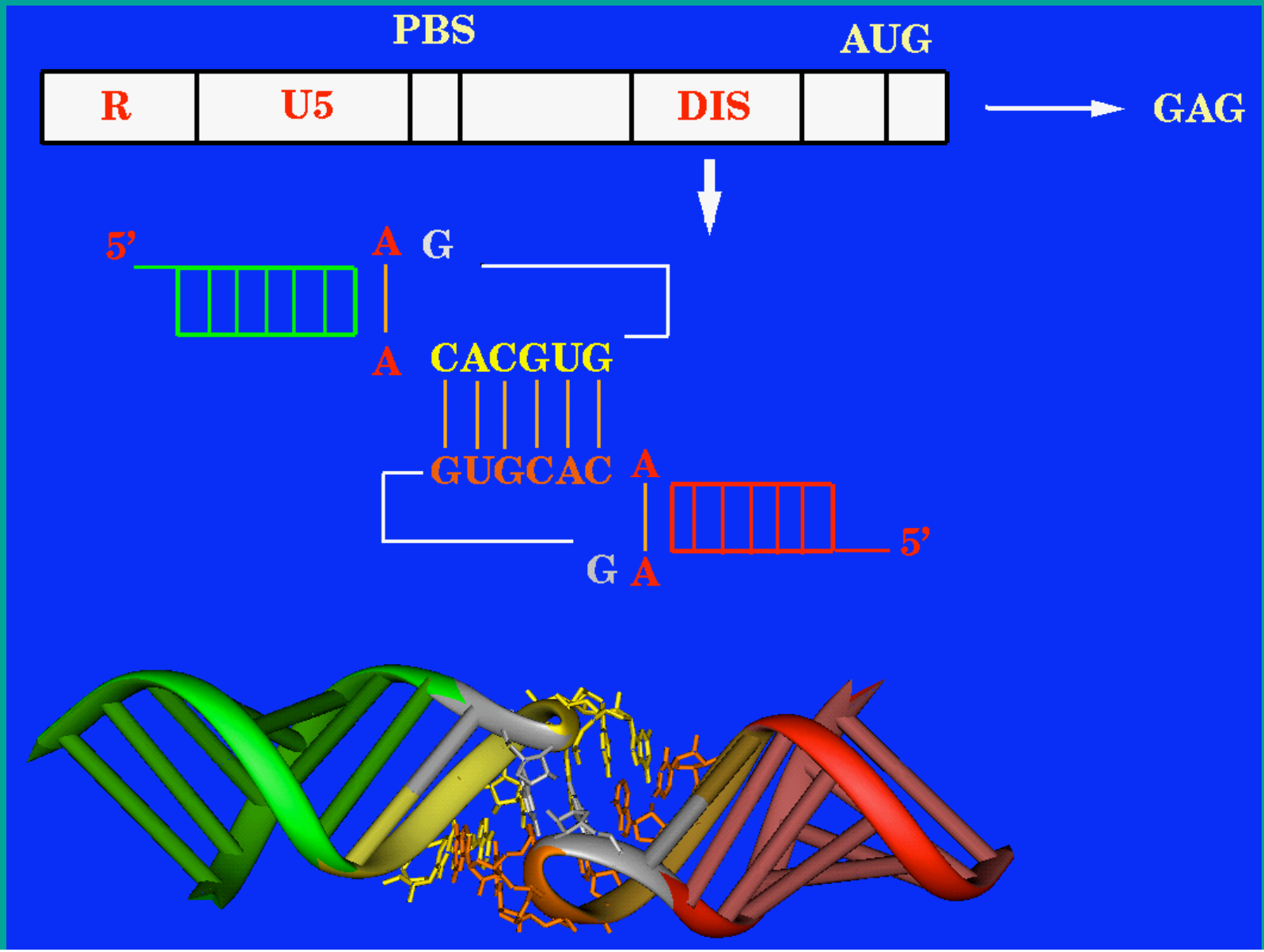
Intermolecular Apical - Internal loop-loop :

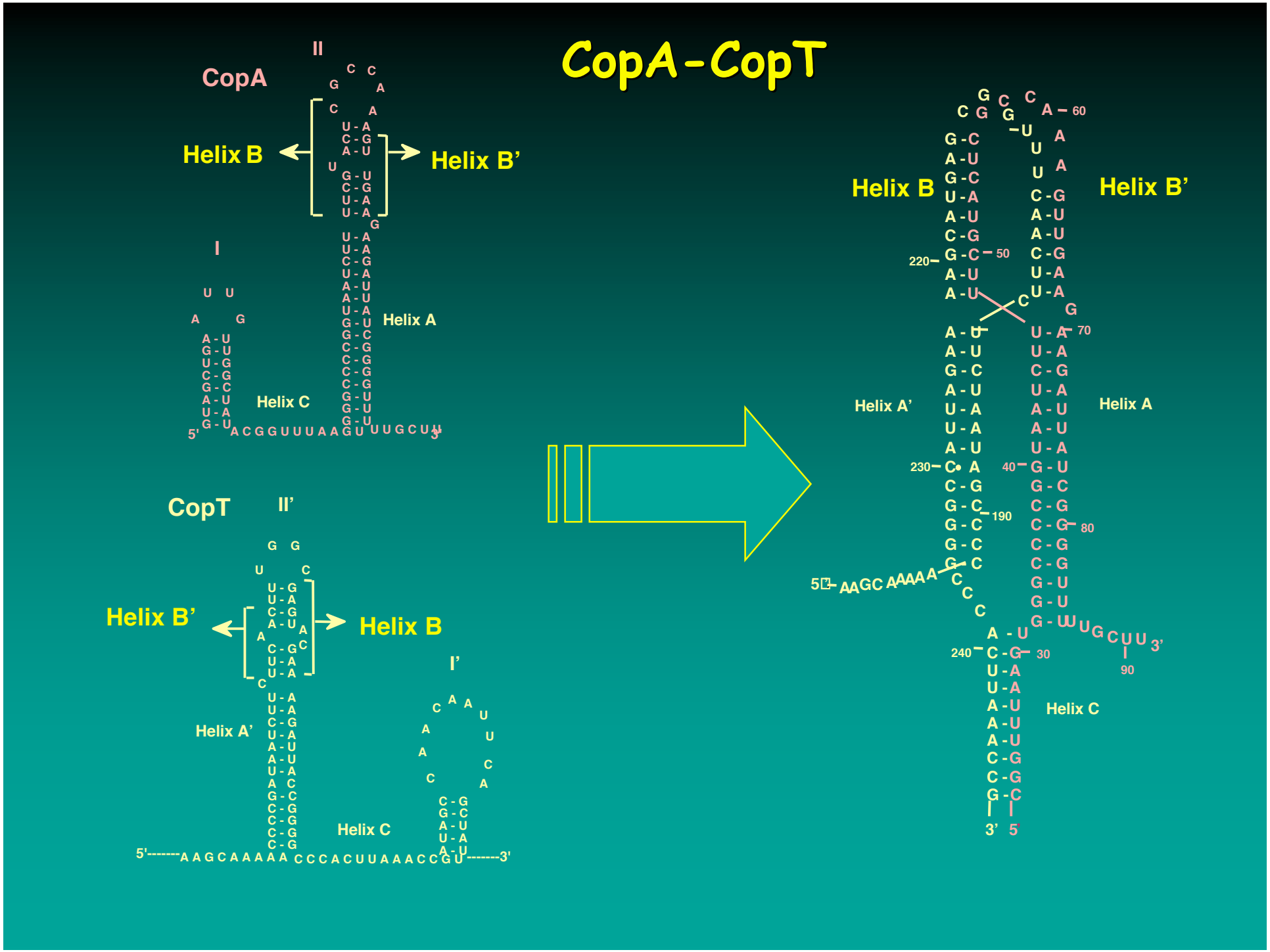
Bicoid mRNA dimerization



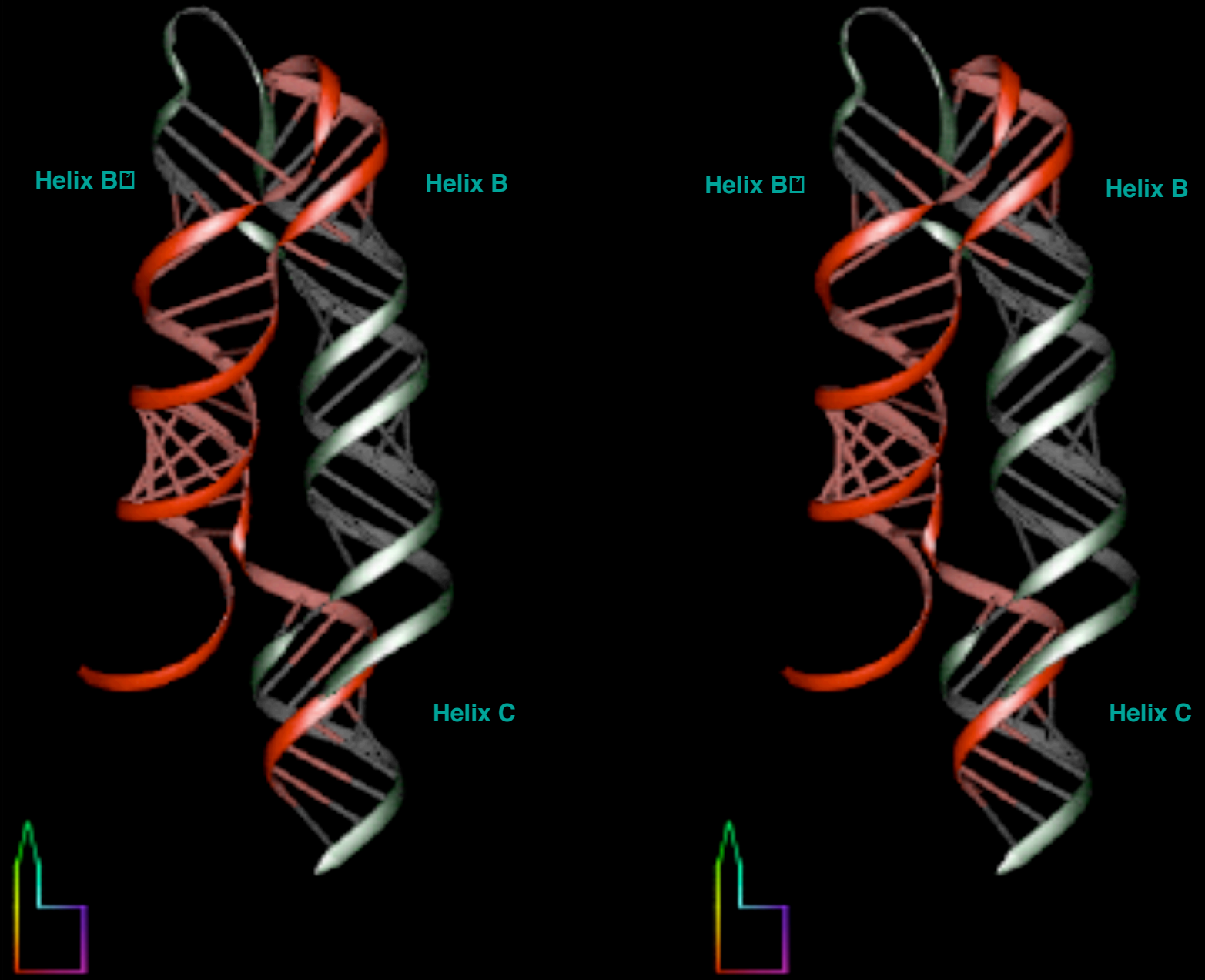


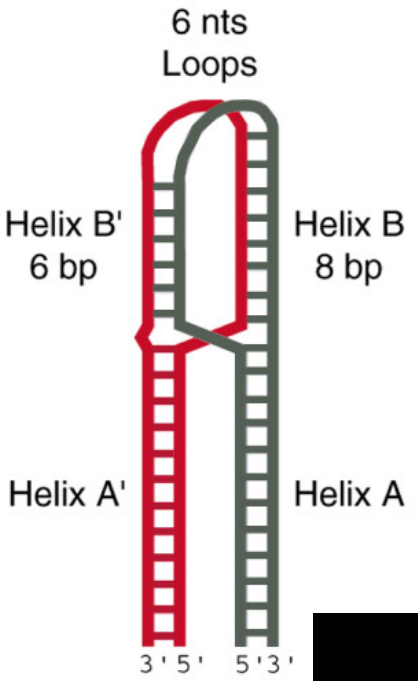
Self-complementary loop-loop : HIV RNA dimerization




CopA-CopT
CopA
II
I
Helix B
Helix B'
Helix A
Helix C
5'
3'
CopT
II'
I'
Helix B'
Helix B
Helix A'
Helix C
5'------AAGCAAAAAA CCCACUUAAACCGU-------3'
Helix B
Helix B'
Helix A'
Helix A
Helix C
5'
3'

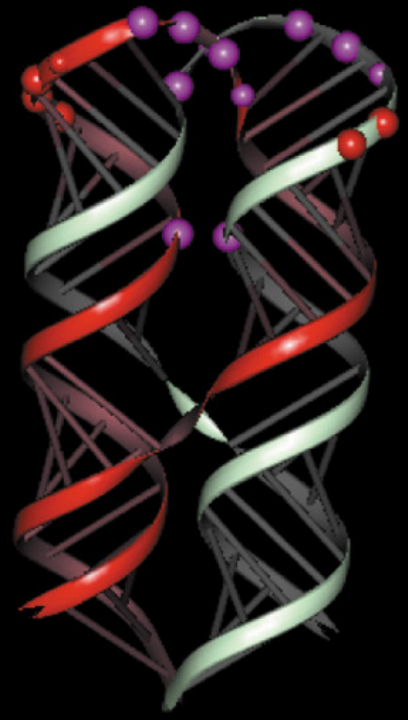
Three-Dimensional Model of the CopA-CopT Complex

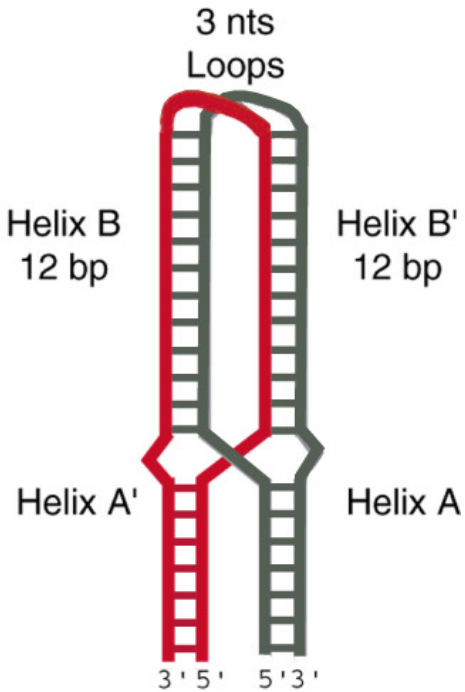




Absence of a knot

Presence of a knot





Active and inactive antisense/target RNA complexes