

GENE EXPRESSION AND PROTEIN SYNTHESIS

Hubert Becker

IPCB, UMR7156 4^{ème} étage

<https://dyps.unistra.fr/enseignement-et-cours/support-de-cours-h-becker/>



RECHERCHE

ENSEIGNEMENT ET COURS

VIE DU LABO

PUBLICATIONS

Équipe Becker / DyPS → Enseignement et cours → Support de cours H. Becker

Support de cours H. Becker

Licence 1 - Biochimie

	H2O et interactions L1S2	pdf	3 Mo
	Les molécules du vivant	pdf	3 Mo
	Biochimie 1 Protéines	pdf	2 Mo
	Glucides	pdf	2 Mo
	Poly2025 Lipides&Biomembranes	pdf	4 Mo
	Acides nucléiques 2025	pdf	4 Mo
	Code génétique 2025	pdf	4 Mo
	Méthodes de séparation et d'analyse des macromolécules biologiques	pdf	3 Mo

Licence 2 - Métabolisme

	Cours_Metabo_2022.pptx	pptx	33 Mo
	CC1 Métabolisme 2022	pdf	289 Ko
	Annales CC1 métabolisme	pdf	443 Ko
	Microsoft Word - CC1 métabo 2025 réponses.docx	pdf	341 Ko
	CC1_Notes_Metabo_final.xlsx	xlsx	157 Ko

Master BGM - OEVE

	cours_OEVE_HB_2018.pdf	pdf	12 Mo
---	------------------------	-----	-------

Master 1 - RAEG

	Cours_RAEG_HB_2020.pptx	pptx	19 Mo
	TD_TBBox_2022.pptx	pptx	483 Ko
	Groupe_RAEG_2023.xlsx	xlsx	10 Ko

Master 1 - EGBP

	EGBP_CC1_annale_1.pdf	pdf	907 Ko
	EGBP_CC1_annale_2.pdf	pdf	622 Ko
	EGBP_Chapter1_2026_1rst_Lecture.pdf	pdf	12 Mo



On the menu...

Chapitre I: Introduction – Genetic Code – RNA Structure

Chapitre II: Ribosomal Steps of mRNA translation

Chapitre III: tRNA – Aminoacyl-tRNA synthetases – Non translational Functions

Chapitre IV: mRNA metabolism

Chapter I: Introduction

I. General mechanisms of the transfer of genetic information

II. The Genetic Code

A. Deciphering & Degeneracy

B. The Wobble pairing

C. Initiation and STOP codons

D. Dynamic reprogramming of genetic information

1. Missense, Nonsense, suppressor

2. STOP codons reprogramming

3. Frameshifts

4. Editing

E. The Genetic Code is not universal

F. Codon frequency

III. General mechanism of translation

IV. RNA structure

A. Primary

B. Secondary

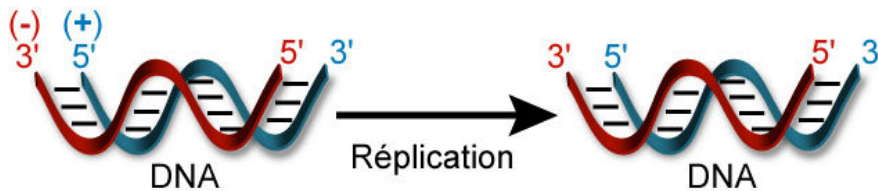
C. Tertiary

D. Quaternary

I. General mechanisms of the transfer of Genetic Information (GI)

- GI rests on the Genetic Code (GC).
- Post-transcriptional steps are those following gene transcription and its product: RNA
- The different processes of the transfer of GI are:

• DNA replication



- Double stranded ADN
- Watson-Crick bases pairing rules: A=T et G≡C
- (+) Strand – forward strand – sense – coding – contains GI – **Watson strand**
- (-) strand – reverse strand – antisense – template for transcription – **Crick Strand**
- (+) & (-) are antiparallel
- Genes are on the (+)
- During replication, the (+) & (-) strands are separated. Each strand of the original DNA molecule then serves as a template for the production of its counterpart, (semiconservative replication). As a result, replication generates 2 double stranded DNA molecules and each of the new helix will be composed of an original DNA strand as well as a newly synthesized strand

Table 1.

Watson- and Crick-strand definitions

Definition	Watson	Crick
Original	cytosine-poor	cytosine-rich
Compositional	pyrimidine-rich	purine-rich
Transcriptional	antisense	sense
Replicational	lagging	leading
Arbitrary	this	that
Database	top/plus	bottom/minus
5' to 3'	left to right (top or left-hand)	right to left (bottom or right-hand)

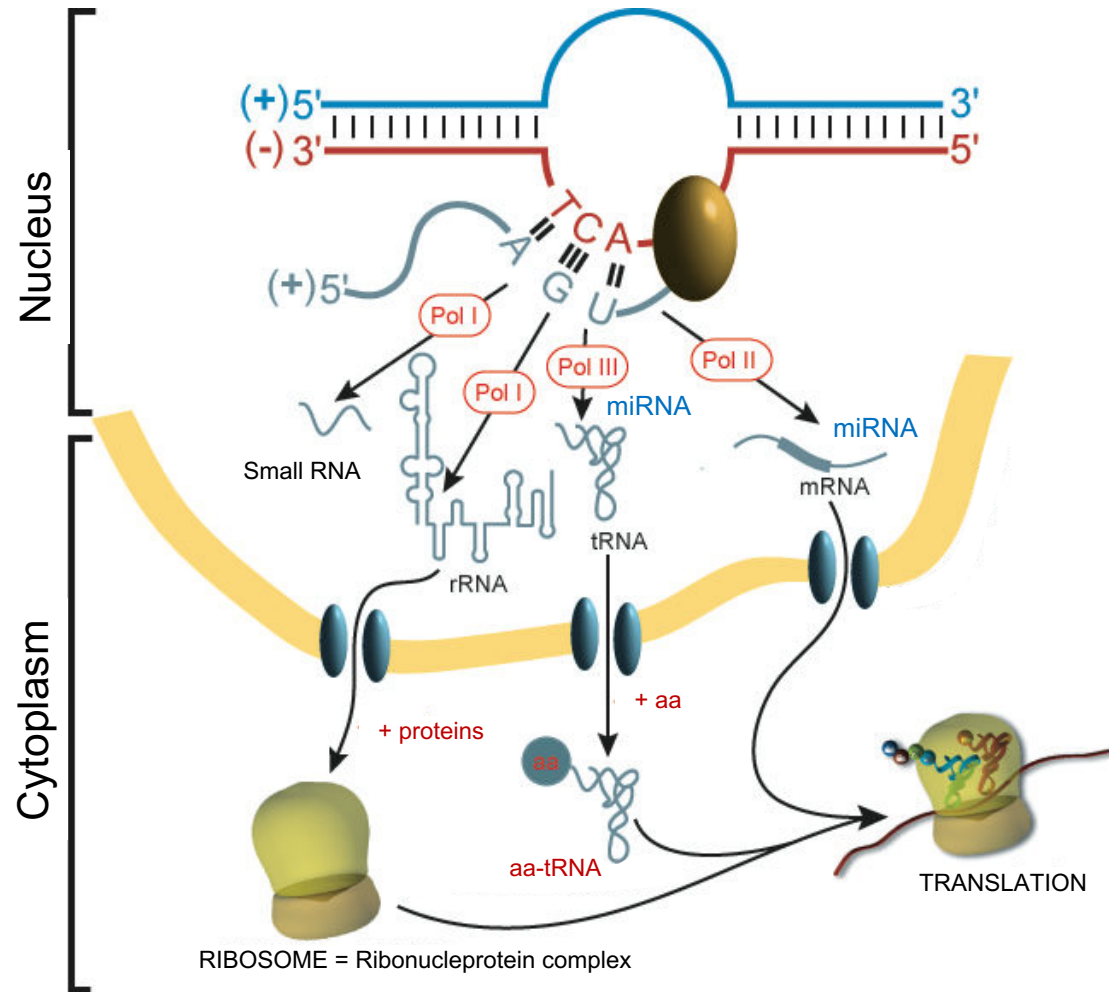
doi: 10.1186/1745-6150-6-7

I. General mechanisms of the transfer of Genetic Information (GI)

- Transcription of DNA

- Catalyzed by DNA-dependent RNA polymérase (RNAPol)
- The (-)/Crick DNA strand is transcribed into sens ARN (+)

RNAPol moves from 3→5' on the Crick/(-) DNA strand. The RNA molecule is synthesized from 5'→3' and has the same sequence than the (+)/Watson strand except T is replace by U



I. General mechanisms of the transfer of Genetic Information (GI)

- The consensual view of the GI flow is that GI is stored into the form of a **DNA** molecule which will be copied into **mRNA** which is the intermediate form of GI that will be translated into **Protein (Prot)**
- All the RNA molecules participating to Protein Synthesis (PSynt) have to fold into an active 3D structure to be fully functional. Therefore one has to know how a given RNA molecule folds to be able to understand PSynt



There are fundamental differences between Eucaryotic and Procaryotic PSynt

- In Proc. there is no nucleus and nuclear membrane (MMB), hence, all steps that take place in the nucleus in Euk. are not uncoupled in Proc. And some steps do not exist: (ex: splicing, nuclear shuttling to the cytoplasm)
- In Proc. there is only one RNAPol for transcription of all sorts of RNA
- In Proc. there is transcription-translation coupling

II. The Genetic Code

A. Deciphering & Degeneracy

♦ Deciphering



II. The Genetic Code

A. Deciphering & Degeneracy

♦ Deciphering

In the years following **1953**, researchers competed to be the first one to crack the GC. To make this competition a bit more interesting, a theoretical physicist and astronomer **George Gamow** invented a club called **“RNA Tie Club”**, in which any member could give his ideas on how a succession of nucleotides (Ntides) could be converted into a succession of aa. This club had 20 members, one for each proteinogenic aa and the members were wearing a tie with the chemical structure of an aa. Marshall Nirenberg did not belong to the club which by the way did not crack the code but generate a lot of interesting theories. One of the main issues that had to be solved was to determine how many Ntides would specify an aa and George Gamow in the late 50s gave a first hint. All researcher at that time knew that the Ntide alphabet was constituted of 4 bases (A, C, G et T/U) therefore, a combination of 2 Ntides would not be sufficient to code for 20 aa: **(4×4) makes 16 combinations** and therefore only 16 aa. It was thus evident that the GC language had to made at least of Ntide triplets **$(4 \times 4 \times 4)$ 64 combinations**. However this is **far more than necessary** for 20 proteinogenic aa



II. The Genetic Code

A. Deciphering & Degeneracy

♦ Deciphering

In the years following **1953**, researchers competed to be the first one to crack the GC. To make this competition a bit more interesting, a theoretical physicist and astronomer **George Gamow** invented a club called **“RNA Tie Club”**, in which any member could give his ideas on how a succession of nucleotides (Ntides) could be converted into a succession of aa. This club had 20 members, one for each proteinogenic aa and the members were wearing a tie with the chemical structure of an aa. Marshall Nirenberg did not belong to the club which by the way did not crack the code but generated a lot of interesting theories. One of the main issues that had to be solved was to determine how many Ntides would specify an aa and George Gamow in the late 50s gave a first hint. All researchers at that time knew that the Ntide alphabet was constituted of 4 bases (A, C, G et T/U) therefore, a combination of 2 Ntides would not be sufficient to code for 20 aa: **(4 × 4) makes 16 combinations** and therefore only 16 aa. It was thus evident that the GC language had to have at least of Ntide triplets **(4 × 4 × 4) 64 combinations**. However this is **far more than necessary** for 20 proteinogenic aa

J.D. Watson et F.H. Crick
Medical Research Council Unit for the Study of the Molecular Structure
of Biological Systems, Cavendish Laboratory, Cambridge.
Submitted April 2nd 1953 ; published April 25th 1953 (Nature 171, page 737).



**LE CODE EST
REDOND-**

DÉGÉNÉRÉ!



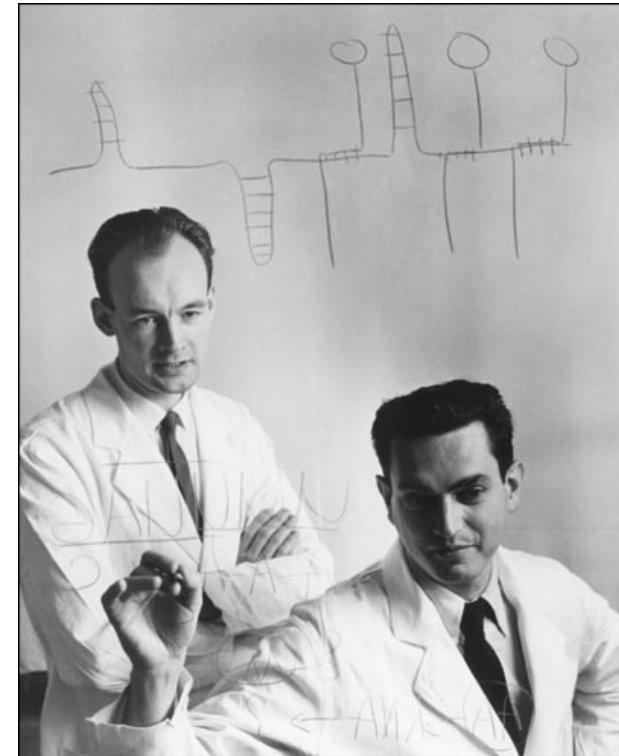
II. The Genetic Code

A. Deciphering & Degeneracy

♦ Deciphering

In 1957, **Marshall Nirenberg** joins NIH as a young PI and succeeds to fund his research. He decides to focus on deciphering the GC G et passe les premières années à démontrer que le RNA est la molécule à partir de laquelle se fait la synthèse protéique (**“would not have to get polynucleotide synthesis very far to break the coding problem. Could crack life's code!”**)

In his lab at NIH (Maryland), Marshall Nirenberg and his post-Doc **Heinrich Matthaei** at the end of 1960 designed the so-called **“The poly-U experiment”** that required an *in vitro* translation system. To do so they harvested *E. coli* cell, broke them by sonication and then centrifugated the exytract at 30000 g to eliminate cell debris and release the cytoplasmic content. This fraction called 30 S was still able to translate a synthetic or natural messenger RNA.



II. The Genetic Code

A. Deciphering & Degeneracy

♦ Deciphering

CHARACTERISTICS AND COMPOSITION OF RNA CODING UNITS*
BY J. HEINRICH MATTHAEI, OLIVER W. JONES, ROBERT G. MARTIN, AND)
MARSHALL W. NIRENBERG
NATIONAL INSTITUTE OF ARTHRITIS AND METABOLIC DISEASES, BETHESDA
Communicated by Richard Roberts, February 27, 1962
Proc. Natl. Acad. Sci. U.S.A. **48**, 666-677

This experiment required Mg^{2+} , ATP, active ribosomes, synthetic mRNA and 20 aa among which 1 was [^{14}C]-radiolabelled



II. The Genetic Code

A. Deciphering & Degeneracy

◆ Deciphering

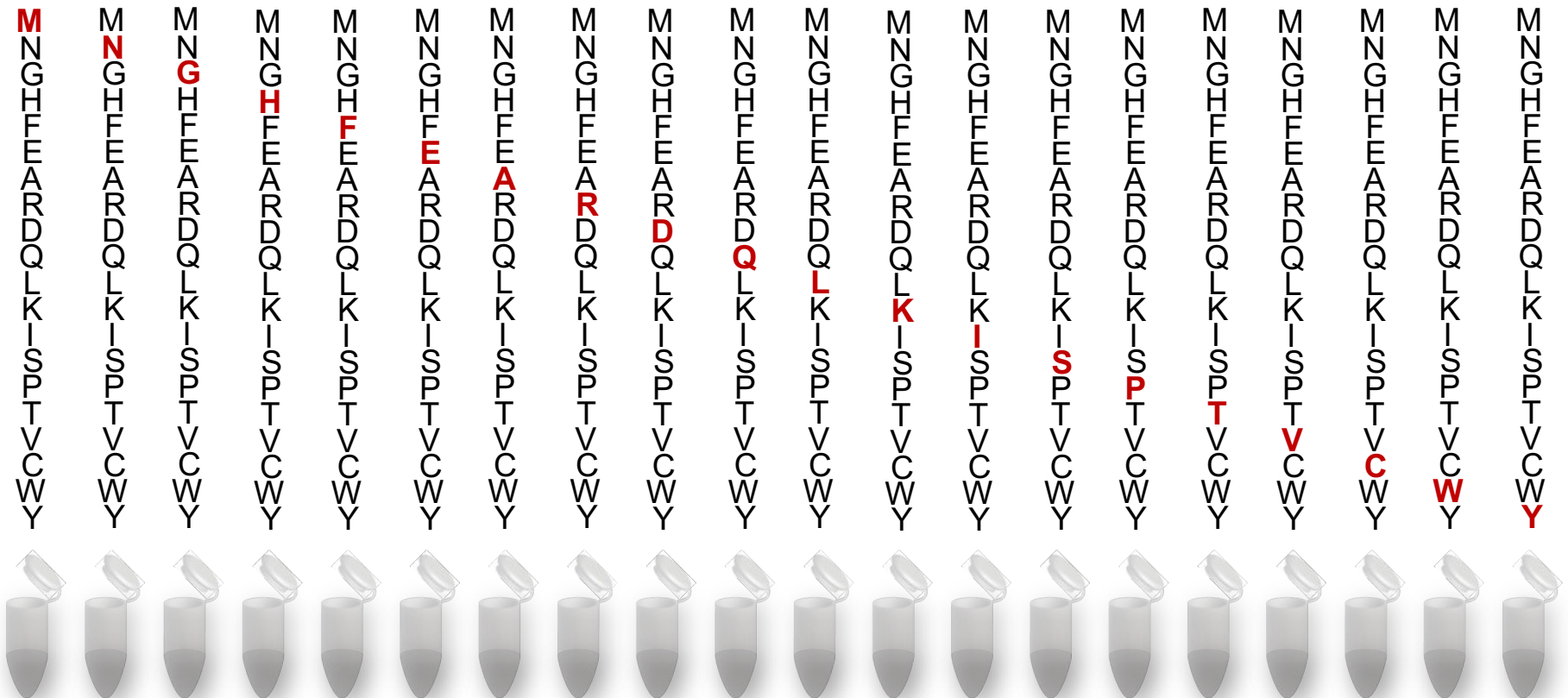
27th of May 1961 3:00 AM

E. coli 30.000 g crude extract

+

5'UUU....UUU3'

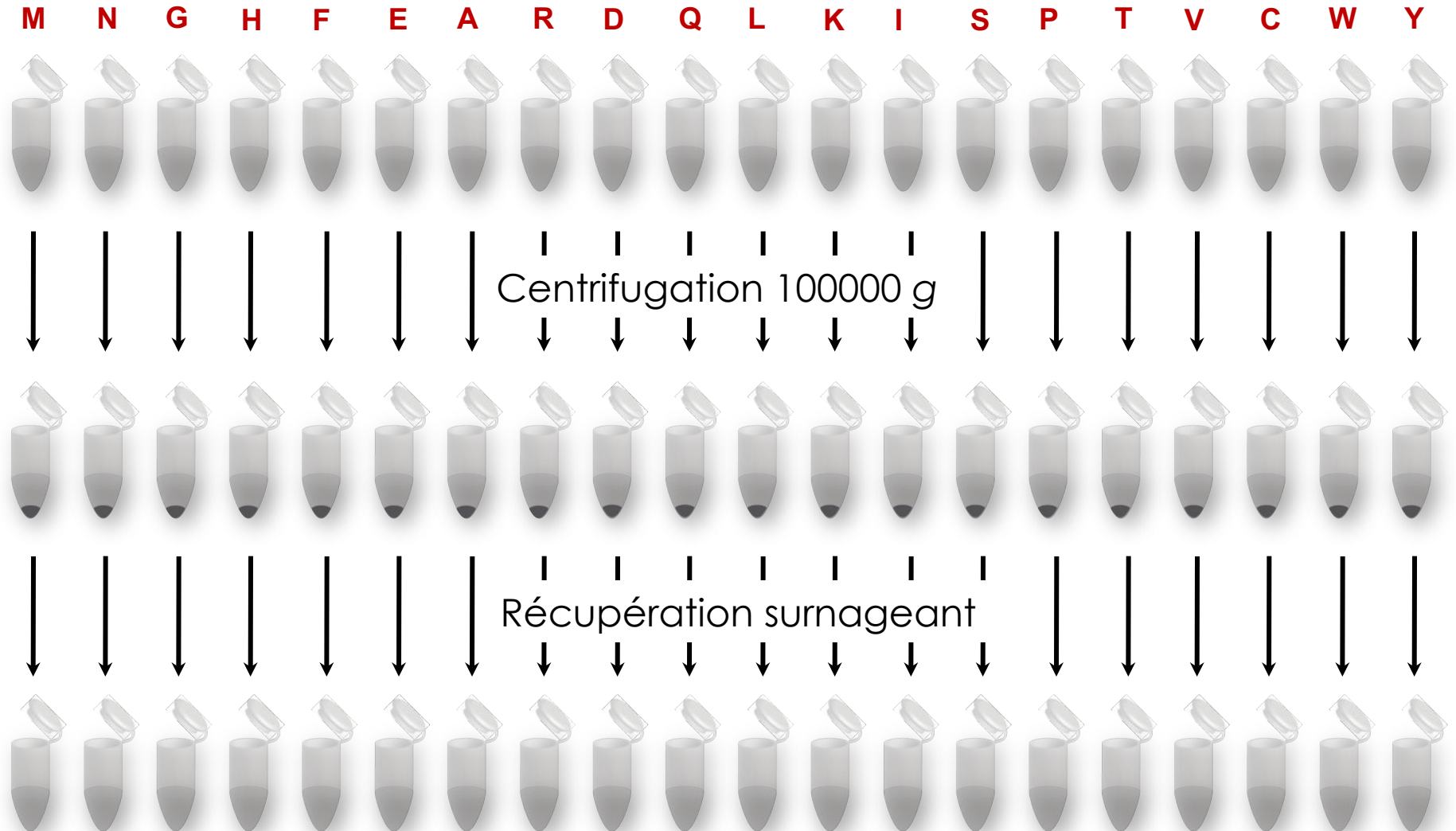
+



II. The Genetic Code

A. Deciphering & Degeneracy

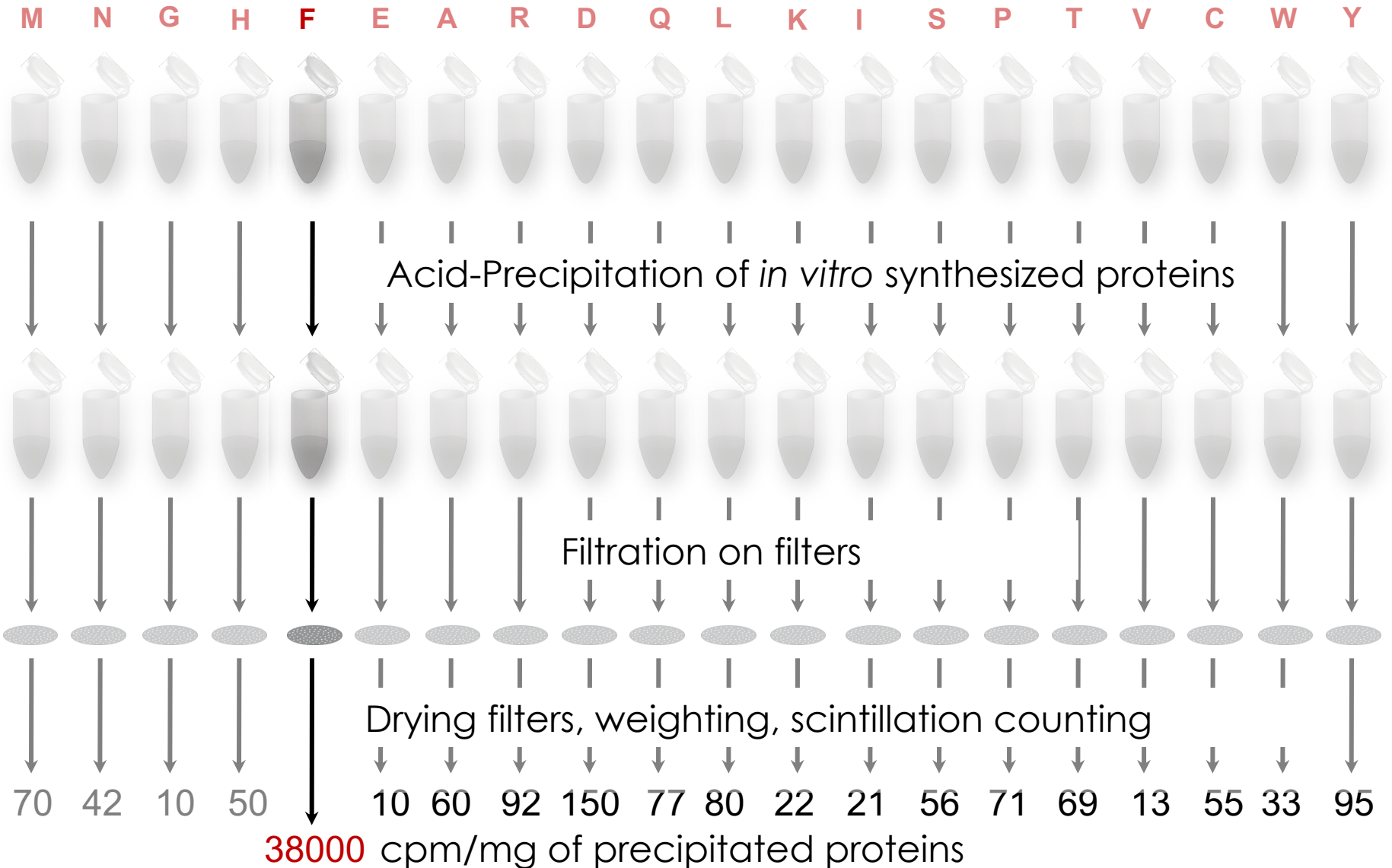
♦ Deciphering



II. The Genetic Code

A. Deciphering & Degeneracy

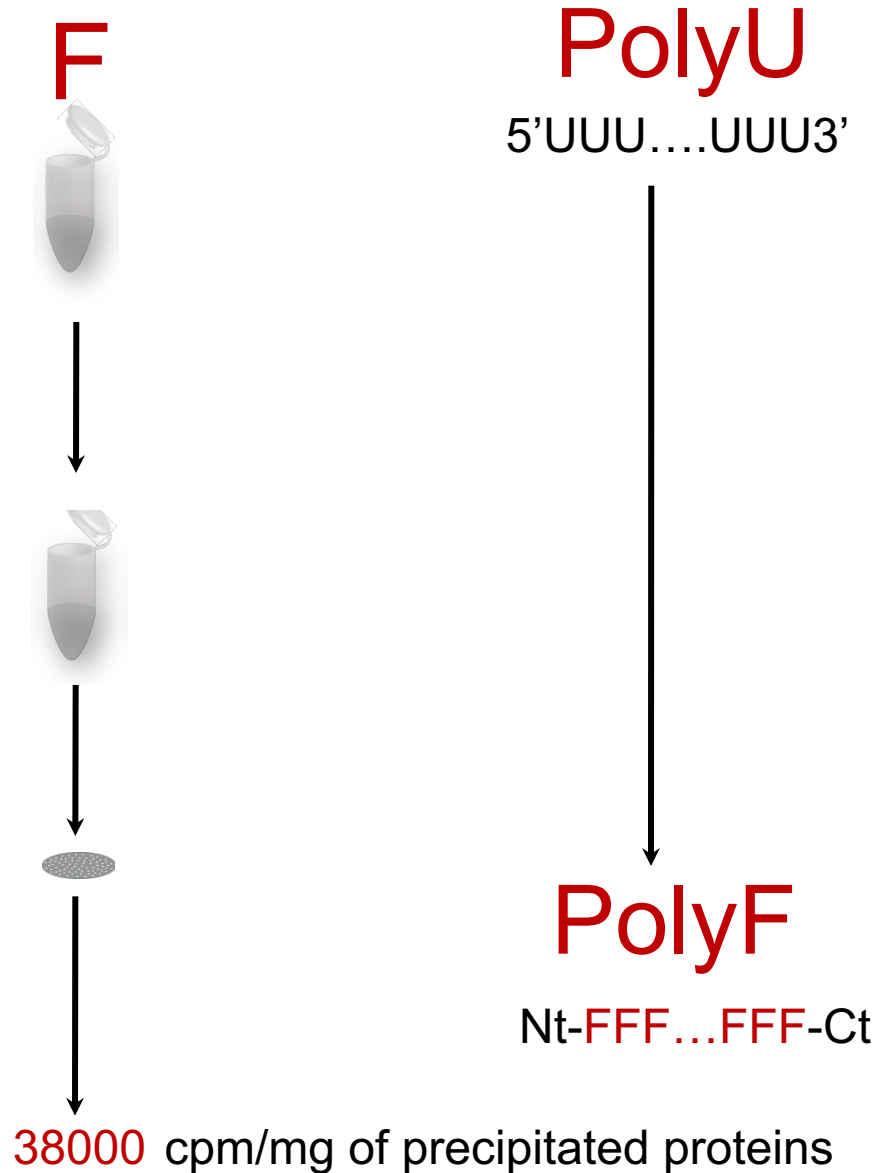
◆ Deciphering



II. The Genetic Code

A. Deciphering & Degeneracy

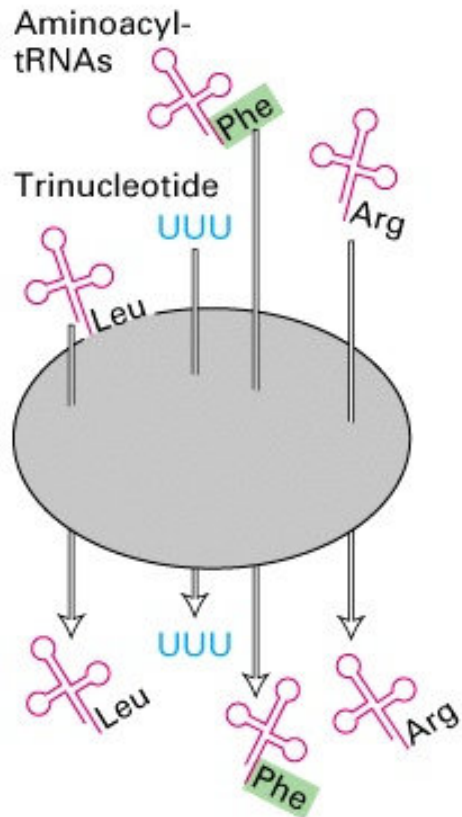
♦ Deciphering



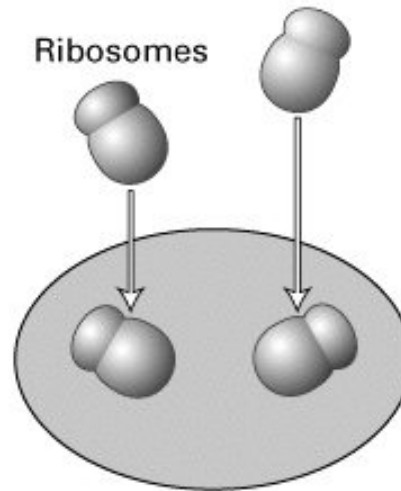
II. The Genetic Code

A. Deciphering & Degeneracy

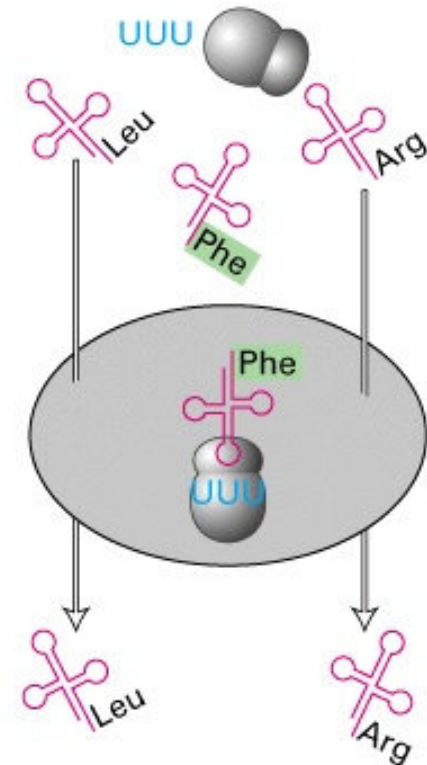
♦ Deciphering



Trinucleotide and all tRNAs pass through filter



Ribosomes stick to filter



Complex of ribosome, UUU, and Phe-tRNA sticks to filter

UUU = Phe

II. The Genetic Code

A. Deciphering & Degeneracy

◆ Deciphering

* Marshall Nirenberg

	C^1	C^2	C^3	H^3	C^4	C^1	C^2	C^3	C^4	C^1	C^2	C^3	C^4	C^1	C^2	C^3	C^4	H^3	C^1	C^2	C^3	C^4
	ALA	ARG	ASP	ASN	CYS	GLU	GLN	GLY	HIS	ILEU	LEU	LYS	MET	VAL	PRO	SER	THR	TRP	TYR	VAL		
UUA	-0.11	-0.30	-0.03	-0.05	-	-0.03	-0.07	-0.03	-0.09	-0.03	-0.24	-0.11	-0.10	1.71	-0.03	-0.26	-0.13	-0.05	0.03	0.01		
UUC	-0.01	-0.24	0.03	-0.10	-	0.01	-0.13	-0.04	-0.04	-0.33	-0.01	-0.03	-0.01	1.59	-0.01	-0.01	-0.01	0.01	0.05	0.02		
UUA	-0.01	-0.10	-0.02	-0.06	-0.01	-0.02	0.03	-	-0.05	-0.02	0.00	-0.03	-0.03	0.00	0.04	-0.03	-0.03	0.05	0.04	0.02		
UUG	-0.20	-0.02	-0.07	-0.05	-	-0.02	-0.08	-0.11	-0.05	0.00	-0.03	-0.06	-0.17	-0.02	-0.07	-0.24	-0.10	-0.05	0.01	0.01		
UUA	0.01	-0.20	0.05	-0.05	-	0.06	-0.18	0.00	0.05	-0.11	-0.07	-0.07	0.02	-0.01	1.37	-0.05	-	-0.02	0.04			
UUC	0.01	-0.32	-0.01	-0.02	-	0.03	-0.16	-0.01	-0.00	-0.06	0.01	-0.03	-0.03	-0.01	0.05	0.03	0.04	0.03	0.01	0.02		
UCA	-0.01	0.08	0.08	0.05	-	0.20	0.17	-	-0.01	-	0.00	-	0.02	-	0.05	0.12	-	-0.02	0.01	-0.06		
UCG	-0.20	0.00	0.00	0.07	-	0.01	-0.23	-0.14	0.01	0.02	0.00	0.04	-0.11	0.03	0.06	1.09	0.01	0.03	0.05	0.03		
UUA	-0.03	-0.11	-0.01	-0.03	-	0.00	-0.11	-0.23	-0.04	0.01	-0.24	-0.09	0.03	-0.37	0.00	-0.03	-0.05	0.01	0.31	0.03		
UAC	-0.01	-0.27	-0.02	0.03	-	-0.02	-0.16	-0.12	-0.04	-0.01	-0.21	-0.06	-0.04	-0.33	0.03	0.00	-0.11	0.02	0.56	0.00		
UAA	-0.03	-0.11	-0.01	-0.02	-	0.02	-0.30	-0.34	-0.08	0.00	0.03	0.10	-0.09	-0.38	0.02	0.03	-0.03	0.00	0.01	0.03		
UAG	-0.07	0.06	0.01	0.12	-	0.04	0.00	-0.01	-0.01	0.02	-0.09	-0.05	-0.02	-0.17	0.00	0.00	-0.15	0.03	0.03	0.01		
UUA	-0.05	-0.08	0.05	0.07	0.03	-0.16	-0.73	-0.01	0.03	-0.08	-0.10	0.02	-0.01	-0.02	-0.02	-0.06	-0.09	0.06	0.02	0.14		
UAC	0.28	-0.18	0.05	0.13	0.04	-0.05	-0.55	-0.03	-0.02	-0.01	0.00	-0.18	0.04	0.03	0.02	-	-0.01	0.03	0.03	0.03		
UUA	-0.12	-0.12	0.07	0.14	-0.10	0.10	-0.13	-0.30	-0.04	-0.11	0.00	-0.37	0.00	0.06	0.06	0.06	-0.08	0.03	0.03	0.05		
UAG	0.04	-0.36	0.01	0.02	-0.01	-0.01	0.28	-	-0.08	-0.03	-0.04	0.00	0.06	0.06	0.14	0.10	-0.08	0.03	0.03	-0.01		
UUA	-0.02	-0.11	0.02	-0.01	-	-0.01	-0.10	-0.22	-0.01	0.00	0.12	0.02	0.05	0.05	-0.01	-0.11	-0.01	0.05	0.04	0.01		
UUC	-0.07	-0.08	-	-	-	-0.03	-0.18	-0.17	0.02	0.01	0.03	0.11	0.05	0.08	-0.16	0.02	-	-	0.03	0.10		
UCA	-0.03	-0.14	0.01	-0.03	0.02	0.00	-0.03	-0.06	-0.07	-0.02	0.03	0.05	0.05	-0.10	-0.18	-0.06	-0.14	-0.04	-	0.02	-0.01	
UUG	-0.14	-0.11	-0.01	0.04	-	0.01	0.11	-1.25	0.03	0.03	0.00	0.03	0.10	0.05	0.03	0.03	-0.06	-0.03	0.01	0.05		
UUA	-0.08	-0.11	-0.01	-0.01	-	-0.02	-0.03	-0.14	-0.03	0.01	-0.11	0.05	0.05	0.03	0.15	-0.11	-0.01	-0.05	0.05	0.02		
UAC	-0.11	-0.01	-0.01	-0.07	-	-0.22	0.03	-0.03	-0.02	0.02	-0.12	0.00	-0.35	0.05	0.00	0.00	0.04	0.02	0.03	0.03		
UCA	-0.15	0.06	-0.03	0.00	-	0.00	0.04	0.01	0.01	0.01	0.02	0.03	0.03	0.10	-0.08	-0.03	0.03	0.03	0.03	0.01		
UUG	-0.08	0.34	0.01	0.05	0.01	-0.02	-0.57	0.06	-0.05	-0.03	-0.02	0.02	0.02	0.01	0.75	-0.04	0.04	-0.03	0.05	0.01		

> Nov III A = 1.56 - 0.132
19. A = 2.10 - 0.144

II. The Genetic Code

A. Deciphering & Degeneracy

♦ Deciphering

Handwritten notes: "Mendel's Nirenberg" and "Ser" (with an arrow pointing to the Ser row in the table).

Vertical label on the left: UCA C G U

Horizontal label on the right: > Nov III A = 1.56 - 2.132
19. A = 2.10 - 2.141

	ALA	ARG	ASP	ASN	CYS	GLU	GLN	GLY	HIS	ILE	LEU	LYS	MET	PHE	PRO	SER	THR	TRP	TYR	VAL
UUA	-0.11	-0.30	-0.03	-0.05	-	-0.03	-0.07	-0.03	-0.09	-0.03	-0.24	-0.11	-0.10	-0.21	-0.03	-0.03	-0.13	-0.05	-0.03	-0.01
UUC	-0.01	-0.24	0.03	-0.10	-	0.01	-0.13	-0.06	-0.01	-0.04	-0.33	-0.04	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.05	-0.02
UUA	-0.01	-0.10	-0.02	-0.06	-0.01	-0.02	0.03	-	-0.05	-0.02	0.00	-0.03	-0.03	-0.04	-0.04	-0.04	-0.03	-0.05	-0.04	-0.02
UUG	-0.20	-0.02	-0.07	-0.05	-	-0.02	-0.08	-0.11	-0.05	0.00	-0.03	-0.06	-0.17	-0.02	-0.07	-0.11	-0.10	-0.05	-0.01	-0.01
UCA	0.01	-0.28	0.05	-0.05	-	0.06	-0.18	0.00	-0.05	-0.11	-0.07	-0.07	0.02	0.01	1.37	-0.01	-	-0.01	-0.04	
UCC	-0.01	-0.33	-0.01	-0.02	-	0.03	-0.16	-0.01	-0.06	-0.09	-0.04	-0.04	-0.04	-0.04	-0.04	-0.04	-0.04	-0.04	-0.04	
UCA	-0.01	0.08	0.08	0.05	-	-	0.20	0.17	-	-0.01	-	0.00	-	0.02	0.02	0.12	-	-0.02	-0.01	
UCG	-0.20	0.00	0.00	0.07	-	0.01	-0.23	-0.44	0.01	0.02	0.00	0.04	-0.11	0.03	0.03	0.09	-0.01	0.05	0.03	
UAU	-0.03	-0.41	-0.21	-0.03	-	0.00	-0.11	-0.23	-0.04	0.01	-0.24	-0.09	0.03	-0.37	0.00	-0.03	-0.05	0.01	0.31	
UAC	-0.01	-0.27	-0.03	0.03	-	-0.02	-0.16	-0.13	-0.04	-0.04	-0.21	-0.06	-0.04	-0.33	0.03	0.00	-0.11	0.02	0.56	
UAA	-0.03	-0.11	-0.01	-0.02	-	0.02	-0.30	-0.34	-0.08	0.00	0.03	0.10	-0.09	-0.38	0.02	0.03	-0.03	0.00	0.01	
UAG	-0.07	0.06	0.01	0.12	-	0.04	0.00	-0.01	-0.01	0.02	-0.09	-0.05	-0.02	-0.17	0.00	0.00	-0.15	0.03	0.03	
UGU	-0.05	-0.03	0.03	0.07	0.03	-0.16	-0.73	-0.01	0.03	-0.03	-0.10	-0.02	-0.01	-0.02	-0.02	-0.02	-0.02	0.02	0.14	
UGC	0.28	-0.18	0.05	0.13	0.04	-0.03	-0.55	-0.03	-0.03	-0.01	0.00	-0.18	0.04	0.03	0.02	-	-0.01	0.03	0.03	
UGA	-0.12	-0.12	0.05	0.14	-0.10	0.10	-0.13	-0.30	-0.03	-0.02	-0.04	-0.11	0.00	-0.37	0.00	0.00	-0.08	0.03	0.03	
UGG	0.04	-0.36	0.01	0.02	-0.03	-0.01	0.28	-	-0.02	-0.03	-0.04	0.00	0.06	0.06	0.14	0.10	-0.08	0.03	-0.01	
CUU	-0.02	-0.11	0.02	-0.01	-	-0.01	-0.10	-0.22	-0.01	0.00	0.12	0.02	0.02	0.00	-0.01	-0.11	-0.01	0.02	0.01	
CUC	-0.07	-0.08	-	-	-	-0.03	-0.18	-0.47	0.02	0.01	0.00	0.11	0.05	0.08	-0.16	0.02	-	0.03	0.10	
CUA	-0.03	-0.14	0.01	-0.03	0.02	0.00	-0.03	-0.66	-0.07	-0.02	0.03	0.00	-0.10	-0.18	-0.06	-0.11	-0.04	-	-0.01	
CUG	-0.14	-0.11	-0.01	0.04	-	0.01	0.11	-1.20	0.03	0.03	0.00	0.03	0.10	0.05	0.03	0.03	-0.06	-0.03	0.05	
GUU	-0.08	-0.11	-0.01	-0.01	-	-0.02	-0.03	-0.14	-0.03	0.01	-0.11	0.05	0.07	0.03	0.15	0.06	-0.07	-0.05	0.02	
GUC	-0.11	-0.01	-0.01	-0.07	-	-0.22	0.03	-0.03	-0.02	0.02	-0.12	0.00	-0.33	0.05	0.00	0.00	-0.04	0.02	0.03	
GUA	-0.15	0.06	-0.03	0.00	-	0.00	0.04	0.01	0.01	0.01	0.02	0.03	0.03	0.10	-0.08	-0.03	-0.03	0.03	0.01	
GUG	-0.08	0.34	0.01	0.05	0.01	-0.02	0.57	0.66	-0.05	-0.03	-0.02	0.02	0.02	0.01	0.75	-0.04	0.04	-0.03	0.01	

II. The Genetic Code

A. Deciphering & Degeneracy

♦ Deciphering

The genetic code has seven main characteristics:

1. It is made up of codons, which are triplets of bases. Each codon specifies a specific amino acid.
2. The codons do not overlap; that is, the sequence GCCAC contains two triplets, "GCC" and "CAC" not counting the "CCC" and other subsequent three-letter sequences.
3. The code includes punctuation in the form of three "stop" codons that do not code for an amino acid: UAA, UAG, and UGA.
4. The genetic code is known as a "degenerate" code. This means that each amino acid is triggered by between one and six codons. (There are only 20 amino acids and 64 possible codon triplets).
5. To read each gene and glean the necessary information to form proteins, cells begin at a fixed and particular starting point on the mRNA strand. The initiation codon is AUG (methionine).
6. The mRNA strand is read from the 5' to the 3' end.
7. If there are mutations or errors in the DNA, the message may be changed and incorrect protein formation results.

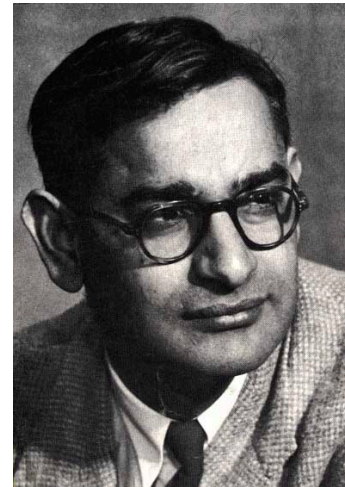
II. The Genetic Code

A. Deciphering & Degeneracy

♦ Deciphering

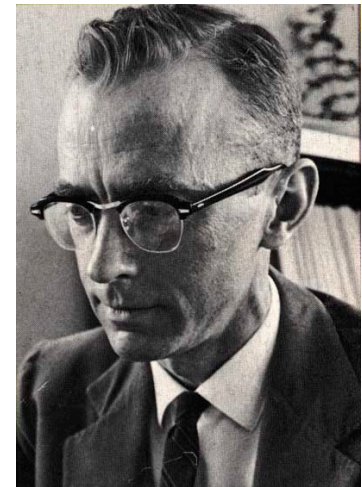


Nobel Prize of Physiology & Medicine
in 1968



[Har Gobind Khorana](#)

RNA synthesis &
deciphering of Leu codons



[Robert W. Holley](#)

tRNA Structure

II. The Genetic Code

A. Deciphering & Degeneracy

♦ Deciphering

CRICK FH, BARNETT L, BRENNER S, WATTS TOBIN RJ. (1961) General nature of the genetic code for proteins. *Nature* **192**:1227-32.

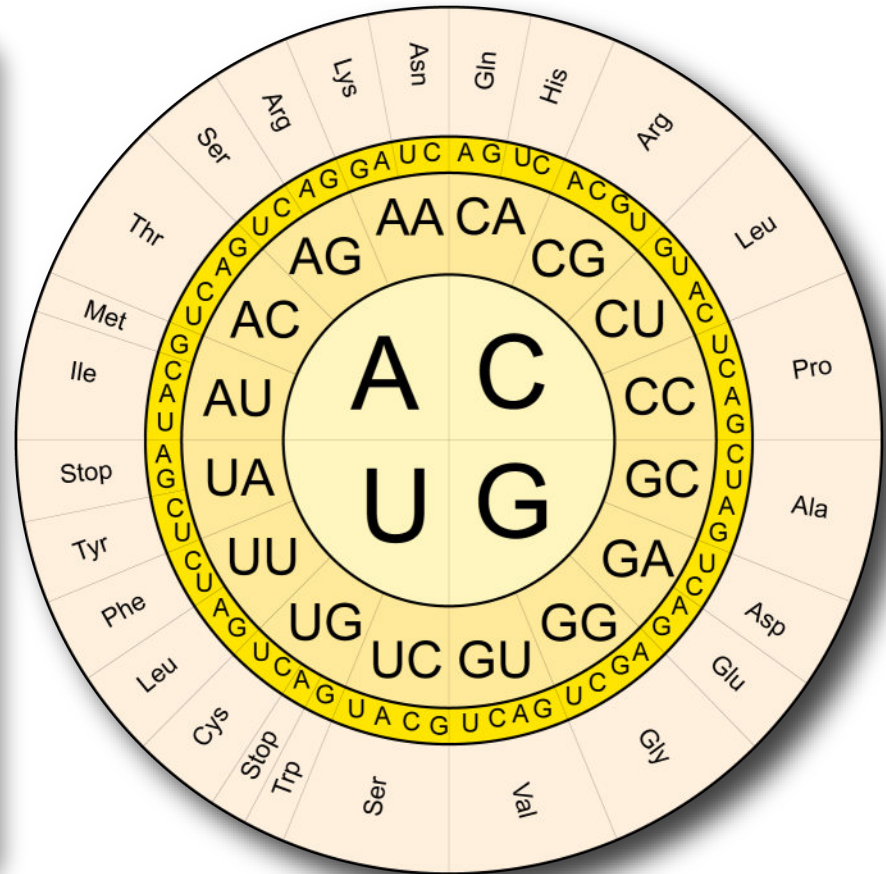
In the experiment, proflavin-induced mutations of the T4 bacteriophage gene, rII_B, were isolated. Proflavin causes mutations by inserting itself between DNA bases, typically resulting in insertion or deletion of a single base pair. The mutants produced by Crick and Brenner could not produce functional rII_B protein because the insertion or deletion of a single nucleotide caused a frameshift mutation. Mutants with two or four nucleotides inserted or deleted were also nonfunctional. However, the mutant strains could be made functional again by using proflavin to insert or delete a total of three nucleotides. This proved that the genetic code uses a codon of three DNA bases that corresponds to an amino acid.

II. The Genetic Code

A. Deciphering & Degeneracy

♦ Deciphering

		Deuxième lettre					
		U	C	A	G		
Première lettre	U	Phe	Ser	Tyr	Cys	Troisième lettre	U
		Phe	Ser	Tyr	Cys		C
		Leu	Ser	Stop	Stop		A
		Leu	Ser	Stop	Trp		G
	C	Leu	Pro	His	Arg		U
		Leu	Pro	His	Arg		C
		Leu	Pro	Gln	Arg		A
		Leu	Pro	Gln	Arg		G
	A	Ile	Thr	Asn	Ser		U
		Ile	Thr	Asn	Ser		C
		Ile	Thr	Lys	Arg		A
		Met	Thr	Lys	Arg		G
	G	Val	Ala	Asp	Gly		U
		Val	Ala	Asp	Gly		C
		Val	Ala	Glu	Gly		A
		Val	Ala	Glu	Gly		G



(Unused letters for the one-letter GC : BJOXZ)

II. The Genetic Code

A. Deciphering & Degeneracy

♦ Deciphering

		Deuxième lettre					
		U	C	A	G		
Première lettre	U	Phe	Ser	Tyr	Cys	U	Troisième lettre
		Phe	Ser	Tyr	Cys	C	
		Leu	Ser	Stop	Stop	A	
		Leu	Ser	Stop	Trp	G	
	C	Leu	Pro	His	Arg	U	
		Leu	Pro	His	Arg	C	
		Leu	Pro	Gln	Arg	A	
		Leu	Pro	Gln	Arg	G	
	A	Ile	Thr	Asn	Ser	U	
		Ile	Thr	Asn	Ser	C	
		Ile	Thr	Lys	Arg	A	
		Met	Thr	Lys	Arg	G	
	G	Val	Ala	Asp	Gly	U	
		Val	Ala	Asp	Gly	C	
		Val	Ala	Glu	Gly	A	
		Val	Ala	Glu	Gly	G	



An aa can be encoded by multiple codons

There are **61** codons to encode 20 natural aa

Most aa is encoded by more than 1 codon

Only **M** and **W** are encoded by 1 single codon

When several codons code for the same aa, the first 2 Nt are invariant and it is the 3rd Nt that varies



Except when the aa are encoded by 6 codons: **L, R and S**

I is the only aa encoded by 3 codons

There are 3 so-called non-coding or STOP codons

UAA (Ochre)

UAG (Amber)

UGA (Opal)

II. The Genetic Code

A. Deciphering & Degeneracy

♦ Deciphering



There are 22 genetically encoded natural AAs:
21st = Selenocysteine (Sec) = (U)
22nd = Pyrrolysine (Pyl)

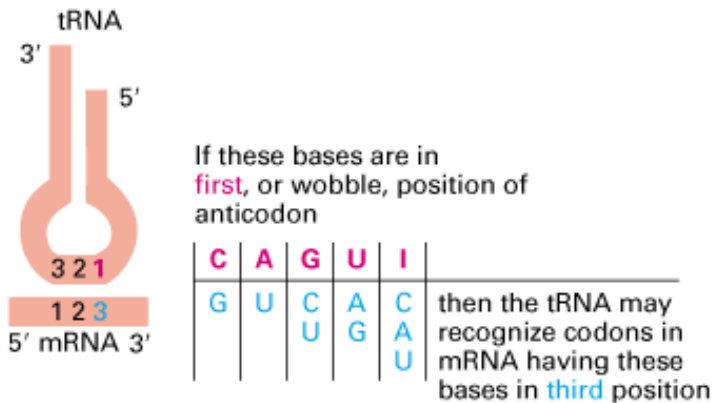


Some STOP codons can be coding

II. The Genetic Code

B. The Wobble pairing

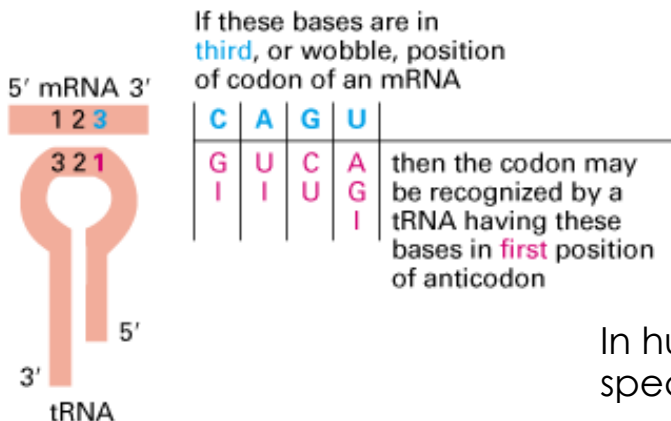
⚠ Since an aa can be encoded by several codons, **are there as many tRNAs as codons? NO**. In prokaryotes there are 30 species of tRNAs to decode 61 codons and in eukaryotes 50 (48 in humans). There is therefore not enough tRNAs. So for some tRNAs the same tRNA must be able to read several codons. This is made possible thanks to **Wobble pairing**



The codon/@codon association is done in an **anti-parallel way**

A single tRNA can read several codons and can **overcome the degeneracy** of the codon (at most 1 tRNA can read 3 ≠ codons)

Some aa are encoded by several codons, so there are up to 4 ≠ tRNAs that decode the same aa we speak of **isoacceptor tRNAs**

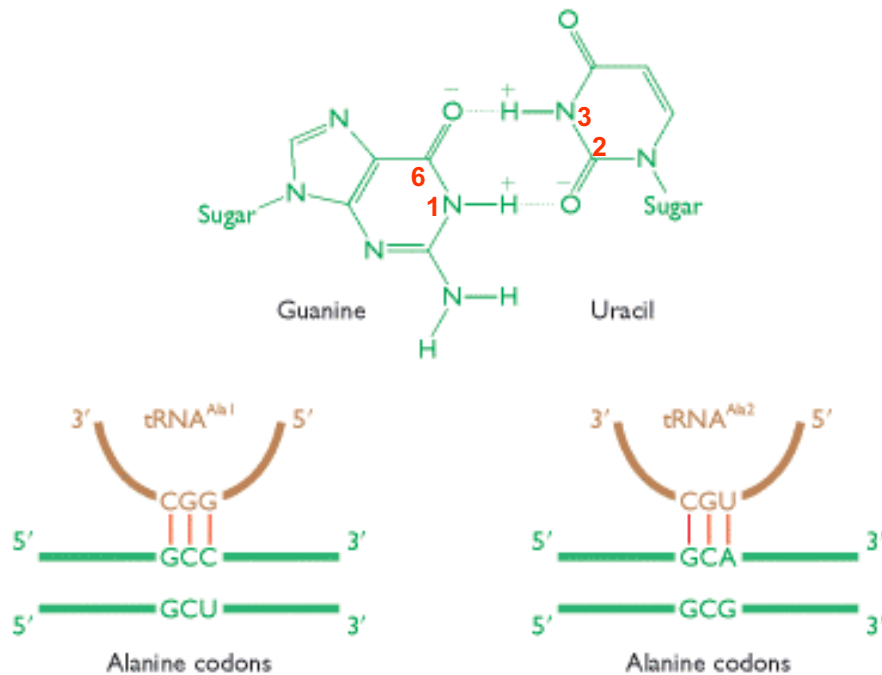


In humans, 16 tRNAs use the wobble while the remaining 32 are specific to a single triplet

II. The Genetic Code

B. The Wobble pairing

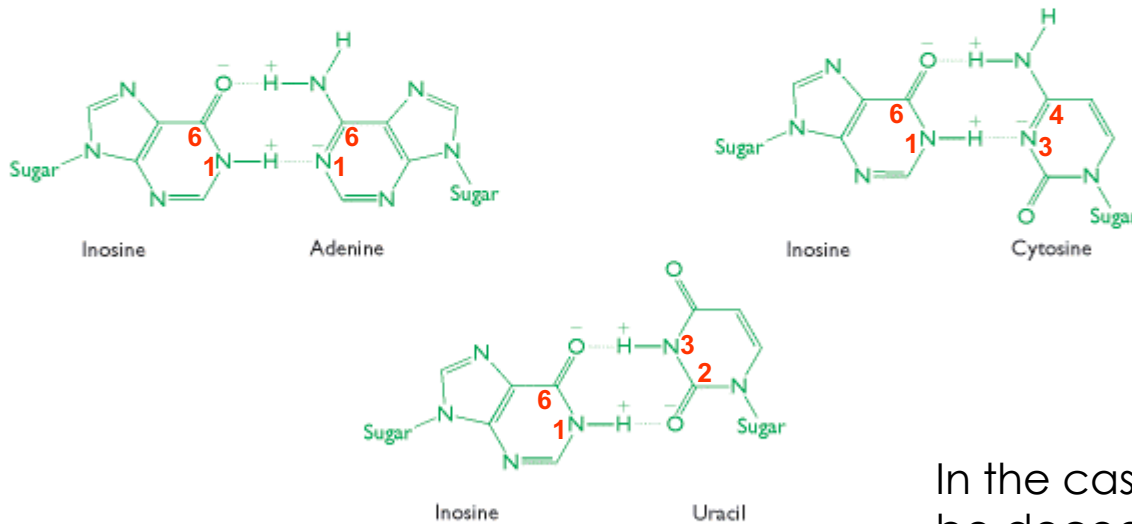
- Wobble pairing is ensured by non-Watson-Crick (WC) base pair (bp) pairings
Ex: Ala is encoded by 4 ≠ codons: GC(U,C,A,G). One of the tRNA^{Ala} has a GGC @codon capable of recognizing GCC and GGU thanks to a non-WC U=G wobble pairing.



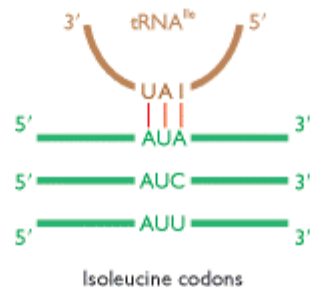
II. The Genetic Code

B. The Wobble pairing

- Nt **Inosine**, which is a precursor of purine bases (A with C=O at position 6), is able to **pair with 3 ≠ bases: A, C, U**. A tRNA with an **I** in the 1st position of the anticodon can therefore recognize 3 ≠ codons. In the case of an aa encoded by 4 codons, 3 can be decoded by a single tRNA → energy saving



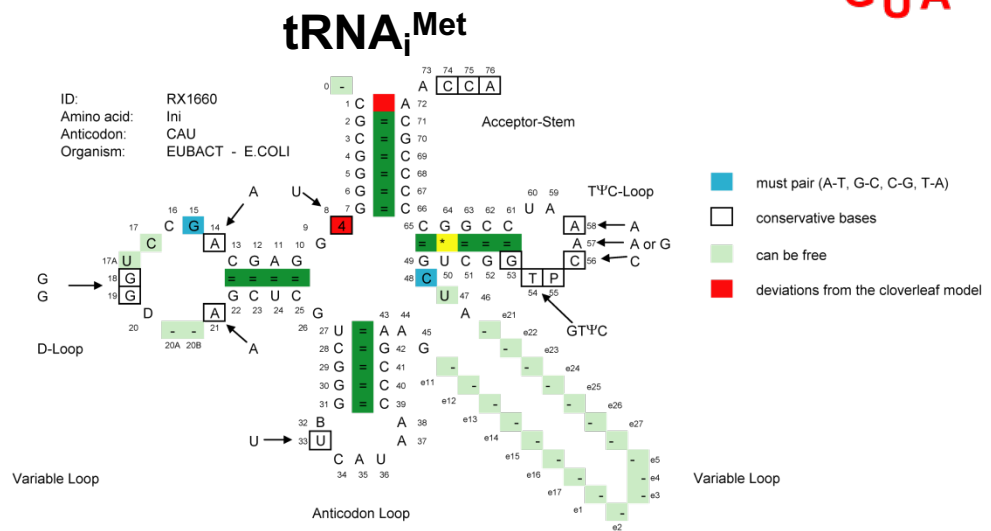
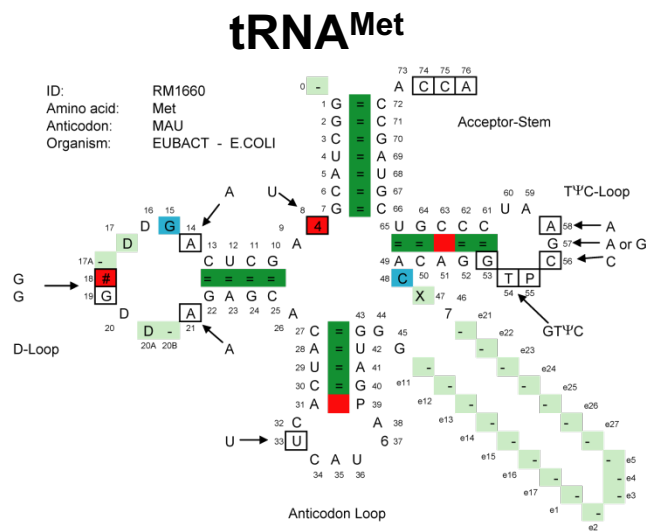
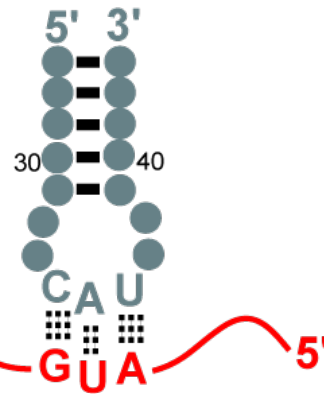
In the case of Ile, the 3 codons can be decoded by a single tRNA.



II. The Genetic Code

C. Initiation & STOP codons

- There is only one Met AUG codon that encodes **2 ≠ Met** :
 - The Met which is in **N-ter of all Prots**, in this case the Met codon is called **initiation codon**
 - The Met, which is **intrachain**, in this case the Met codon is called **elongator codon**
- To decode the initiation codon, there is only one @codon **5'CAU3'**
- However, in the cell there are 2 tRNA^{Met}**



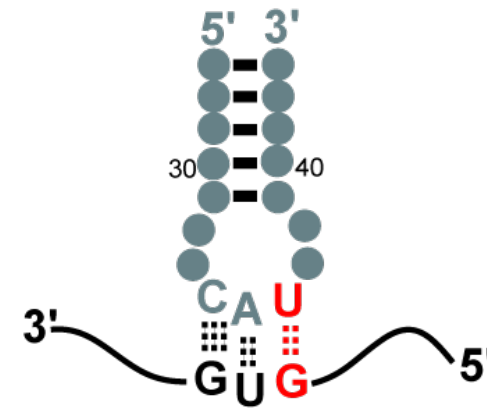
For the 2 tRNA^{Met} the Wobble pairing does not exist, any idea why ?

II. The Genetic Code

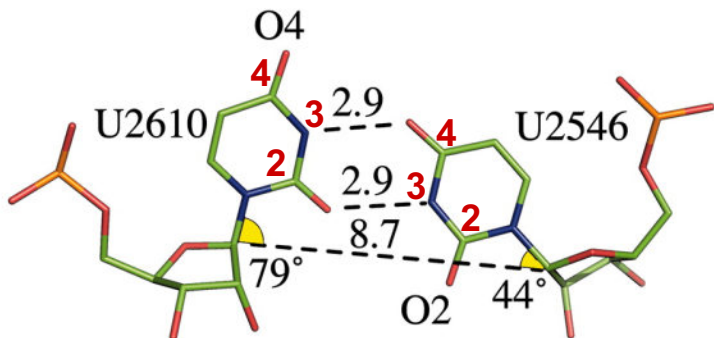
C. Initiation & STOP codons

- **However**, initiation can occur at the level of a **GUG codon instead of AUG** and sometimes even at the level of a **UUG codon**, but always by the initiator **tRNA^{Met}**

		Deuxième lettre				
		U	C	A	G	
Première lettre	U	Phe	Ser	Tyr	Cys	U
		Phe	Ser	Tyr	Cys	C
		Leu	Ser	Stop	Stop	A
		Leu	Ser	Stop	Trp	G
	C	Leu	Pro	His	Arg	U
		Leu	Pro	His	Arg	C
		Leu	Pro	Gln	Arg	A
		Leu	Pro	Gln	Arg	G
	A	Ile	Thr	Asn	Ser	U
		Ile	Thr	Asn	Ser	C
		Ile	Thr	Lys	Arg	A
		Met	Thr	Lys	Arg	G
	G	Val	Ala	Asp	Gly	U
		Val	Ala	Asp	Gly	C
		Val	Ala	Glu	Gly	A
		Val	Ala	Glu	Gly	G



⚠ For the recognition of the GUG codon, a G=U pair (but not wobble pairing) is required



Initiation efficiency: AUG > GUG > UUG (6:2:1)

2 carbonyl (4 & 2)-amino(3) pairings

II. The Genetic Code

C. Initiation & STOP codons

- In all organisms there are 3 STOP codons that are not recognized by any aa-tRNA (except for code deviation and **redefinition of a STOP codon**)

		Deuxième lettre					
		U	C	A	G		
Première lettre	U	Phe	Ser	Tyr	Cys	U	Troisième lettre
		Phe	Ser	Tyr	Cys	C	
		Leu	Ser	Stop	Stop	A	
		Leu	Ser	Stop	Trp	G	
	C	Leu	Pro	His	Arg	U	
		Leu	Pro	His	Arg	C	
		Leu	Pro	Gln	Arg	A	
		Leu	Pro	Gln	Arg	G	
	A	Ile	Thr	Asn	Ser	U	
		Ile	Thr	Asn	Ser	C	
		Ile	Thr	Lys	Arg	A	
		Met	Thr	Lys	Arg	G	
	G	Val	Ala	Asp	Gly	U	
		Val	Ala	Asp	Gly	C	
		Val	Ala	Glu	Gly	A	
		Val	Ala	Glu	Gly	G	

UAA: Ochre)

UGA: Opal)

UAG: Amber)

II. The Genetic Code

D. General information on the reprogramming and dynamics of the genetic code

- The GI can be modified at two levels:
 - At the DNA level through mutations
 - At the RNA level by post-transcriptional modifications

1. Mutations: Missens, Nonsense, Suppressives

● Missens

- It changes the nature of the codon. E.g., the GGA codon encoding the G, a mutation would be located on the first Nt of the codon, e.g.: GGA→AGA, which encodes the R. This codon will no longer be decoded by the tRNA^{Gly(UCC)} but by the tRNA^{Arg(UCU)}.

There will be incorporation of R instead of G → without consequences or dramatic if aa of the active site of an enzyme (loss of activity or ↓ of activity or ↑ of activity).

● Nonsense

- It is a mutation that transforms a codon specifying an aa into a STOP codon that does not code for any aa. E.g.: if the 3rd position of the UAC codon encoding Y is mutated to G, UAC→UAG = STOP, there will be a stop in protein synthesis and a truncated protein will be obtained. Depending on where the SP stops, the prot can be inactive.

II. The Genetic Code

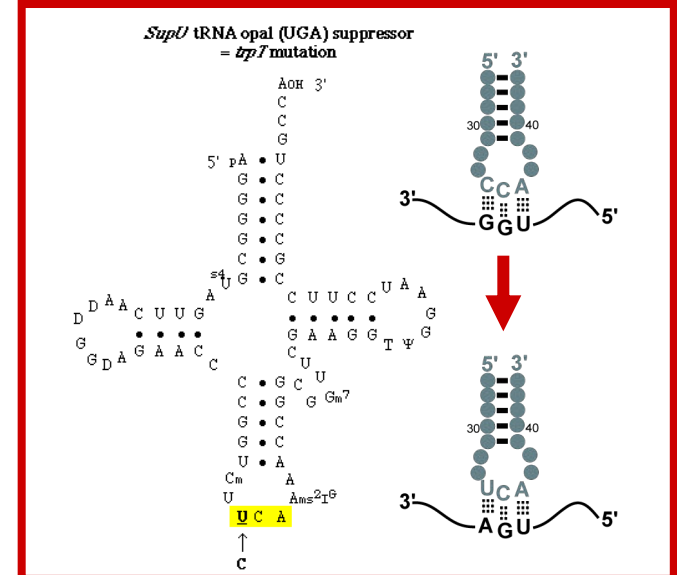
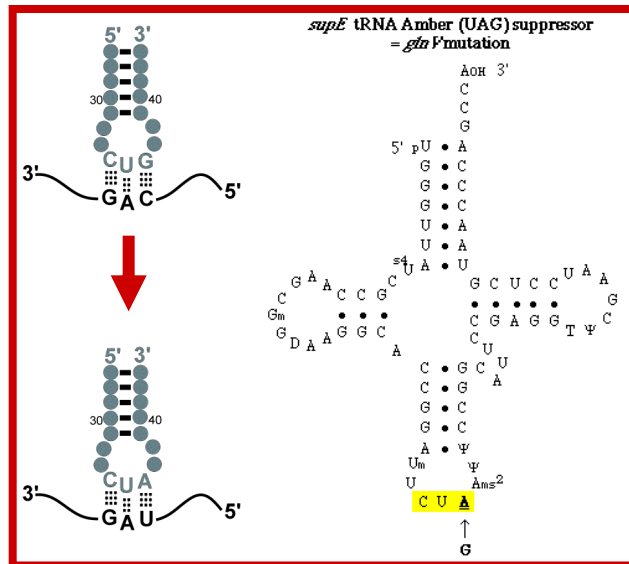
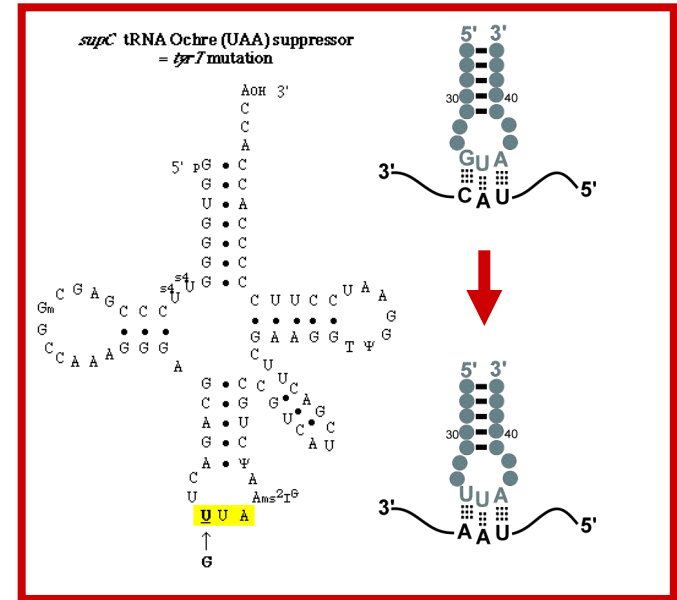
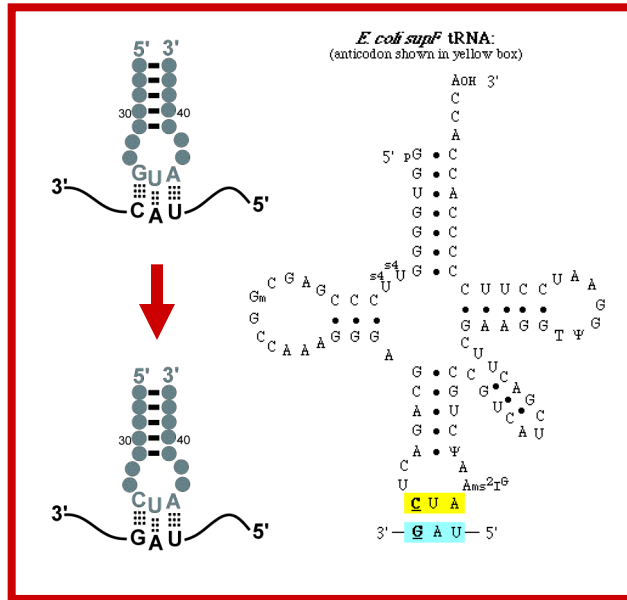
D. General information on the reprogramming and dynamics of the genetic code

● **Suppressive**

- A cell must defend itself against the mutations it undergoes (UV radiation, ionizing radiation, chemical compounds, etc.). One of the ways to reverse mutations is the **suppressive mutation**.
 - It is a mutation on the tRNA that corrects a **missense or nonsense mutation** by returning to the correct reading of the codon, so **it affects the @codon** of the tRNA
 - The tRNA that has been the subject of a suppressive mutation is called a **suppressor tRNA** (in the case of a nonsense mutation)
 - There are natural suppressor tRNAs that the cell has acquired during evolution to compensate these nonsense mutations, e.g. the tRNA^{Tyr} in yeast which exists in small quantities and which allows to correct the UAC→UAG nonsense mutation.
- **The tRNAs do not allow to correct all STOP codons because the natural STOP codons are placed in a particular structural environment** that allows the implementation of the translation termination steps.

II. The Genetic Code

D. General information on the reprogramming and dynamics of the genetic code

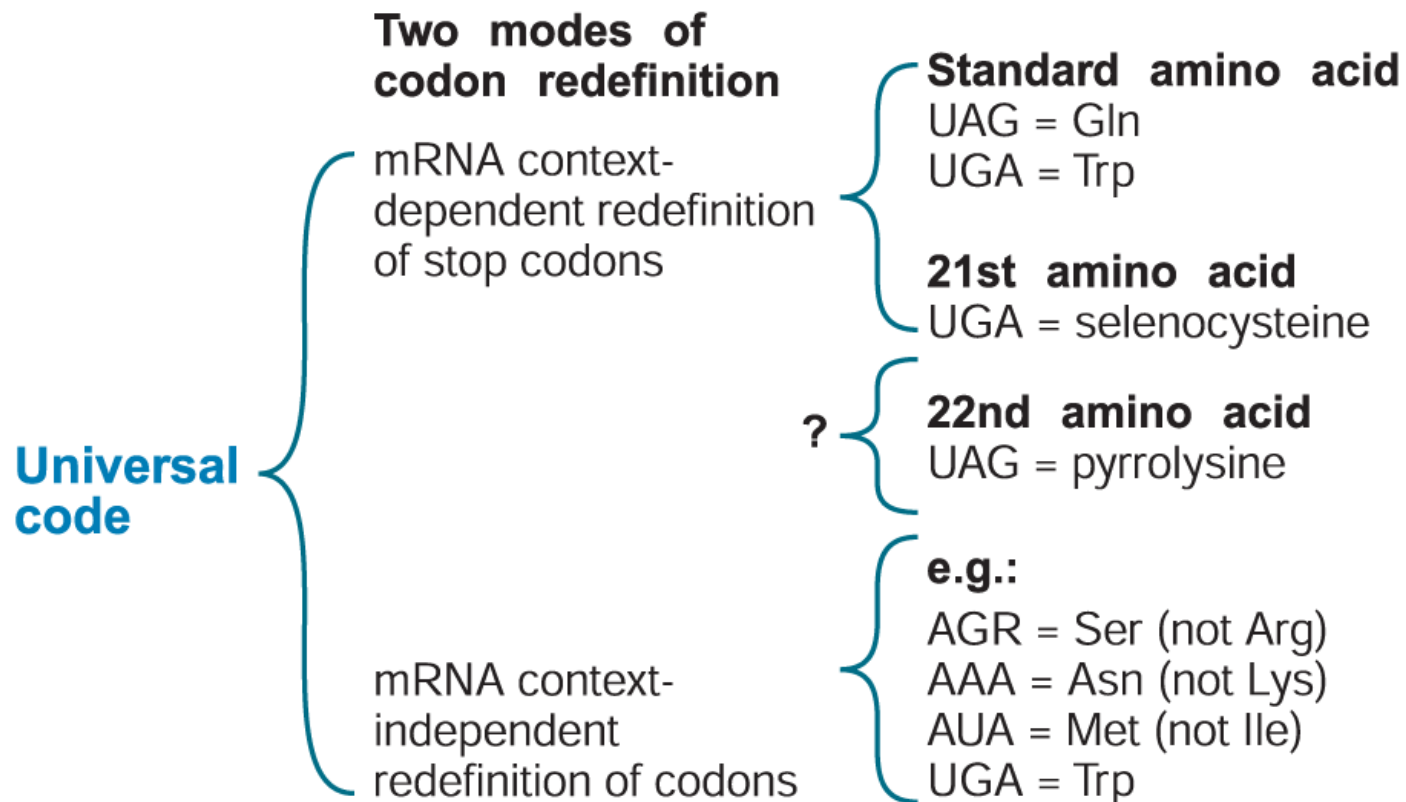


II. The Genetic Code

D. General information on the reprogramming and dynamics of the genetic code

2. Reassignment/reprogramming of STOP codons

- There are STOP codons that are located inside coding regions. Protein synthesis does not stop at these STOP codons because they are decoded by an aa. There are **2 modes of codon reprogramming**

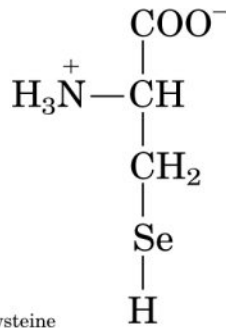
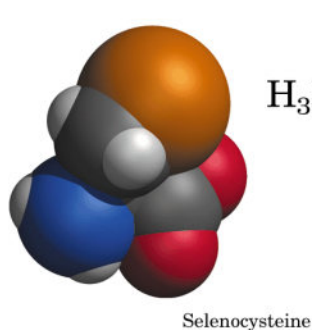


II. The Genetic Code

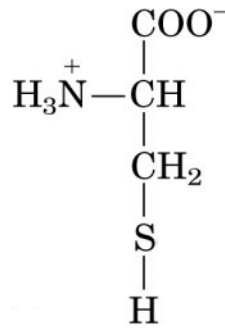
D. General information on the reprogramming and dynamics of the genetic code

2. Reassignment/reprogramming of STOP codons

- At present, 2 aa are encoded by context-dependent reprogramming of the STOP codons, these are the **21st aa of the GC: selenocysteine (Sec)** and the **22nd aa: Pyrrolysine (Pyl)**.

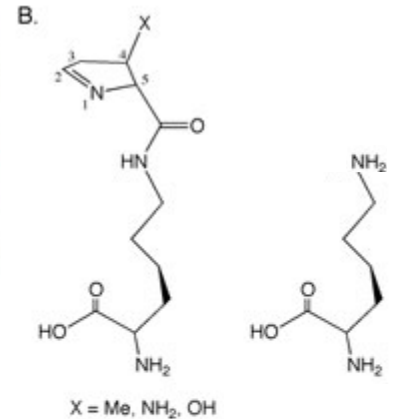
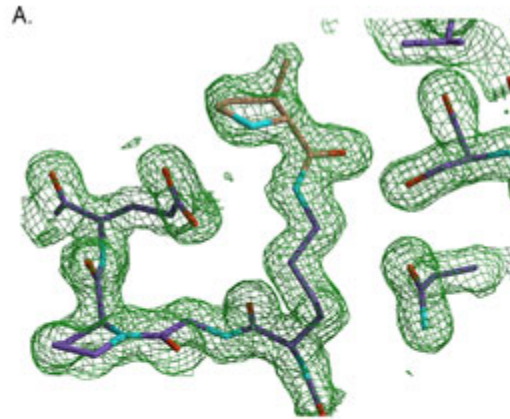


Sec



Cys

UGA (Opal)



Pyl

Lys

UAG (amber)



In this case, the STOP codon is always a STOP **except when the mRNA of a gene has a secondary and/or tertiary structure indicating** to the ribosome that this STOP codon encodes a new aa

II. The Genetic Code

D. General information on the reprogramming and dynamics of the genetic code

2. Reassignment/reprogramming of STOP codons



This reprogramming of the STOP codon = expansion of the genetic code cf repertoire in aa ↑

- Why ↑ the repertoire?

Allows for new chemical groups (gpt) to make new enzymatic reactions

Sec: In active site of many redox enzymes (Ez)

Pyl: In active site of methanogenic archaea methyl-transferases

II. The Genetic Code

D. General information on the reprogramming and dynamics of the genetic code

2. Reassignment/reprogramming of STOP codons

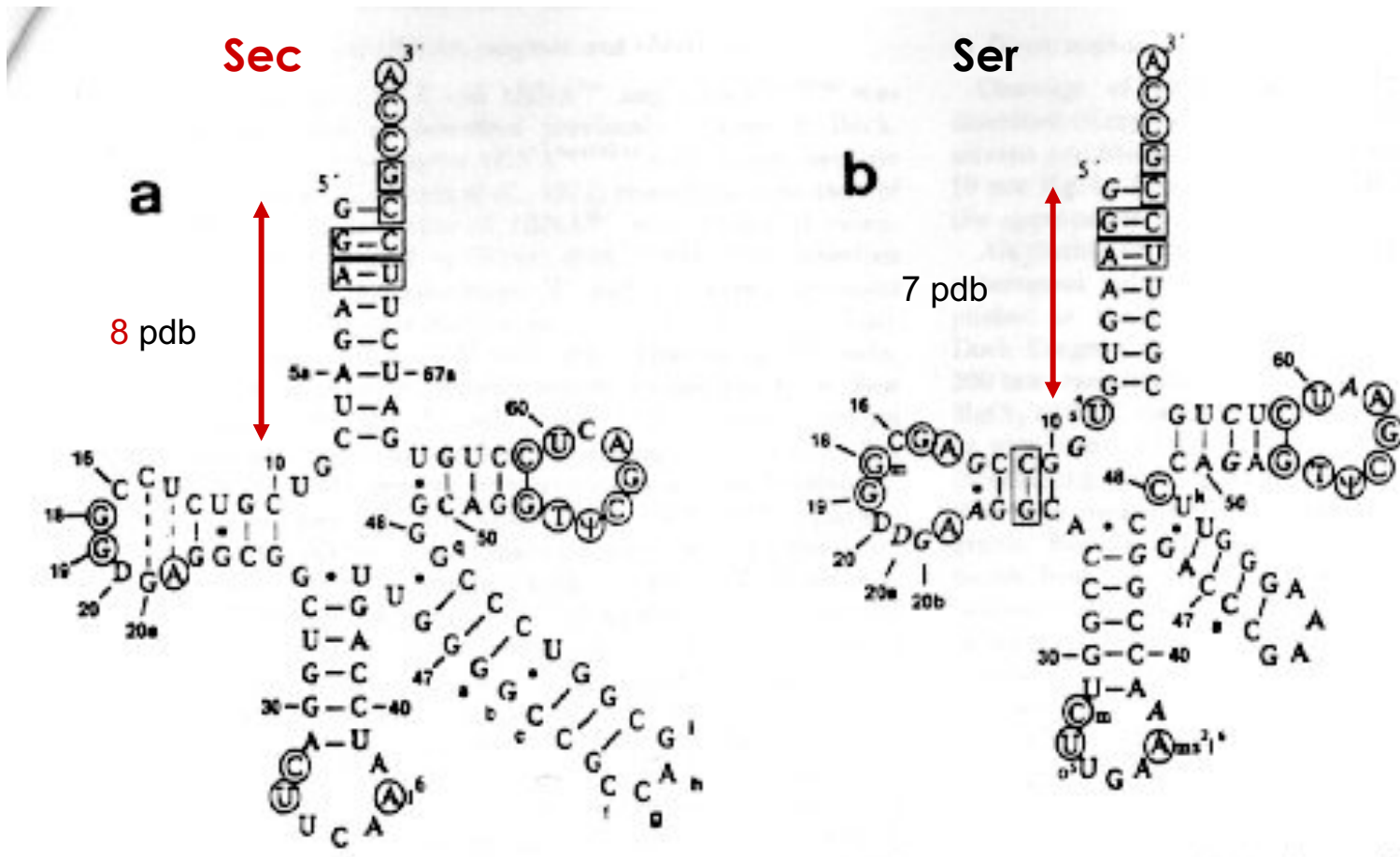
- A **new AA** to be introduced in the GC
- A **suppressor tRNA** that will be specific to this new aa
- An **aminoacyl-tRNA synthetase** (aaRS) loading this new aa onto the tRNA^{Sup}
- A **message on the mRNA** telling the ribosome that it is not a stop codon but that it encodes the new aa
- A specific **elongation factor** of the new aa-tRNA

II. The Genetic Code

D. General information on the reprogramming and dynamics of the genetic code

2. Reassignment/reprogramming of STOP codons: the 21st genetically-encoded aa: Sec

- A suppressor tRNA that will be specific to this new aa



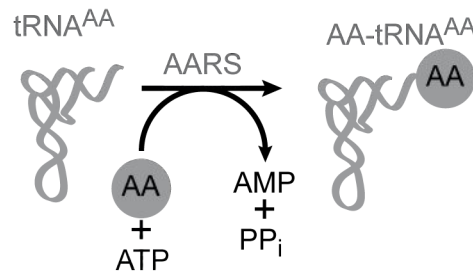
II. The Genetic Code

D. General information on the reprogramming and dynamics of the genetic code

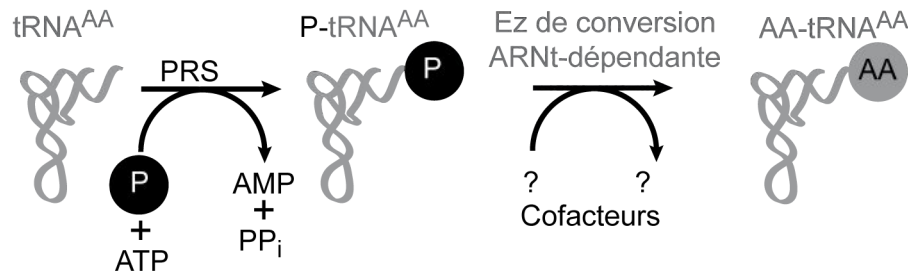
2. Reassignement/reprogramming of STOP codons

- An aminoacyl-tRNA synthetase (aaRS) charging this new aa onto the tRNA^{Sup}

Direct Pathway



Indirect Pathway



P = précurseur métabolique de **AA**

II. The Genetic Code

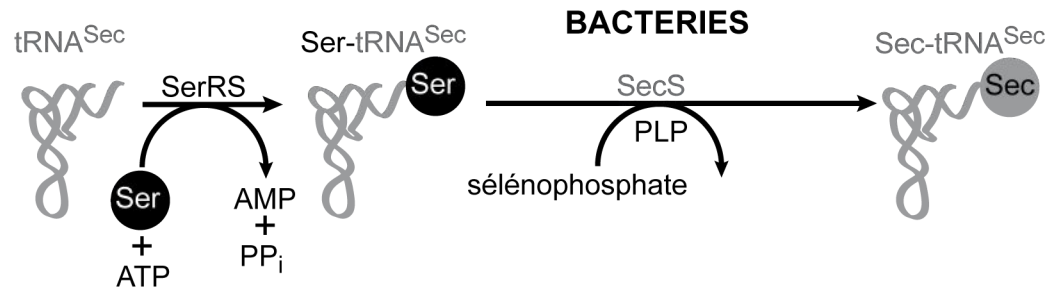
D. General information on the reprogramming and dynamics of the genetic code

2. Reassignment/reprogramming of STOP codons: the 21st genetically-encoded aa: Sec

- An aminoacyl-tRNA synthetase (aaRS) charging this new aa onto the tRNA^{Sup}



There is no selenocysteinyI-tRNA synthetase (SecRS), the synthesis of Sec-tRNA is done by **Ser-tRNA^{Sec}-dependent modification**. The indirect pathway generating Sec-tRNA^{Sec} is specific and different for each kingdom of life



II. The Genetic Code

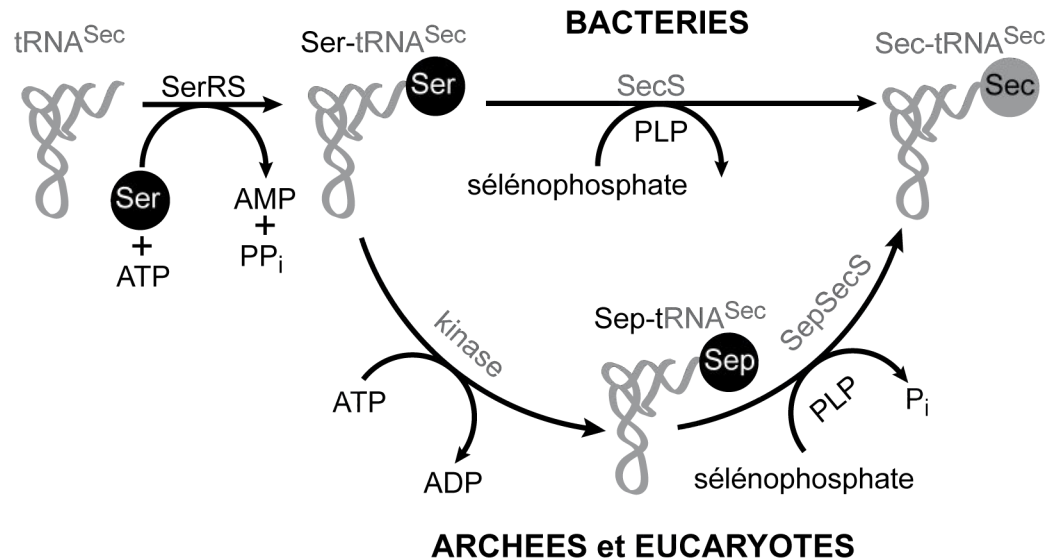
D. General information on the reprogramming and dynamics of the genetic code

2. Reassignment/reprogramming of STOP codons: the 21st genetically-encoded aa: Sec

- An aminoacyl-tRNA synthetase (aaRS) charging this new aa onto the tRNA^{Sup}



There is no selenocysteinyI-tRNA synthetase (SecRS), the synthesis of Sec-tRNA is done by **Ser-tRNA^{Sec}-dependent modification**. The indirect pathway generating Sec-tRNA^{Sec} is specific and different for each kingdom of life

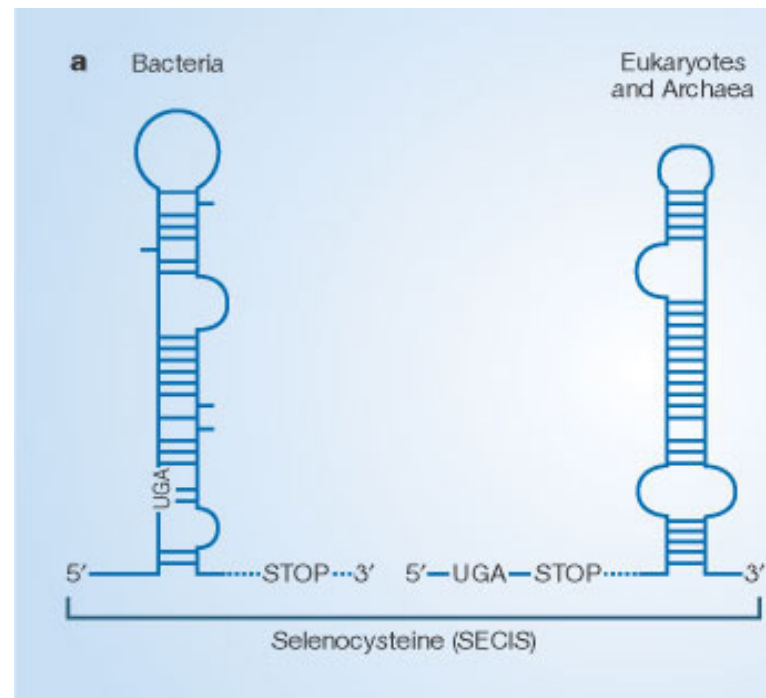
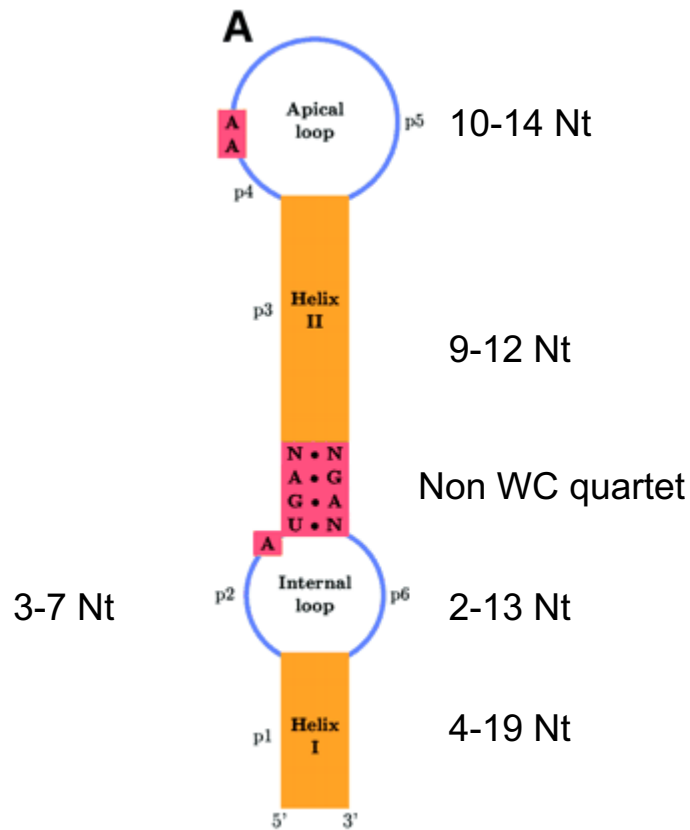


II. The Genetic Code

D. General information on the reprogramming and dynamics of the genetic code

2. Reassignment/reprogramming of STOP codons: the 21st genetically-encoded aa: Sec

- A structural message on the mRNA telling the ribosome that it is not a stop codon but that it encodes the new aa

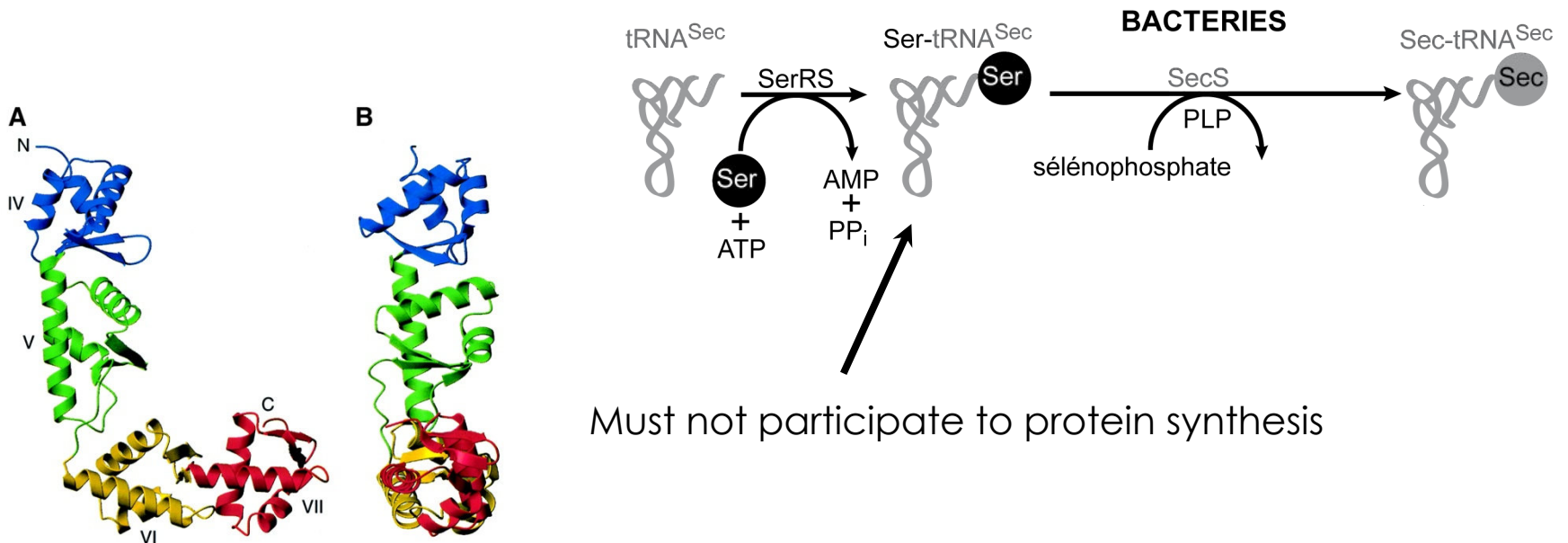


II. The Genetic Code

D. General information on the reprogramming and dynamics of the genetic code

2. Reassignment/reprogramming of STOP codons: the 21st genetically-encoded aa: Sec

- A specific elongation factor of the new aa-tRNA



3D structure of SelB-C (C-terminal part binding to SECIS)

⚠ The addition of Sec to the code repertoire requires a specific elongation factor: **SelB** in Bacteria and Archaea, and **EF-Sec** in eukaryotes

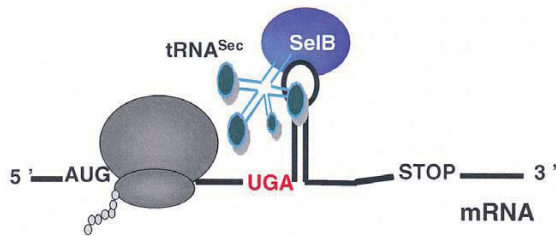
II. The Genetic Code

D. General information on the reprogramming and dynamics of the genetic code

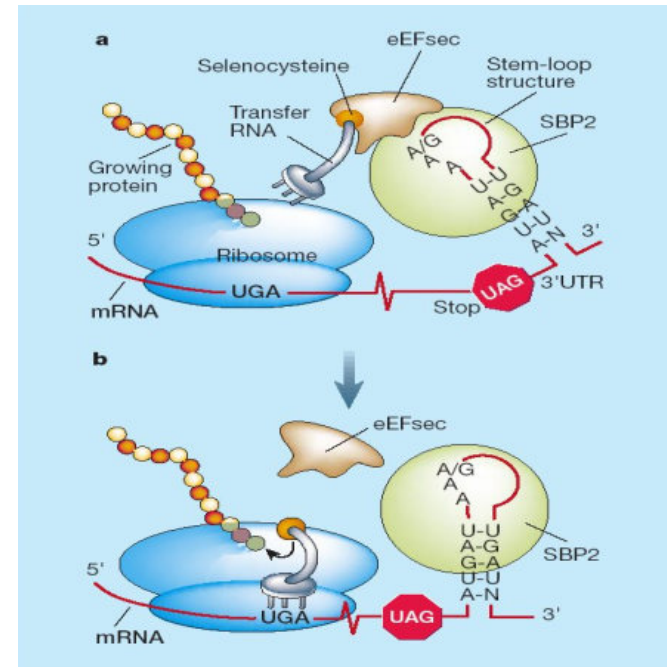
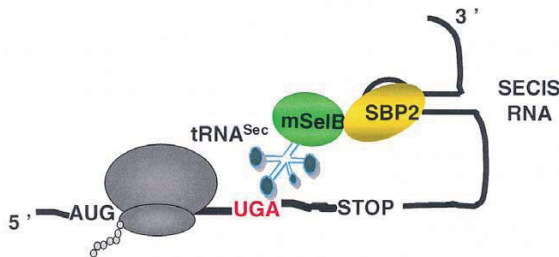
2. Reassignment/reprogramming of STOP codons: the 21st genetically-encoded aa: Sec

- A specific elongation factor of the new aa-tRNA

Eubacteria: One protein, two functions



Eukaryotes: Two proteins, more than two functions ?



Selenoprotein Synthesis



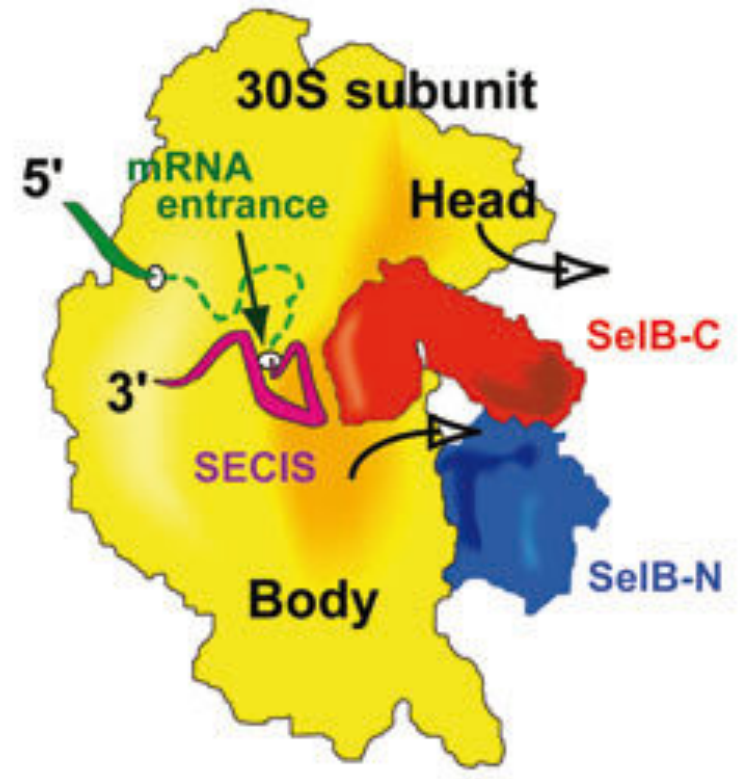
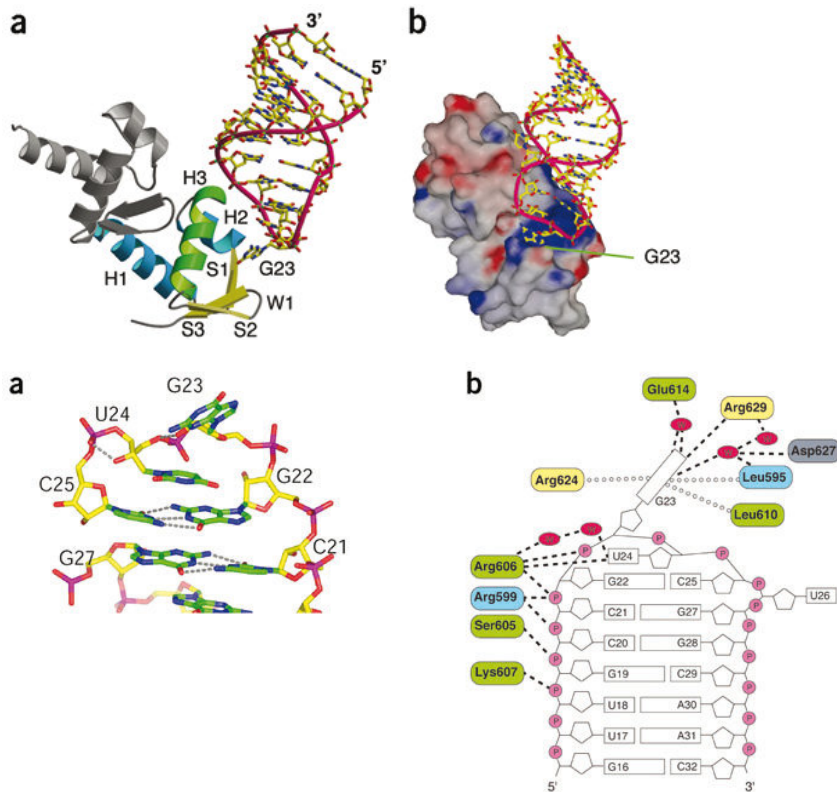
The addition of Sec to the code repertoire requires a specific elongation factor: **SelB** in Bacteria and Archaea, and **EF-Sec** in eukaryotes

II. The Genetic Code

D. General information on the reprogramming and dynamics of the genetic code

2. Reassignment/reprogramming of STOP codons: the 21st genetically-encoded aa: Sec

- A specific elongation factor of the new aa-tRNA



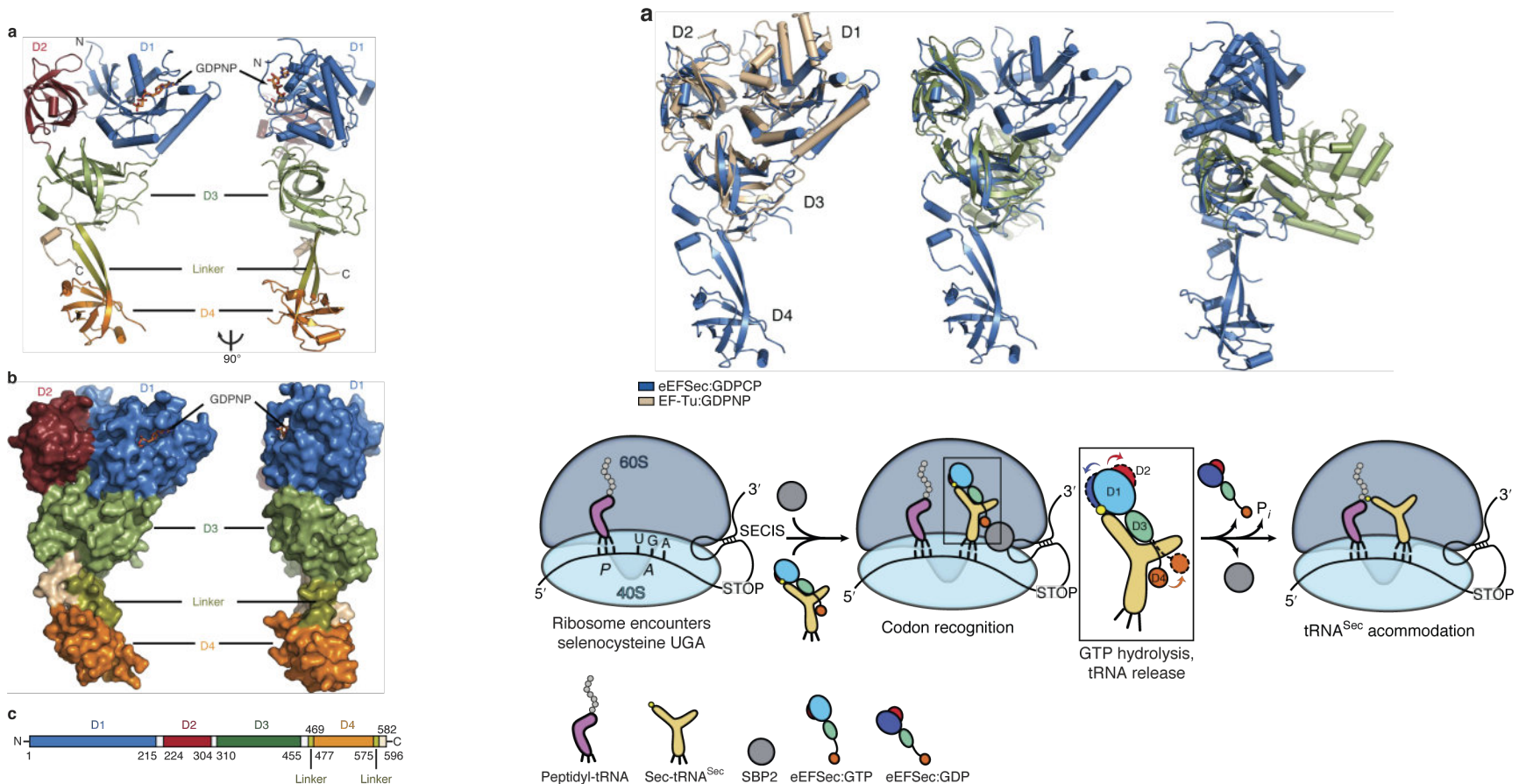
The addition of Sec to the code repertoire requires a specific elongation factor: **SelB** in Bacteria and Archaea, and **EF-Sec** in eukaryotes

II. The Genetic Code

D. General information on the reprogramming and dynamics of the genetic code

2. Reassignment/reprogramming of STOP codons: the 21st genetically-encoded aa: Sec

- A specific elongation factor of the new aa-tRNA: 3D-st *H. sapiens* EF-Sec



II. The Genetic Code

D. General information on the reprogramming and dynamics of the genetic code

2. Reassignment/reprogramming of STOP codons: the 22nd genetically-encoded aa : Pyl

- A **new AA** to be introduced in the GC
- A **suppressor tRNA** that will be specific to this new aa
- An **aminoacyl-tRNA synthetase** (aaRS) loading this new aa onto the tRNA^{Sup}
- A **message on the mRNA** telling the ribosome that it is not a stop codon but that it encodes the new aa
- A specific **elongation factor** of the new aa-tRNA

II. The Genetic Code

D. General information on the reprogramming and dynamics of the genetic code

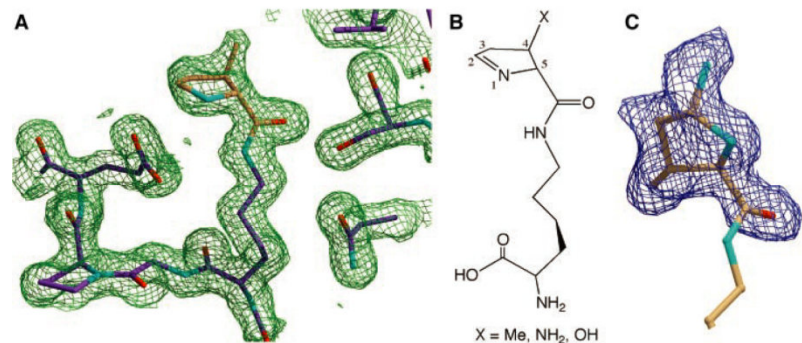
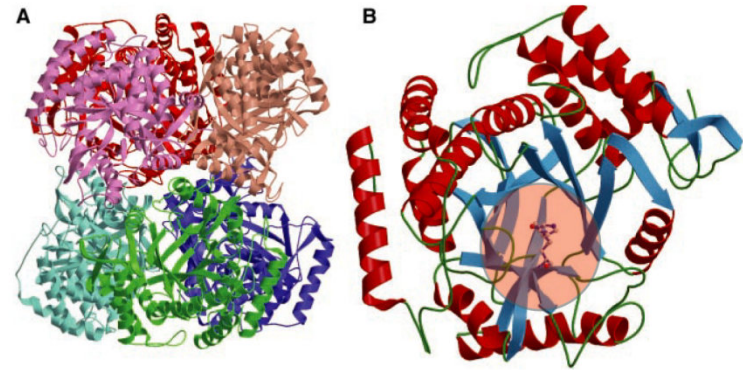
2. Reassignment/reprogramming of STOP codons: the 22nd genetically-encoded aa : Pyl

- A new AA to be introduced in the GC

A New UAG-Encoded Residue in the Structure of a Methanogen Methyltransferase

Bing Hao,¹ Weimin Gong,¹ Tsuneo K. Ferguson,²
Carey M. James,² Joseph A. Krzycki,^{1*} Michael K. Chan^{1*}

24 MAY 2002 VOL 296 SCIENCE



II. The Genetic Code

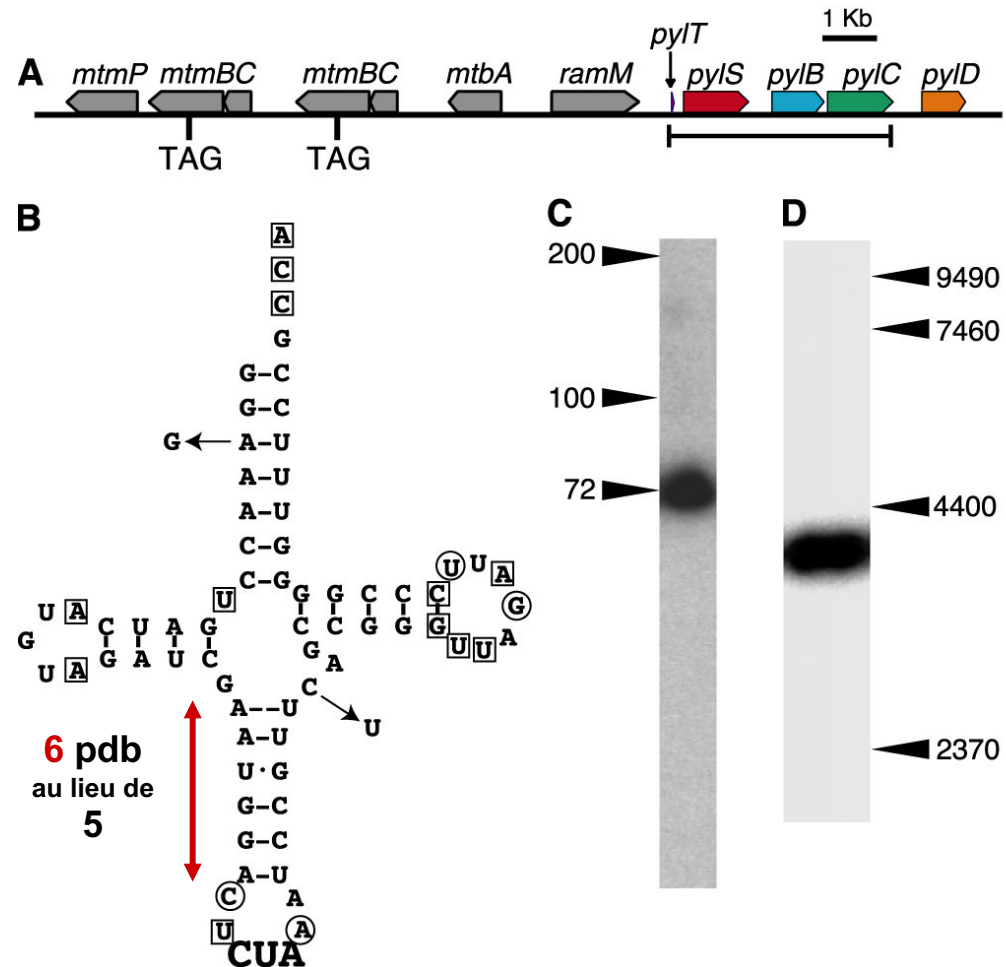
D. General information on the reprogramming and dynamics of the genetic code

2. Reassignment/reprogramming of STOP codons: the 22nd genetically-encoded aa : Pyl

- A suppressor tRNA that will be specific to this new aa

Pyrrolysine Encoded by UAG in
Archaea: Charging of a
UAG-Decoding Specialized tRNA

Gayathri Srinivasan, Carey M. James, Joseph A. Krzycki*
SCIENCE VOL 296 24 MAY 2002



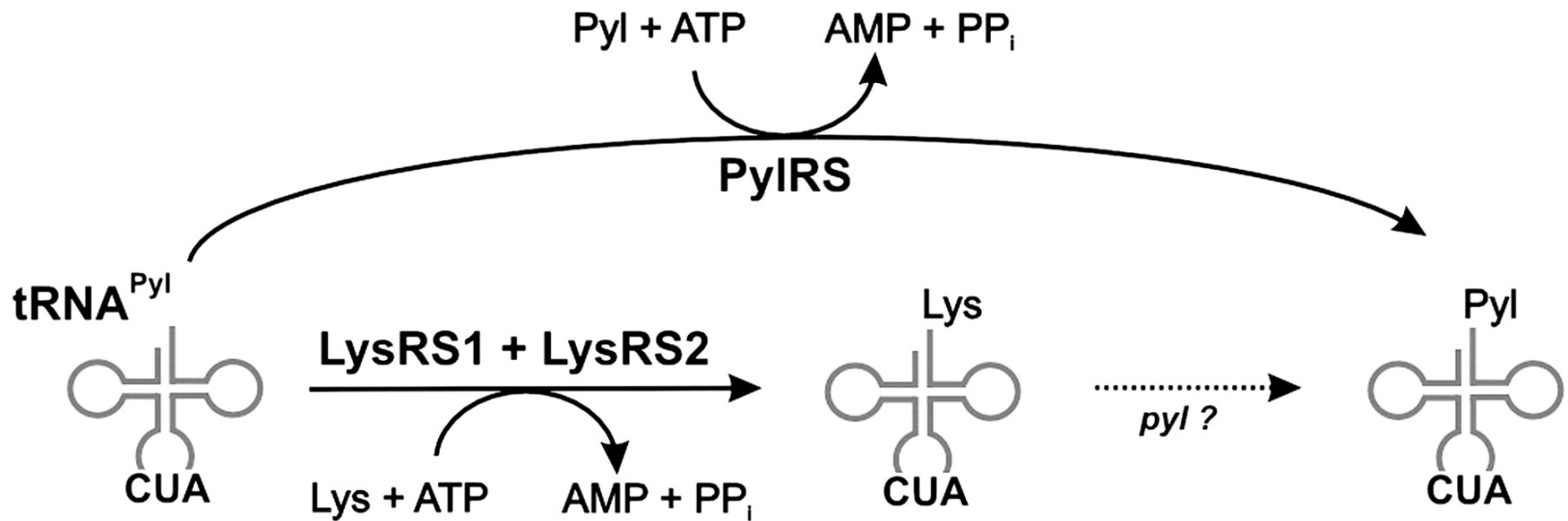
II. The Genetic Code

D. General information on the reprogramming and dynamics of the genetic code

2. Reassignment/reprogramming of STOP codons: the 22nd genetically-encoded aa : Pyl

- An aminoacyl-tRNA synthetase (aaRS) loading this new aa onto the tRNA^{Pyl}

⚠ There is a **pyrrolysyl-tRNA synthetase or PylRS**

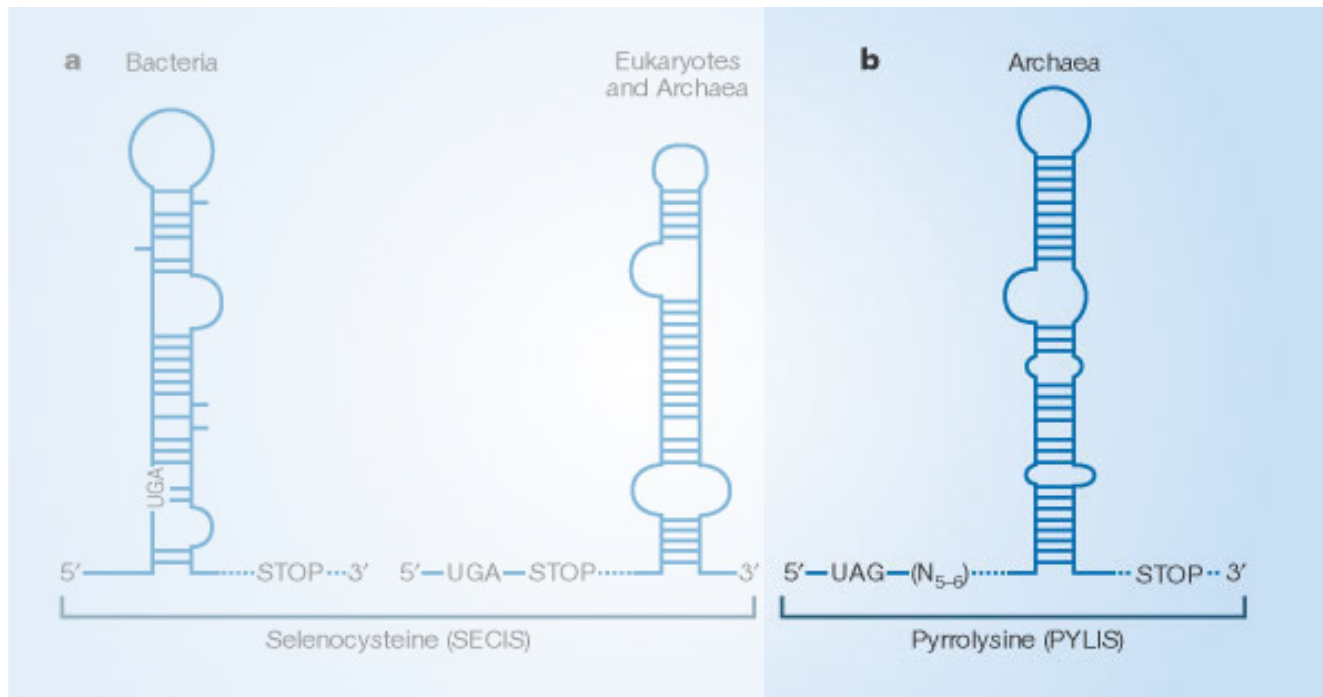


II. The Genetic Code

D. General information on the reprogramming and dynamics of the genetic code

2. Reassignment/reprogramming of STOP codons: the 22nd genetically-encoded aa : Pyl

- A structural message on the mRNA indicating that it is not a stop codon but that it encodes the new aa



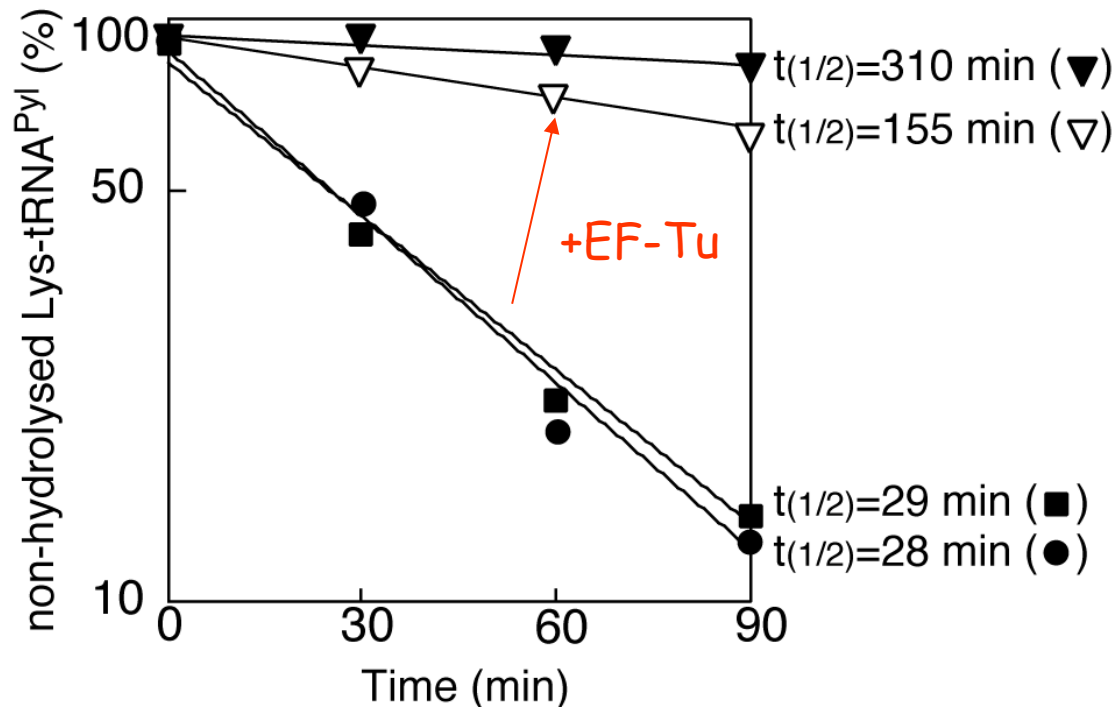
II. The Genetic Code

D. General information on the reprogramming and dynamics of the genetic code

2. Reassignment/reprogramming of STOP codons: the 22nd genetically-encoded aa : Pyl

- A specific elongation factor of the new aa-tRNA

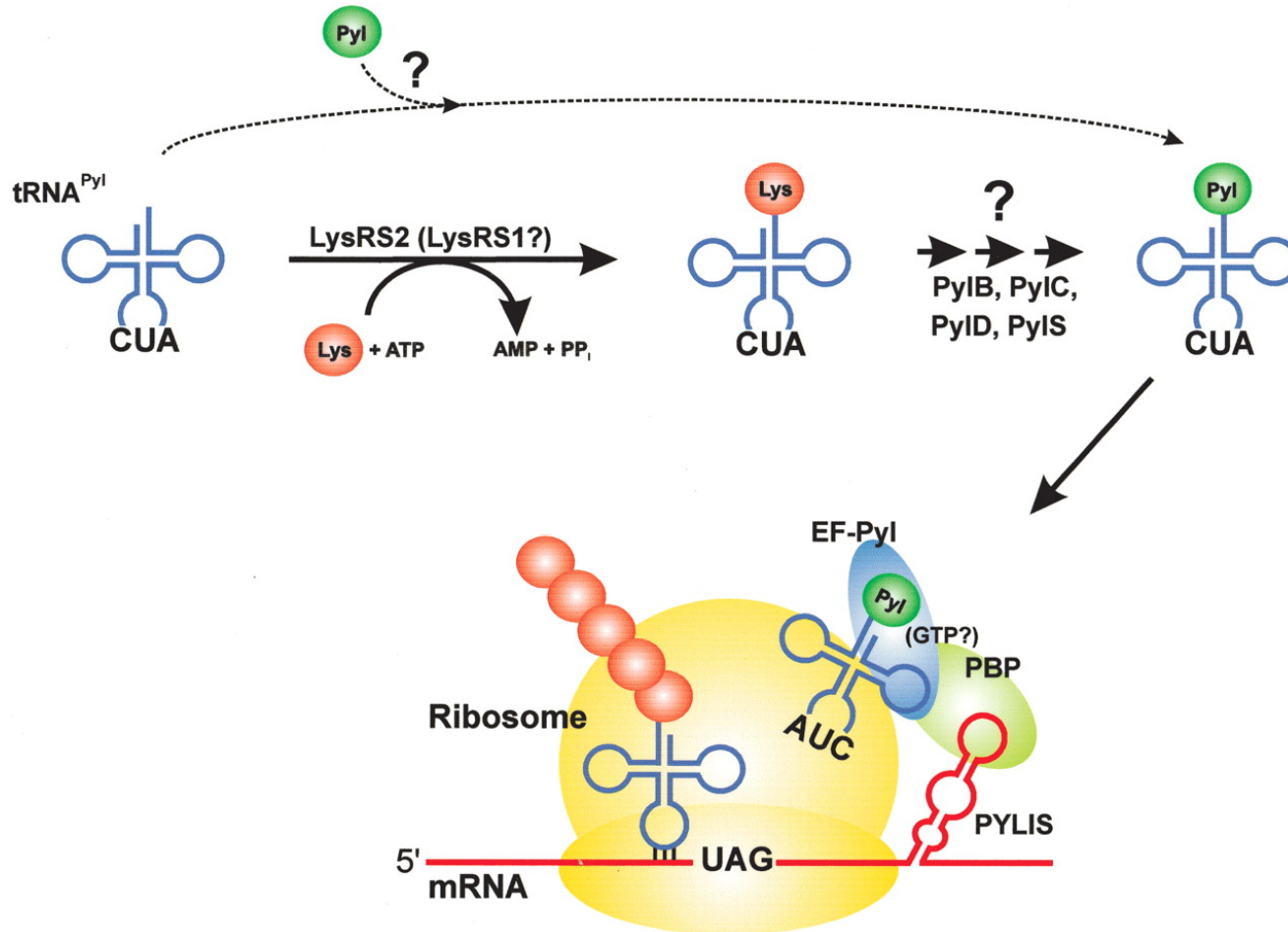
⚠ EF-Tu carries Pyl-tPyl RNA to the ribosome



II. The Genetic Code

D. General information on the reprogramming and dynamics of the genetic code

2. Reassignment/reprogramming of STOP codons: the 22nd genetically-encoded aa : Pyl



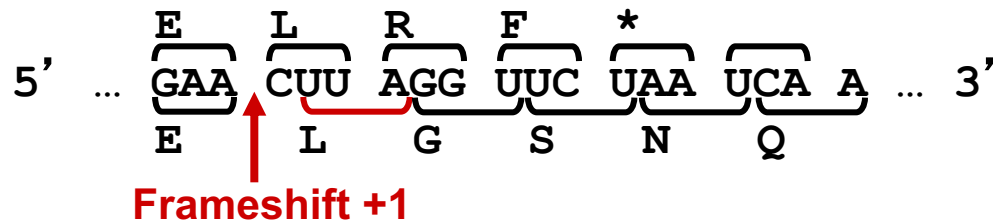
Hypothetical mechanism based on that of Sec

II. The Genetic Code

D. General information on the reprogramming and dynamics of the genetic code

3. Frameshift

- An mRNA is decrypted codon by codon, from +1 to the STOP codon
- +1 or -1 jump error → frameshift



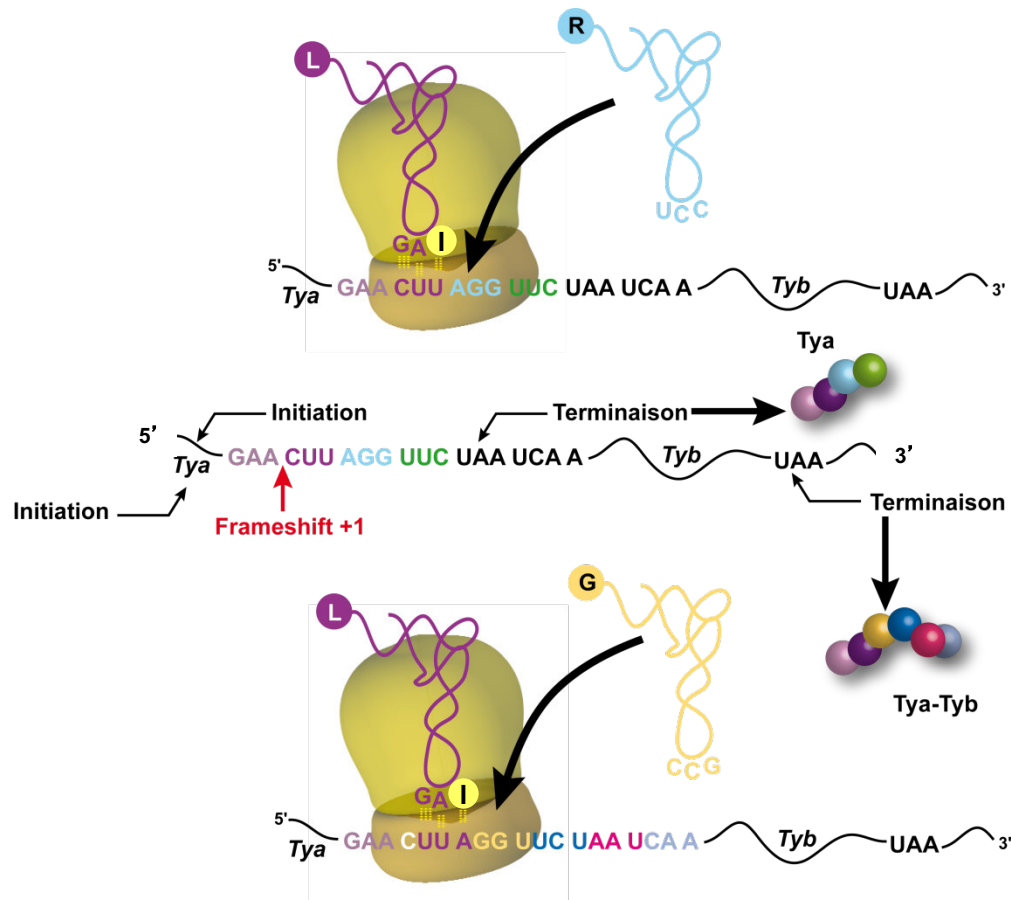
- These jumps are not random

II. The Genetic Code

D. General information on the reprogramming and dynamics of the genetic code

3. Frameshift

- The +1 frameshift is required for the expression of the yeast Ty transposon tyb gene

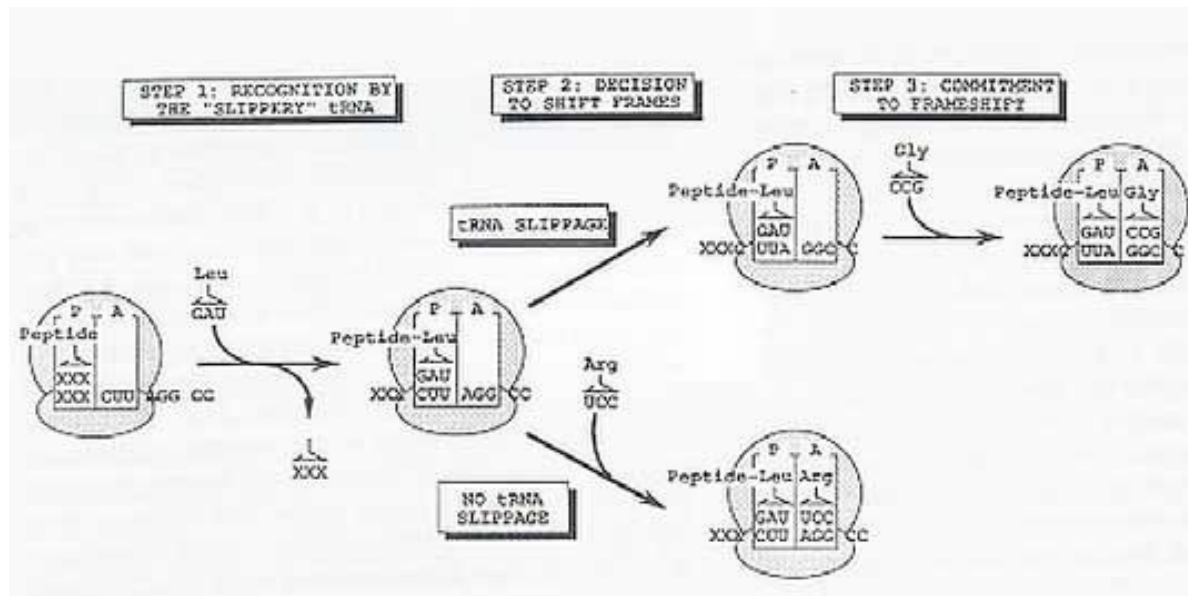


II. The Genetic Code

D. General information on the reprogramming and dynamics of the genetic code

3. Frameshift

- The +1 frameshift is required for the expression of the yeast Ty transposon tyb gene

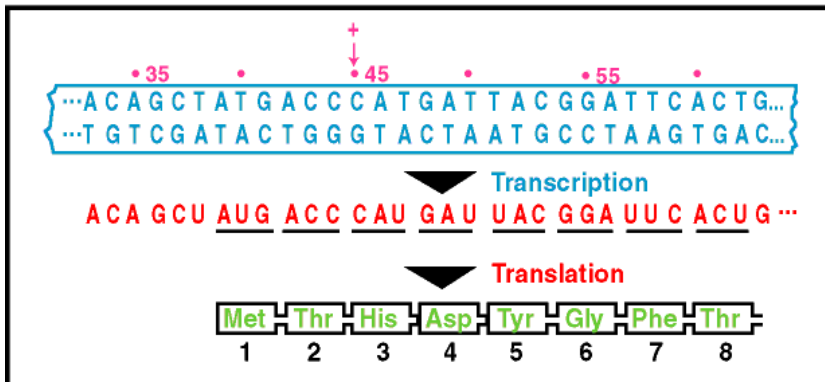
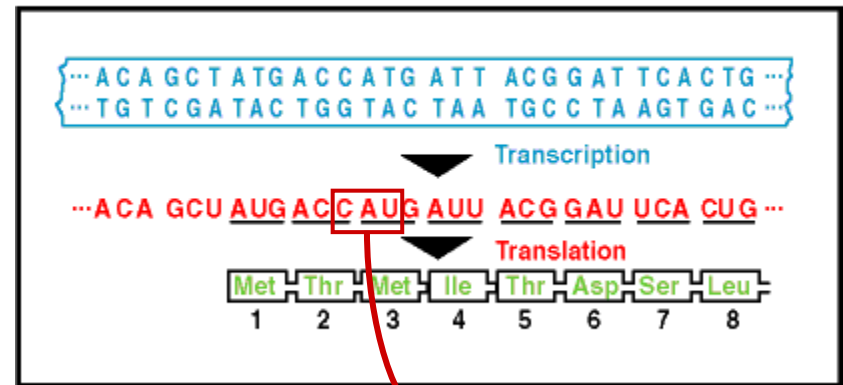
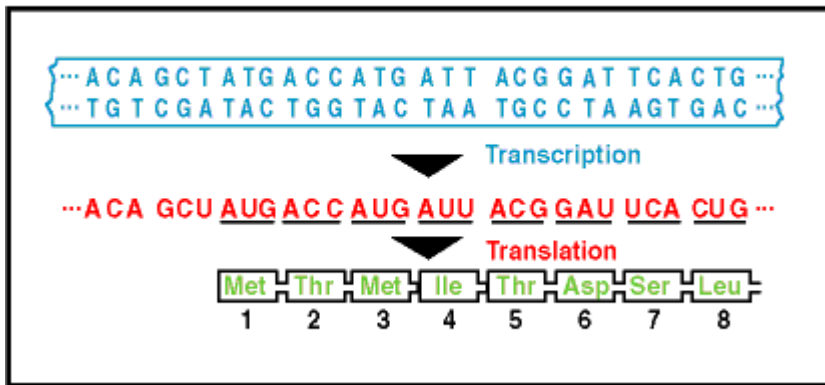


II. The Genetic Code

D. General information on the reprogramming and dynamics of the genetic code

3. Frameshift

- Ribosome-induced frameshift -1 also exists



The ribosome slides 1 nt backward

II. The Genetic Code

D. General information on the reprogramming and dynamics of the genetic code

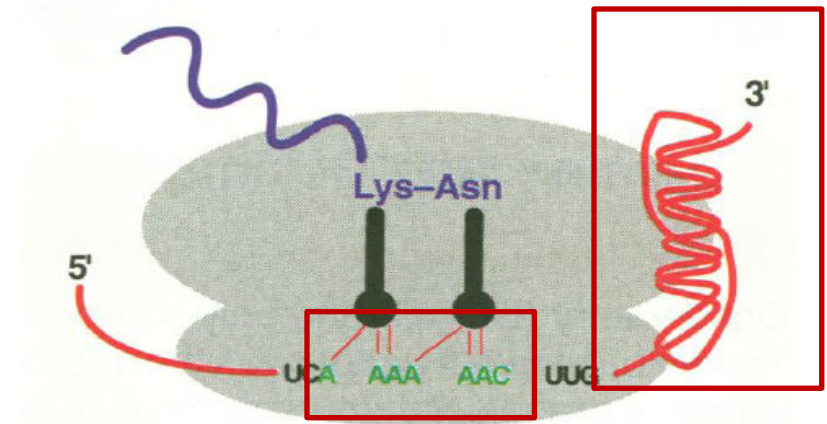
3. Frameshift

- Ribosome-induced frameshift -1 also exists

The mRNA must contain 2 important signals

A **slippery sequence**: this is where the frameshift event happens.

A **3'-stimulatory RNA structure**: it forces the ribosome to pause allowing the -1 slip.

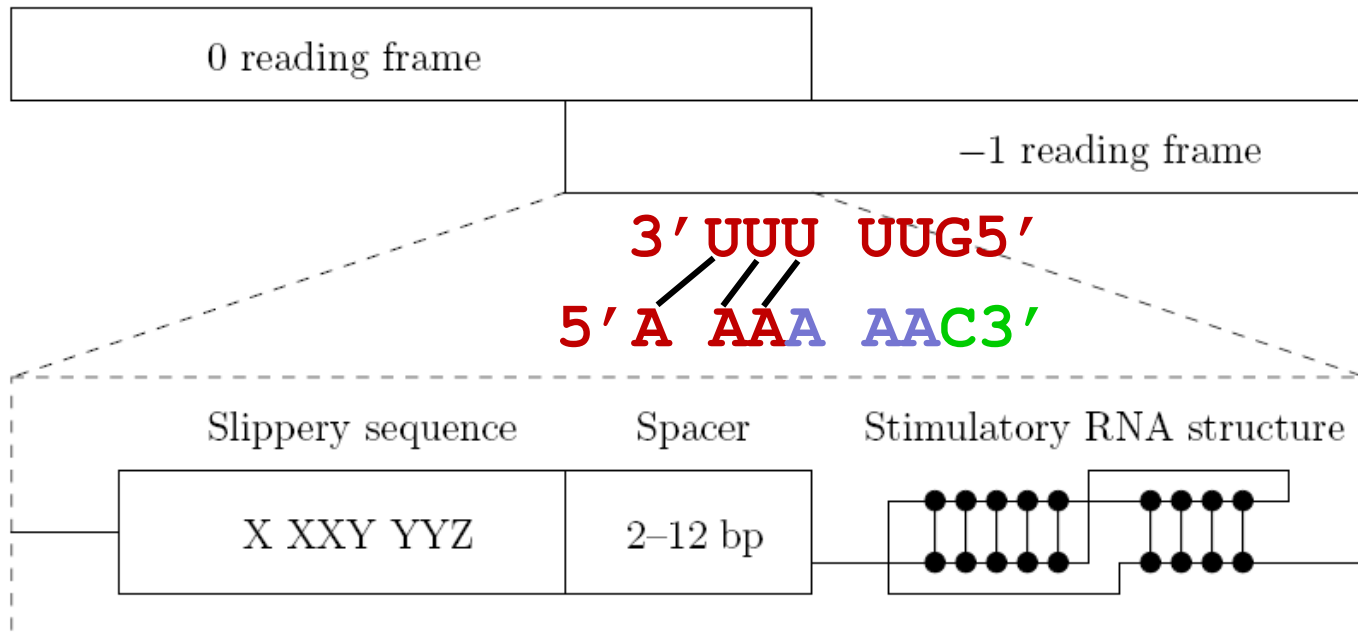


II. The Genetic Code

D. General information on the reprogramming and dynamics of the genetic code

3. Frameshift

- Ribosome-induced frameshift -1 also exists



hepta-nucleotide (7-mer) **X XXY YYZ**.

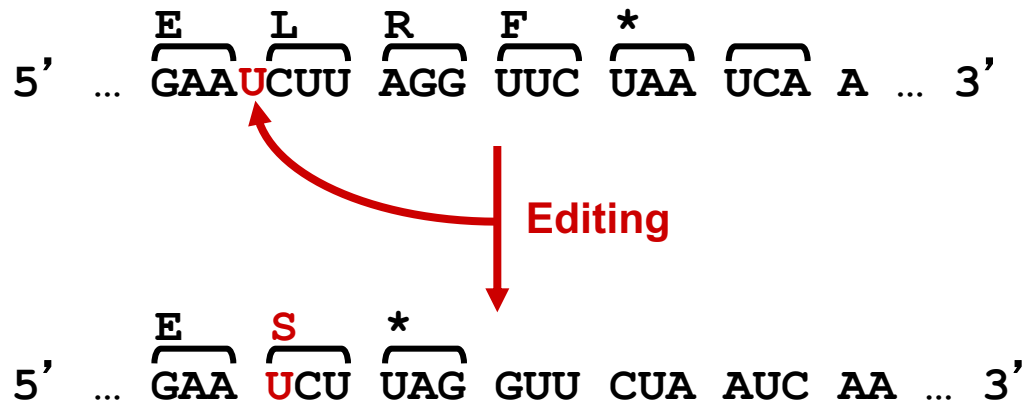
The spaces in X, XXY, YYZ separate the codons in the normal phase of reading, and X and Z can be any nucleotide, while **Y is either A or U**.

II. The Genetic Code

D. General information on the reprogramming and dynamics of the genetic code

4. Re-edition or Editing in trypanosoma mitochondria

- Insertion/deletion editing of an Nt will change the sequence (seq) and very often there is premature appearance of a STOP codon



⚠ Through this editing mechanism, there can be up to 10 U added in the same region one after the other



II. The Genetic Code

D. General information on the reprogramming and dynamics of the genetic code

4. Re-edition or Editing in trypanosoma mitochondria

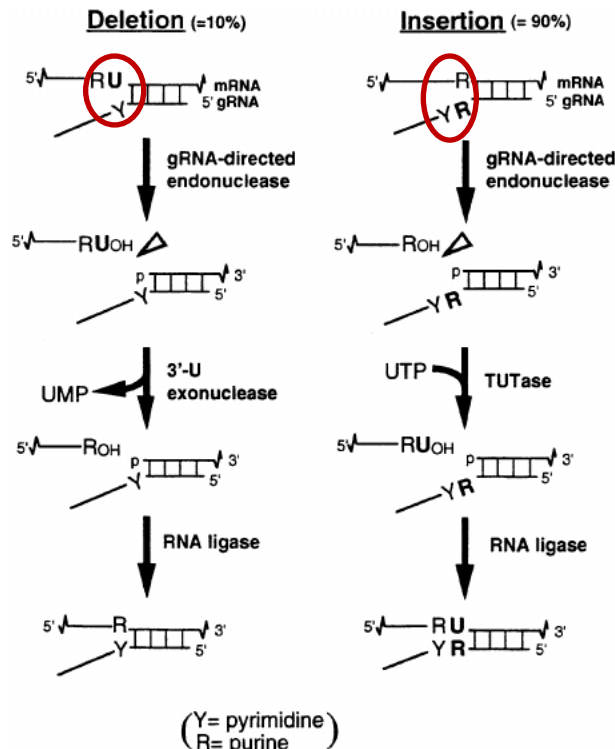
- Insertion/deletion editing of an Nt will change the sequence (seq) and very often there is premature appearance of a STOP codon

The EMBO Journal Vol. 20 No. 17 pp. 4694-4704, 2001

Roles for ligases in the RNA editing complex of
Trypanosoma brucei: band IV is needed for
U-deletion and RNA repair

Structural context

Catherine E.Huang, Jorge Cruz-Reyes,
Alevtina G.Zhelonkina, Sean O'Hearn,
Elizabeth Wirtz¹ and Barbara Sollner-Webb²
Department of Biological Chemistry, The Johns Hopkins University
School of Medicine, 725 North Wolfe Street, Baltimore, MD 21205
and ¹Laboratory of Molecular Parasitology, The Rockefeller
University, New York, NY 10021, USA
²Corresponding author
e-mail: bsolln@jhmi.edu

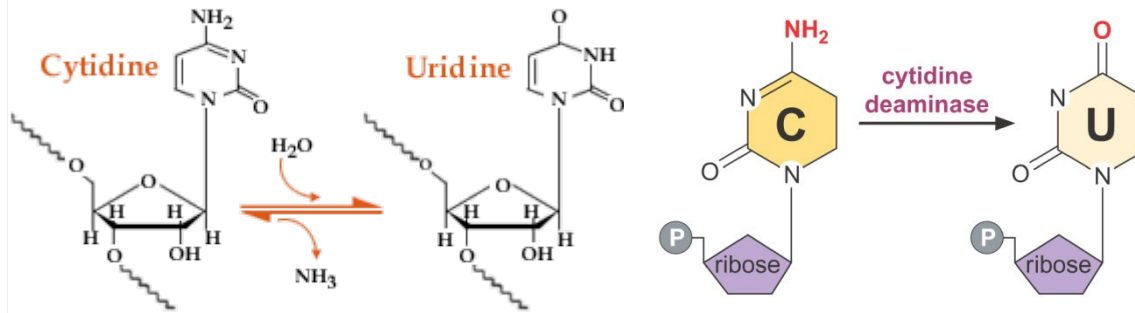


II. The Genetic Code

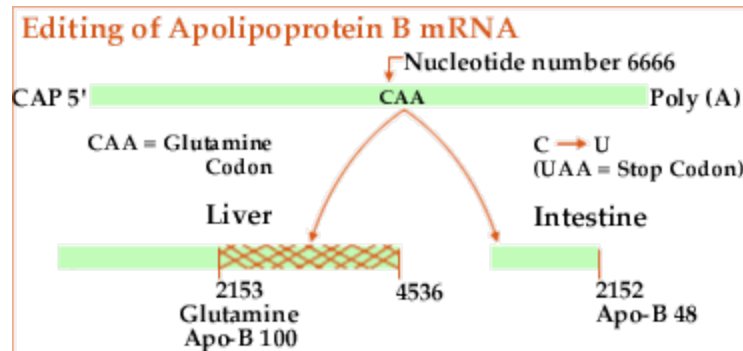
D. General information on the reprogramming and dynamics of the genetic code

4. Re-edition or Editing in trypanosoma mitochondria

- Editing by chemical transformation of an Nt



Copyright © 2004 Pearson Education, Inc., publishing as Benjamin Cummings



RNA editing of the apo-B gene. In mammals, the apo-B gene is expressed in both hepatocytes (liver cells) and intestinal epithelial cells. However, **in liver** cells, its product is a 500 kD protein called Apo-B100 whereas in **intestine** cells its product is a smaller protein called Apo-B48. The Apo-B100 is produced without RNA editing, but the Apo-B48 is synthesized from an mRNA whose sequence has been altered by a specific enzyme. This enzyme changes a codon, CAA, in the middle of the original mRNA to the stop codon UAA, thereby causing early termination of the protein synthesis.

II. The Genetic Code

E. Universality of the genetic code

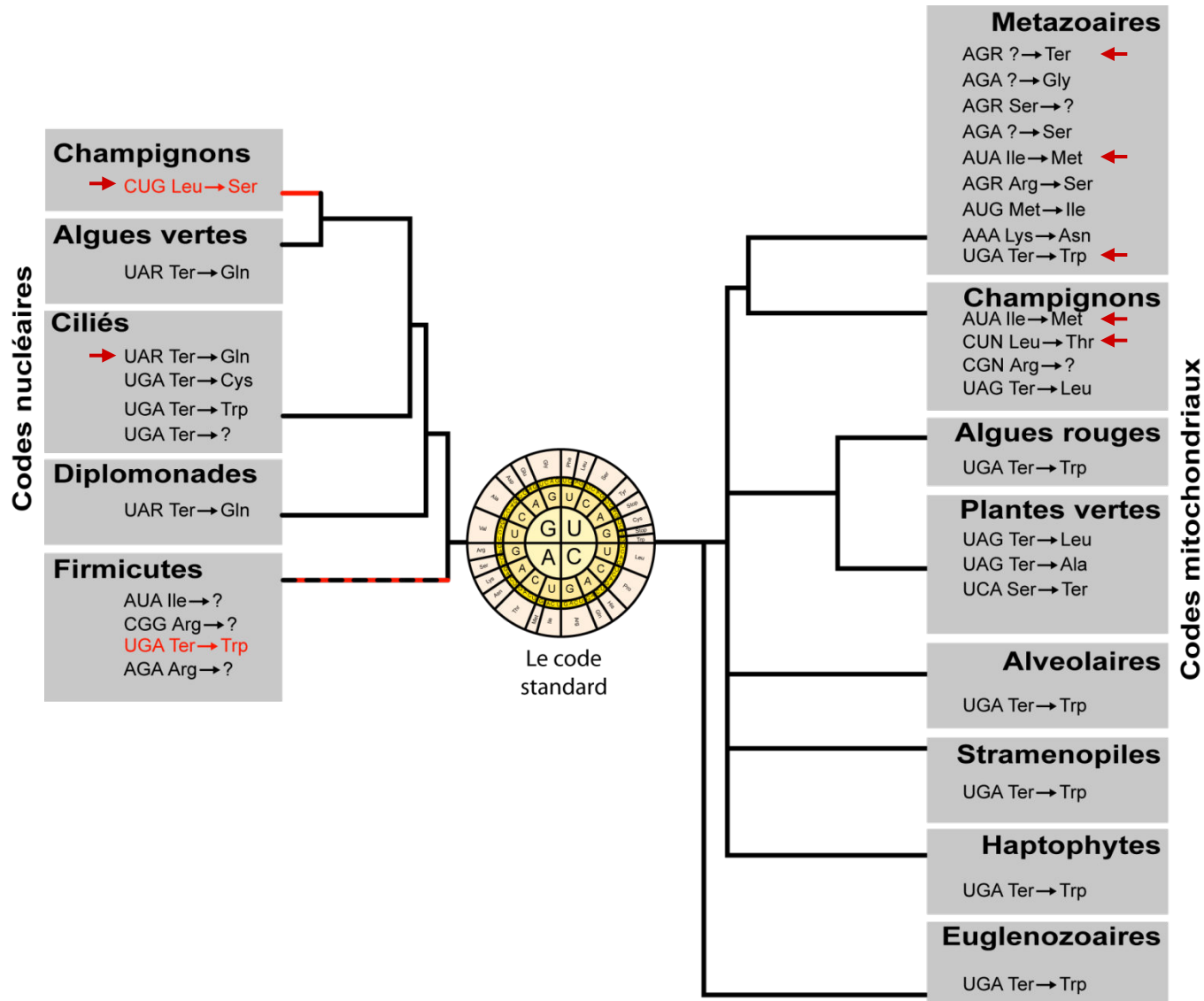
Crick, F. H. (1968). The origin of the genetic code. J. Mol. Biol. 38, 367-379.

The genetic code discovered by Marshal W. Nirenberg (Nirenberg and Matthaei, 1961) in the 1960s has long been considered a "frozen accident" (Crick, 1968), incapable of evolving. The immutability of the genetic code originated from the simple logical assumption that any change in this code would lead to the alteration of the meaning of the codons and thus cause the introduction of errors in the translation of each messenger RNA.

 **GC is almost universal, the main deviations are uncommonly observed in mitochondrial genomes but there are also deviations in nuclear codes**

II. The Genetic Code

E. Universality of the genetic code



F. Codon usage

! Can there be 6 codons that actually code for an aa?

The answer is yes, all codons are always in use

! Can some codons never be used?

The answer is almost always no, all codons are usually used (but it can happen for a STOP)

! Are some codons used more than others?

The answer is yes, it is the use of codons and it is specific to each organism

You go to the website and enter the names of the following organisms: Escherichia coli; Thermus thermophilus, Saccharomyces cerevisiae and Homo Sapiens (1 at a time)

<http://www.kazusa.or.jp/codon/>

Codon Usage Database

Data source

[NCBI-GenBank](#) Flat File Release 160.0 [June 15 2007].

Data amount

35,799 organisms
3,027,973 complete protein coding genes (CDS's)

[Announcement](#)

QUERY Box for search with Latin name of organism

Case: ☒ sensitive ☐ insensitive

Input a scientific name (or its [regular expression](#)) for an organism and press "Submit" or return key. Use **Latin name** such as "Marchantia polymorpha", "Saccharomyces cerevisiae" etc., not "liverwort", "yeast" etc.

F. Codon usage

You get the following 4 tables:

Escherichia coli K12 [gbtct]: 14 CDS's (5122 codons)

fields: [triplet] [frequency: **per thousand**] ([number])

UUU 19.7(101)	UCU 5.7(29)	UAU 16.8(86)	UGU 5.9(30)
UUC 15.0(77)	UCC 5.5(28)	UAC 14.6(75)	UGC 8.0(41)
UUA 15.2(78)	UCA 7.8(40)	UAA 1.8(9)	UGA 1.0(5)
UUG 11.9(61)	UCG 8.0(41)	UAG 0.0(0)	UGG 10.7(55)

CUU 11.9(61)	CCU 8.4(43)	CAU 15.8(81)	CGU 21.1(108)
CUC 10.5(54)	CCC 6.4(33)	CAC 13.1(67)	CGC 26.0(133)
CUA 5.3(27)	CCA 6.6(34)	CAA 12.1(62)	CGA 4.3(22)
CUG 46.9(240)	CCG 26.7(137)	CAG 27.7(142)	CGG 4.1(21)

AUU 30.5(156)	ACU 8.0(41)	AAU 21.9(112)	AGU 7.2(37)
AUC 18.2(93)	ACC 22.8(117)	AAC 24.4(125)	AGC 16.6(85)
AUA 3.7(19)	ACA 6.4(33)	AAA 33.2(170)	AGA 1.4(7)
AUG 24.8(127)	ACG 11.5(59)	AAG 12.1(62)	AGG 1.6(8)

GUU 16.8(86)	GCU 10.7(55)	GAU 37.9(194)	GGU 21.3(109)
GUC 11.7(60)	GCC 31.6(162)	GAC 20.5(105)	GGC 33.4(171)
GUA 11.5(59)	GCA 21.1(108)	GAA 43.7(224)	GGA 9.2(47)
GUG 26.4(135)	GCG 38.5(197)	GAG 18.4(94)	GGG 8.6(44)

Coding GC 52.35% 1st letter GC 60.82% 2nd letter GC 40.61% 3rd letter GC 55.62%

Thermus thermophilus HB8 [gbtct]: 2238 CDS's (669747 codons)

fields: [triplet] [frequency: **per thousand**] ([number])

UUU 6.7(4471)	UCU 0.7(495)	UAU 1.2(830)	UGU 0.2(125)
UUC 30.9(20692)	UCC 15.3(10233)	UAC 27.5(18407)	UGC 3.7(2493)
UUA 1.1(744)	UCA 0.3(210)	UAA 0.6(431)	UGA 1.4(970)
UUG 9.6(6428)	UCG 3.8(2561)	UAG 1.2(837)	UGG 13.8(9274)

CUU 16.9(11332)	CCU 4.5(3009)	CAU 0.9(582)	CGU 1.5(1006)
CUC 72.9(48852)	CCC 48.0(32153)	CAC 17.7(11824)	CGC 29.0(19436)
CUA 3.3(2184)	CCA 1.4(969)	CAA 3.1(2076)	CGA 1.1(766)
CUG 41.0(27438)	CCG 11.6(7757)	CAG 21.2(14228)	CGG 36.8(24675)

AUU 2.4(1594)	ACU 0.4(291)	AAU 0.4(273)	AGU 0.4(263)
AUC 22.3(14902)	ACC 27.0(18095)	AAC 15.1(10097)	AGC 13.8(9217)
AUA 1.2(789)	ACA 0.4(298)	AAA 2.9(1930)	AGA 0.8(533)
AUG 14.4(9621)	ACG 9.5(6356)	AAG 32.6(21825)	AGG 16.3(10939)

GUU 2.5(1706)	GCU 2.6(1751)	GAU 1.7(1112)	GGU 2.2(1456)
GUC 27.0(18068)	GCC 84.7(56708)	GAC 34.1(22864)	GGC 37.0(24778)
GUA 1.5(992)	GCA 1.7(1134)	GAA 10.6(7089)	GGA 5.5(3662)
GUG 50.1(33558)	GCG 26.2(17562)	GAG 75.3(50430)	GGG 48.3(32366)

Coding GC 69.67% 1st letter GC 72.19% 2nd letter GC 45.02% 3rd letter GC 91.78%

Saccharomyces cerevisiae [gbpln]: 14411 CDS's (6534504 codons)

fields: [triplet] [frequency: **per thousand**] ([number])

UUU 26.1(170666)	UCU 23.5(153557)	UAU 18.8(122728)	UGU 8.1(52903)
UUC 18.4(120510)	UCC 14.2(92923)	UAC 14.8(96596)	UGC 4.8(31095)
UUA 26.2(170884)	UCA 18.7(122028)	UAA 1.1(6913)	UGA 0.7(4447)
UUG 27.2(177573)	UCG 8.6(55951)	UAG 0.5(3312)	UGG 10.4(67789)

CUU 12.3(80076)	CCU 13.5(88263)	CAU 13.6(89007)	CGU 6.4(41791)
CUC 5.4(35545)	CCC 6.8(44309)	CAC 7.8(50785)	CGC 2.6(16993)
CUA 13.4(87619)	CCA 18.3(119641)	CAA 27.3(178251)	CGA 3.0(19562)
CUG 10.5(68494)	CCG 5.3(34597)	CAG 12.1(79121)	CGG 1.7(11351)

AUU 30.1(196893)	ACU 20.3(132522)	AAU 35.7(233124)	AGU 14.2(92466)
AUC 17.2(112176)	ACC 12.7(83207)	AAC 24.8(162199)	AGC 9.8(63726)
AUA 17.8(116254)	ACA 17.8(116084)	AAA 41.9(273618)	AGA 21.3(139081)
AUG 20.9(136805)	ACG 8.0(52045)	AAG 30.8(201361)	AGG 9.2(60289)

GUU 22.1(144243)	GCU 21.2(138358)	GAU 37.6(245641)	GGU 23.9(156109)
GUC 11.8(76947)	GCC 12.6(82357)	GAC 20.2(132048)	GGC 9.8(63903)
GUA 11.8(76927)	GCA 16.2(105910)	GAA 45.6(297944)	GGA 10.9(71216)
GUG 10.8(70337)	GCG 6.2(40358)	GAG 19.2(125717)	GGG 6.0(39359)

Coding GC 39.77% 1st letter GC 44.58% 2nd letter GC 36.64% 3rd letter GC 38.10%

Homo sapiens [gbpri]: 93487 CDS's (40662582 codons)

fields: [triplet] [frequency: **per thousand**] ([number])

UUU 17.6(714298)	UCU 15.2(618711)	UAU 12.2(495699)	UGU 10.6(430311)
UUC 20.3(824692)	UCC 17.7(718892)	UAC 15.3(622407)	UGC 12.6(513028)
UUA 7.7(311881)	UCA 12.2(496448)	UAA 1.0(40285)	UGA 1.6(63237)
UUG 12.9(525688)	UCG 4.4(179419)	UAG 0.8(32109)	UGG 13.2(535595)

CUU 13.2(536515)	CCU 17.5(713233)	CAU 10.9(441711)	CGU 4.5(184609)
CUC 19.6(796638)	CCC 19.8(804620)	CAC 15.1(613713)	CGC 10.4(423516)
CUA 7.2(290751)	CCA 16.9(688038)	CAA 12.3(501911)	CGA 6.2(250760)
CUG 39.6(1611801)	CCG 6.9(281570)	CAG 34.2(1391973)	CGG 11.4(464485)

AUU 16.0(650473)	ACU 13.1(533609)	AAU 17.0(689701)	AGU 12.1(493429)
AUC 20.8(846466)	ACC 18.9(768147)	AAC 19.1(776603)	AGC 19.5(791383)
AUA 7.5(304565)	ACA 15.1(614523)	AAA 24.4(993621)	AGA 12.2(494682)
AUG 22.0(896005)	ACG 6.1(246105)	AAG 31.9(1295568)	AGG 12.0(486463)

GUU 11.0(448607)	GCU 18.4(750096)	GAU 21.8(885429)	GGU 10.8(437126)
GUC 14.5(588138)	GCC 27.7(1127679)	GAC 25.1(1020595)	GGC 22.2(903565)
GUA 7.1(287712)	GCA 15.8(643471)	GAA 29.0(1177632)	GGA 16.5(669873)
GUG 28.1(1143534)	GCG 7.4(299495)	GAG 39.6(1609975)	GGG 16.5(669768)

Coding GC 52.27% 1st letter GC 55.72% 2nd letter GC 42.54% 3rd letter GC 58.55%

F. Codon usage

If I compare 2 bacteria, 1 mesophilic: *E. coli* (37 ° C) and 1 thermophilic *T. thermophilus* (75 ° C)

***Escherichia coli* K12 [gbt]: 14 CDS's (5122 codons)**

fields: [triplet] [frequency: per thousand] ([number])

UUU 19.7(101)	UCU 5.7(29)	UAU 16.8(86)	UGU 5.9(30)
UUC 15.0(77)	UCC 5.5(28)	UAC 14.6(75)	UGC 8.0(41)
UUA 15.2(78)	UCA 7.8(40)	UAA 1.8(9)	UGA 1.0(5)
UUG 11.9(61)	UCG 8.0(41)	UAG 0.0(0)	UGG 10.7(55)
CUU 11.9(61)	CCU 8.4(43)	CAU 15.8(81)	CGU 21.1(108)
CUC 10.5(54)	CCC 6.4(33)	CAC 13.1(67)	CGC 26.0(133)
CUA 5.3(27)	CCA 6.6(34)	CAA 12.1(62)	CGA 4.3(22)
CUG 46.9(240)	CCG 26.7(137)	CAG 27.7(142)	CGG 4.1(21)
AUU 30.5(156)	ACU 8.0(41)	AAU 21.9(112)	AGU 7.2(37)
AUC 18.2(93)	ACC 22.8(117)	AAC 24.4(125)	AGC 16.6(85)
AUA 3.7(19)	ACA 6.4(33)	AAA 33.2(170)	AGA 1.4(7)
AUG 24.8(127)	ACG 11.5(59)	AAG 12.1(62)	AGG 1.6(8)
GUU 16.8(86)	GCU 10.7(55)	GAU 37.9(194)	GGU 21.3(109)
GUC 11.7(60)	GCC 31.6(162)	GAC 20.5(105)	GGC 33.4(171)
GUA 11.5(59)	GCA 21.1(108)	GAA 43.7(224)	GGA 9.2(47)
GUG 26.4(135)	GCG 38.5(197)	GAG 18.4(94)	GGG 8.6(44)

Coding GC 52.35% 1st letter GC 60.82% 2nd letter GC 40.61% 3rd letter GC 55.62%

Different use due to physiological temperature

Codons less frequent in Tth than in Eco:
those ending in A or U since Tth has 75% G-C

Since *E. coli* is often used for the overexpression of proteins by transformation with an overexpression plasmid, it is important to know which Tth codons are very much more used by Tth than by Eco = rare codons of Eco vs* Tth

Rare codons of Eco vs Tth

A rare codon is decoded by a minor isoacceptor tRNA, which means that the amount (concentration) of this tRNA is low in the cell.

If I want to overexpress a Tth protein in Eco, I have to make sure that Eco has the use of the codons closest to Tth.

***Thermus thermophilus* HB8 [gbt]: 2238 CDS's (669747 codons)**

fields: [triplet] [frequency: per thousand] ([number])

UUU 6.7(4471)	UCU 0.7(495)	UAU 1.2(830)	UGU 0.2(125)
UUC 30.9(20692)	UCC 15.3(10233)	UAC 27.5(18407)	UGC 3.7(2493)
UUA 1.1(744)	UCA 0.3(210)	UAA 0.6(431)	UGA 1.4(970)
UUG 9.6(6428)	UCG 3.8(2561)	UAG 1.2(837)	UGG 13.8(9274)
CUU 16.9(11332)	CCU 4.5(3009)	CAU 0.9(582)	CGU 1.5(1006)
CUC 72.9(48852)	CCC 48.0(32153)	CAC 17.7(11824)	CGC 29.0(19436)
CUA 3.3(2184)	CCA 1.4(969)	CAA 3.1(2076)	CGA 1.1(766)
CUG 41.0(27438)	CCG 11.6(7757)	CAG 21.2(14228)	CGG 36.8(24675)
AUU 2.4(1594)	ACU 0.4(291)	AAU 0.4(273)	AGU 0.4(263)
AUC 22.3(14902)	ACC 27.0(18095)	AAC 15.1(10097)	AGC 13.8(9217)
AUA 1.2(789)	ACA 0.4(298)	AAA 2.9(1930)	AGA 0.8(533)
AUG 14.4(9621)	ACG 9.5(6356)	AAG 32.6(21825)	AGG 16.3(10939)
GUU 2.5(1706)	GCU 2.6(1751)	GAU 1.7(1112)	GGU 2.2(1456)
GUC 27.0(18068)	GCC 84.7(56708)	GAC 34.1(22864)	GGC 37.0(24778)
GUA 1.5(992)	GCA 1.7(1134)	GAA 10.6(7089)	GGA 5.5(3662)
GUG 50.1(33558)	GCG 26.2(17562)	GAG 75.3(50430)	GGG 48.3(32366)

Coding GC 69.67% 1st letter GC 72.19% 2nd letter GC 45.02% 3rd letter GC 91.78%

F. Codon usage

If I compare 2 bacteria, 1 mesophilic: *E. coli* (37 ° C) and 1 thermophilic *T. thermophilus* (75 ° C)

Escherichia coli K12 [gbt]: 14 CDS's (5122 codons)

fields: [triplet] [frequency: per thousand] ([number])

UUU 19.7(101)	UCU 5.7(29)	UAU 16.8(86)	UGU 5.9(30)
UUC 15.0(77)	UCC 5.5(28)	UAC 14.6(75)	UGC 8.0(41)
UUA 15.2(78)	UCA 7.8(40)	UAA 1.8(9)	UGA 1.0(5)
UUG 11.9(61)	UCG 8.0(41)	UAG 0.0(0)	UGG 10.7(55)
CUU 11.9(61)	CCU 8.4(43)	CAU 15.8(81)	CGU 21.1(108)
CUC 10.5(54)	CCC 6.4(33)	CAC 13.1(67)	CGC 26.0(133)
CUA 5.3(27)	CCA 6.6(34)	CAA 12.1(62)	CGA 4.3(22)
CUG 46.9(240)	CCG 26.7(137)	CAG 27.7(142)	CGG 4.1(21)
AUU 30.5(156)	ACU 8.0(41)	AAU 21.9(112)	AGU 7.2(37)
AUC 18.2(93)	ACC 22.8(117)	AAC 24.4(125)	AGC 16.6(85)
AUA 3.7(19)	ACA 6.4(33)	AAA 33.2(170)	AGA 1.4(7)
AUG 24.8(127)	ACG 11.5(59)	AAG 12.1(62)	AGG 1.6(8)
GUU 16.8(86)	GCU 10.7(55)	GAU 37.9(194)	GGU 21.3(109)
GUC 11.7(60)	GCC 31.6(162)	GAC 20.5(105)	GGC 33.4(171)
GUA 11.5(59)	GCA 21.1(108)	GAA 43.7(224)	GGA 9.2(47)
GUG 26.4(135)	GCG 38.5(197)	GAG 18.4(94)	GGG 8.6(44)

Coding GC 52.35% 1st letter GC 60.82% 2nd letter GC 40.61% 3rd letter GC 55.62%

The 3 most problematic codons are:

STOP amber UAG: 837 times + in Tth than Eco

Leu (L) CUC: 7 times + in Tth que Eco

Arg (R) AGG: 10 times + in Tth than Eco

Solution:

Use *E. coli* BL21-CodonPlus (DE3)-RIL strain

<https://www.ncbi.nlm.nih.gov/pubmed/10910733>

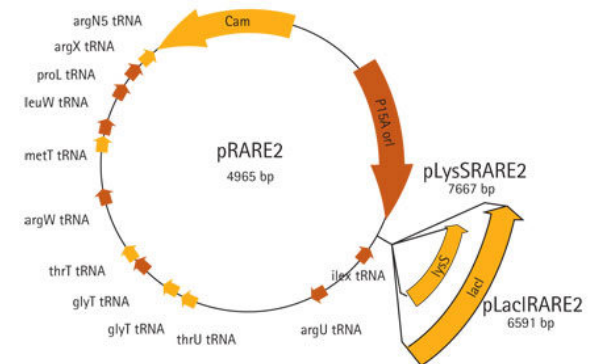
E. coli BL21-CodonPlus (DE3)-RIL contains a pRARE plasmid overexpressing minor tRNA^{Arg} (R), tRNA^{Ile} (I) and tRNA^{Leu} (L) that decodes the rare R I L codons of Eco. As a result, these codons are no longer rare

Thermus thermophilus HB8 [gbt]: 2238 CDS's (669747 codons)

fields: [triplet] [frequency: per thousand] ([number])

UUU 6.7(4471)	UCU 0.7(495)	UAU 1.2(830)	UGU 0.2(125)
UUC 30.9(20692)	UCC 15.3(10233)	UAC 27.5(18407)	UGC 3.7(2493)
UUA 1.1(744)	UCA 0.3(210)	UAA 0.6(431)	UGA 1.4(970)
UUG 9.6(6428)	UCG 3.8(2561)	UAG 1.2(837)	UGG 13.8(9274)
CUU 16.9(11332)	CCU 4.5(3009)	CAU 0.9(582)	CGU 1.5(1006)
CUC 72.9(48852)	CCC 48.0(32153)	CAC 17.7(11824)	CGC 29.0(19436)
CUA 3.3(2184)	CCA 1.4(969)	CAA 3.1(2076)	CGA 1.1(766)
CUG 41.0(27438)	CCG 11.6(7757)	CAG 21.2(14228)	CGG 36.8(24675)
AUU 2.4(1594)	ACU 0.4(291)	AAU 0.4(273)	AGU 0.4(263)
AUC 22.3(14902)	ACC 27.0(18095)	AAC 15.1(10097)	AGC 13.8(9217)
AUA 1.2(789)	ACA 0.4(298)	AAA 2.9(1930)	AGA 0.8(533)
AUG 14.4(9621)	ACG 9.5(6356)	AAG 32.6(21825)	AGG 16.3(10939)
GUU 2.5(1706)	GCU 2.6(1751)	GAU 1.7(1112)	GGU 2.2(1456)
GUC 27.0(18068)	GCC 84.7(56708)	GAC 34.1(22864)	GGC 37.0(24778)
GUA 1.5(992)	GCA 1.7(1134)	GAA 10.6(7089)	GGA 5.5(3662)
GUG 50.1(33558)	GCG 26.2(17562)	GAG 75.3(50430)	GGG 48.3(32366)

Coding GC 69.67% 1st letter GC 72.19% 2nd letter GC 45.02% 3rd letter GC 91.78%



https://www.merckmillipore.com/INTERSHOP/static/WFS/Merck-Site/-/Merck/en_US/Freestyle/BI-Bioscience/Genomic-Analysis/pRARE2.jpg

F. Codon usage

You get the following 4 tables:

Escherichia coli K12 [gbtct]: 14 CDS's (5122 codons)

fields: [triplet] [frequency: per thousand] ([number])

UUU 19.7(101)	UCU 5.7(29)	UAU 16.8(86)	UGU 5.9(30)
UUC 15.0(77)	UCC 5.5(28)	UAC 14.6(75)	UGC 8.0(41)
UUA 15.2(78)	UCA 7.8(40)	UAA 1.8(9)	UGA 1.0(5)
UUG 11.9(61)	UCG 8.0(41)	UAG 0.0(0)	UGG 10.7(55)
CUU 11.9(61)	CCU 8.4(43)	CAU 15.8(81)	CGU 21.1(108)
CUC 10.5(54)	CCC 6.4(33)	CAC 13.1(67)	CGC 26.0(133)
CUA 5.3(27)	CCA 6.6(34)	CAA 12.1(62)	CGA 4.3(22)
CUG 46.9(240)	CCG 26.7(137)	CAG 27.7(142)	CGG 4.1(21)
AUU 30.5(156)	ACU 8.0(41)	AAU 21.9(112)	AGU 7.2(37)
AUC 18.2(93)	ACC 22.8(117)	AAC 24.4(125)	AGC 16.6(85)
AUA 3.7(19)	ACA 6.4(33)	AAA 33.2(170)	AGA 1.4(7)
AUG 24.8(127)	ACG 11.5(59)	AAG 12.1(62)	AGG 1.6(8)
GUU 16.8(86)	GCU 10.7(55)	GAU 37.9(194)	GGU 21.3(109)
GUC 11.7(60)	GCC 31.6(162)	GAC 20.5(105)	GGC 33.4(171)
GUA 11.5(59)	GCA 21.1(108)	GAA 43.7(224)	GGA 9.2(47)
GUG 26.4(135)	GCG 38.5(197)	GAG 18.4(94)	GGG 8.6(44)

Coding GC 52.35% 1st letter GC 60.82% 2nd letter GC 40.61% 3rd letter GC 55.62%

Saccharomyces cerevisiae [gbpln]: 14411 CDS's (6534504 codons)

fields: [triplet] [frequency: per thousand] ([number])

UUU 26.1(170666)	UCU 23.5(153557)	UAU 18.8(122728)	UGU 8.1(52903)
UUC 18.4(120510)	UCC 14.2(92923)	UAC 14.8(96596)	UGC 4.8(31095)
UUA 26.2(170884)	UCA 18.7(122028)	UAA 1.1(6913)	UGA 0.7(4447)
UUG 27.2(177573)	UCG 8.6(55951)	UAG 0.5(3312)	UGG 10.4(67789)
CUU 12.3(80076)	CCU 13.5(88263)	CAU 13.6(89007)	CGU 6.4(41791)
CUC 5.4(35545)	CCC 6.8(44309)	CAC 7.8(50785)	CGC 2.6(16993)
CUA 13.4(87619)	CCA 18.3(119641)	CAA 27.3(178251)	CGA 3.0(19562)
CUG 10.5(68494)	CCG 5.3(34597)	CAG 12.1(79121)	CGG 1.7(11351)
AUU 30.1(196893)	ACU 20.3(132522)	AAU 35.7(233124)	AGU 14.2(92466)
AUC 17.2(112176)	ACC 12.7(83207)	AAC 24.8(162199)	AGC 9.8(63726)
AUA 17.8(116254)	ACA 17.8(116084)	AAA 41.9(273618)	AGA 21.3(139081)
AUG 20.9(136805)	ACG 8.0(52045)	AAG 30.8(201361)	AGG 9.2(60289)
GUU 22.1(144243)	GCU 21.2(138358)	GAU 37.6(245641)	GGU 23.9(156109)
GUC 11.8(76947)	GCC 12.6(82357)	GAC 20.2(132048)	GGC 9.8(63903)
GUA 11.8(76927)	GCA 16.2(105910)	GAA 45.6(297944)	GGA 10.9(71216)
GUG 10.8(70337)	GCG 6.2(40358)	GAG 19.2(125717)	GGG 6.0(39359)

Coding GC 39.77% 1st letter GC 44.58% 2nd letter GC 36.64% 3rd letter GC 38.10%

The rarity of some R I L codons is also the main difference between the use of codons in prokaryotes vs eukaryotes even both species live at the same physiological temperature, e.g. Eco vs Sce and Hsa.

rare codons of Eco vs Sce and Hsa

The most effective version of BL21 RIL is called: Rosetta 2 and overexpresses in addition to other minor tRNAs: Met, Thr, Pro, Gly...

Homo sapiens [gbpri]: 93487 CDS's (40662582 codons)

fields: [triplet] [frequency: per thousand] ([number])

UUU 17.6(714298)	UCU 15.2(618711)	UAU 12.2(495699)	UGU 10.6(430311)
UUC 20.3(824692)	UCC 17.7(718892)	UAC 15.3(622407)	UGC 12.6(513028)
UUA 7.7(311881)	UCA 12.2(496448)	UAA 1.0(40285)	UGA 1.6(63237)
UUG 12.9(525688)	UCG 4.4(179419)	UAG 0.8(32109)	UGG 13.2(535595)
CUU 13.2(536515)	CCU 17.5(713233)	CAU 10.9(441711)	CGU 4.5(184609)
CUC 19.6(796638)	CCC 19.8(804620)	CAC 15.1(613713)	CGC 10.4(423516)
CUA 7.2(290751)	CCA 16.9(688038)	CAA 12.3(501911)	CGA 6.2(250760)
CUG 39.6(1611801)	CCG 6.9(281570)	CAG 34.2(1391973)	CGG 11.4(464485)
AUU 16.0(650473)	ACU 13.1(533609)	AAU 17.0(689701)	AGU 12.1(493429)
AUC 20.8(846466)	ACC 18.9(768147)	AAC 19.1(776603)	AGC 19.5(791383)
AUA 7.5(304565)	ACA 15.1(614523)	AAA 24.4(993621)	AGA 12.2(494682)
AUG 22.0(896005)	ACG 6.1(246105)	AAG 31.9(1295568)	AGG 12.0(486463)
GUU 11.0(448607)	GCU 18.4(750096)	GAU 21.8(885429)	GGU 10.8(437126)
GUC 14.5(588138)	GCC 27.7(1127679)	GAC 25.1(1020595)	GGC 22.2(903565)
GUA 7.1(287712)	GCA 15.8(643471)	GAA 29.0(1177632)	GGA 16.5(669873)
GUG 28.1(1143534)	GCG 7.4(299495)	GAG 39.6(1609975)	GGG 16.5(669768)

Coding GC 52.27% 1st letter GC 55.72% 2nd letter GC 42.54% 3rd letter GC 58.55%

III. RNA structure

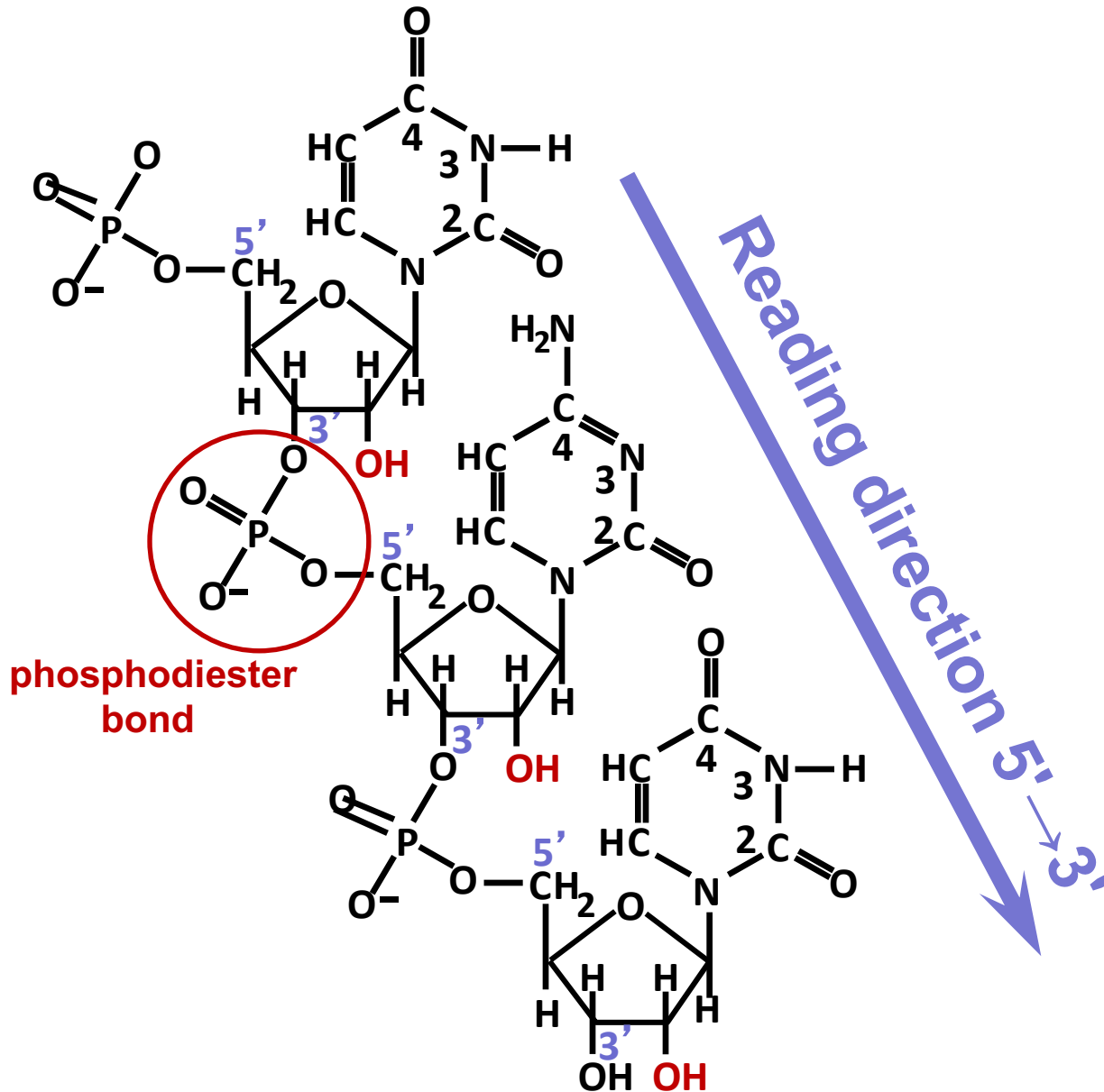


It is the 3D structure of an RNA as well as the dynamics of this structure that gives its function to an RNA molecule

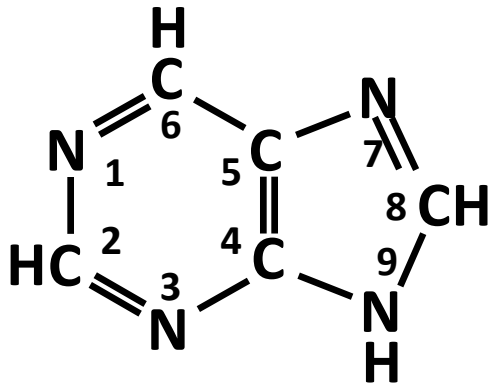
A. Primary structure

III. RNA structure

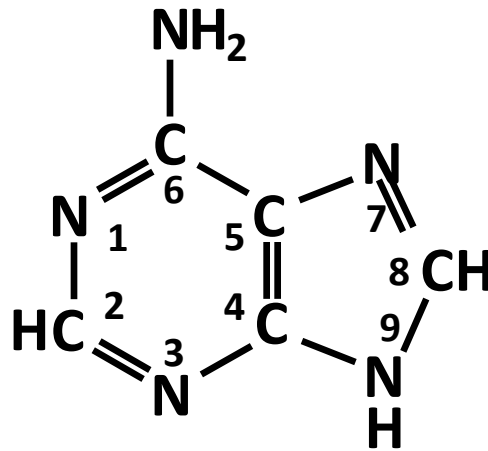
A. Primary structure



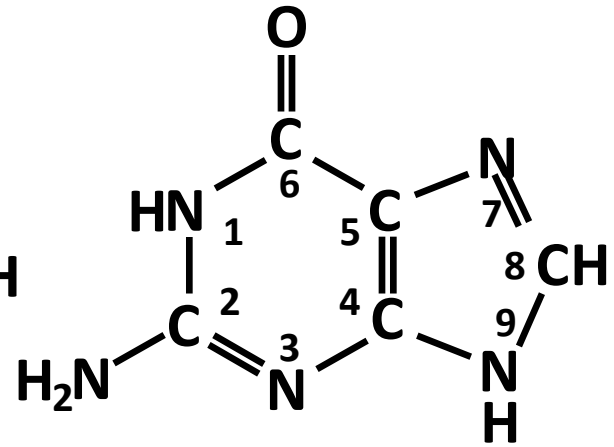
Bases / nucleobases / nitrogen bases



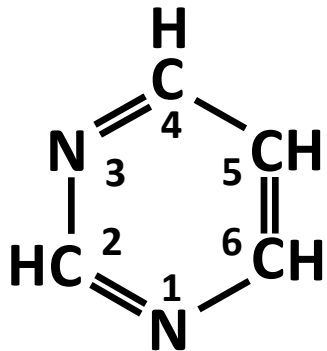
Purine



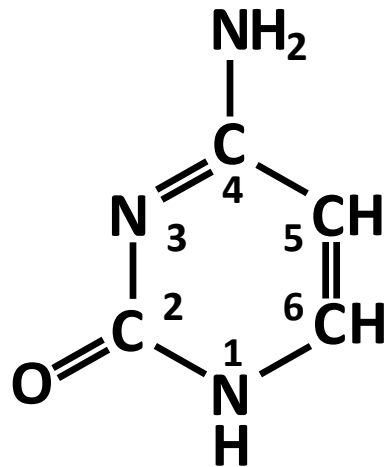
Adenine



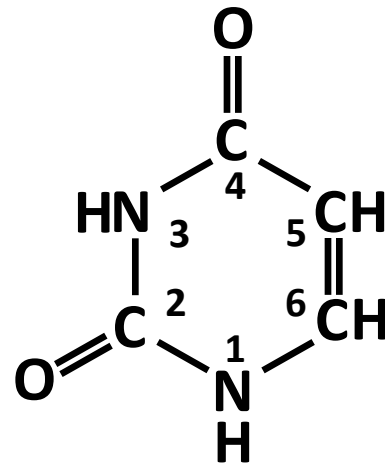
Guanine



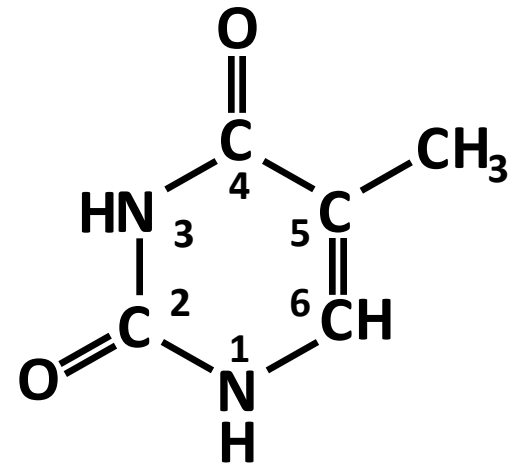
Pyrimidine



Cytosine



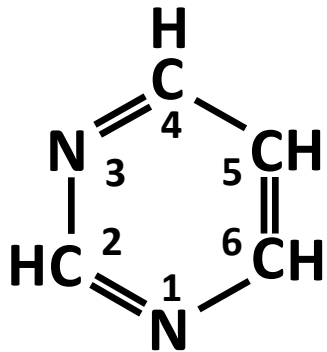
Uracil



Thymine
(5-methyl-uracil)

Bases / nucleobases / nitrogen bases

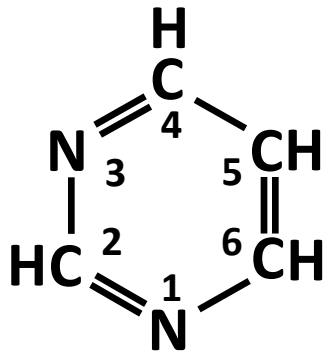
- The pyrimidic ring is a hexagon



Pyrimidine

Bases / nucleobases / nitrogen bases

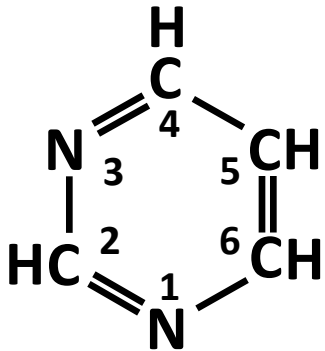
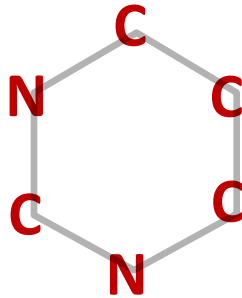
- The pyrimidic ring is a hexagon
- The 2 nitrogens are not side by side



Pyrimidine

Bases / nucleobases / nitrogen bases

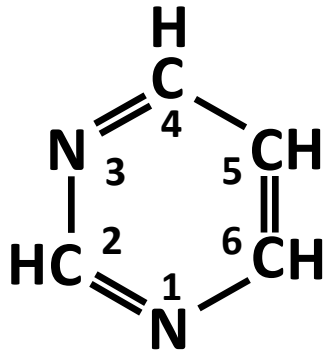
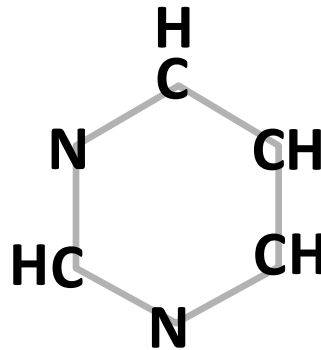
- The pyrimidic ring is a hexagon
- The 2 nitrogens are not side by side



Pyrimidine

Bases / nucleobases / nitrogen bases

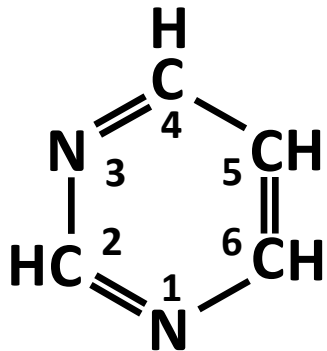
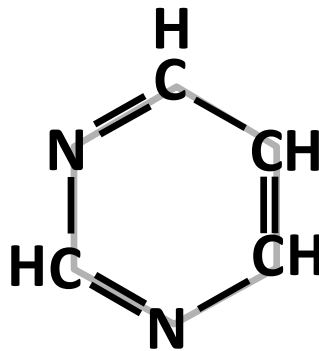
- The pyrimidic ring is a hexagon
- The 2 nitrogens are not side by side
- Nitrogens are imino (N) and carbons CH



Pyrimidine

Bases / nucleobases / nitrogen bases

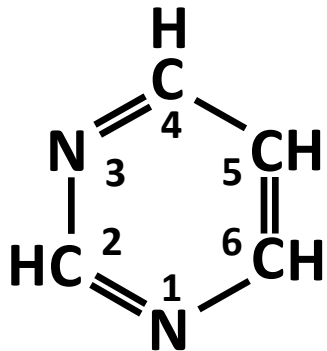
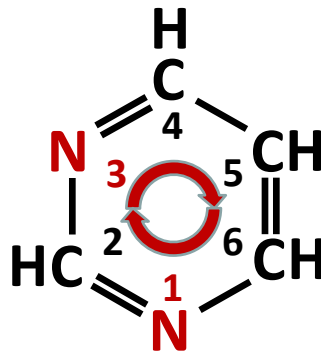
- The pyrimidic ring is a hexagon
- The 2 nitrogens are not side by side
- Nitrogens are imino (N) and carbons CH
- Nitrogen is trivalent, carbon tetra



Pyrimidine

Bases / nucleobases / nitrogen bases

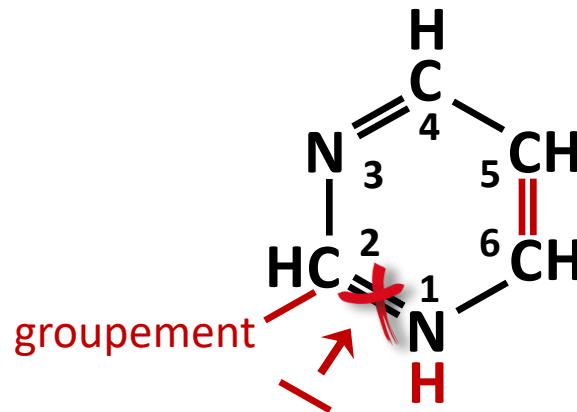
- The pyrimidic ring is a hexagon
- The 2 nitrogens are not side by side
- Nitrogens are imino (N) and carbons CH
- Nitrogen is trivalent, carbon tetra
- We number from the bottom, in the clockwise direction – N1,N3



Pyrimidine

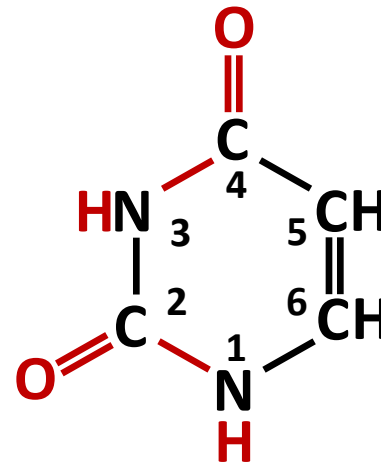
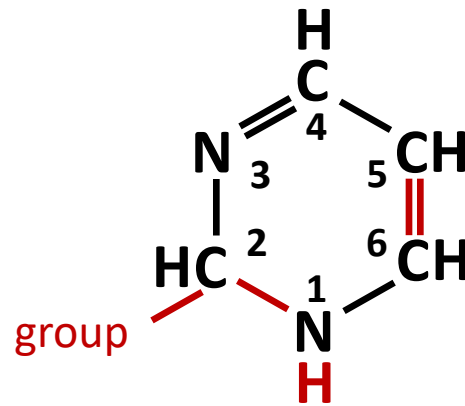
Bases / nucleobases / nitrogen bases

- Substituted groups = carbonyl and amino
- The N1 must be amino to be able to make the β -glycosidic bond
- We never touch the C5=C6 double bond and no substitutions on C5, C6



Bases / nucleobases / nitrogen bases

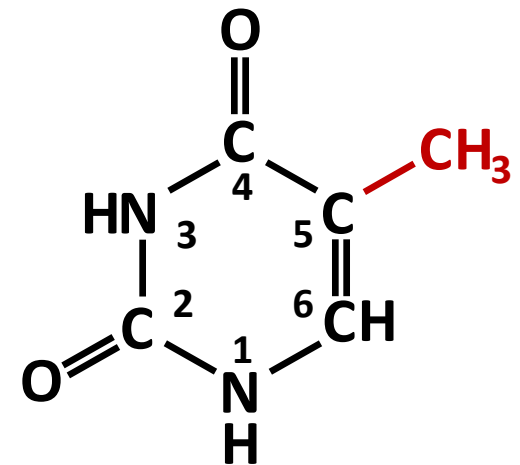
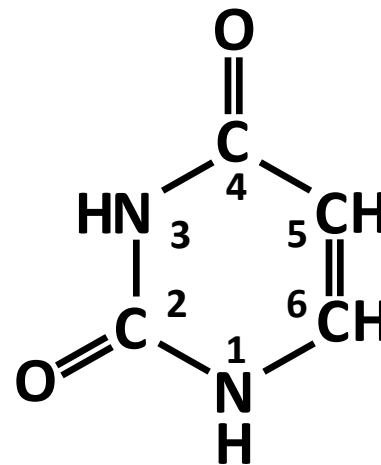
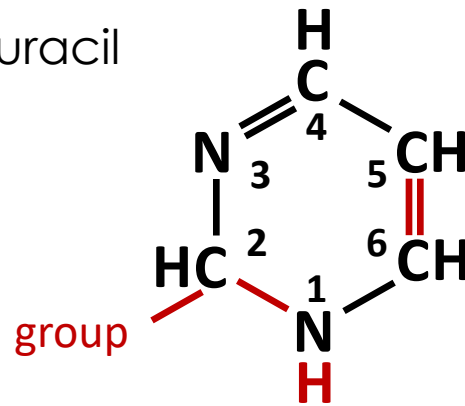
- Substituted groups = carbonyl and amino
- The N1 must be amino to be able to make the β -glycosidic bond
- We never touch the C5=C6 double bond and no substitutions on C5, C6
- Uracil is the simplest: 2 carbonyls (2 and 4), since C is tetravalent the double bond 3=4 has to go



Uracile

Bases / nucleobases / nitrogen bases

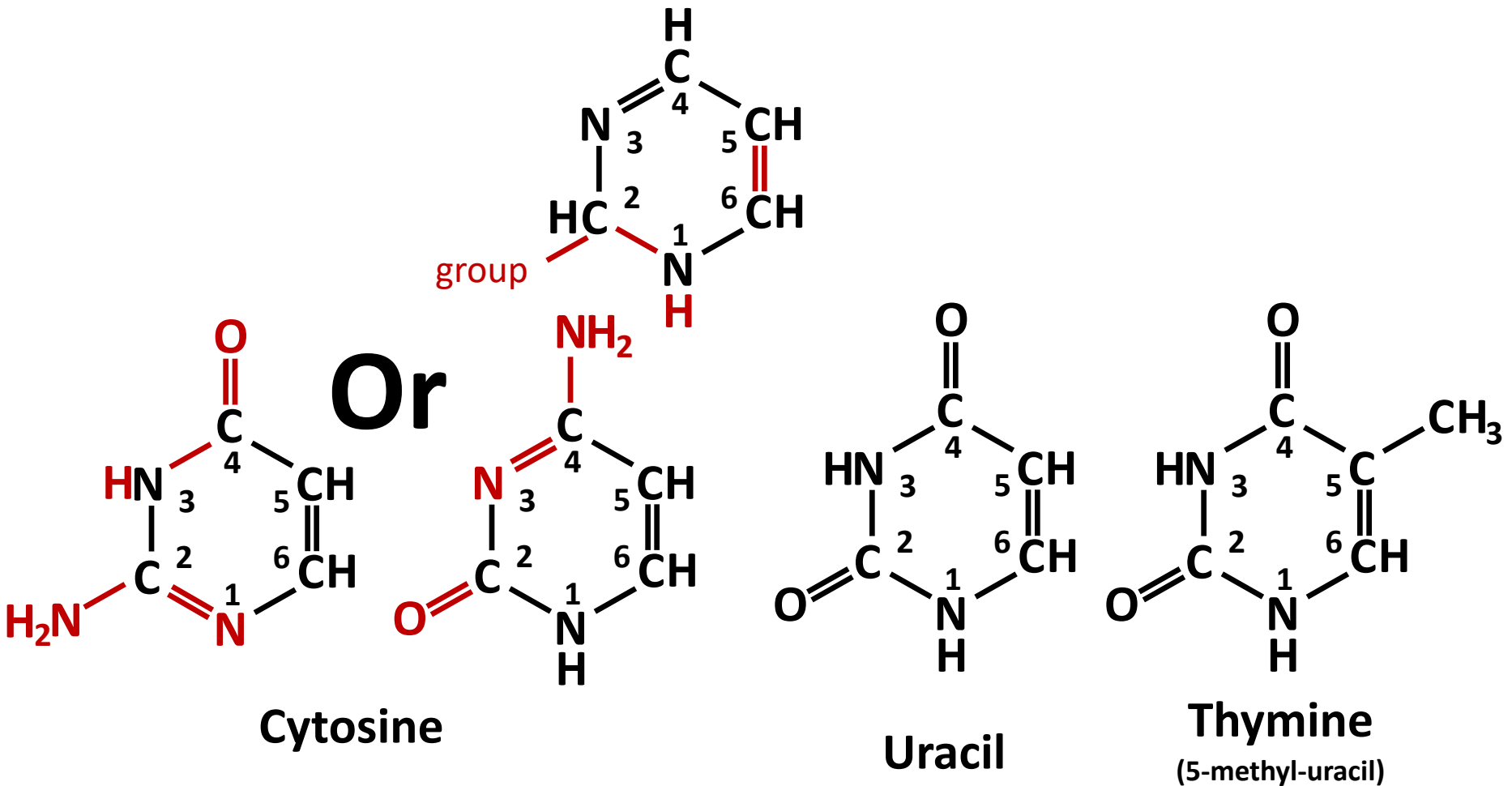
- Substituted groups = carbonyl and amino
- The N1 must be amino to be able to make the β -glycosidic bond
- We never touch the C5=C6 double bond and no substitutions on C5, C6
- Uracil is the simplest: 2 carbonyls (2 and 4), since C is tetravalent the double bond 3=4 has to go
- Thymine = 5-methyl-uracil



(5-methyl-uracil)

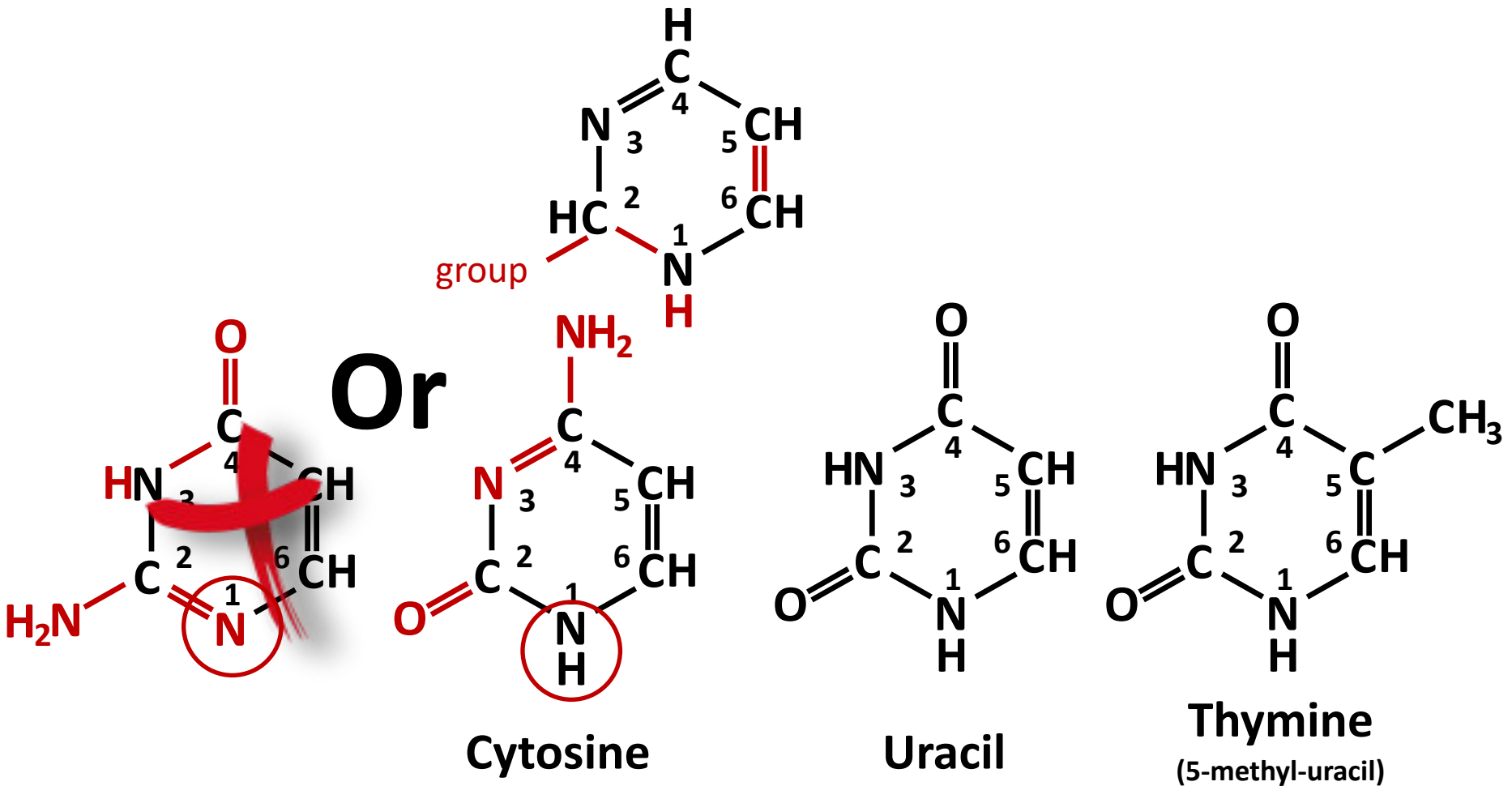
Bases / nucleobases / nitrogen bases

- For cytosine, 1 carbonyl and 1 amino, so 2 possibilities:
carbonyl 2, amino 4 or vice-versa

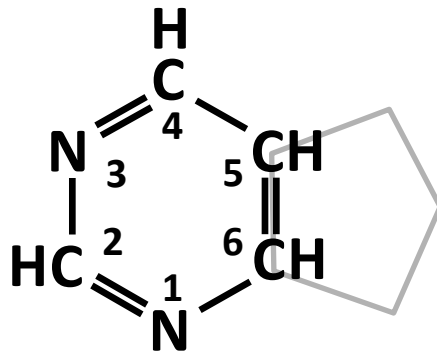


Bases / nucleobases / nitrogen bases

- For cytosine, 1 carbonyl and 1 amino, so 2 possibilities:
carbonyl 2, amino 4 or vice-versa
- The N1 must be amino to be able to make the β -glycosidic bond

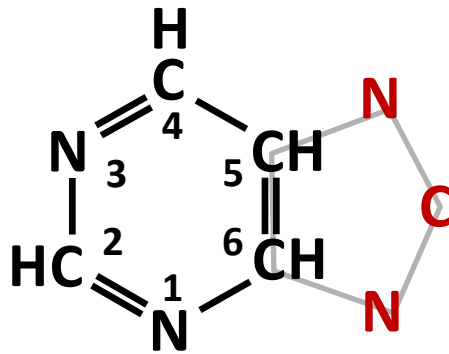


Bases / nucleobases / nitrogen bases



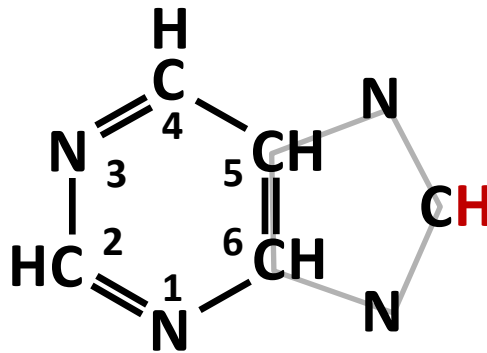
- For the purine, we add a pentagon

Bases / nucleobases / nitrogen bases



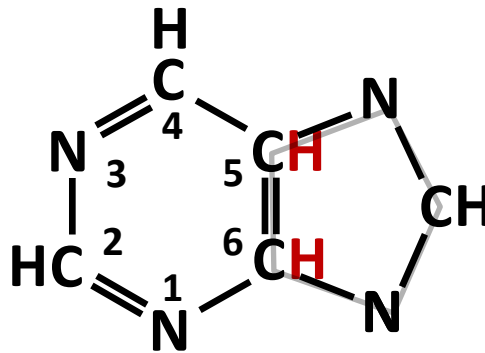
- For the purine, we add a pentagon
- The 2 nitrogens are not side by side

Bases / nucleobases / nitrogen bases



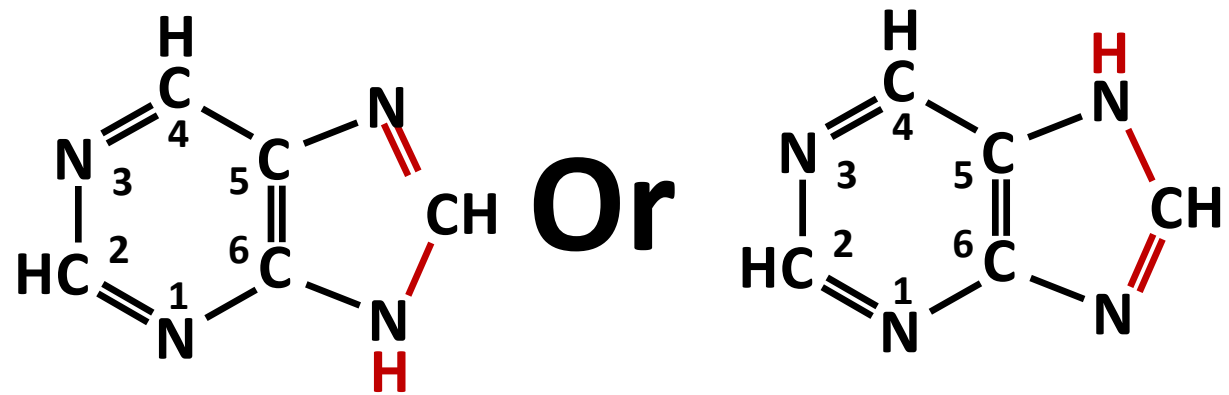
- For the purine, we add a pentagon
- The 2 nitrogens are not side by side
- Carbon is CH

Bases / nucleobases / nitrogen bases



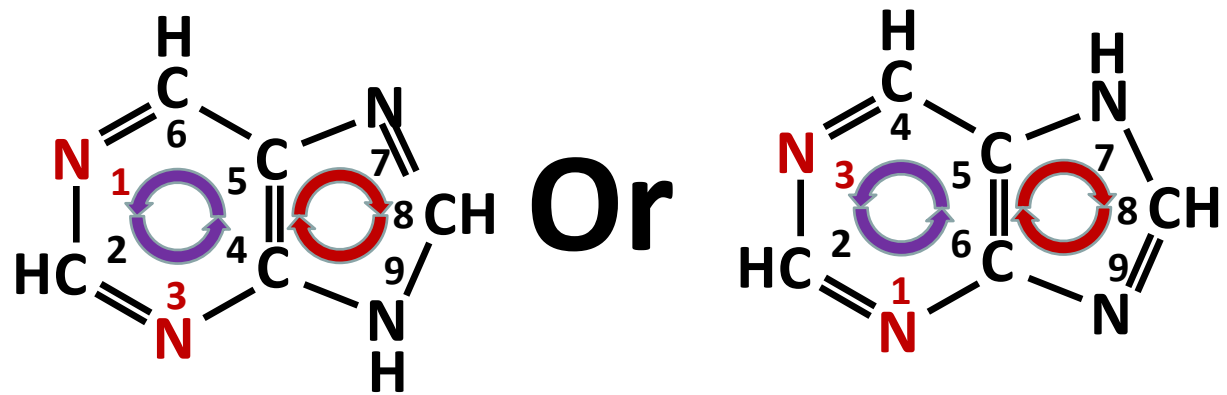
- For the purine, we add a pentagon
- The 2 nitrogens are not side by side
- Carbon is CH
- Carbon is tetravalent

Bases / nucleobases / nitrogen bases



- For the purine, we add a pentagon
- The 2 nitrogens are not side by side
- Carbon is CH
- Carbon is tetravalent
- Carbon is always tetravalent and nitrogen trivalent so 2 possibilities

Bases / nucleobases / nitrogen bases

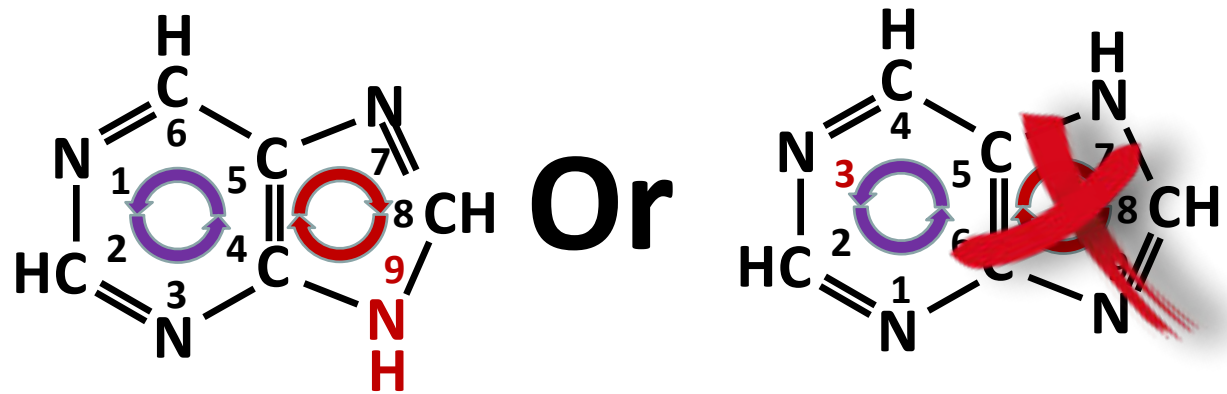


- We change the direction of the numbering: Counterclockwise in the hexagonal ring plus we start with the other nitrogen **BUT WE KEEP N1, N3** and we change again numbering direction in the pentagon

- For the purine, we add a pentagon
- The 2 nitrogens are not side by side
- Carbon is CH
- Carbon is tetravalent
- Carbon is always tetravalent and nitrogen trivalent so 2 possibilities

Bases / nucleobases / nitrogen bases

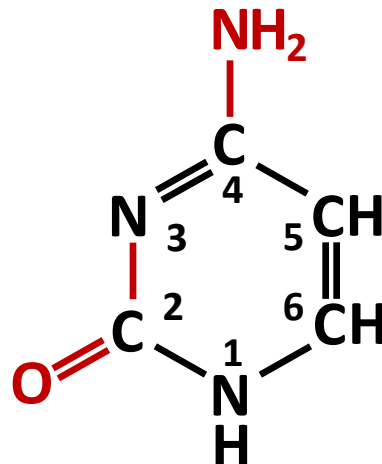
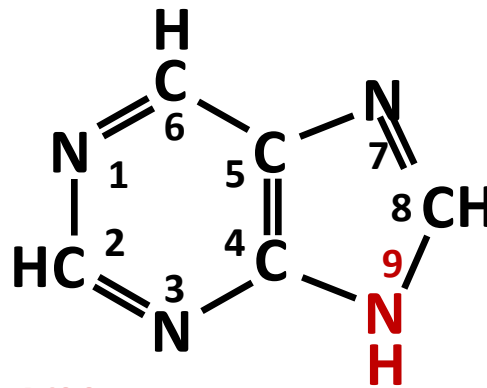
- The pentagon's amine is 9 because it's nitrogen has to make the β -glycosidic bond



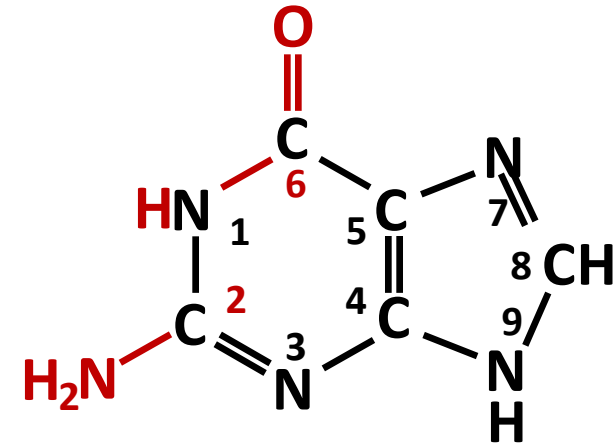
- We change the direction of the numbering: Counterclockwise in the hexagonal ring plus we start with the other nitrogen **BUT WE KEEP N1, N3** and we change again numbering direction in the pentagon

- For the purine, we add a pentagon
- The 2 nitrogens are not side by side
- Carbon is CH
- Carbon is tetravalent
- Carbon is always tetravalent and nitrogen trivalent so 2 possibilities

Bases / nucleobases / nitrogen bases



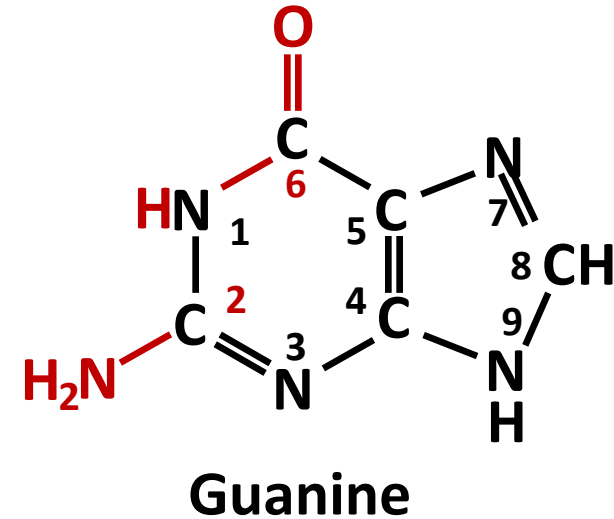
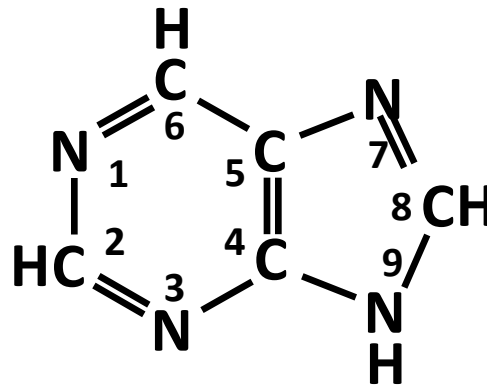
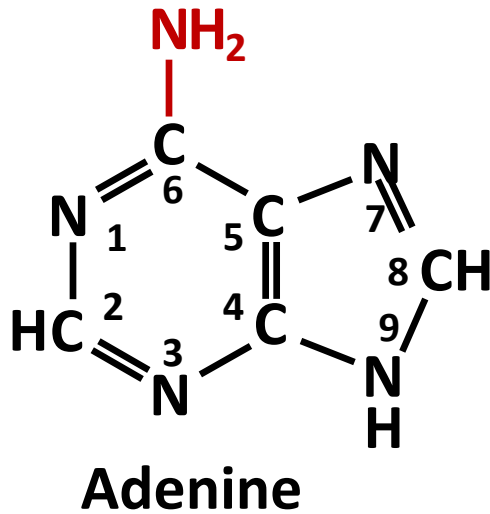
Cytosine



Guanine

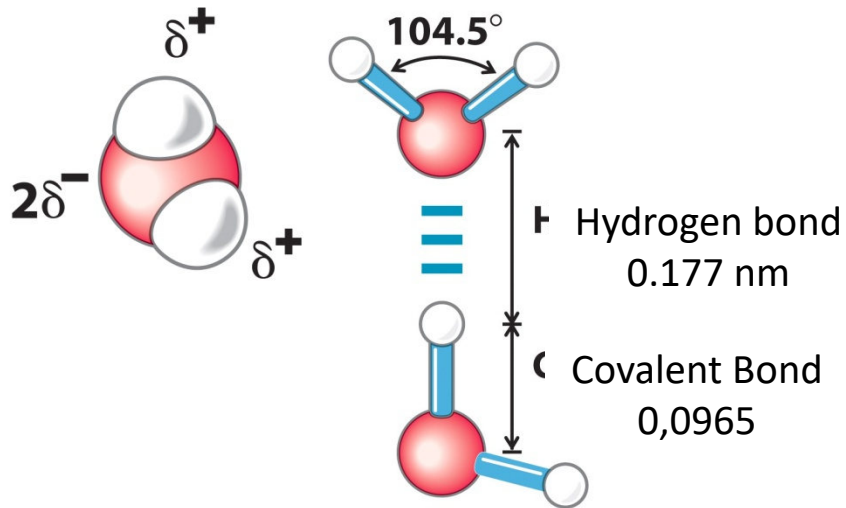
- Unlike pyrimidine, the N that makes the connection to the ribose is already amine because we do not substitute not the imidazole ring (pentagon)
- The substituting groups = carbonyl and amino
- Guanine pairs with cytosine and it is the reverse (complementary) cytosine substitutions

Bases / nucleobases / nitrogen bases



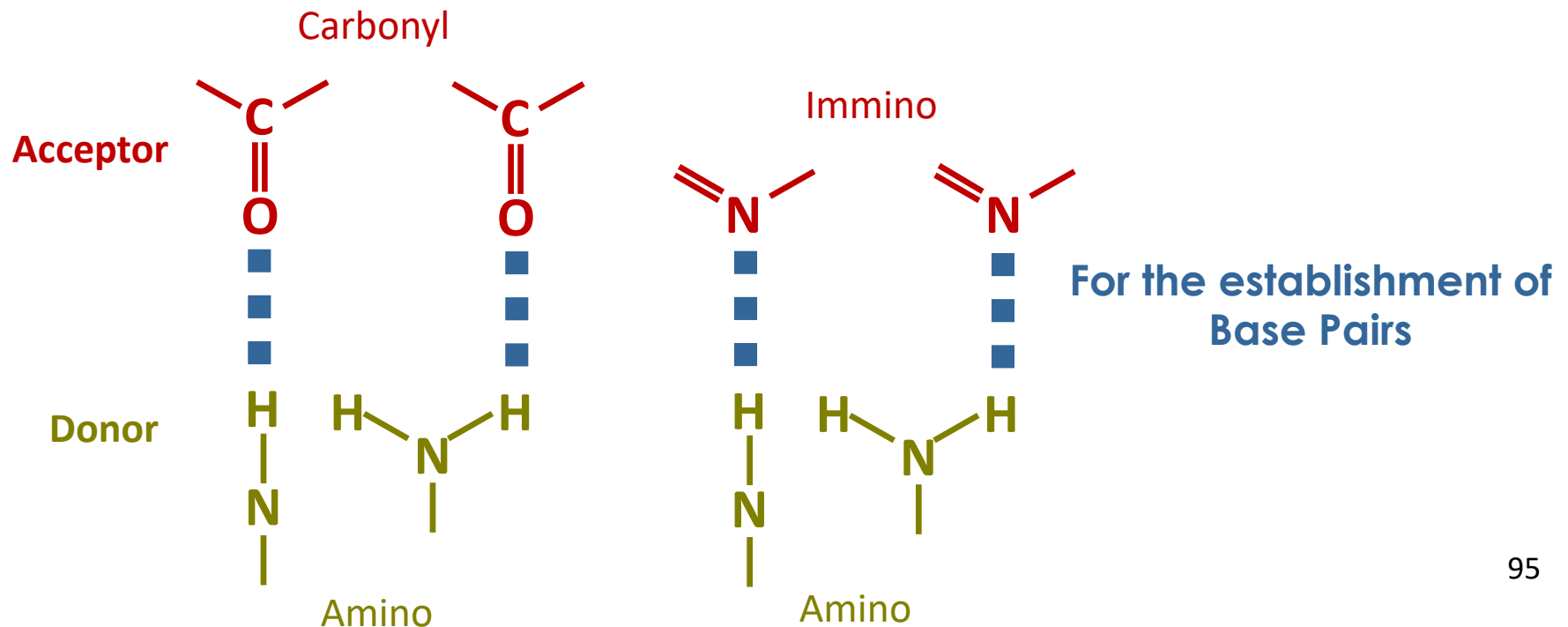
- Adenine, a single substitution: NH_2 and not in the same position as guanine
or
6-amino-purine
- Unlike pyrimidine, the N that makes the connection to the ribose is already amine because we do not substitute not the imidazole ring (pentagon)
- The substituting groups = carbonyl and amino
- Guanine pairs with cytosine and it is the reverse (complementary) cytosine substitutions

Base pairs



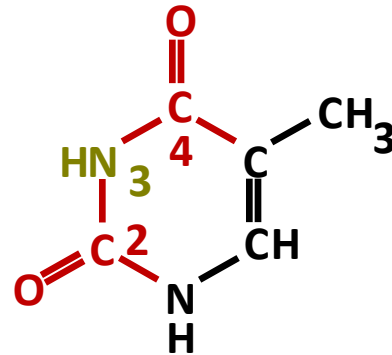
The hydrogen bond is a non-covalent physical bond, of the dipole-dipole type. It is of low energy (twenty times weaker than conventional covalent bond), and connects molecules by involving a hydrogen atom. A hydrogen bonding donor and an acceptor are required:

- the donor is a compound: nitrogen, oxygen, fluorine, carrying a hydrogen atom (as in amines, alcohols, thiols);
- The acceptor is a compound (only nitrogen, oxygen or fluorine) carrying free pairs.

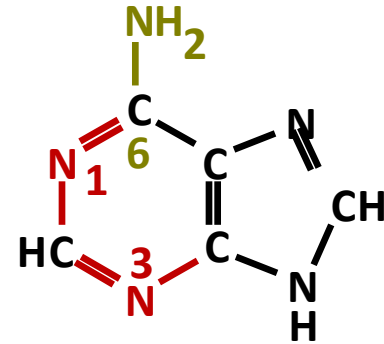


Base pairs

Thymine

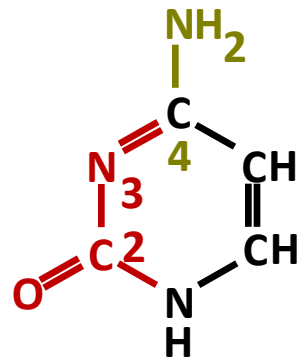


Adenine

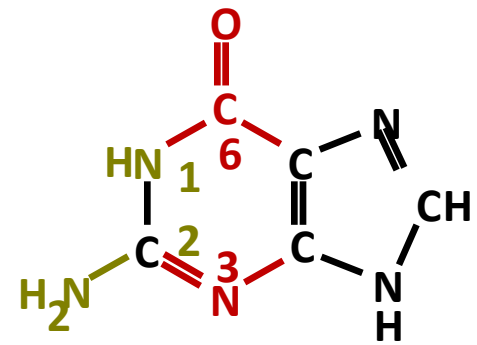


Groups potentially involved

Donor
Acceptor of H bond



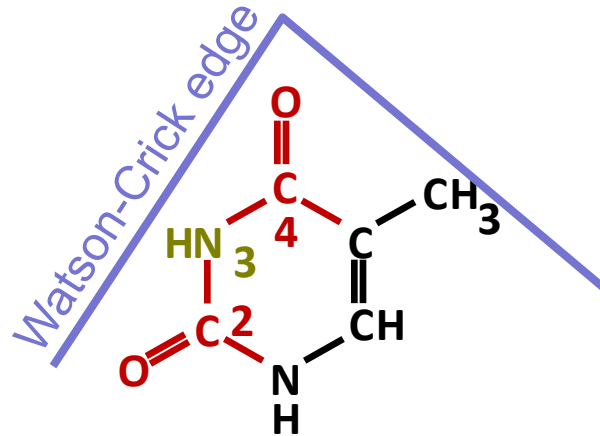
Cytosine



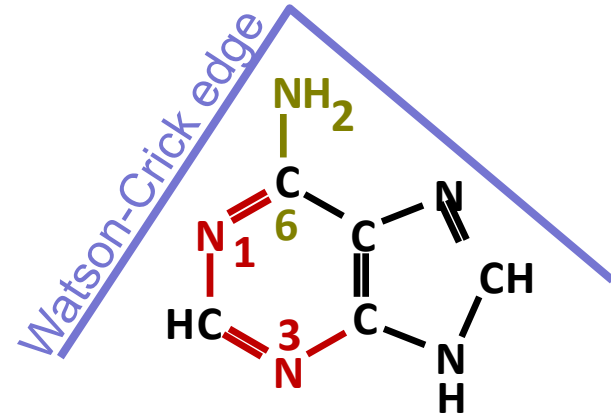
Guanine

Base pairs

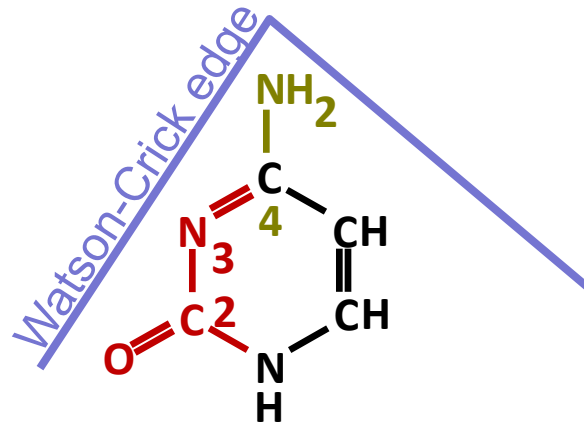
Thymine



Adenine

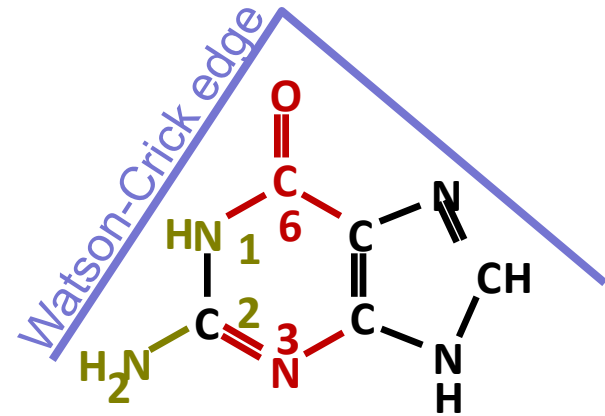


- The Watson-Crick edge



Cytosine

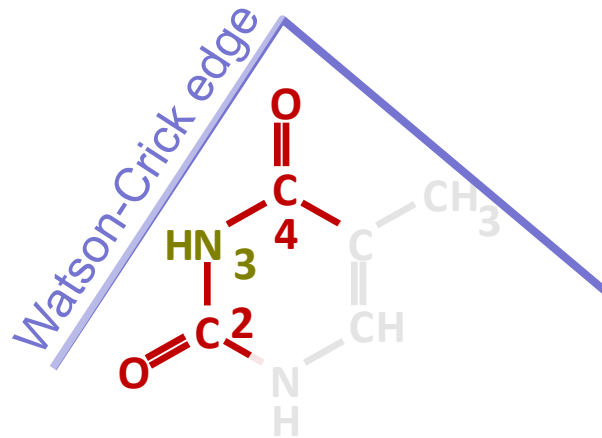
Donor
Acceptor of H bond



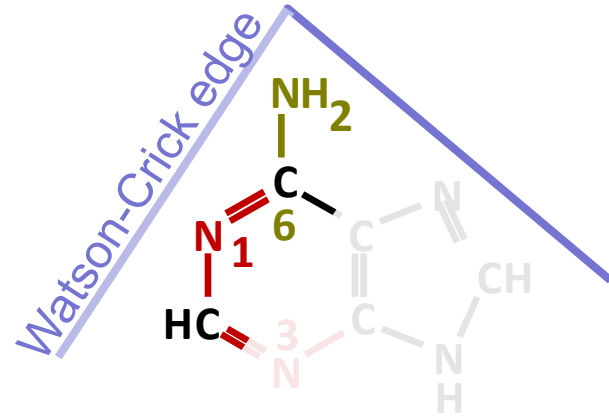
Guanine

Base pairs

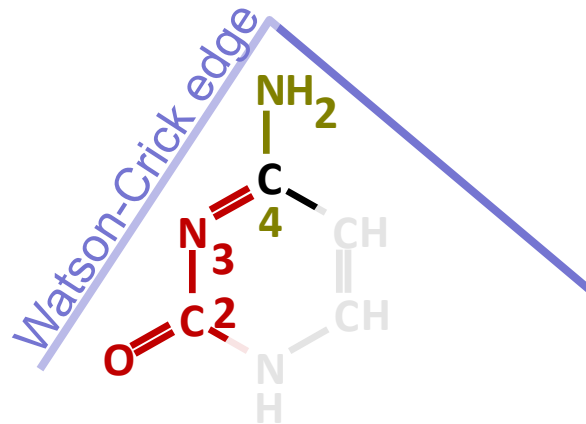
Thymine



Adenine

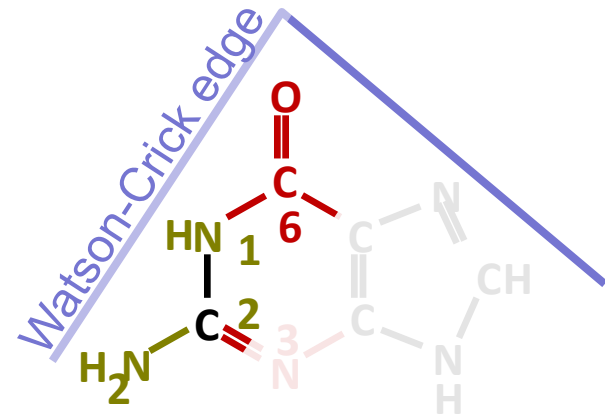


- The Watson-Crick edge



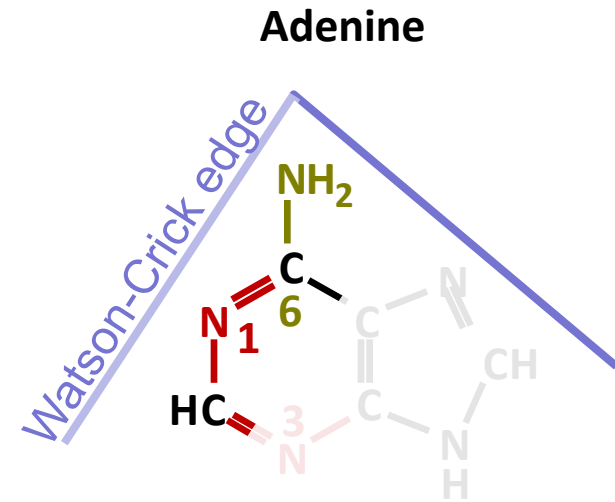
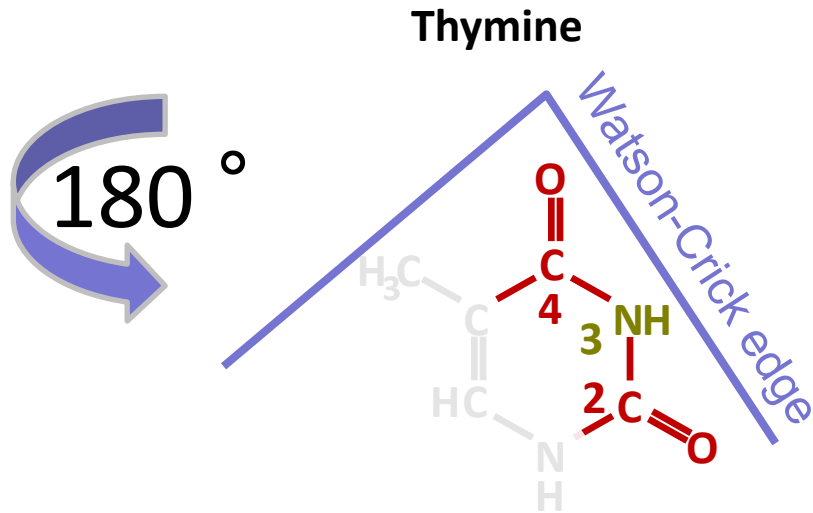
Cytosine

Donor
Acceptor of H bond

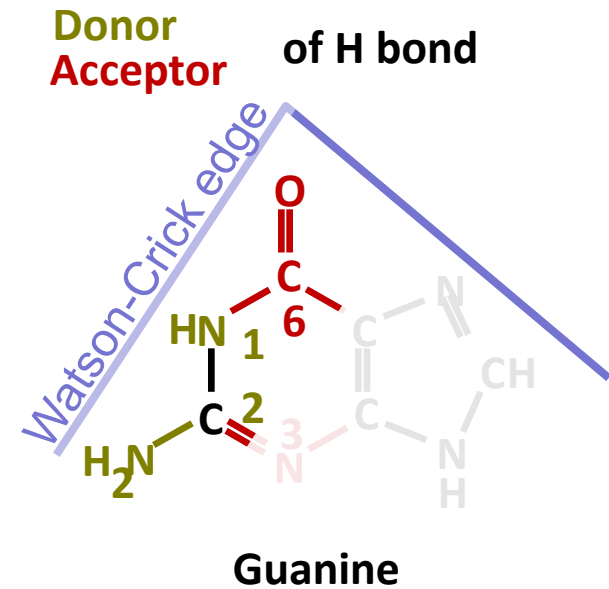
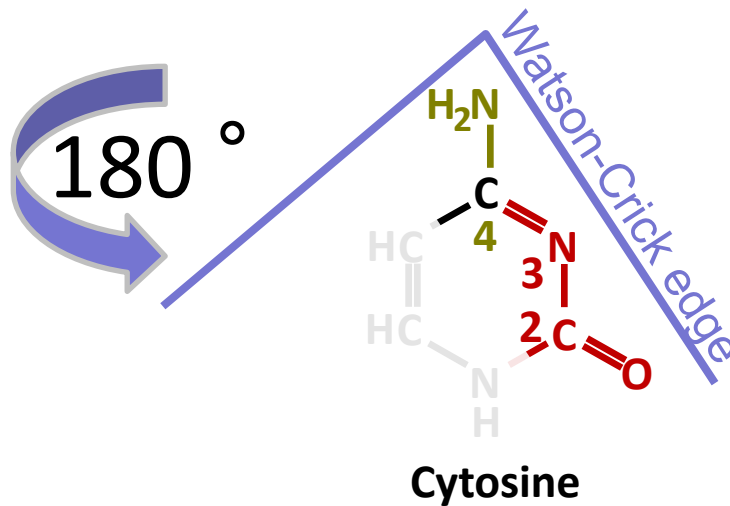


Guanine

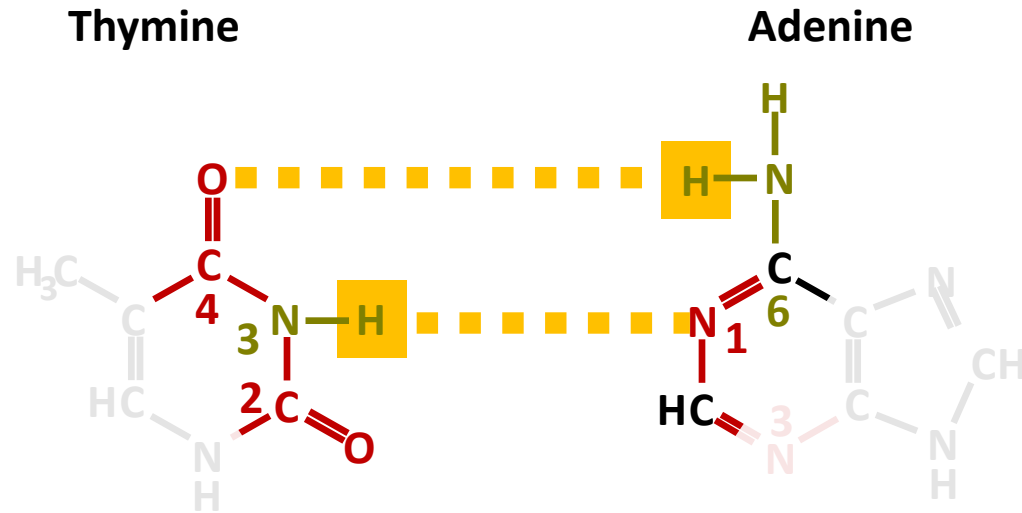
Base pairs



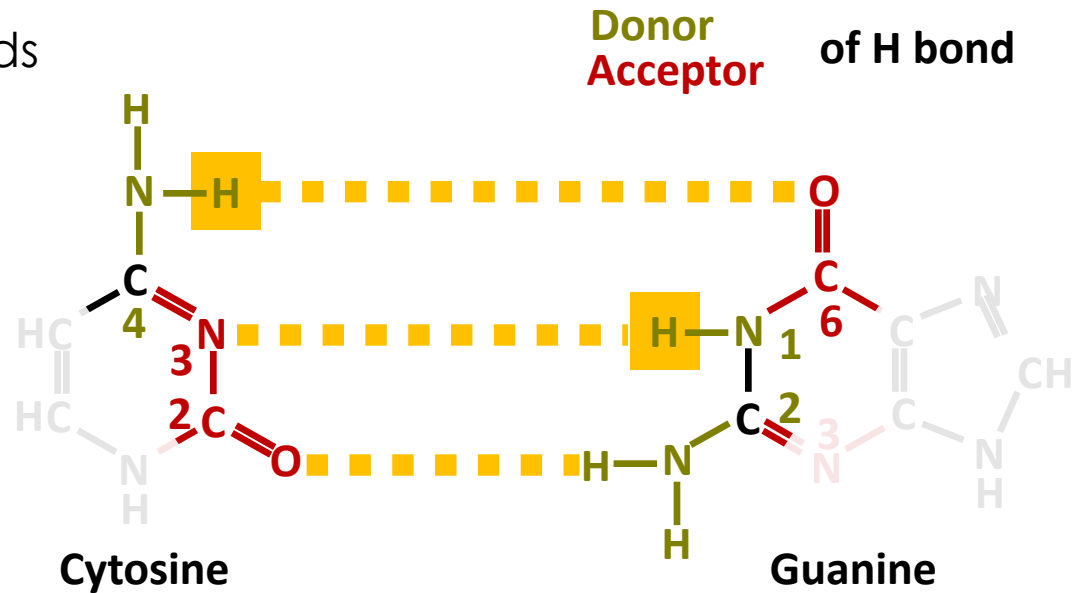
- The Watson-Crick edge



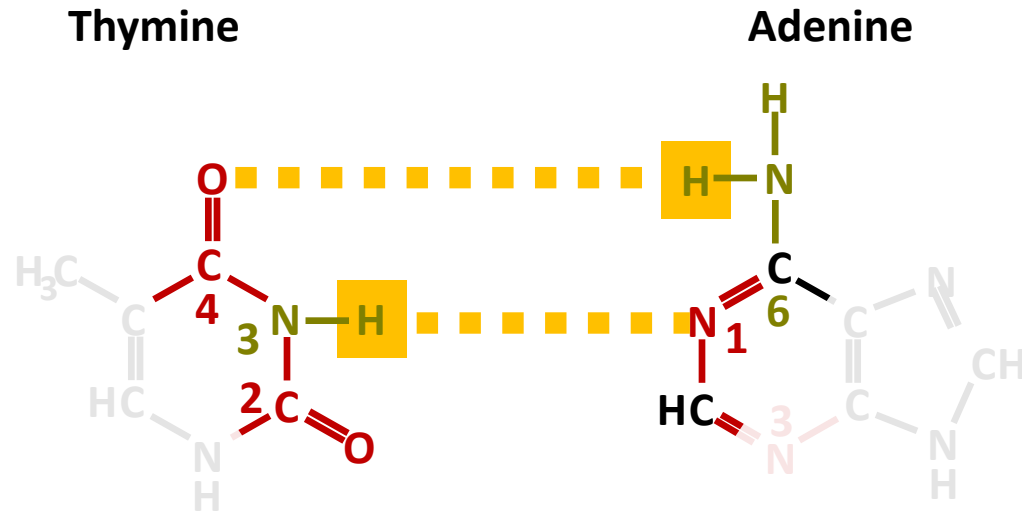
Base pairs



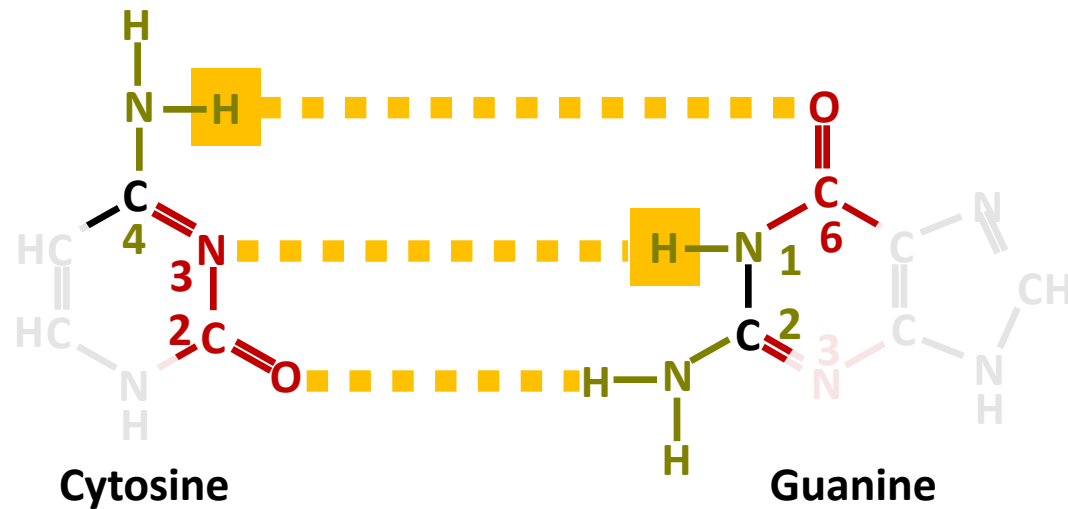
- Possible H-bonds



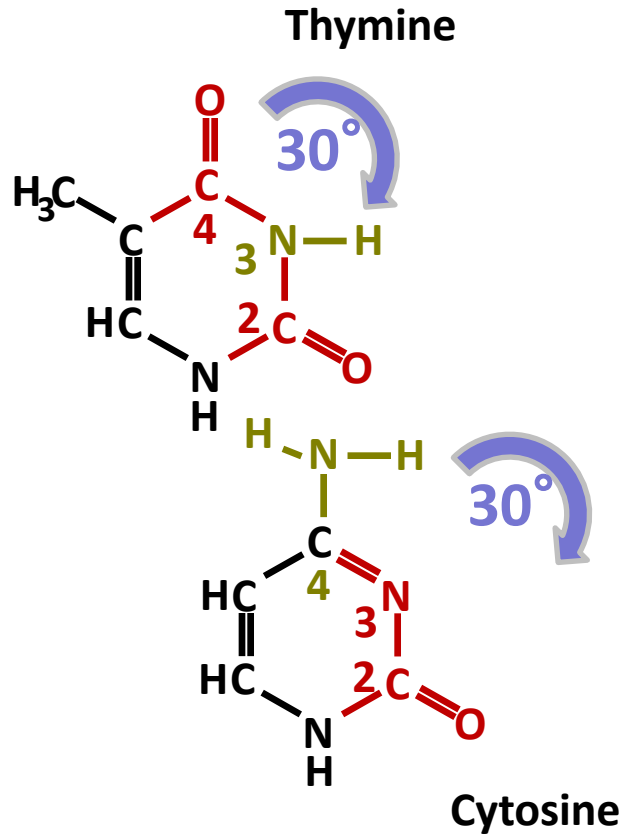
Base pairs



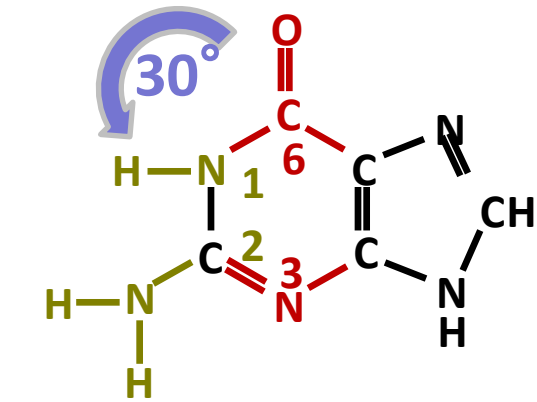
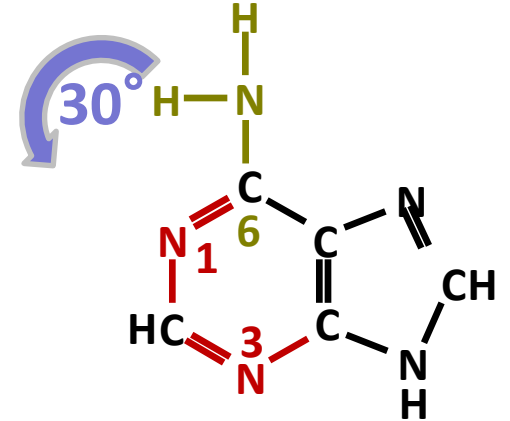
Always N1-H-N3



Base pairs

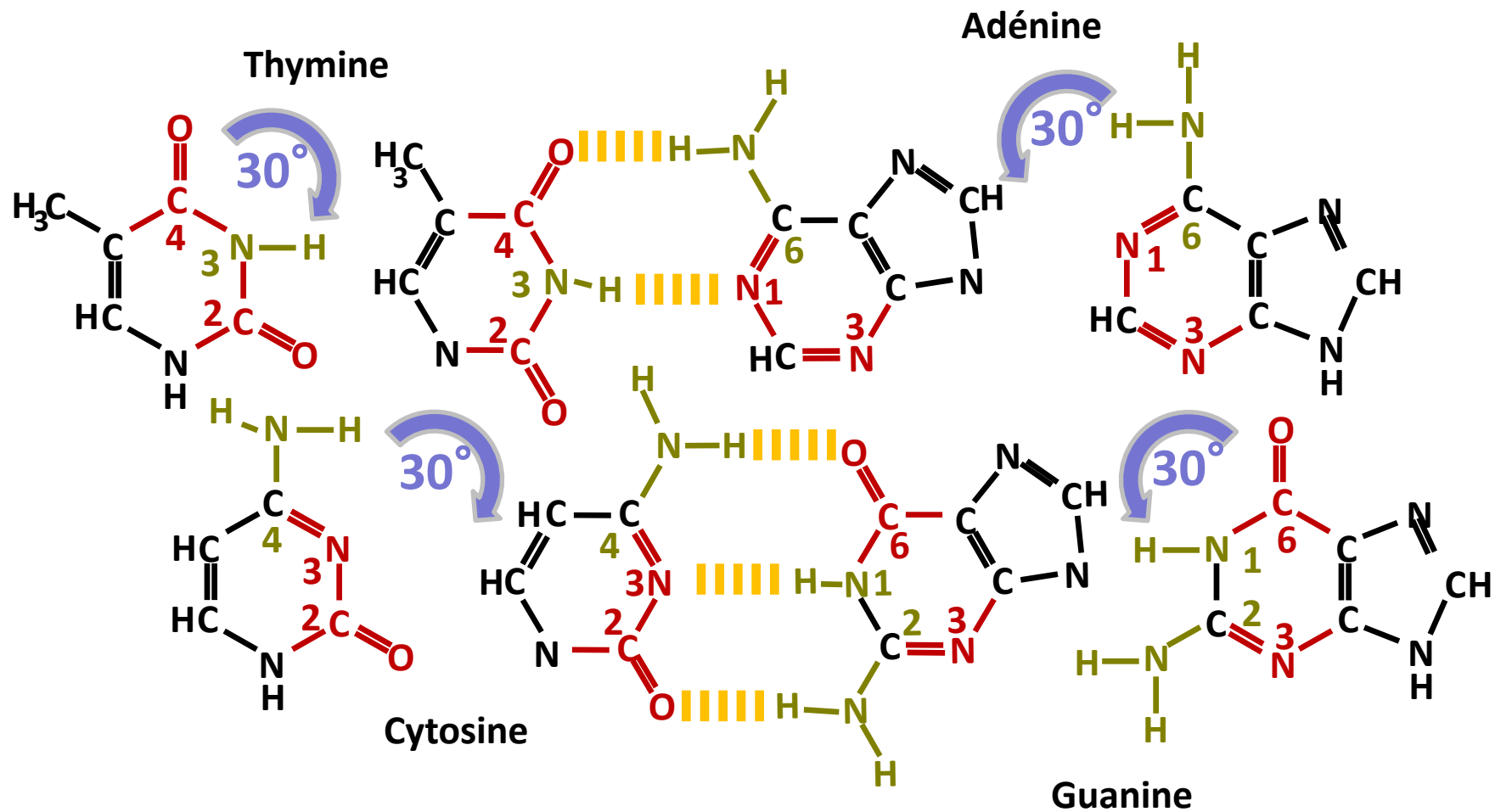


Adenine

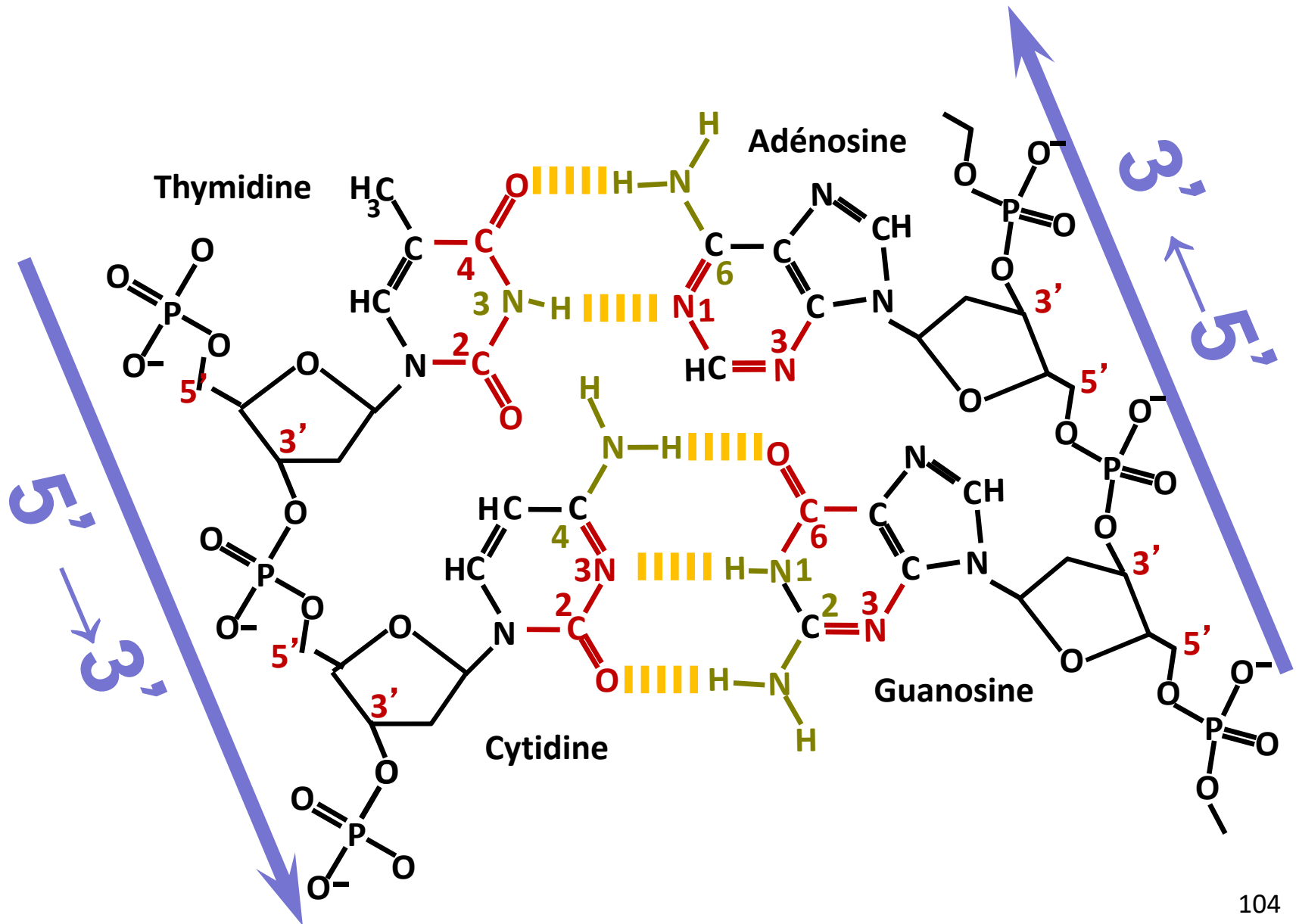


Guanine

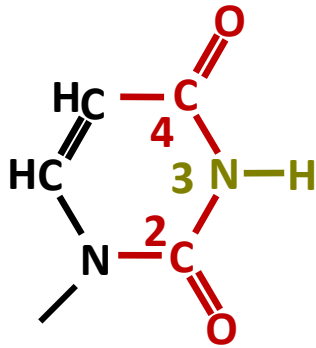
Base pairs



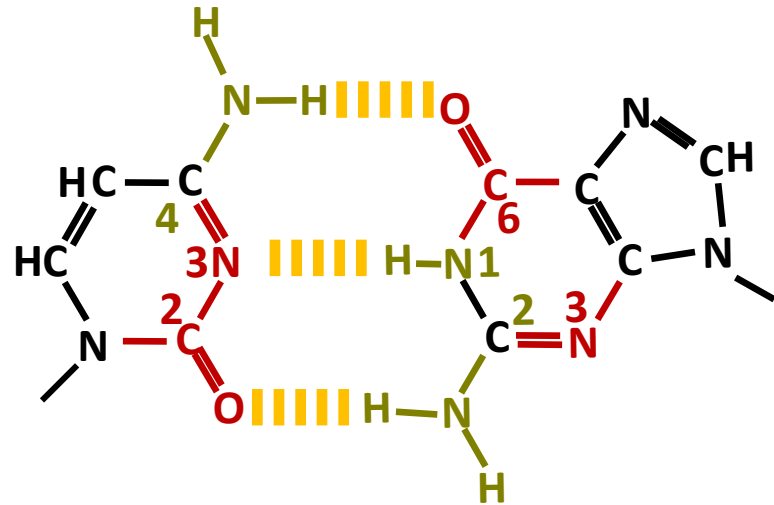
Base pairs



Base pairs

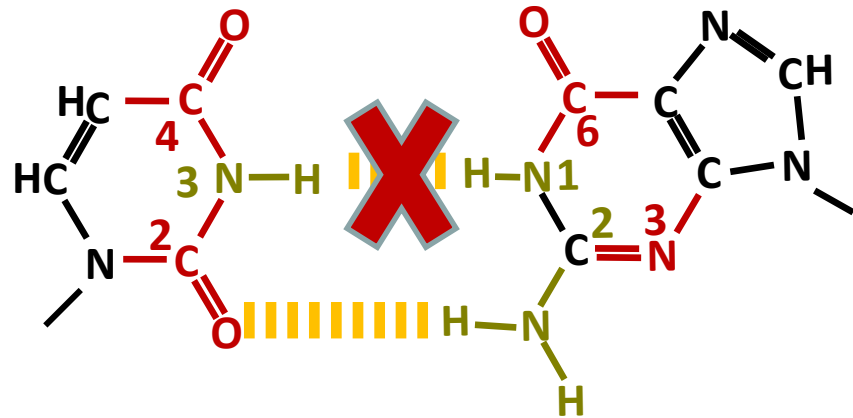


U



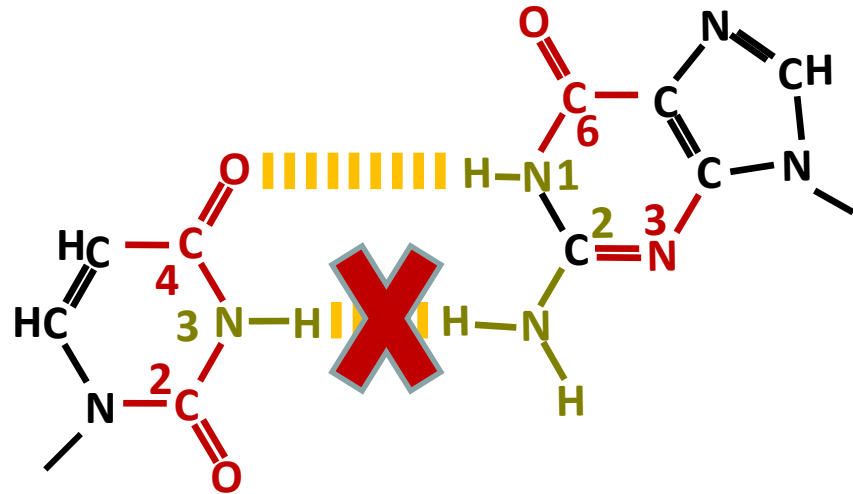
C ≡ G

Base pairs



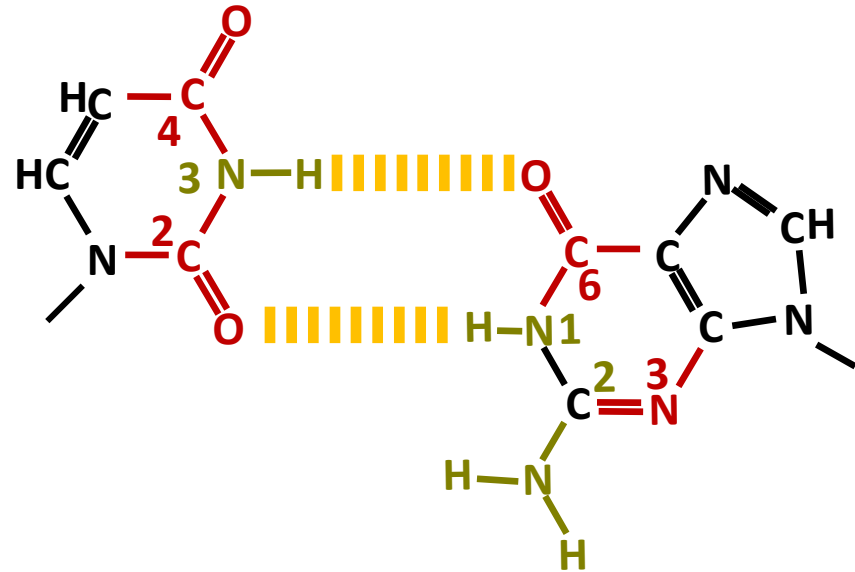
U  G

Base pairs



U G

Base pairs

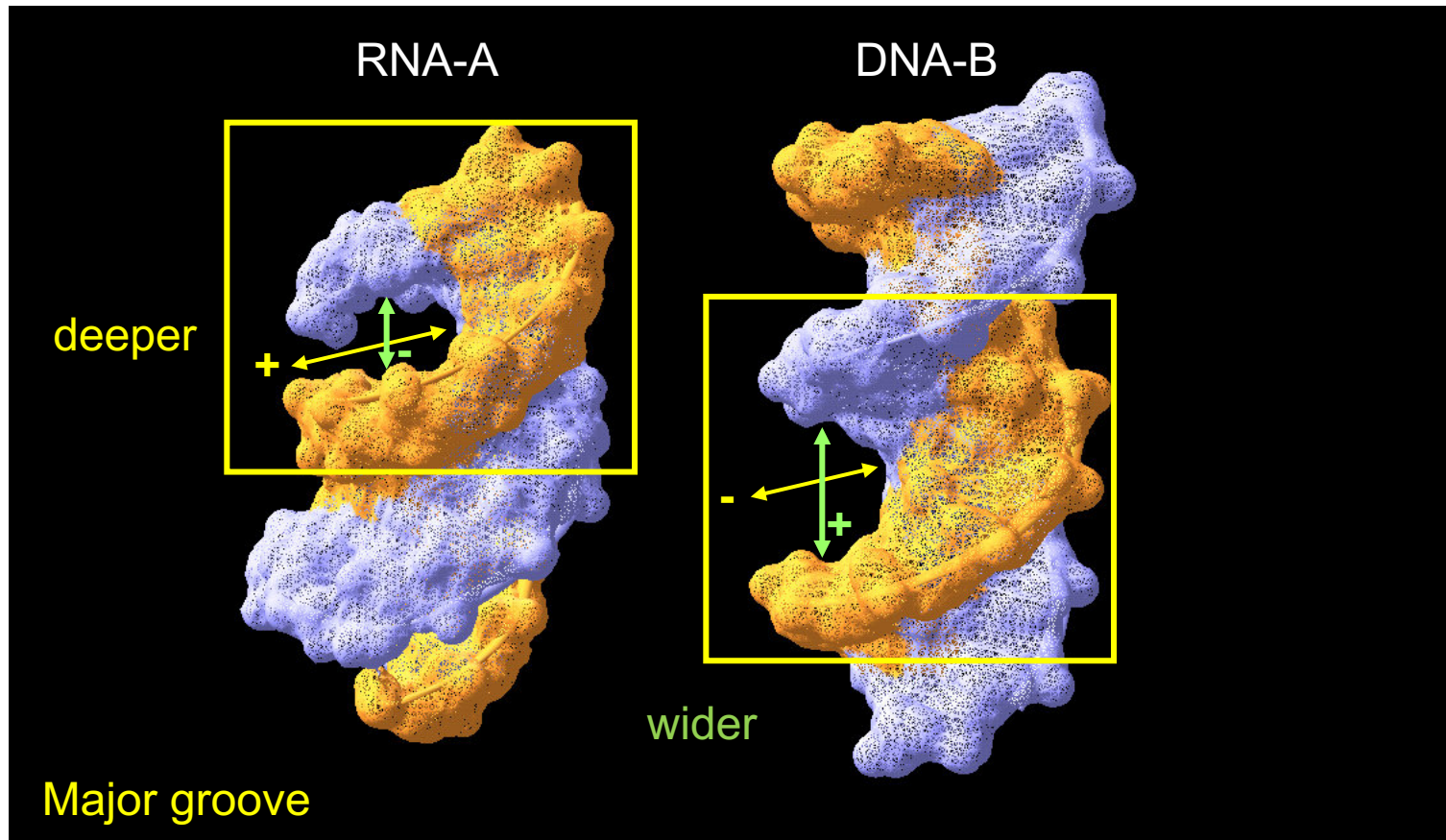


U G

III. RNA structure

B. Secondary structure

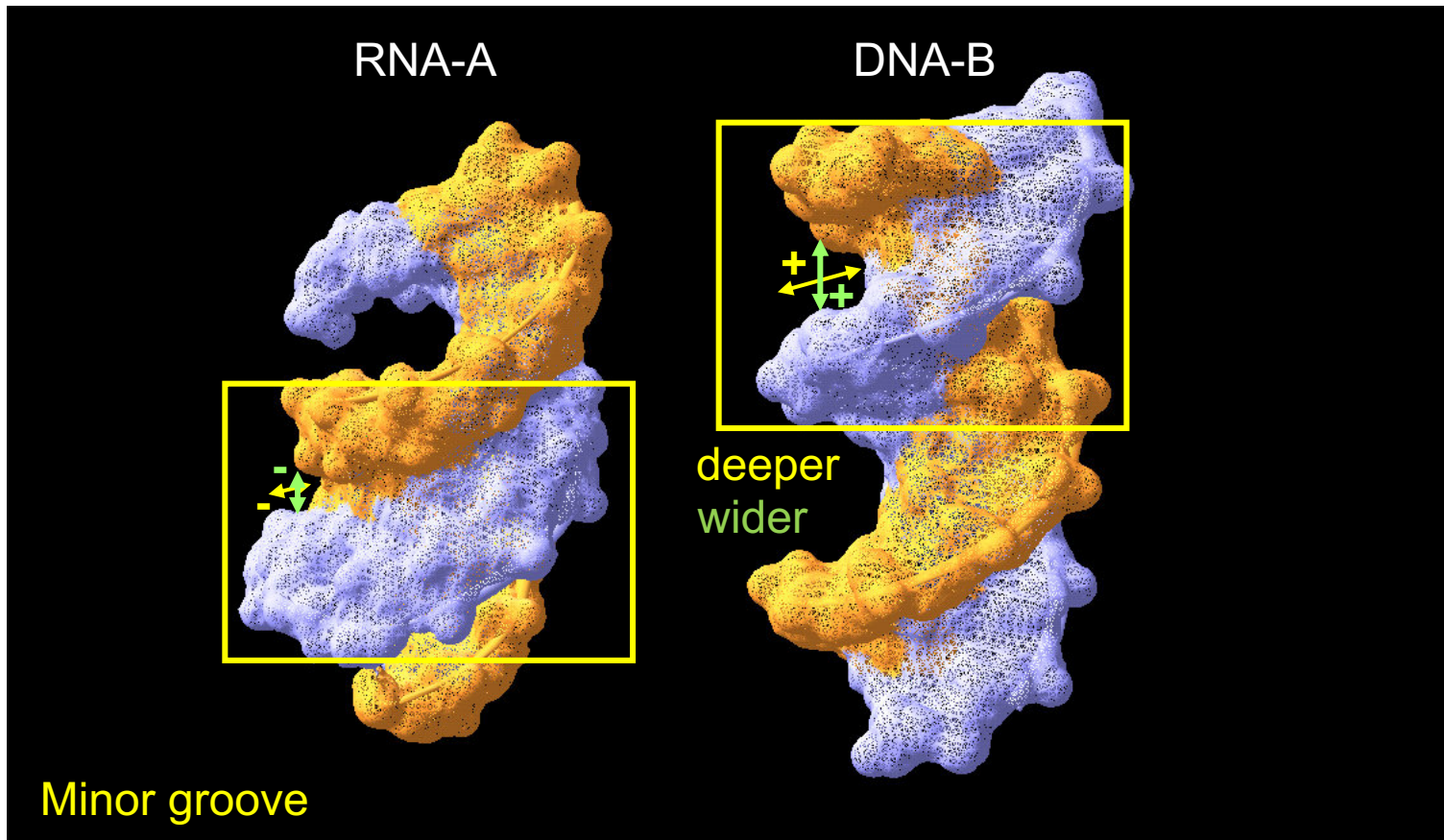
- Comparison between DNA Helix B and RNA Helix A



III. RNA structure

B. Secondary structure

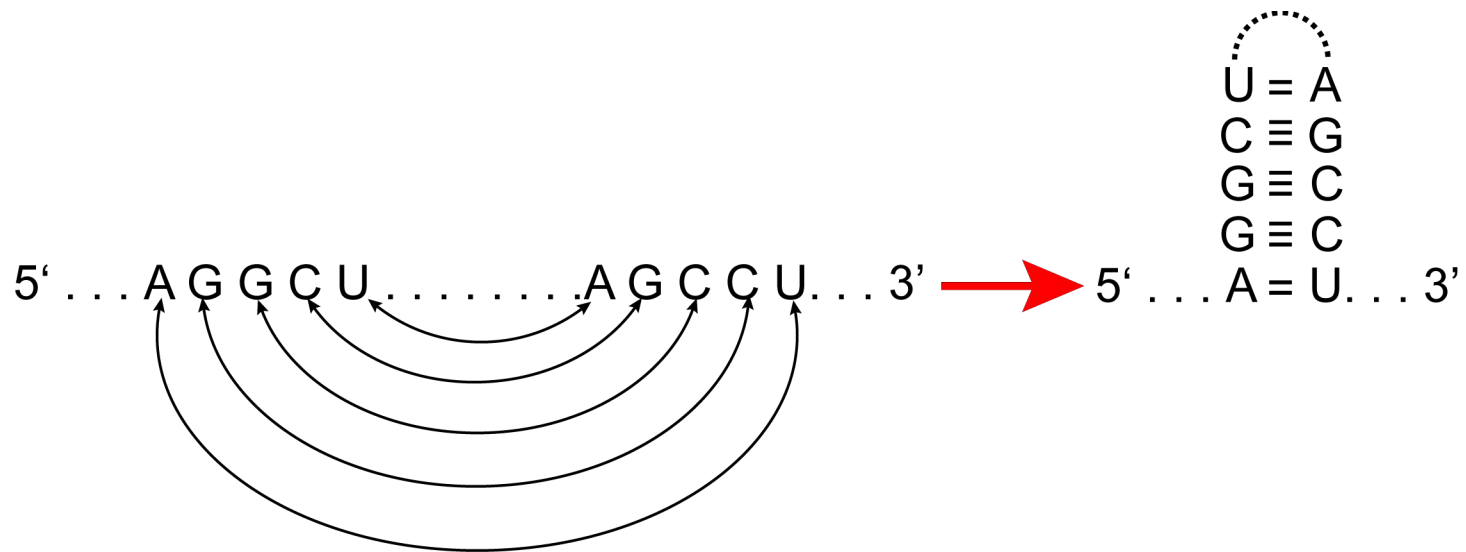
- Comparison between DNA Helix B and RNA Helix A



III. RNA structure

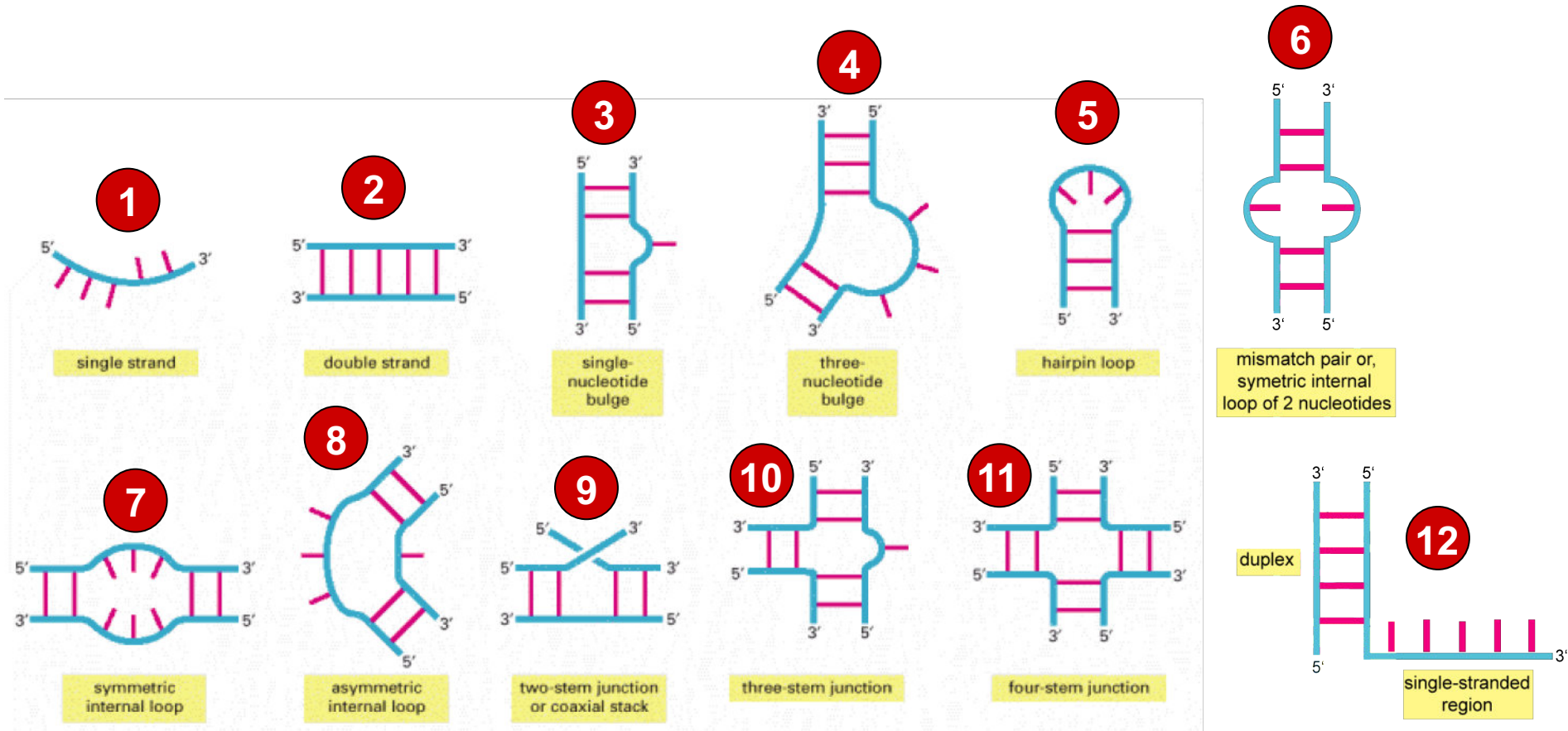
B. Secondary structure

- The RNA chain folds in space with a so-called spongy structure



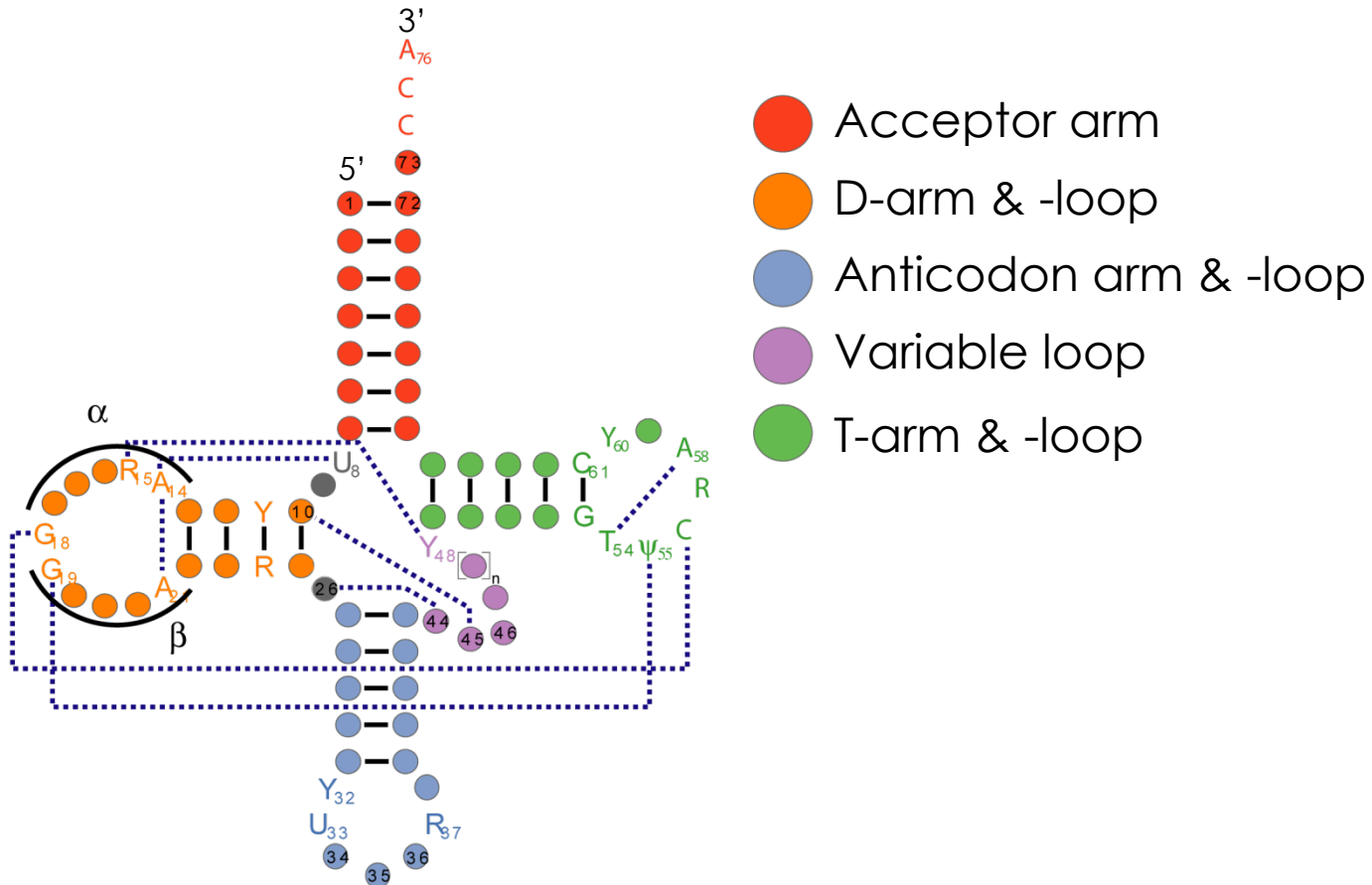
III. RNA structure

B. Secondary structure



III. RNA structure

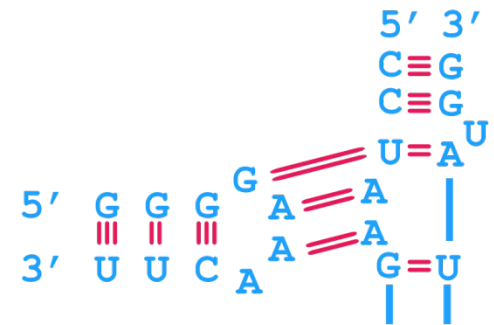
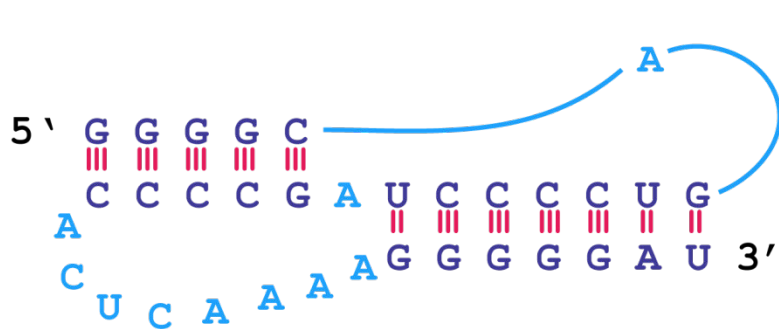
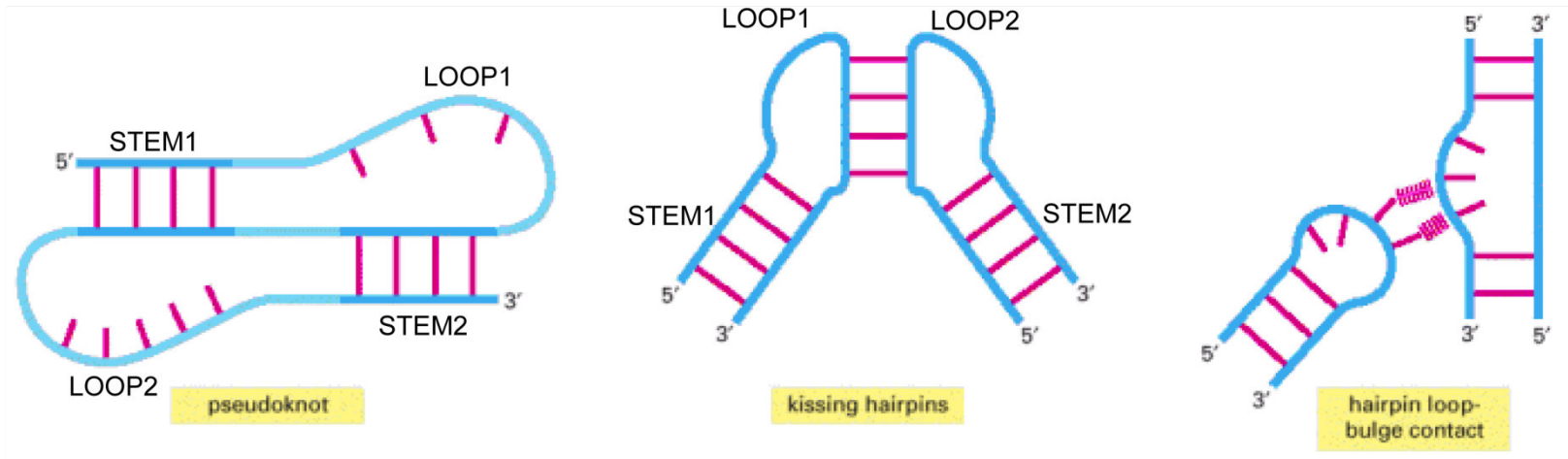
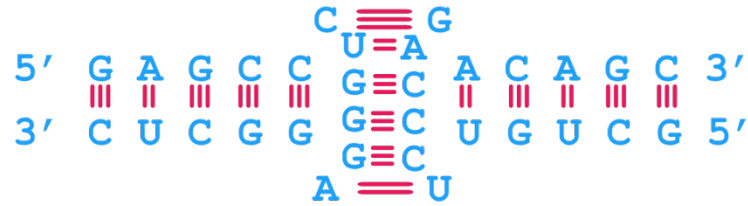
B. Secondary structure



Secondary cloverleaf structure of tRNA. Numbering of Nt is according to Steinberg et al., 1993. Nt highlighted in the structure are conserved accross all tRNA sequences. Abbreviations: Y: pyrimidine, R: purine, A: adenosine, C: cytidine, G: guanosine, U: uracil., T: thymidine, Ψ: pseudouridine. WC base pairs are schematized by black lines and tertiary interactions allowing folding of the cloverleaf structure into the tertiary L-shape structure are dashed.

III. RNA structure

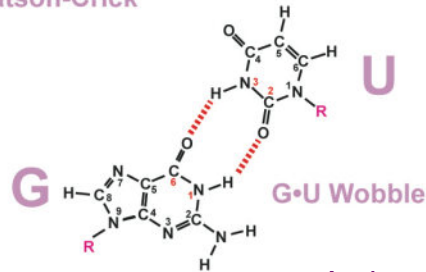
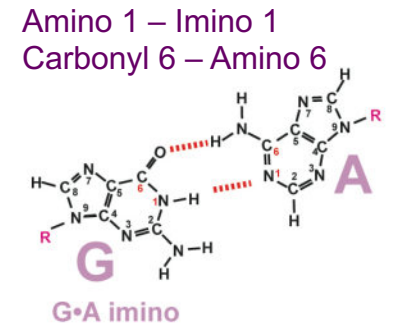
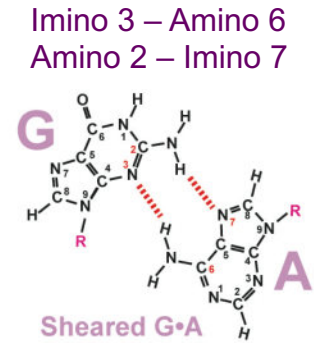
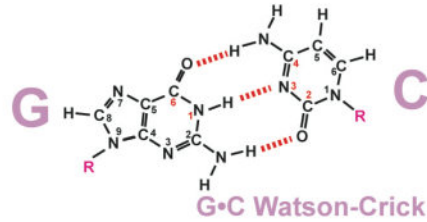
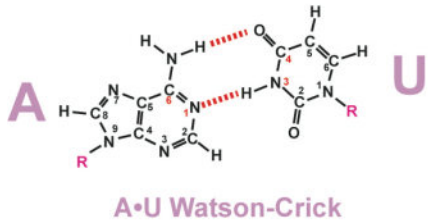
C. Tertiary structure



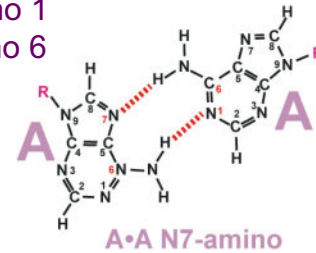
III. RNA structure

C. Tertiary structure

- The formation of these tertiary structural elements calls for many interactions between bp that are not classical, they are non-canonical pairings.



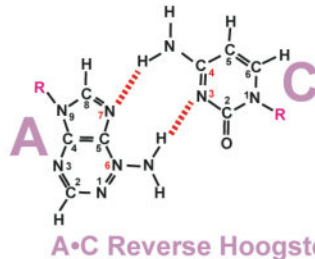
Amino 6 – Imino 1
Imino 1 – Amino 6



Amino 2 – Carbonyl 2
Amino 1 – Imino 7



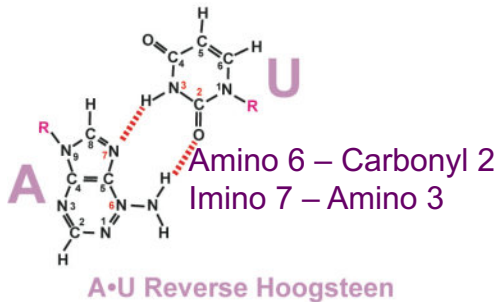
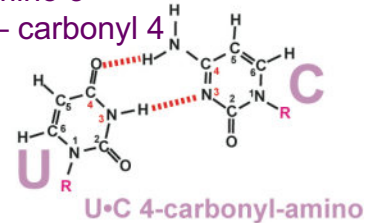
Amino 6 – Imino 3
Imino 7 – Amino 4



Carbonyl 2 – Amino 3
Amino 3 – carbonyl 2



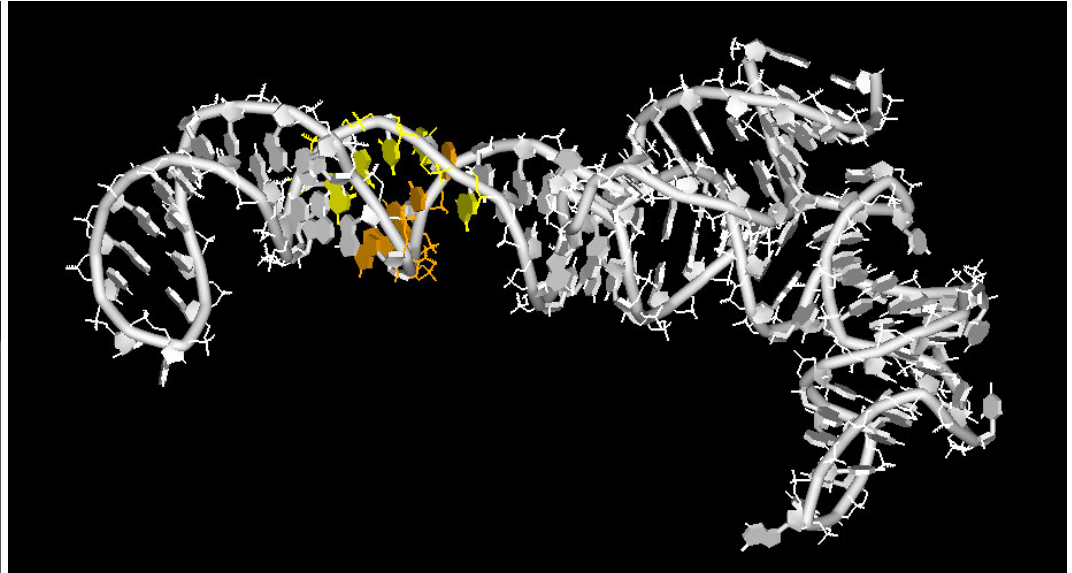
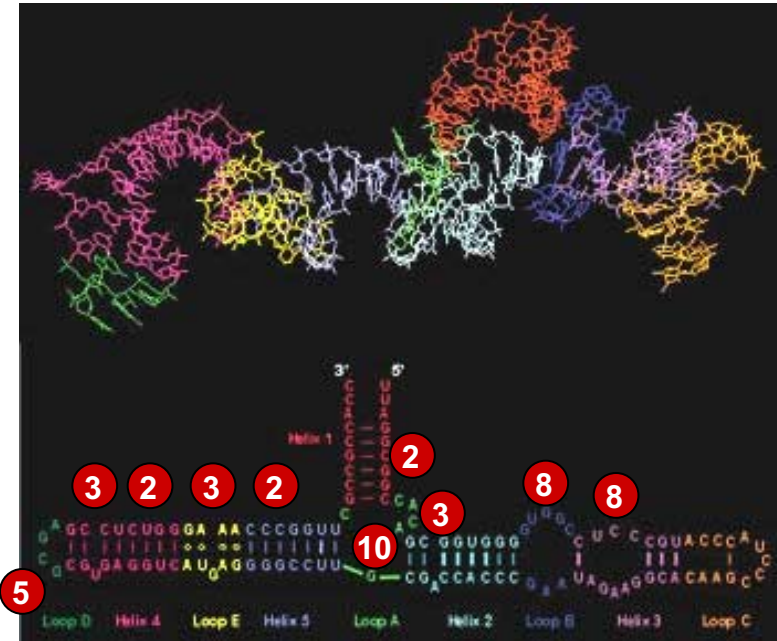
Amino 2 – Imino 3
Carbonyl 4 – carbonyl 4



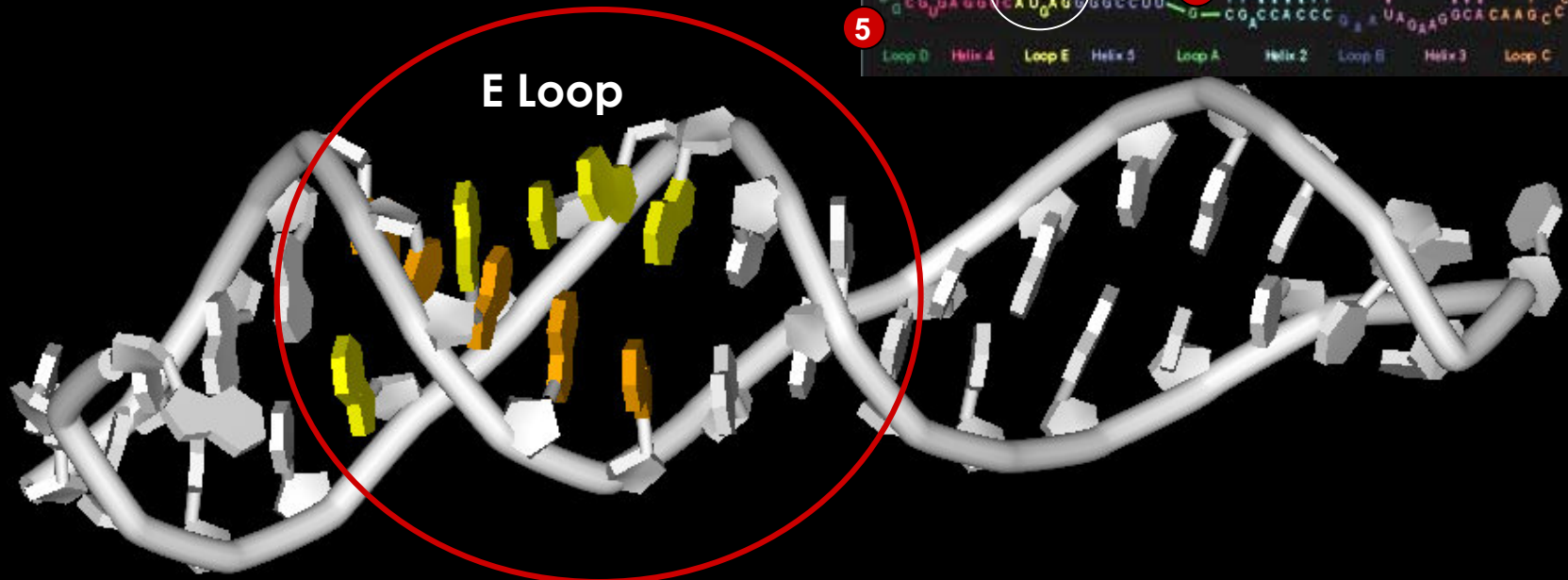
III. RNA structure

C. Tertiary structure

- Tertiary St of *E. coli* 5S RNA



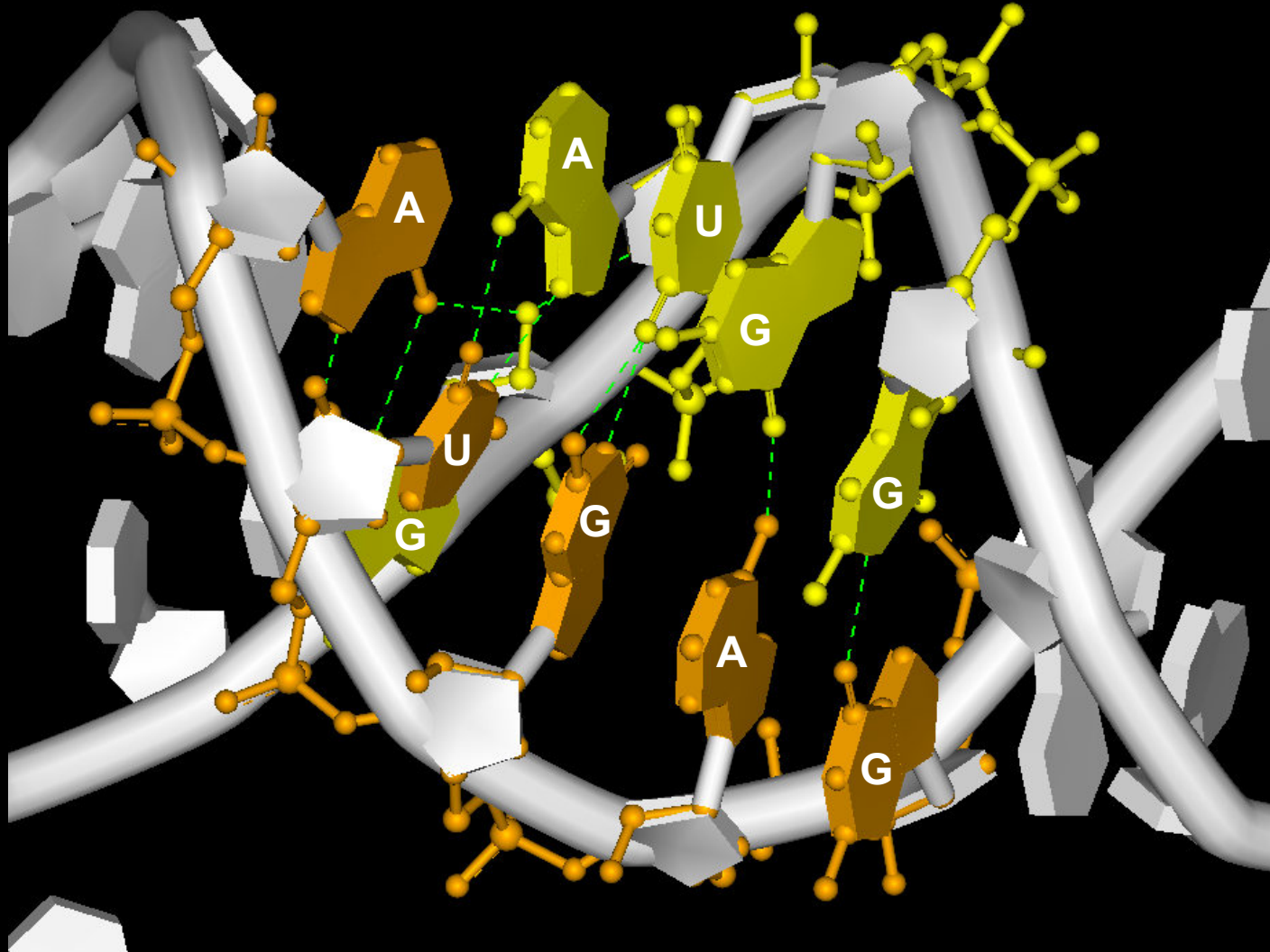
- It is a 120 Nt RNA whose tertiary structure presents many elements of tertiary St that we have just seen



III. RNA structure

C. Tertiary structure

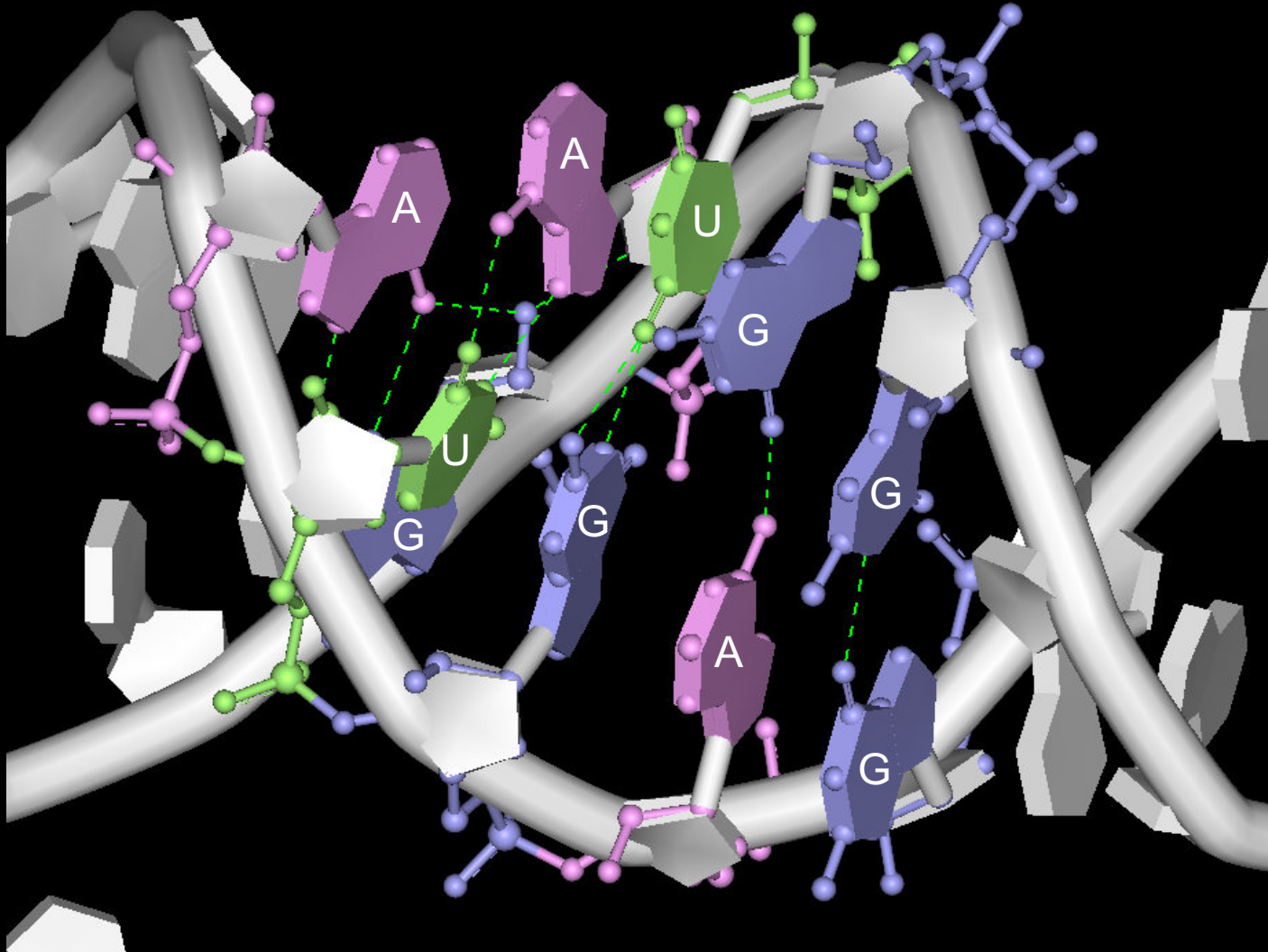
- A region of this RNA is formed by 4 purine-purine pairings : E loop



III. RNA structure

C. Tertiary structure

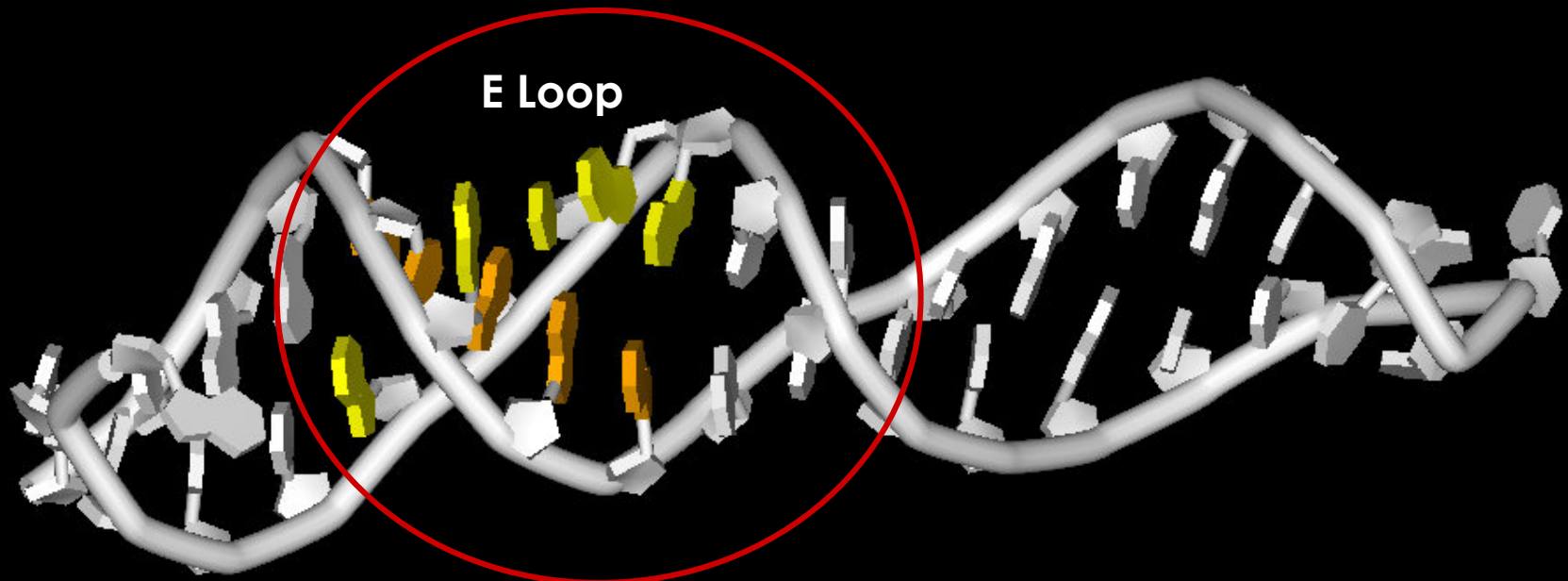
- A region of this RNA is formed by 4 purine-purine pairings : E loop



III. RNA structure

C. Tertiary structure

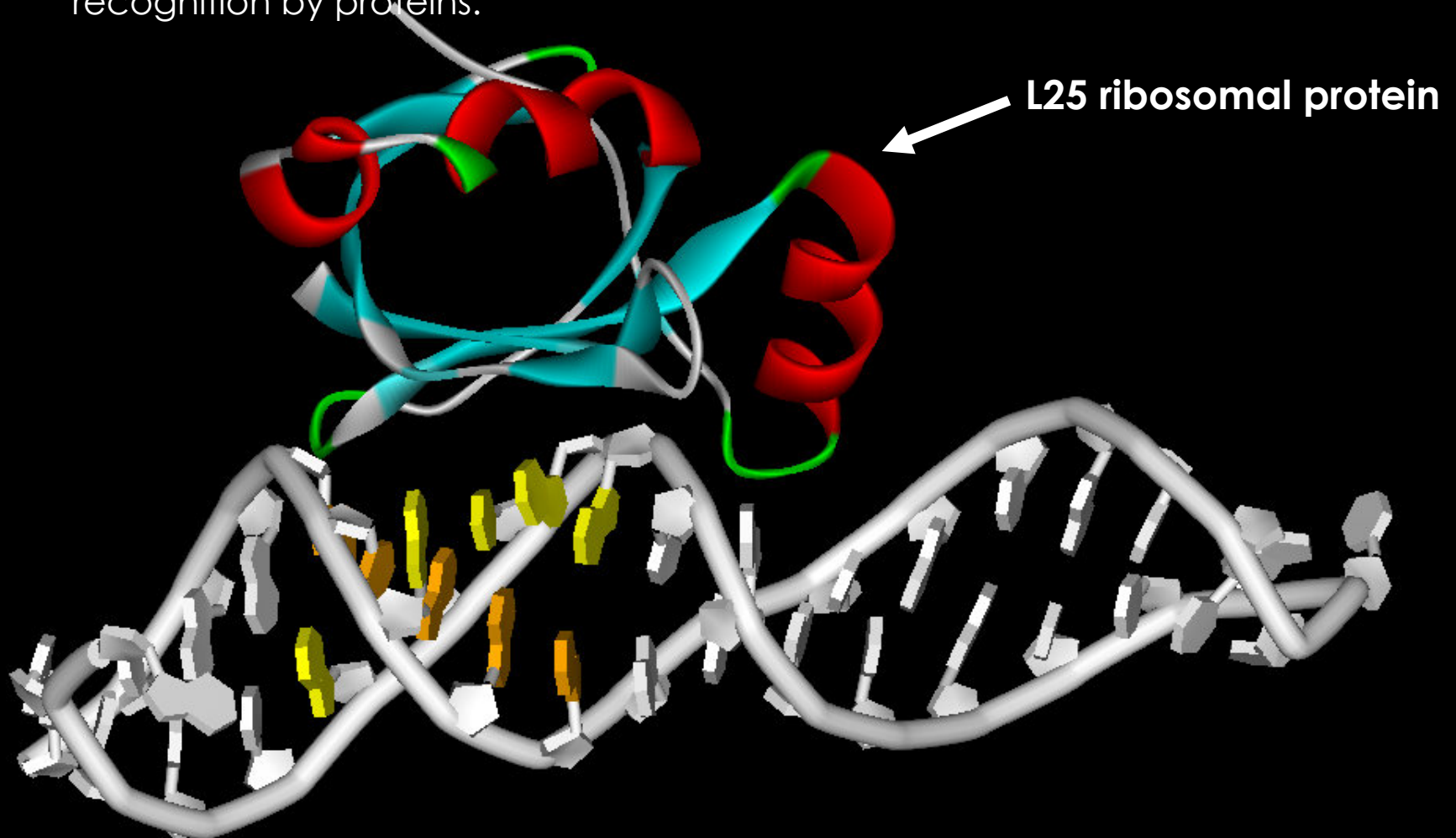
- In this region, the helix is deformed and the structure is different from an RNA helix A. It is often this type of deformation that constitutes the sites of recognition by proteins.



III. RNA structure

C. Tertiary structure

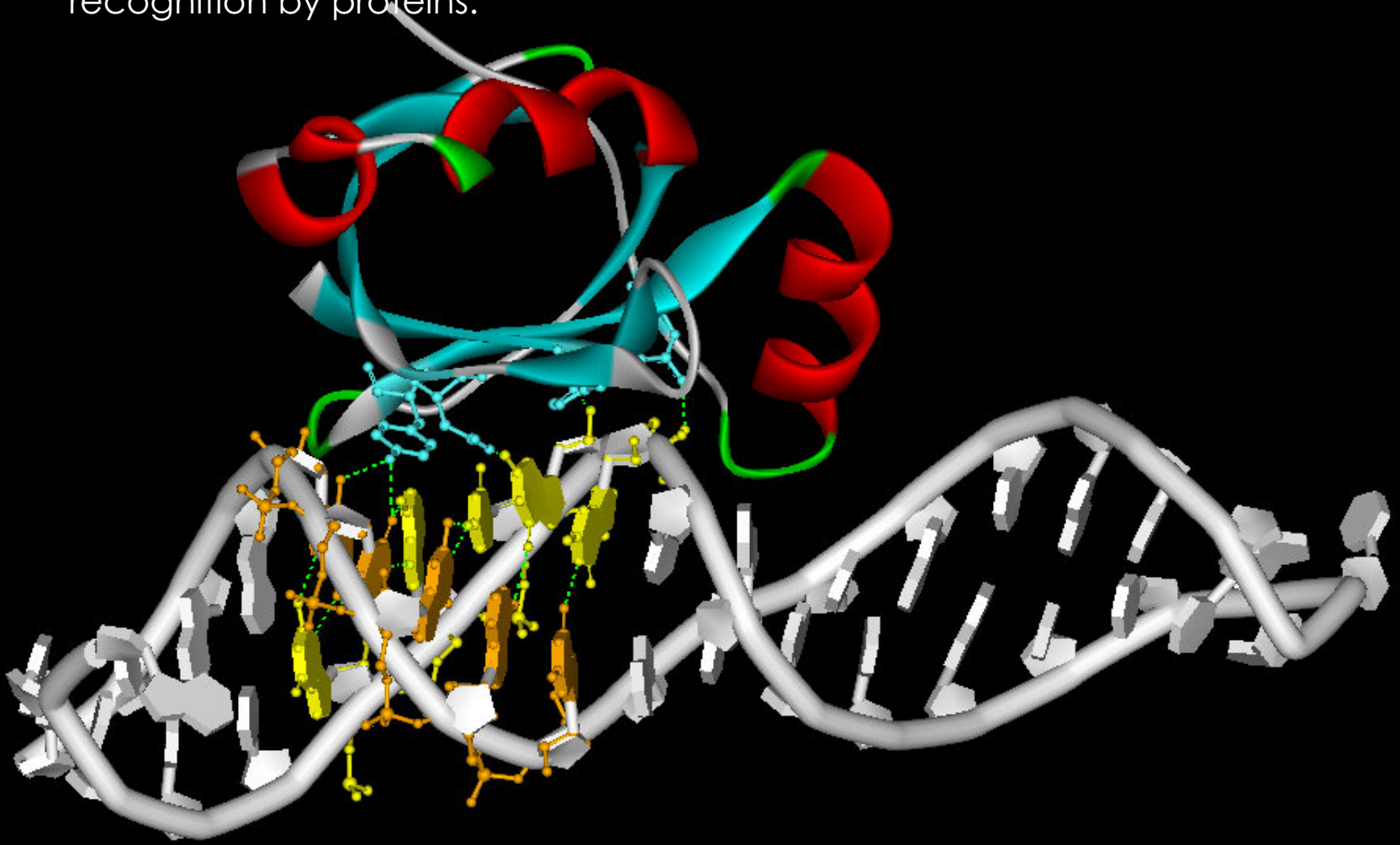
- In this region, the helix is deformed and the structure is different from an RNA helix A. It is often this type of deformation that constitutes the sites of recognition by proteins.



III. RNA structure

C. Tertiary structure

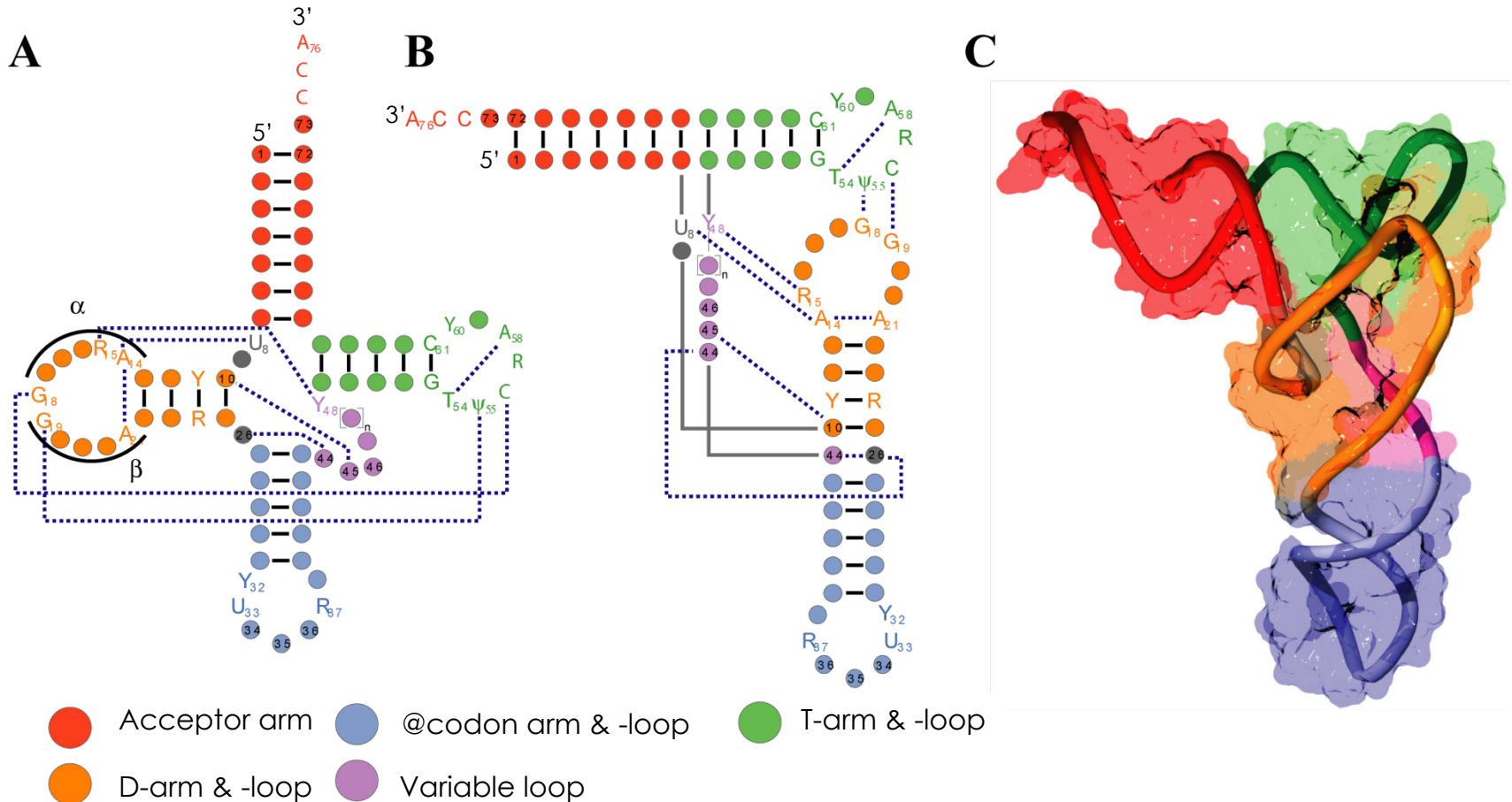
- In this region, the helix is deformed and the structure is different from an RNA helix A. It is often this type of deformation that constitutes the sites of recognition by proteins.



III. RNA structure

C. Tertiary structure

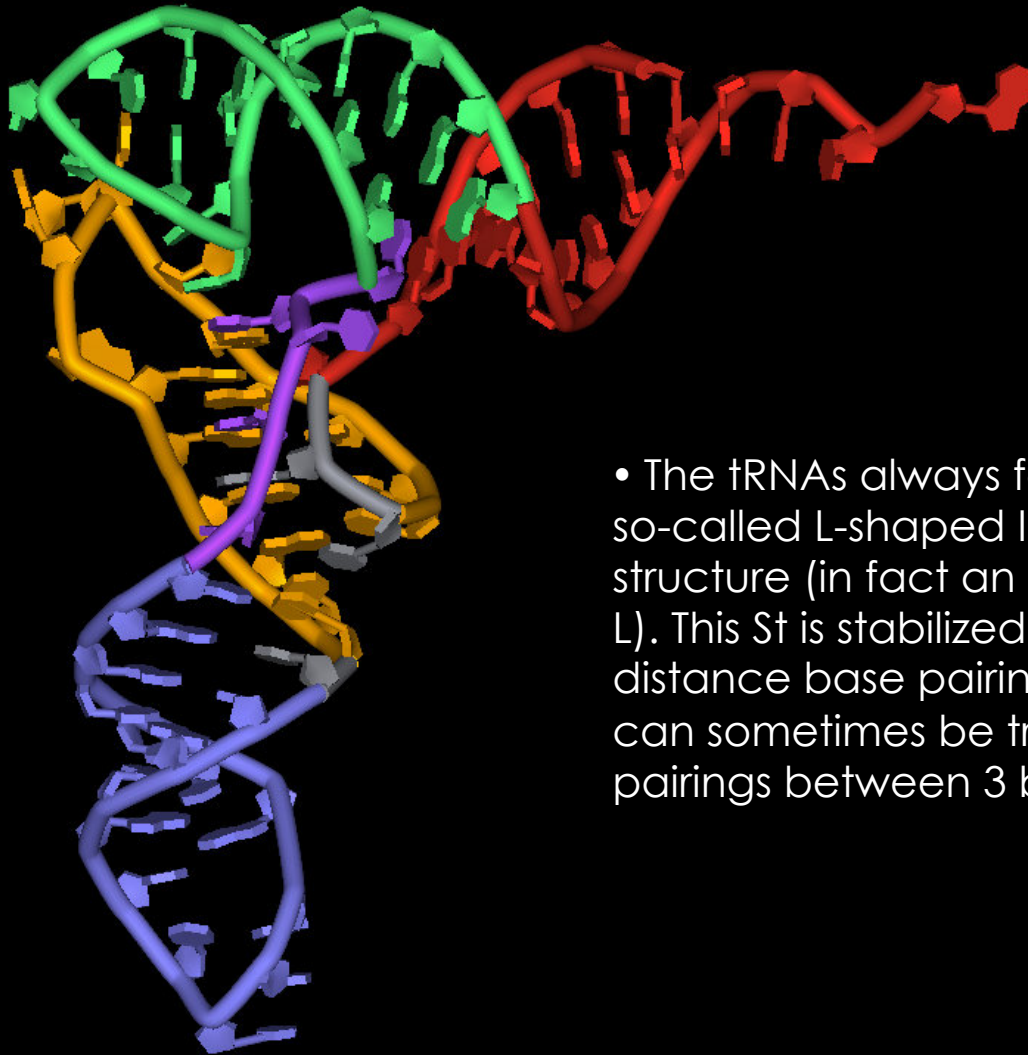
- Tertiary St of a tRNA (60 to 90 Nt)



Secondary cloverleaf structure of tRNA. Numbering of Nt is according to Steinberg et al., 1993. Nt highlighted in the structure are conserved accross all tRNA sequences. Abbreviations: Y: pyrimidine, R: purine, A: adenosine, C: cytidine, G: guanosine, U: uracil., T: thymidine, Ψ: pseudouridine. WC base pairs are schematized by black lines and tertiary interactions allowing folding of the cloverleaf structure into the tertiary L-shape structure are dashed.

III. RNA structure

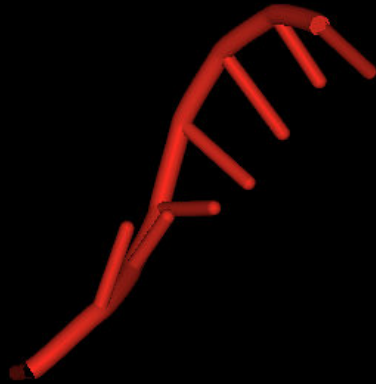
C. Tertiary structure



- The tRNAs always fold into a so-called L-shaped tertiary structure (in fact an inverted L). This structure is stabilized by long-distance base pairings that can sometimes be triple pairings between 3 bases

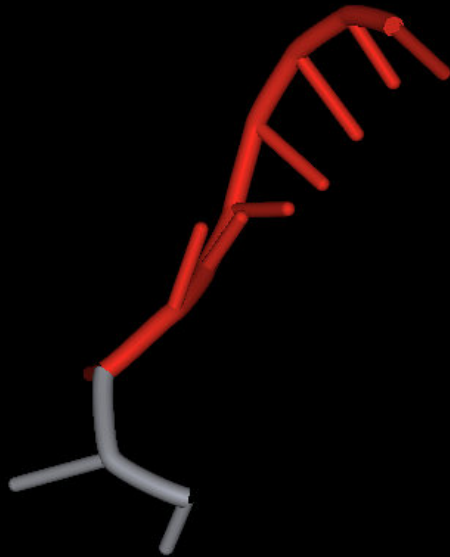
III. RNA structure

C. Tertiary structure



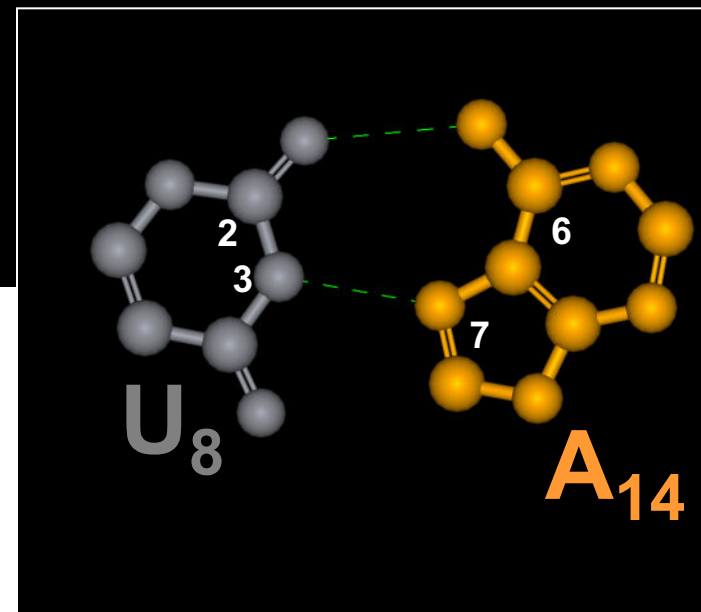
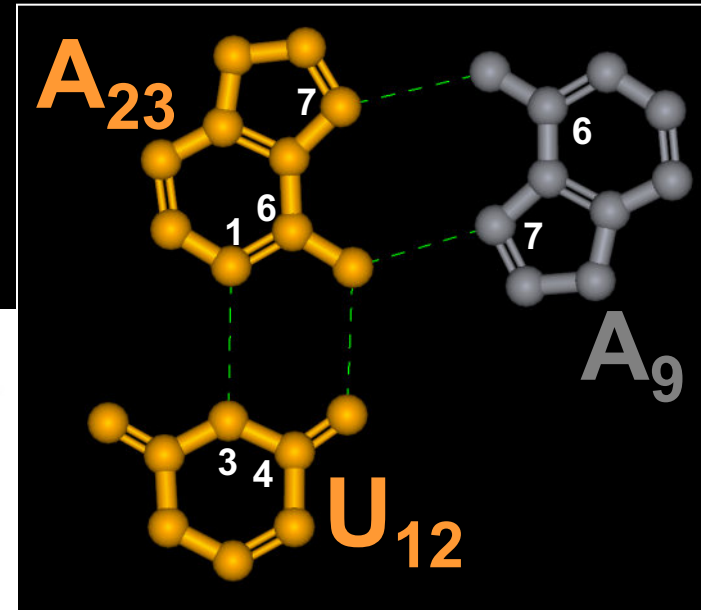
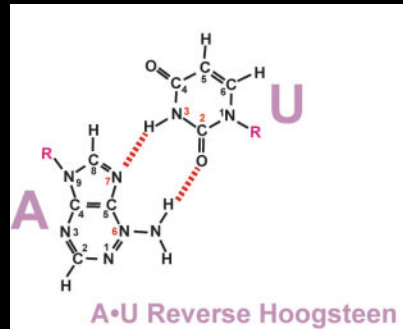
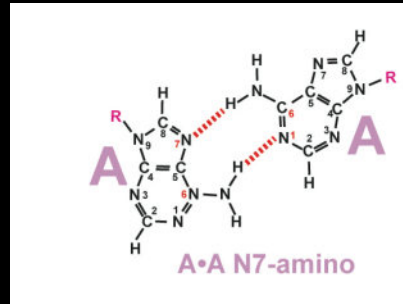
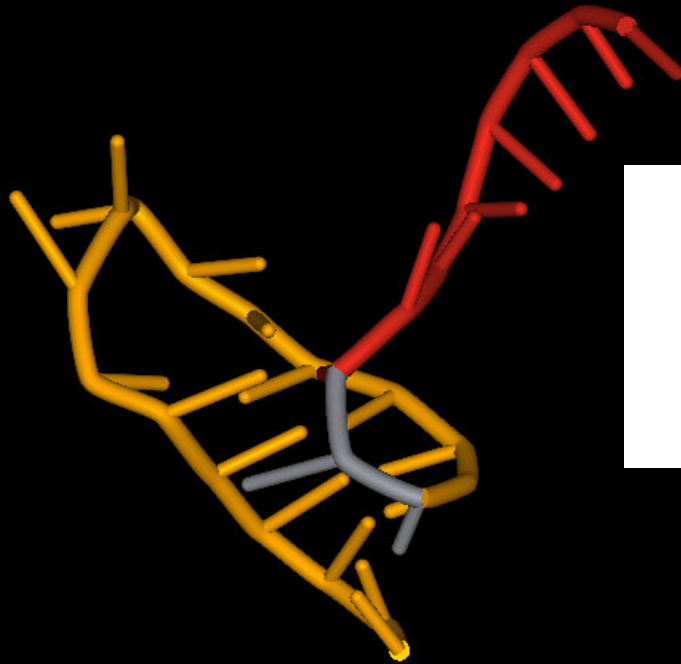
III. RNA structure

C. Tertiary structure



III. RNA structure

C. Tertiary structure



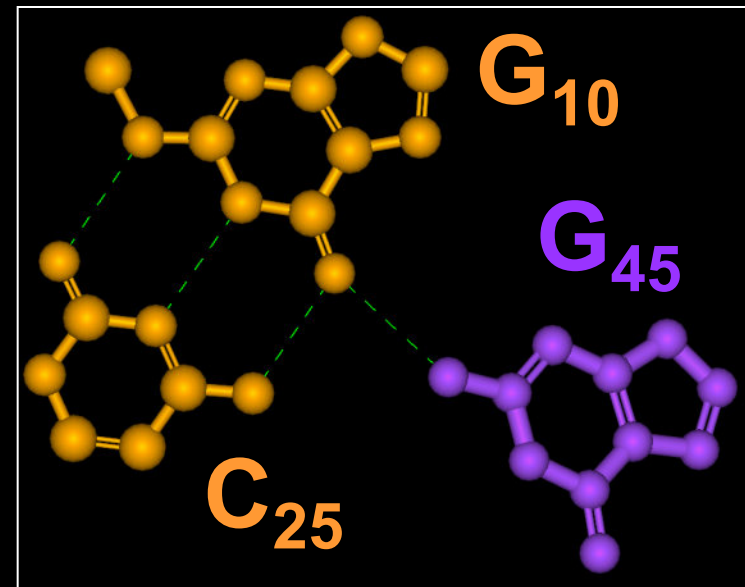
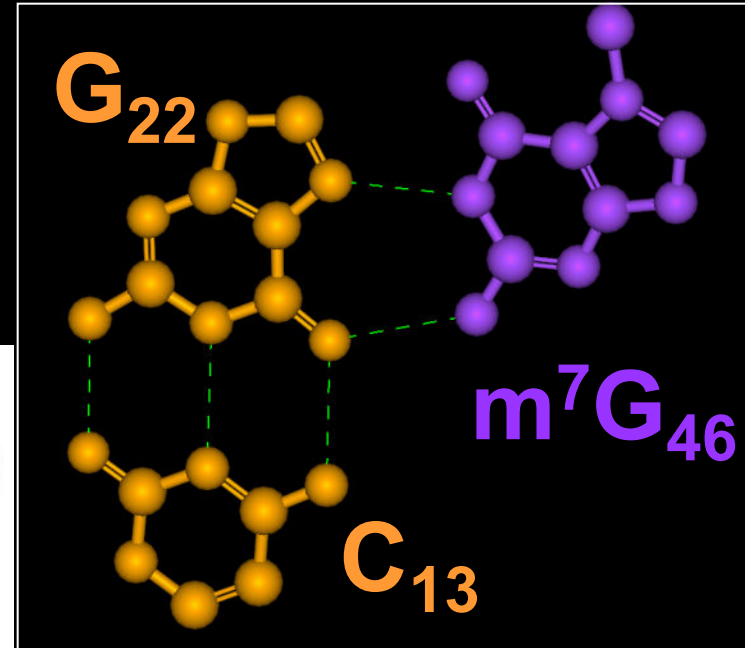
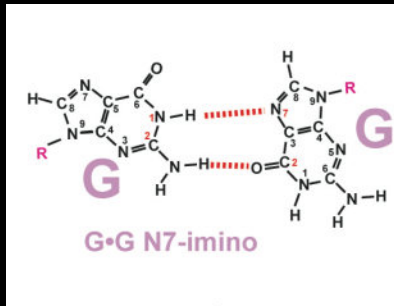
III. RNA structure

C. Tertiary structure



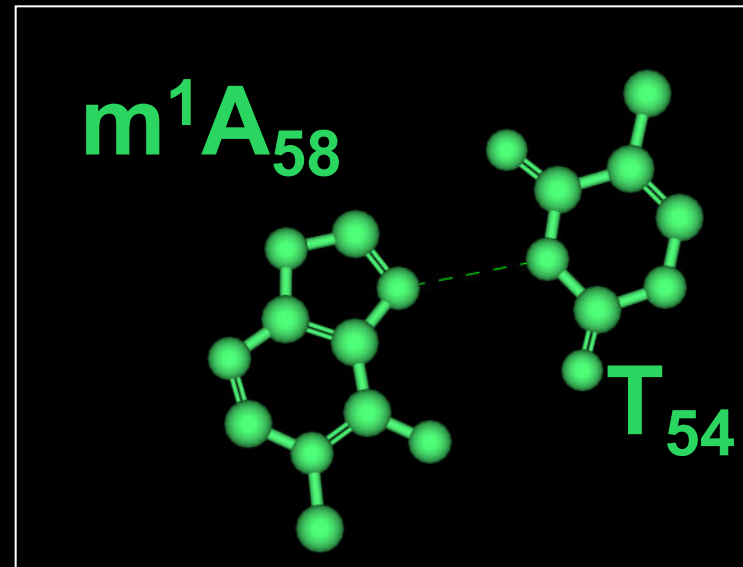
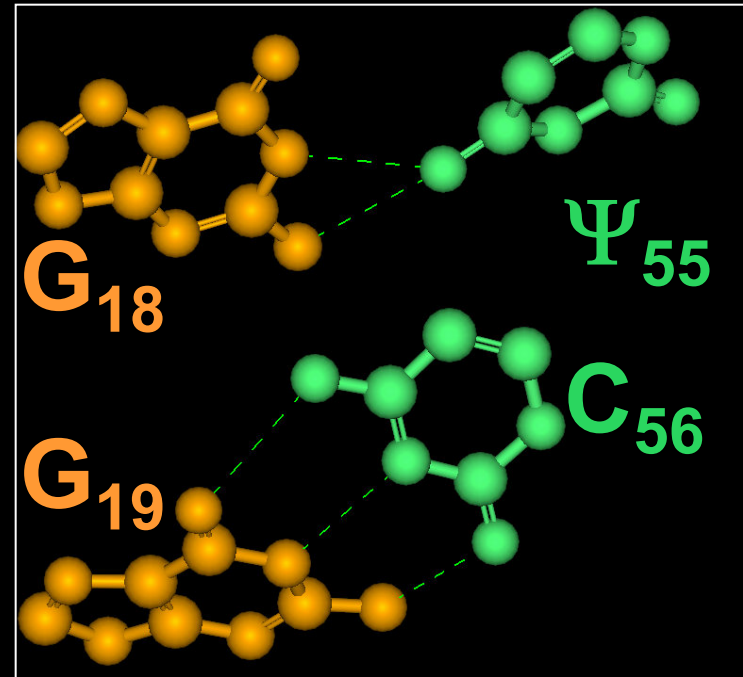
III. RNA structure

C. Tertiary structure



III. RNA structure

C. Tertiary structure



III. RNA structure

C. Tertiary structure



