

GENE EXPRESSION AND PROTEIN SYNTHESIS

Hubert Becker

IPCB, UMR7156 4^{ème} étage

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



Équipe de recherche **Hubert Becker**

Dynamiques & plasticité des synthétases | DyPS

Université de Strasbourg



RECHERCHE ▾

ENSEIGNEMENT ET COURS ▾

VIE DU LABO ▾

PUBLICATIONS ▾



ACCÈS DIRECTS ▾

FR EN

Accueil GMGM → Équipes de recherche → Équipe Becker | DyPS → Enseignement et cours → Support de cours H. Becker

Support de cours H. Becker

Licence 1 - Biochimie

↓	ANiques_2020.ppt	ppt	21 Mo
↓	Code_Genetique_et_Traduction_2020.pptx	pptx	14 Mo
↓	L1S2_Glucides_2020.pptx	pptx	2 Mo

Licence 2 - Métabolisme

↓	Cours_Metabo_2022.pptx	pptx	33 Mo
---	------------------------	------	-------

Master BGM - OEVE

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

Master BGM -RAEG

↓	Cours_RAEG_HB_2020.pptx	pptx	19 Mo
↓	TD_TBox_2022.pptx	pptx	483 Ko

→ Enseignement et cours

↳ Support de cours H. Becker

Actualités

Chaires juniors dans le domaine de recherches sur les mitochondries

Le Laboratoire d'excellence (LabEx) MitoCross annonce l'appel aux candidatures de deux chaires juniors dans le domaine... →

Publications

Publication Equipe Becker, Friant et collaborateurs

Assigning mitochondrial localization of dual localized

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

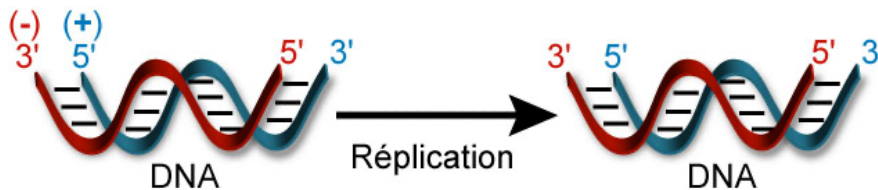
C. Tertiary

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)

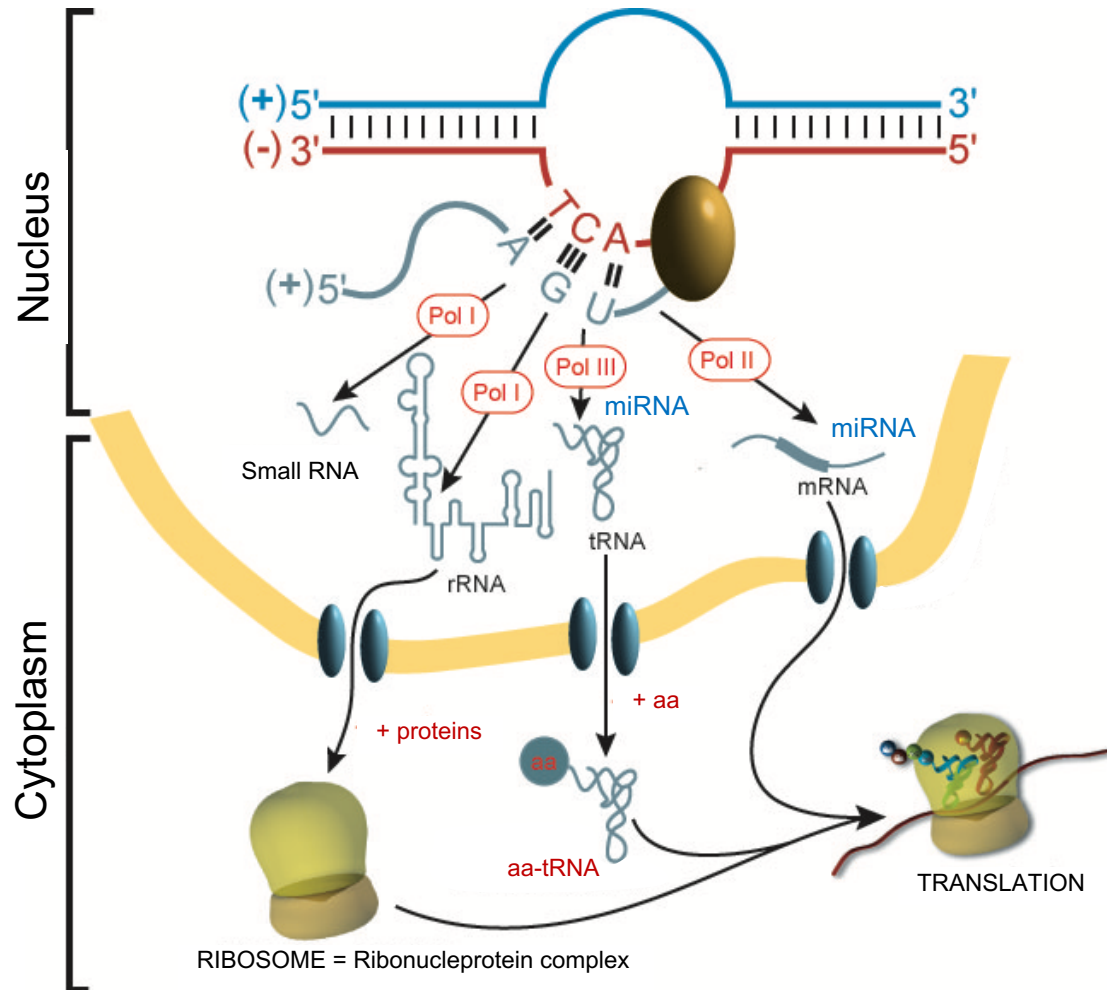
doi: 10.1186/1745-6150-6-7

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 50 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



**LE CODE EST
REDOND-**

DÉGÉNÉRÉ!



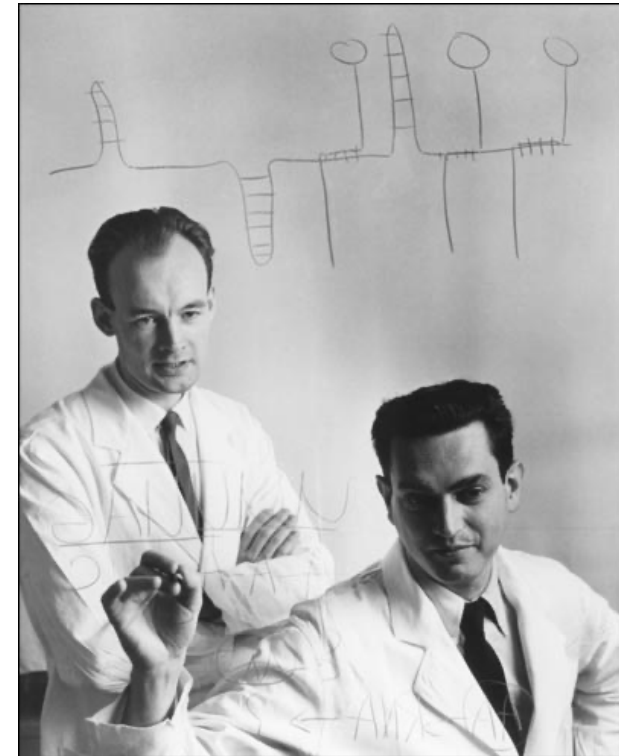
II. The Genetic Code

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 U7NITS*
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]-radilabelled



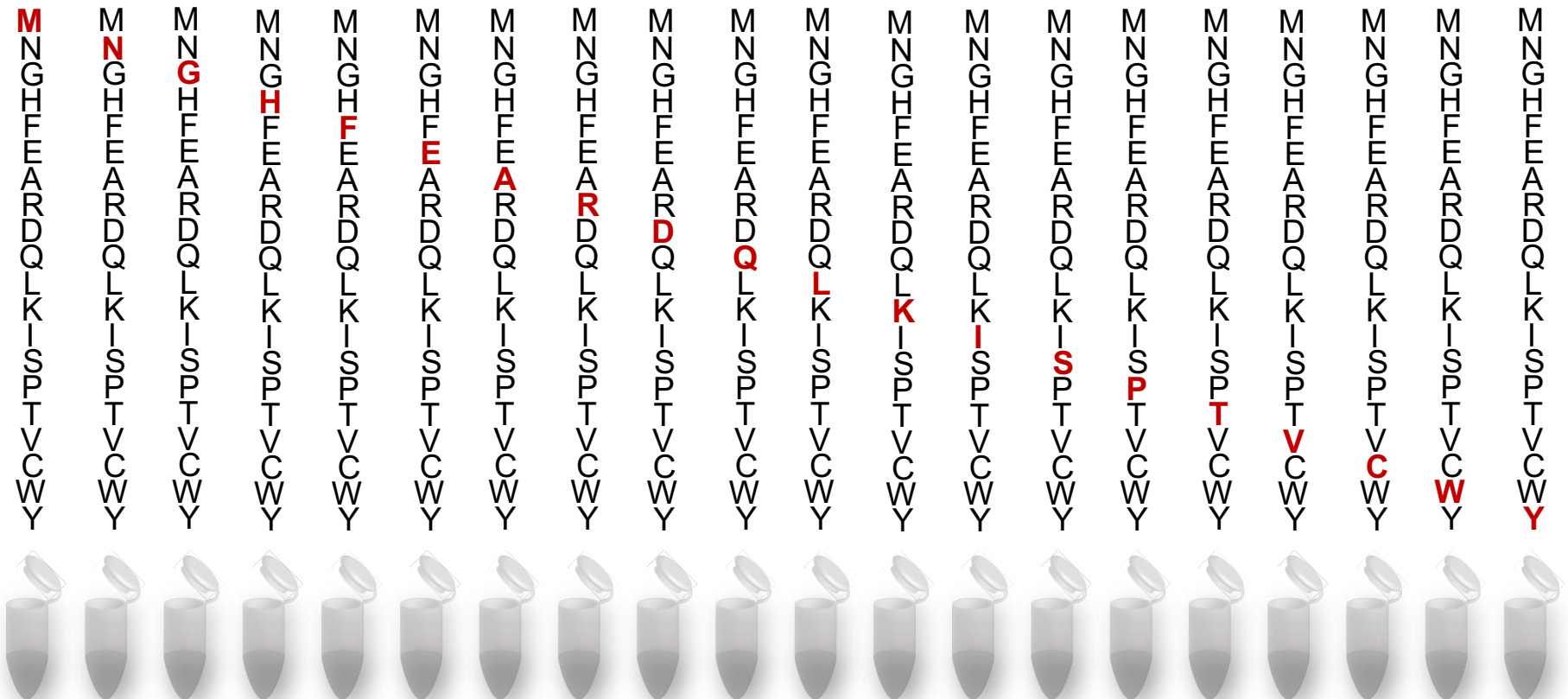
II. The Genetic Code

A. Deciphering & Degeneracy

♦ Deciphering

27th of May 1961 3:00 AM

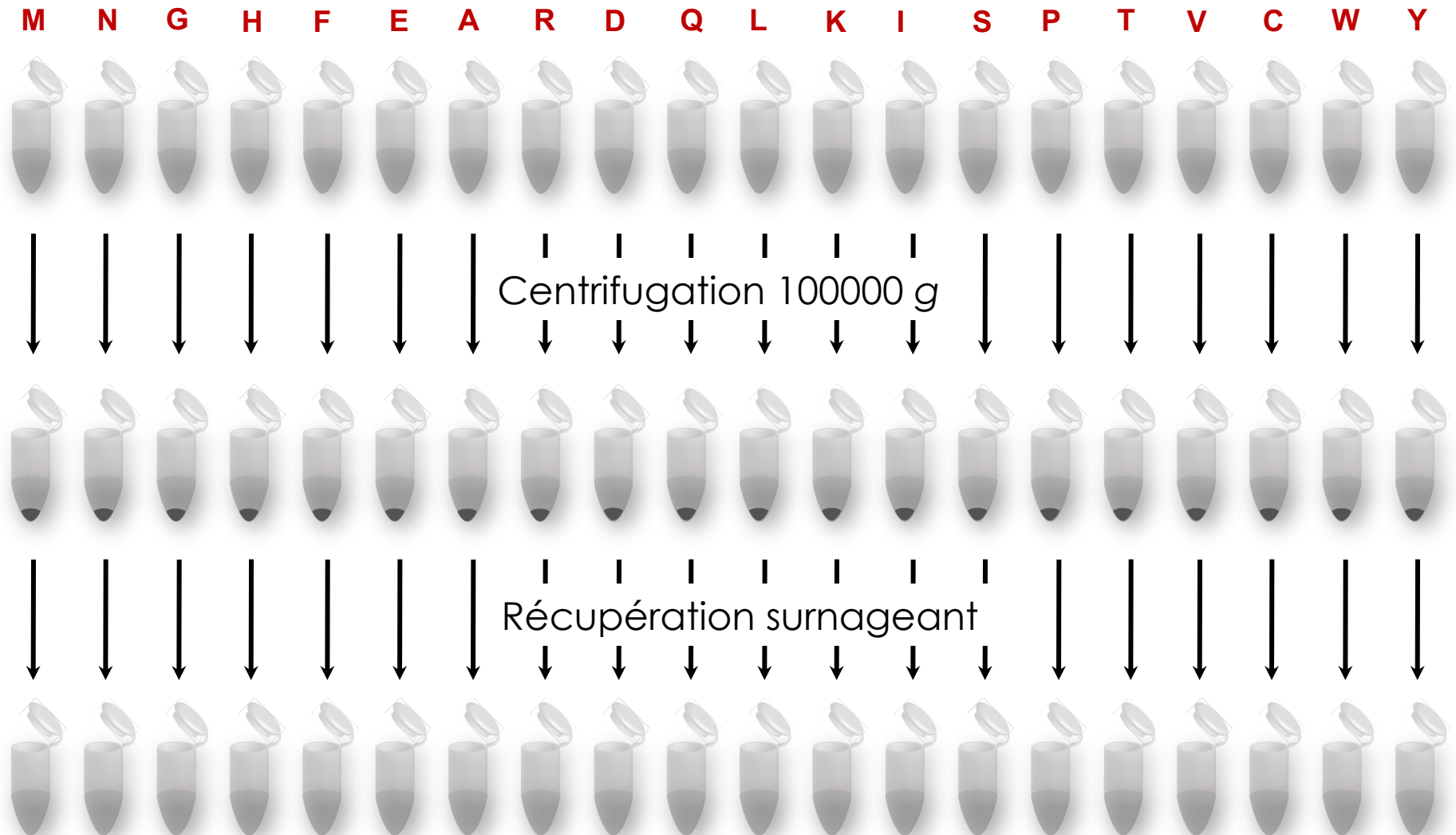
E. coli 30 S
+
5'UUU....UUU3'
+



II. The Genetic Code

A. Deciphering & Degeneracy

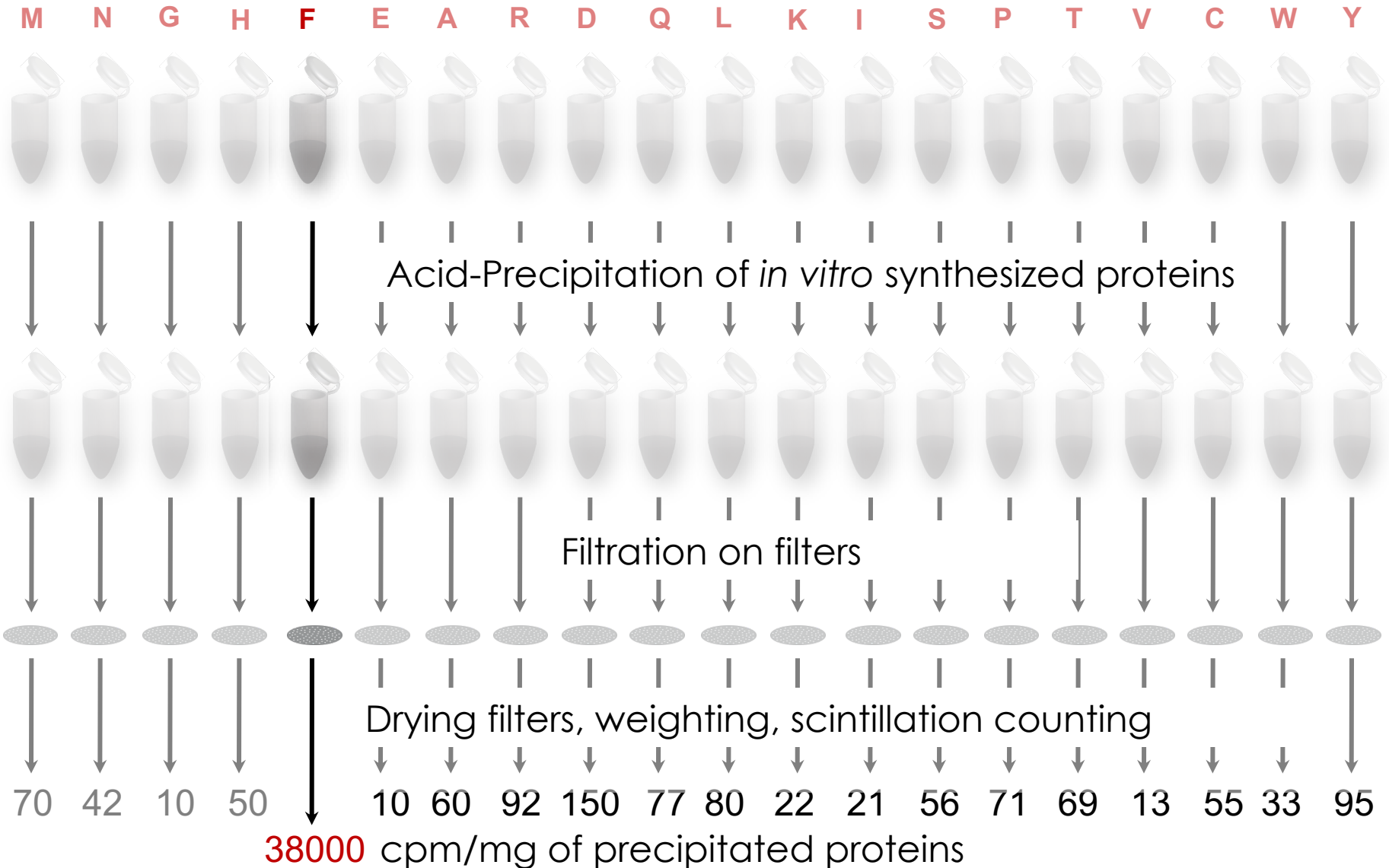
♦ Deciphering



II. The Genetic Code

A. Deciphering & Degeneracy

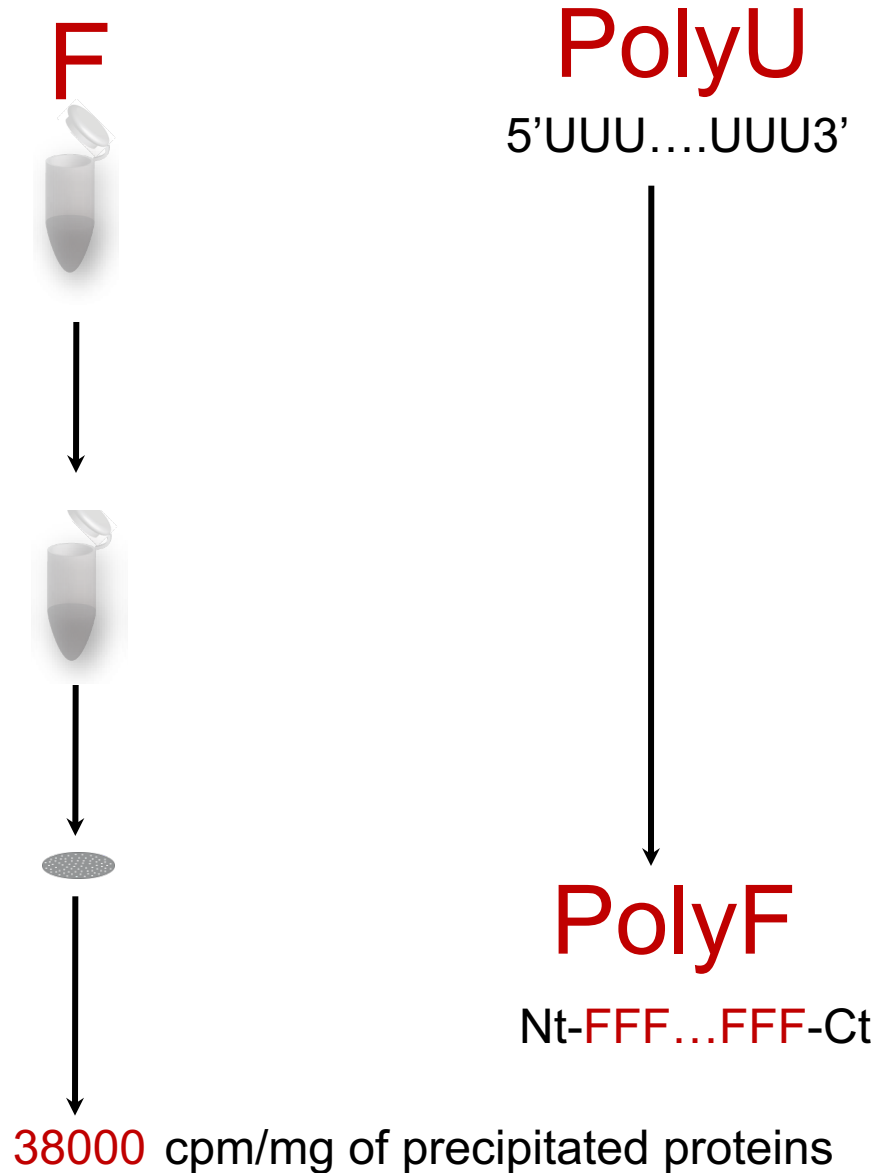
◆ Deciphering



II. The Genetic Code

A. Deciphering & Degeneracy

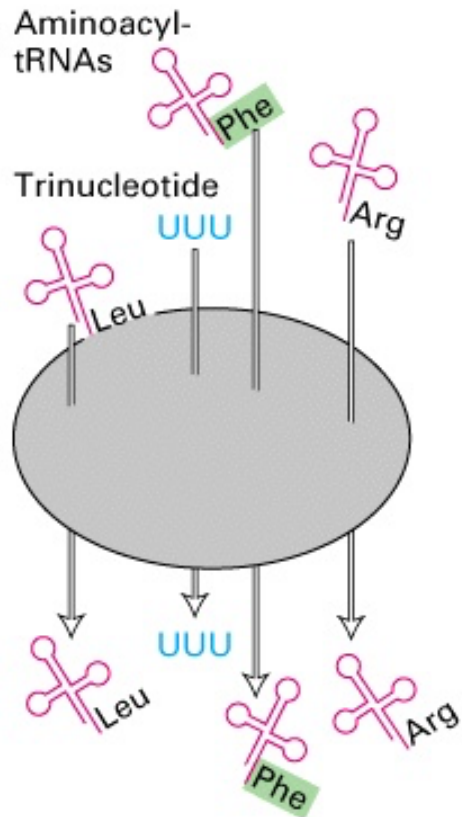
♦ Deciphering



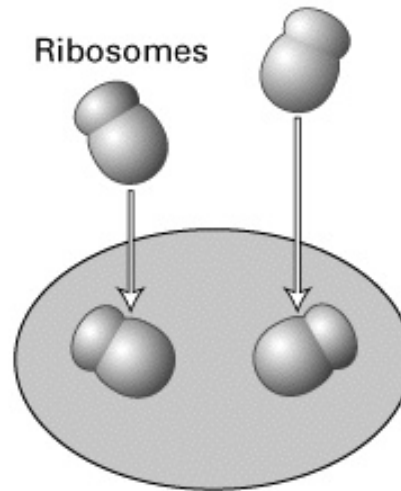
II. The Genetic Code

A. Deciphering & Degeneracy

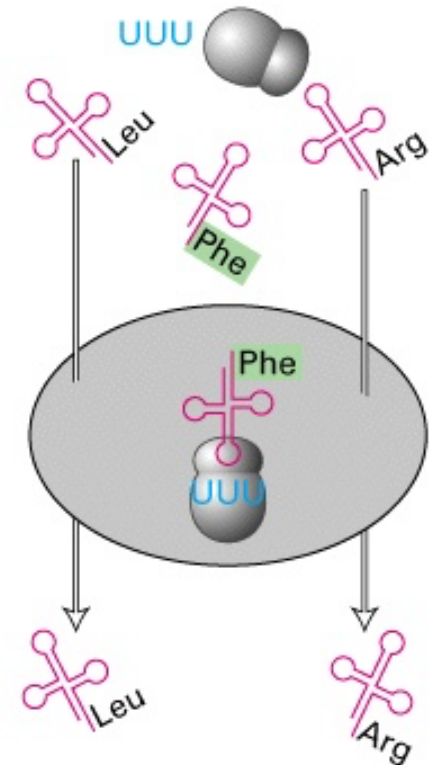
♦ Deciphering



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	C^4	H^3	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	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	-0.03	-0.01	-0.03	-0.03	-0.01	-0.01				
UUC	-0.01	-0.24	0.03	-0.10	-	0.01	-0.13	-0.04	-0.04	-0.03	-0.33	-0.04	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01				
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	-0.03	-0.03	-0.03	-0.03	-0.03				
UUG	-0.20	-0.02	-0.07	-0.05	-	-0.02	-0.08	-0.01	-0.05	0.00	-0.03	-0.06	-0.17	-0.02	-0.07	-0.24	-0.10	-0.05	-0.01				
UUA	0.01	-0.20	-0.05	-0.05	-	0.06	-0.18	-0.01	0.00	-0.11	-0.07	-0.07	0.02	-0.07	1.37	-0.01	-0.01	-0.01	-0.01				
UUC	0.01	-0.32	-0.01	-0.02	-	0.03	-0.16	-0.01	0.00	-0.06	0.01	-0.07	-0.02	-0.01	0.05	-0.01	-0.01	-0.01	-0.01				
UUA	-0.01	-0.08	-0.08	-0.05	-	-0.20	-0.17	-	-0.01	-	-0.01	-	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01				
UUG	-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				
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	-0.03	0.01	0.31				
UUC	-0.01	-0.27	-0.03	0.03	-	-0.03	-0.16	-0.13	-0.04	-0.01	-0.21	-0.06	-0.04	-0.33	0.03	0.00	-0.11	0.02	0.56				
UUA	-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				
UUG	-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	-0.05	0.03	0.03				
UUA	-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	-0.06	-0.09	0.06	0.02	0.14				
UUC	0.28	-0.18	0.05	0.13	0.04	-0.03	-0.55	-0.03	-0.03	-0.01	-0.03	0.00	-0.18	0.04	0.03	0.02	-	0.04	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	-0.08	0.03	0.03	0.05				
UUG	0.04	-0.36	0.01	0.02	-0.03	-0.01	0.28	-	-0.03	-0.03	-0.04	0.00	0.06	0.06	0.14	0.10	-0.08	0.03	0.03				
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	-0.01	-0.01	-0.01	0.00	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				
UUA	-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				
UUG	-0.14	-0.11	-0.01	0.04	-	0.01	0.11	-1.25	0.03	0.03	0.03	0.03	0.10	0.05	0.03	0.03	-0.06	-0.03	0.01				
UUA	-0.08	-0.11	-0.01	-0.01	-	-0.03	-0.03	-0.14	-0.03	0.01	-0.11	-0.05	0.03	0.03	0.15	-0.01	-0.01	-0.03	0.03				
UUC	-0.11	-0.01	-0.01	-0.07	-	-0.22	0.03	-0.03	-0.02	0.02	-0.12	-0.03	0.00	-0.35	0.03	0.00	0.00	0.04	0.02				
UUA	-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				
UUG	-0.08	-0.34	0.01	0.05	0.01	-0.02	-0.57	0.06	-0.03	-0.02	-0.02	-0.02	-0.01	0.75	-0.04	0.04	-0.03	-0.03	0.01				

> Nov III A = 1.56 - 0.132
13. A = 2.10 - 0.141

II. The Genetic Code

A. Deciphering & Degeneracy

♦ Deciphering

Handwritten notes: "Mendel's Nirenberg" and "Ser" with an arrow pointing to the Ser codon 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	-0.04	-0.04	-0.04	-0.33	-0.04	-0.01	-0.04	-0.04	-0.04	-0.04	-0.04	-0.04	-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	-0.10	-0.10	-0.05	-0.01	-0.01
UCA	-0.01	-0.24	-0.05	-0.05	-	-0.06	-0.13	-0.04	-0.04	-0.04	-0.11	-0.04	-0.04	-0.04	-0.04	-0.04	-0.04	-0.04	-0.04	-0.02
UCU	-0.01	-0.24	-0.05	-0.05	-	-0.06	-0.13	-0.04	-0.04	-0.04	-0.11	-0.04	-0.04	-0.04	-0.04	-0.04	-0.04	-0.04	-0.04	-0.02
UCG	-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	-0.10	-0.10	-0.05	-0.01	-0.01
UAU	-0.03	-0.11	-0.01	-0.03	-	0.00	-0.11	-0.03	0.01	-0.24	-0.09	0.03	-0.37	0.00	-0.03	-0.05	0.01	0.31	0.03	0.03
UAC	-0.01	-0.27	-0.03	0.03	-	-0.03	-0.16	-0.13	-0.04	-0.04	-0.21	-0.04	-0.33	0.03	0.00	-0.11	0.02	0.56	0.00	0.00
UAA	-0.03	-0.11	-0.01	-0.03	-	0.00	-0.11	-0.03	0.01	-0.24	-0.09	0.03	-0.37	0.00	-0.03	-0.05	0.01	0.31	0.03	0.03
UAG	-0.07	-0.06	0.01	0.12	-	0.04	0.00	-0.05	-0.01	-0.02	-0.09	-0.05	-0.02	-0.17	0.00	0.00	-0.15	0.03	0.03	-0.01
UGU	-0.05	-0.03	0.03	0.03	0.03	-0.16	-0.13	-0.04	0.03	-0.03	-0.10	0.02	-0.01	-0.02	-0.06	-0.09	0.06	0.02	0.14	0.03
UGC	0.28	-0.18	0.05	0.13	0.04	-0.03	-0.05	-0.03	-0.03	-0.01	0.00	-0.18	0.04	0.03	0.02	-0.01	0.01	0.03	0.03	0.03
UGA	-0.12	-0.12	0.07	0.14	-0.10	-0.10	-0.13	-0.30	-0.03	-0.04	-0.11	0.00	-0.37	0.00	0.00	-0.08	0.03	0.03	0.03	0.03
UGG	0.04	-0.36	0.01	0.02	-0.03	-0.01	0.28	-	-0.03	-0.03	-0.04	0.00	0.06	0.06	0.14	0.10	-0.08	0.03	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	-0.03	-0.01	-0.11	0.01	0.02	0.04	0.01
CUC	-0.07	-0.08	-	-	-	-0.03	-0.18	-0.17	0.02	0.01	0.00	0.11	0.05	0.05	-0.16	0.02	-	0.03	0.10	0.01
CUA	-0.03	-0.14	0.01	-0.03	0.02	0.00	-0.03	-0.06	-0.07	-0.02	0.03	0.00	-0.10	-0.18	-0.06	-0.11	-0.04	-	0.02	-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	0.01	0.05
CCU	-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	0.05	0.02
CCC	-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	0.03
CCA	-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
CCG	-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	0.05	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.

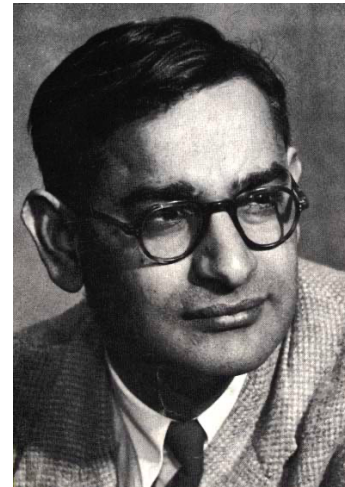
II. The Genetic Code

A. Deciphering & Degeneracy

♦ Deciphering

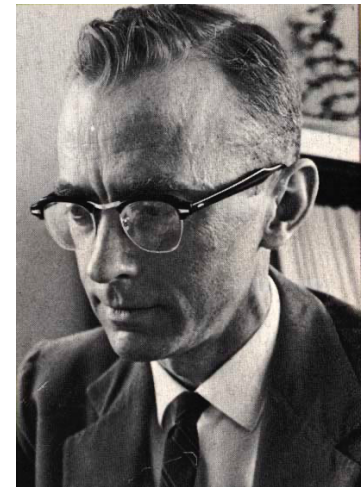


Nobel Prize of Physiology & Medicine
in 1968



[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.

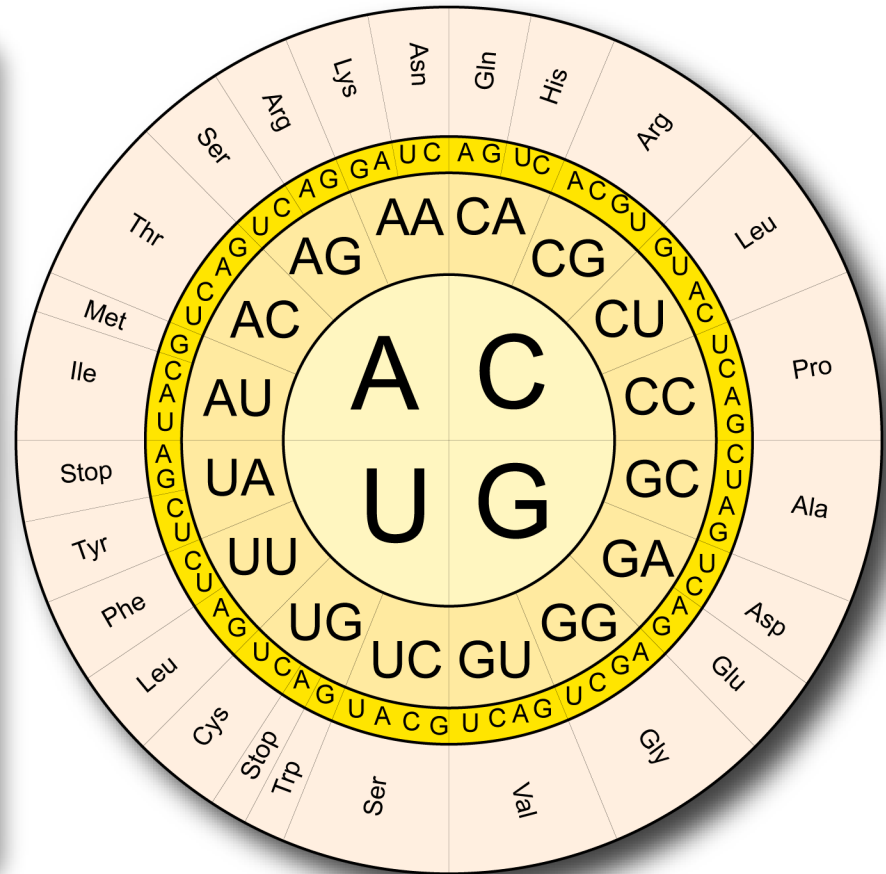
In the experiment, proflavin-induced mutations of the T4 bacteriophage gene, rII_B, were isolated. Proflavin causes mutations by inserting itself between DNA bases, typically resulting in insertion or deletion of a single base pair. The mutants produced by Crick and Brenner could not produce functional 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 : XJOBZ)

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)

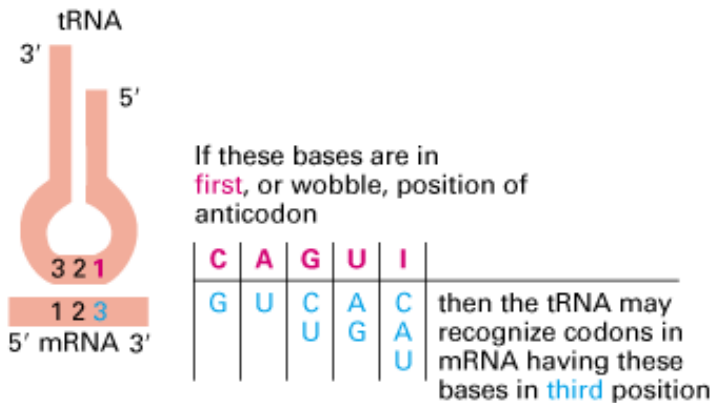


Some STOP codons can be coding

II. The Genetic Code

B. The Wobble pairing

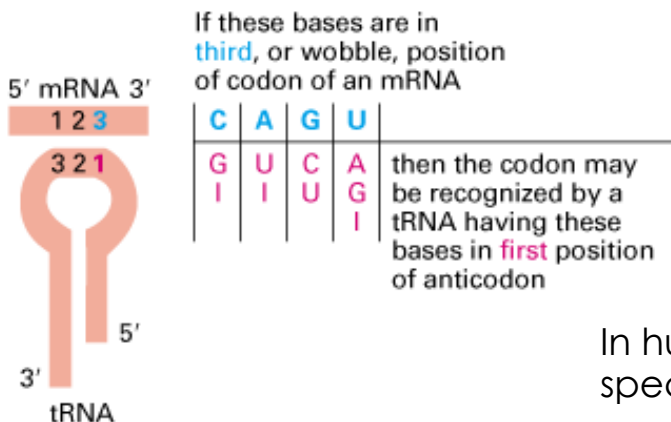
⚠ Since an aa can be encoded by several codons, **are there as many tRNAs as codons? NO**. In prokaryotes there are 30 species of tRNAs to decode 61 codons and in eukaryotes 50 (48 in humans). There is therefore not enough tRNAs. So for some tRNAs the same tRNA must be able to read several codons. This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

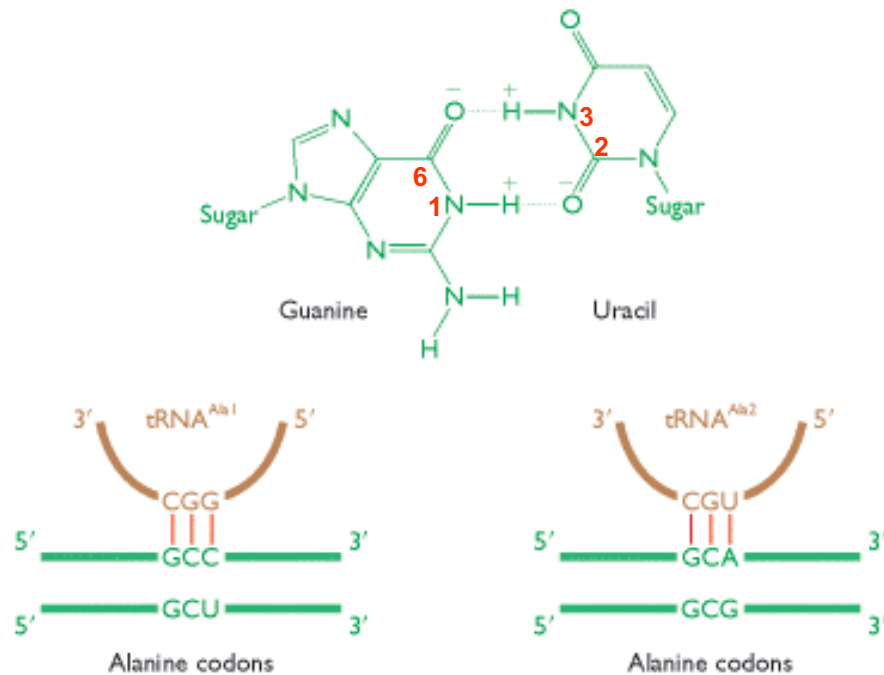


In humans, 16 tRNAs use the wobble while the remaining 32 are specific to a single triplet

II. The Genetic Code

B. The Wobble pairing

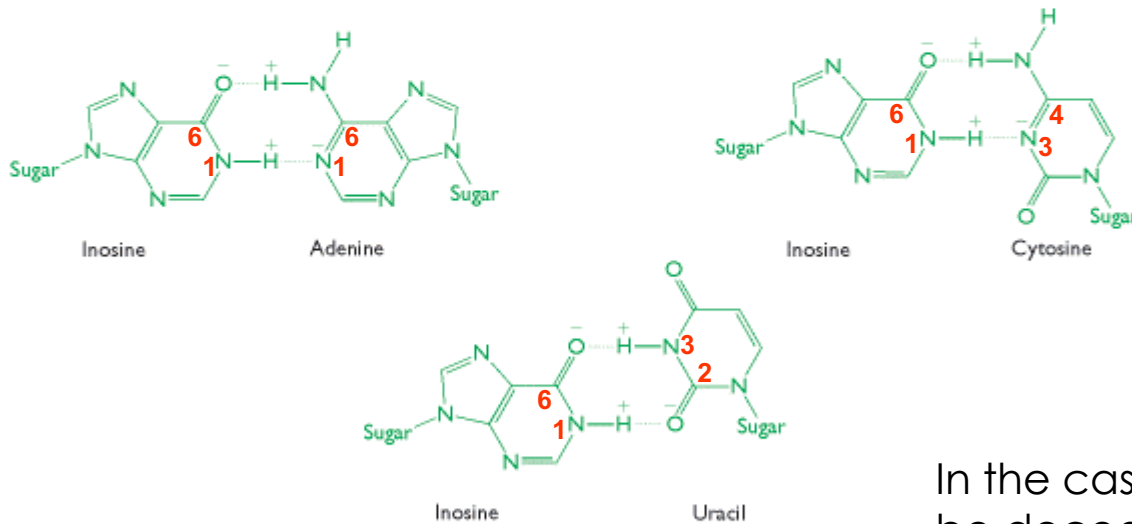
- Wobble pairing is ensured by non-Watson-Crick (WC) base pair (bp) pairings
Ex: Ala is encoded by 4 ≠ codons: GC(U,C,A,G). One of the tRNA^{Ala} has a GGC @codon capable of recognizing GCC and GGU thanks to a non-WC U=G wobble pairing.



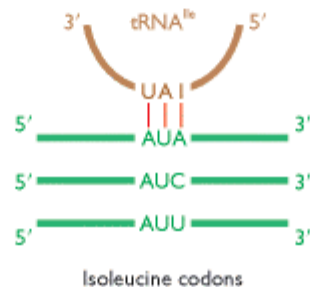
II. The Genetic Code

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



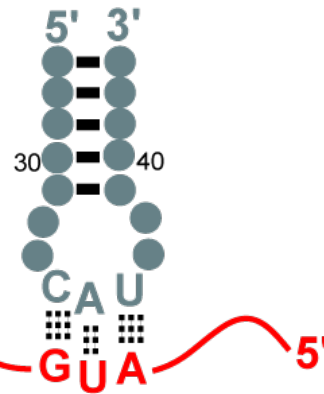
In the case of Ile, the 3 codons can be decoded by a single tRNA.



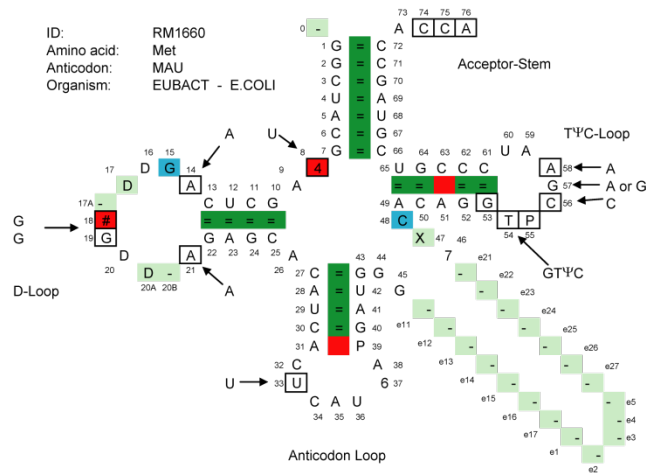
II. The Genetic Code

C. Initiation & STOP codons

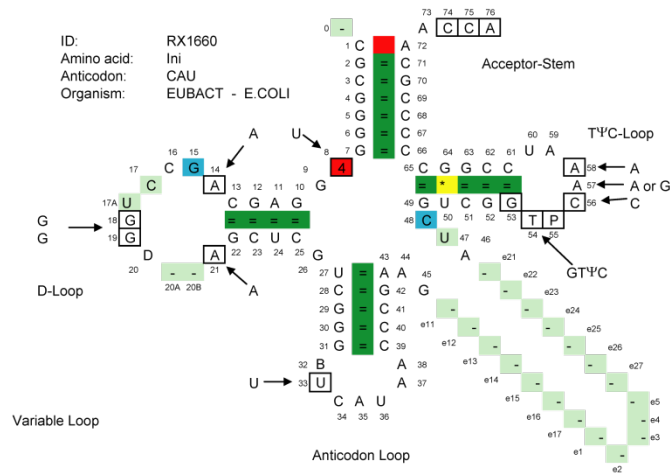
- There is only one Met AUG codon that encodes **2 ≠ Met** :
 - The Met which is in **N-ter of all Prots**, in this case the Met codon is called **initiation codon**
 - The Met, which is **intrachain**, in this case the Met codon is called **elongator codon**
- To decode the initiation codon, there is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



tRNA^{Met}



tRNA_i^{Met}



- must pair (A-T, G-C, C-G, T-A)
- conservative bases
- can be free
- deviations from the cloverleaf model



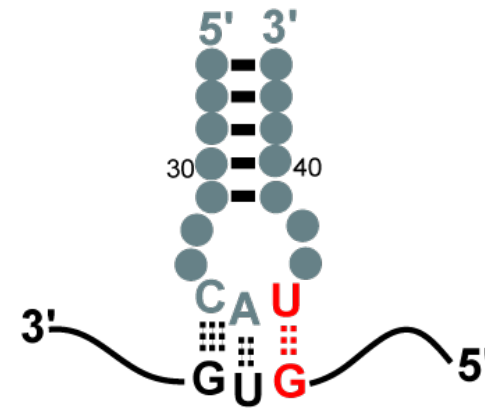
For the 2 tRNA^{Met} the Wobble pairing does not exist

II. The Genetic Code

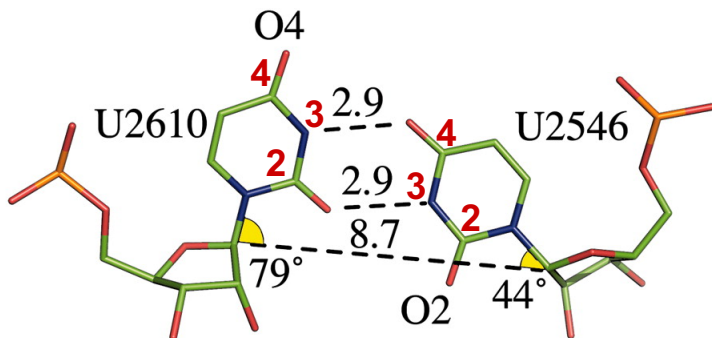
C. Initiation & STOP codons

- **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)

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

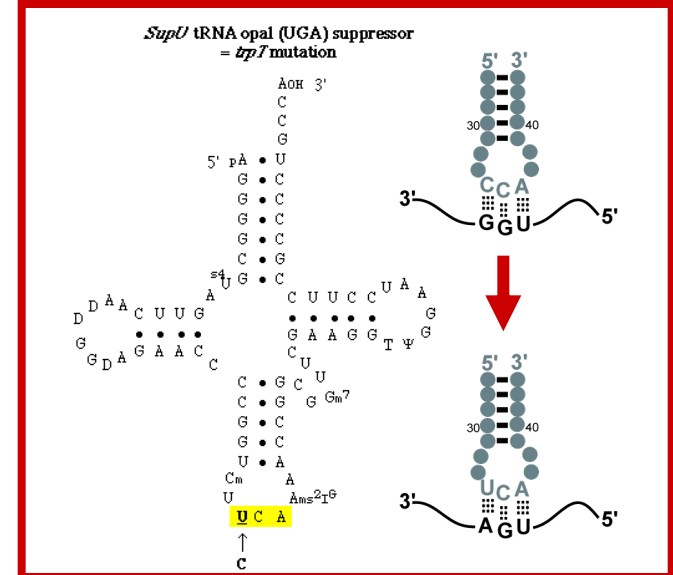
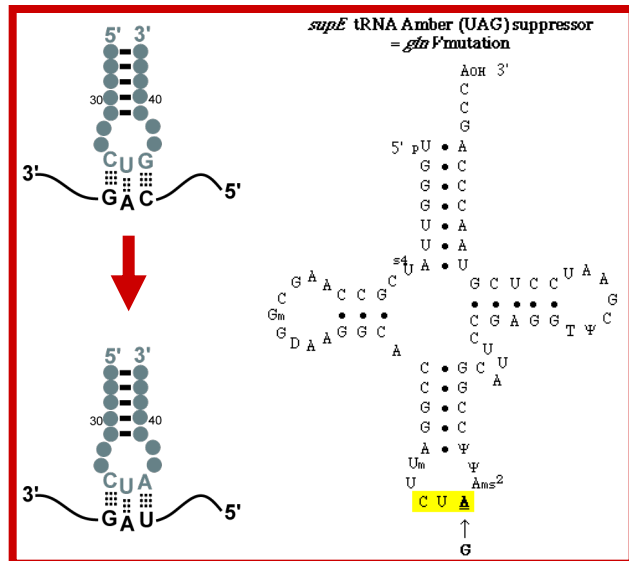
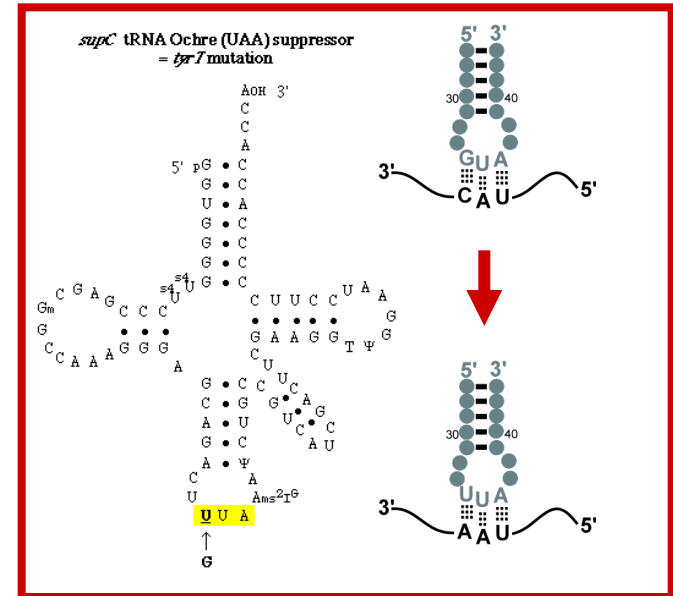
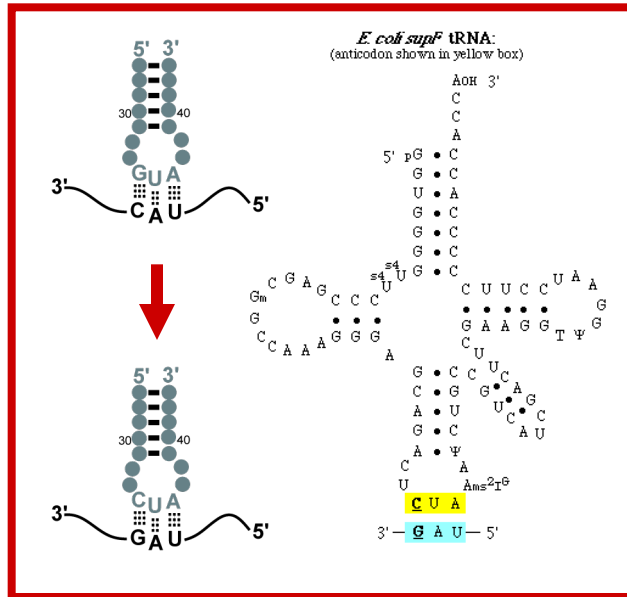
D. General information on the reprogramming and dynamics of the genetic code

● **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

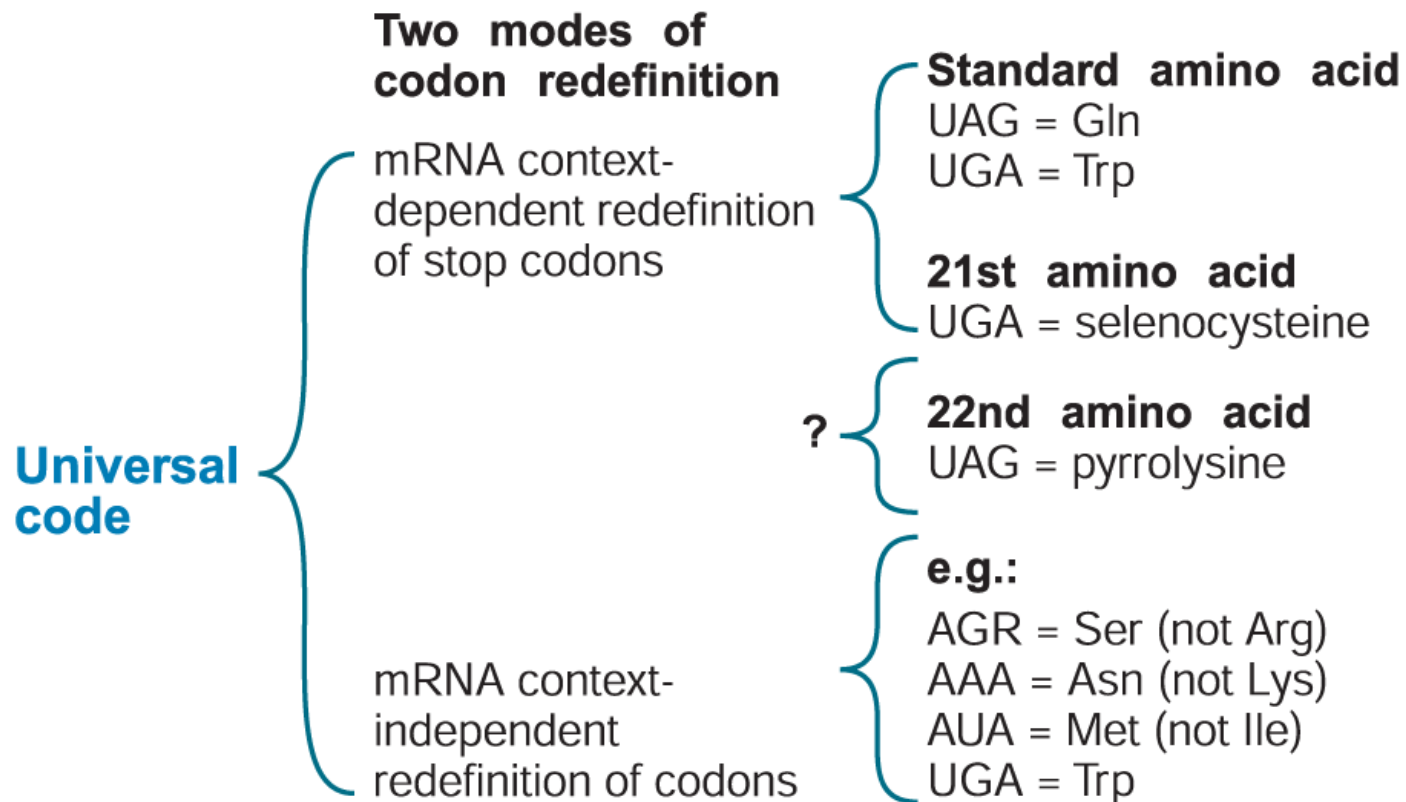


II. The Genetic Code

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**

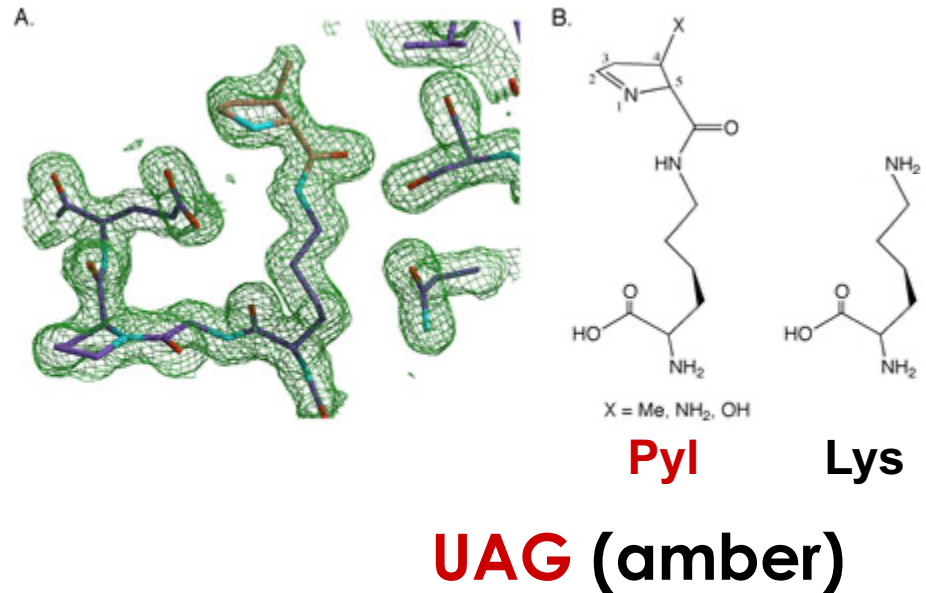
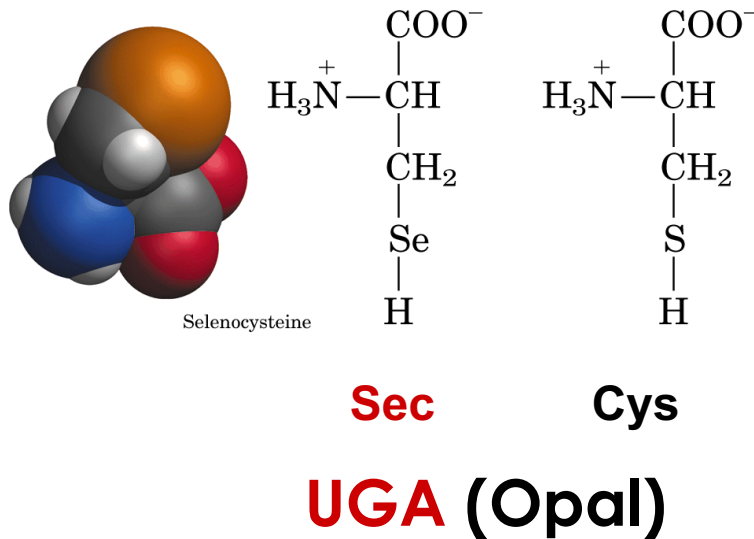


II. The Genetic Code

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)**.



 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

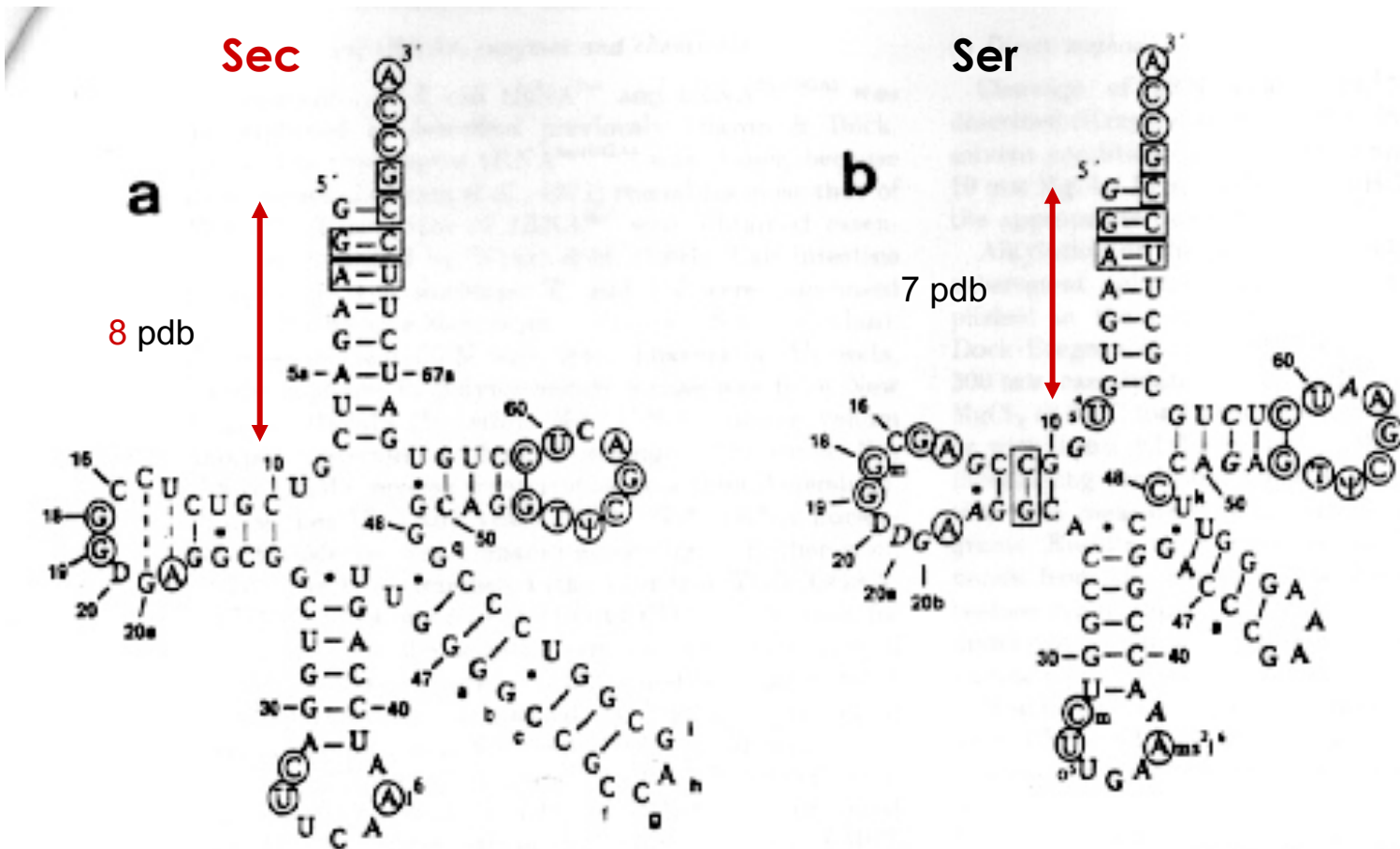
- A **new AA** to be introduced in the GC
- A **suppressor tRNA** that will be specific to this new aa
- An **aminoacyl-tRNA synthetase** (aaRS) loading this new aa onto the tRNA^{Sup}
- A **message on the mRNA** telling the ribosome that it is not a stop codon but that it encodes the new aa
- A specific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



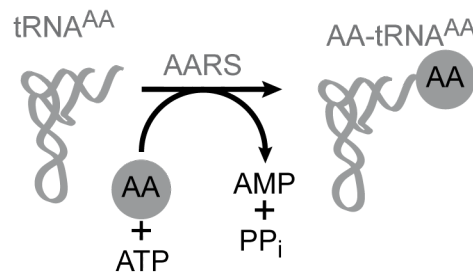
II. The Genetic Code

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

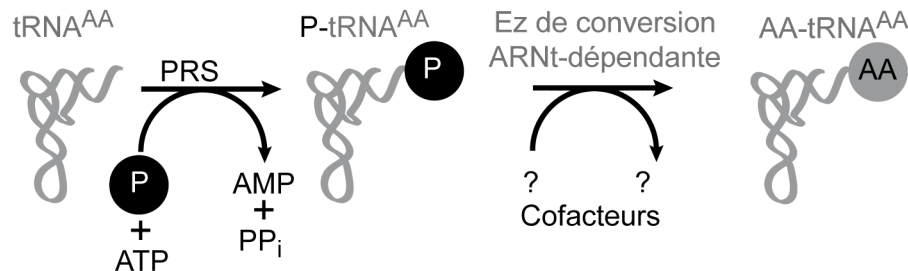
2. Reassignement/reprogramming of STOP codons

- 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

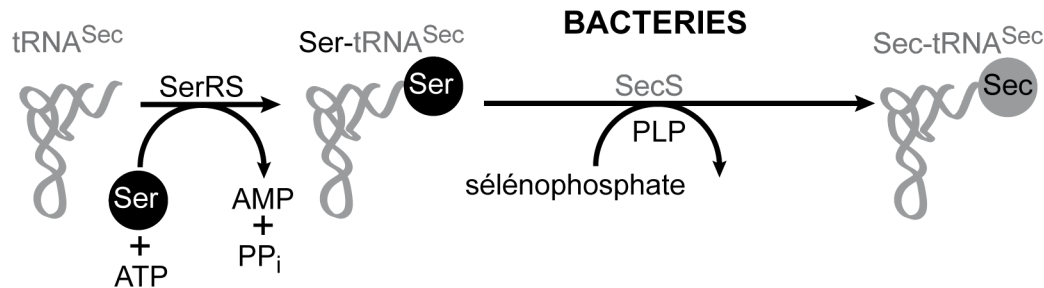
D. General information on the reprogramming and dynamics of the genetic code

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

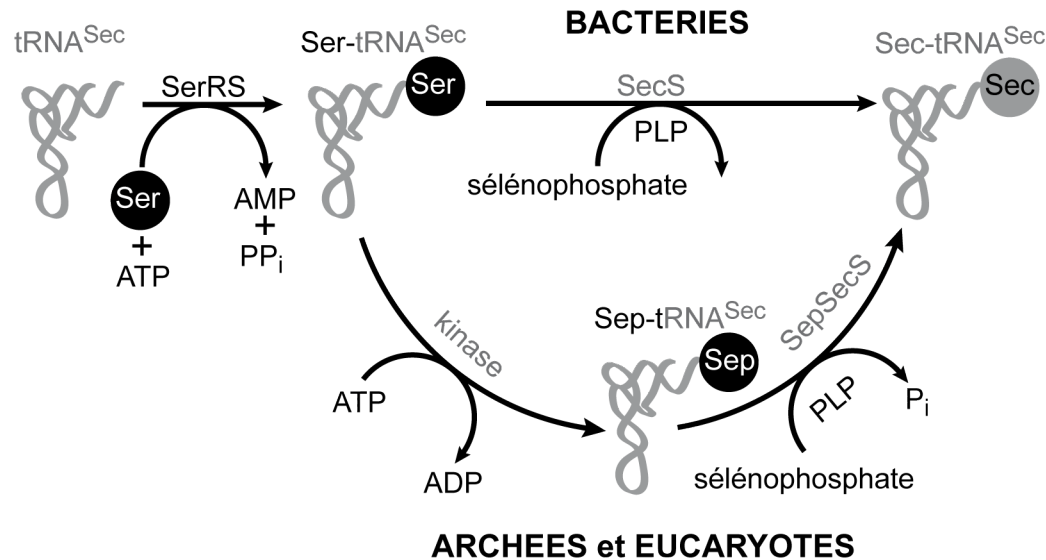
D. General information on the reprogramming and dynamics of the genetic code

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

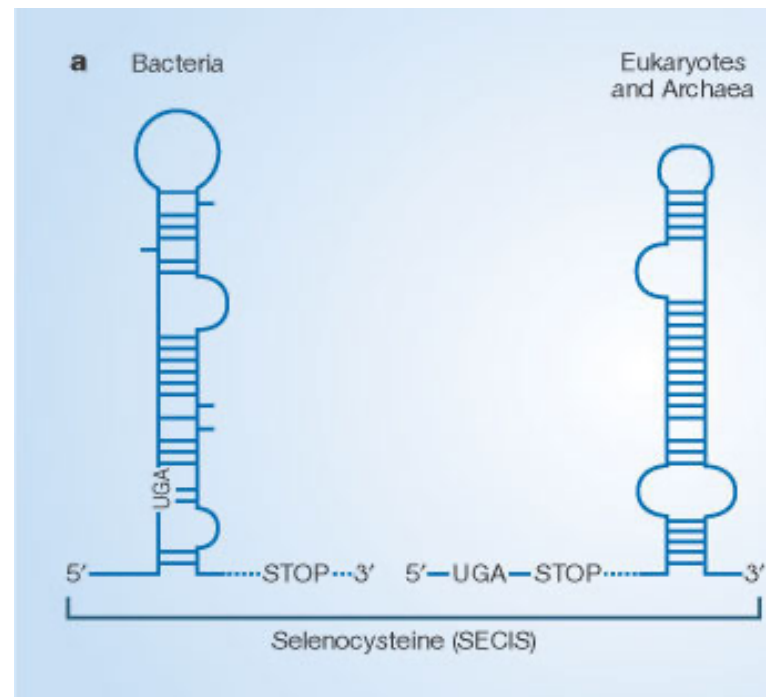
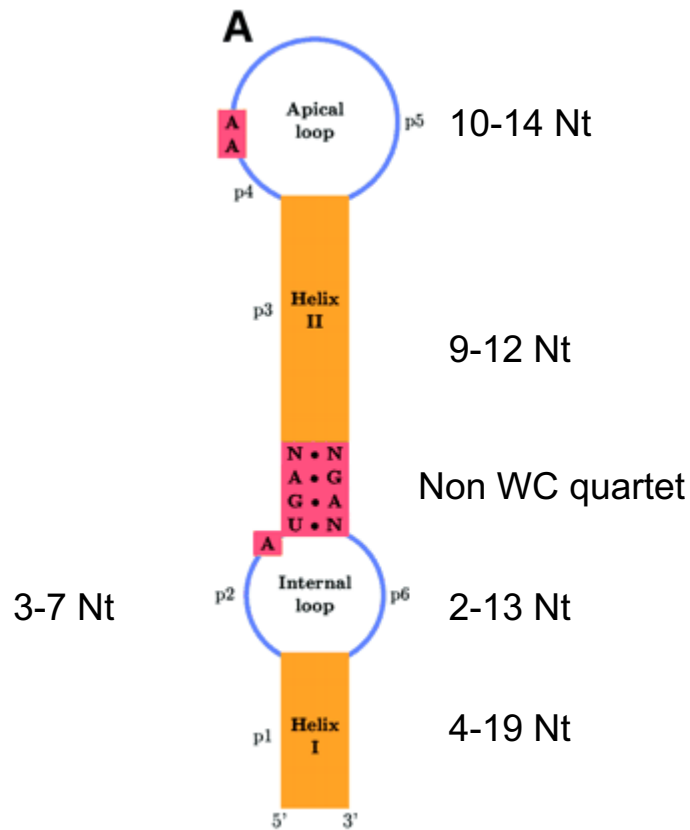


II. The Genetic Code

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

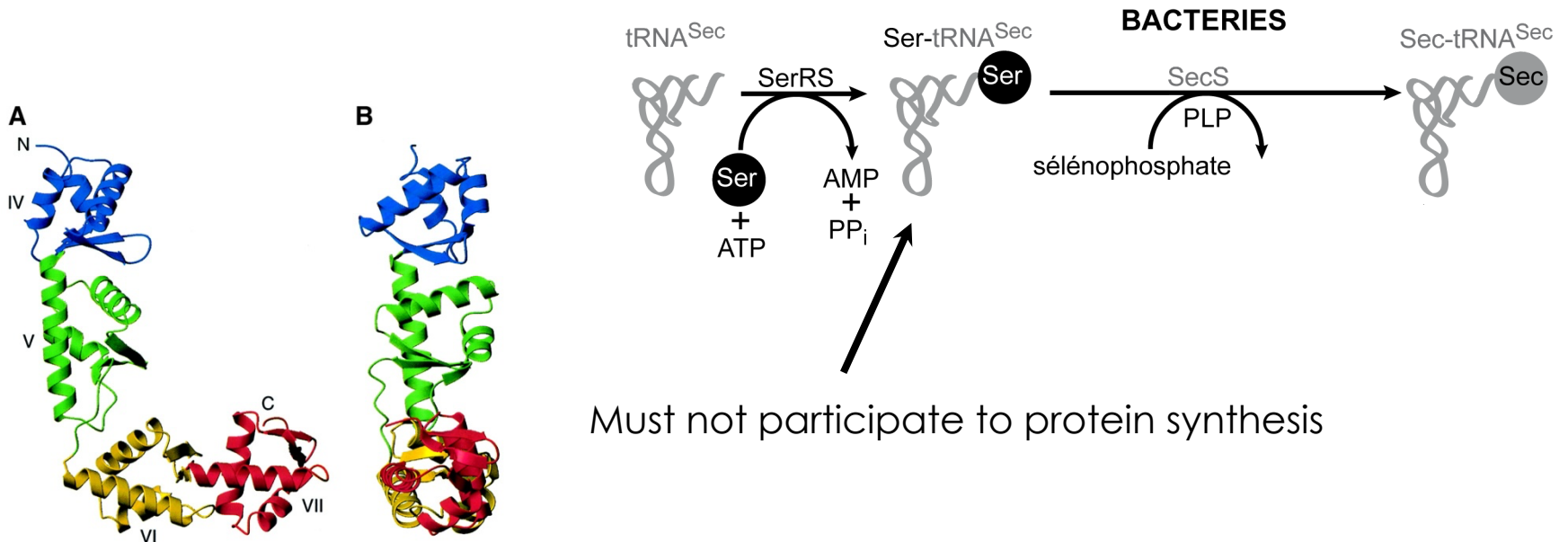


II. The Genetic Code

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

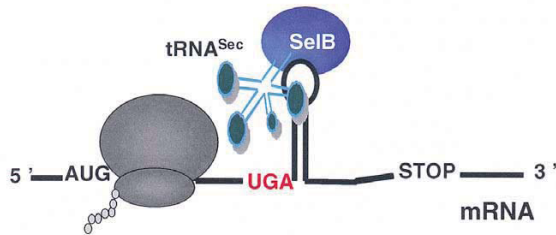
II. The Genetic Code

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

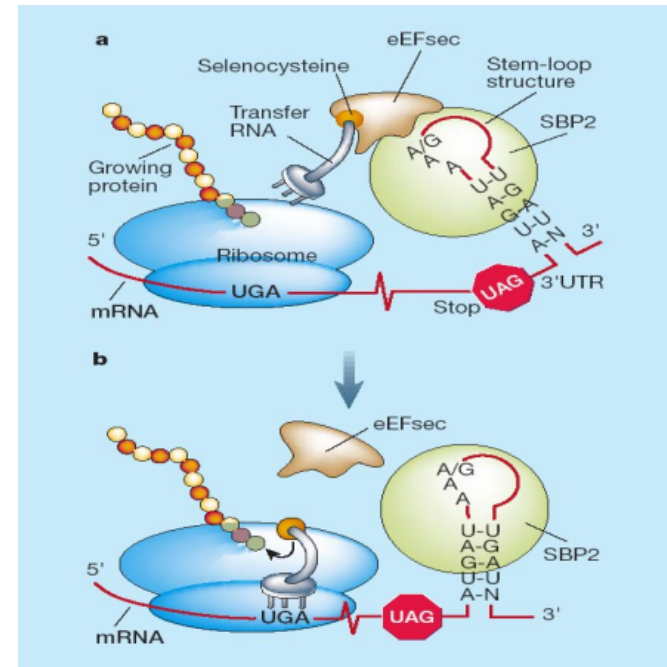
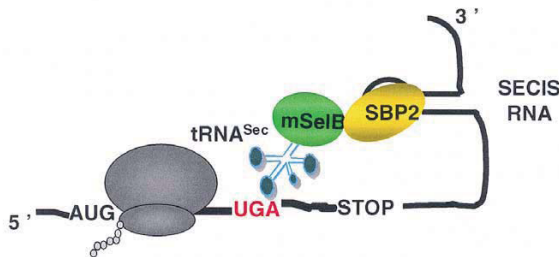
2. Reassignment/reprogramming of STOP codons: the 21st genetically-encoded aa: Sec

- A specific elongation factor of the new aa-tRNA

Eubacteria: One protein, two functions



Eukaryotes: Two proteins, more than two functions ?



Selenoprotein Synthesis

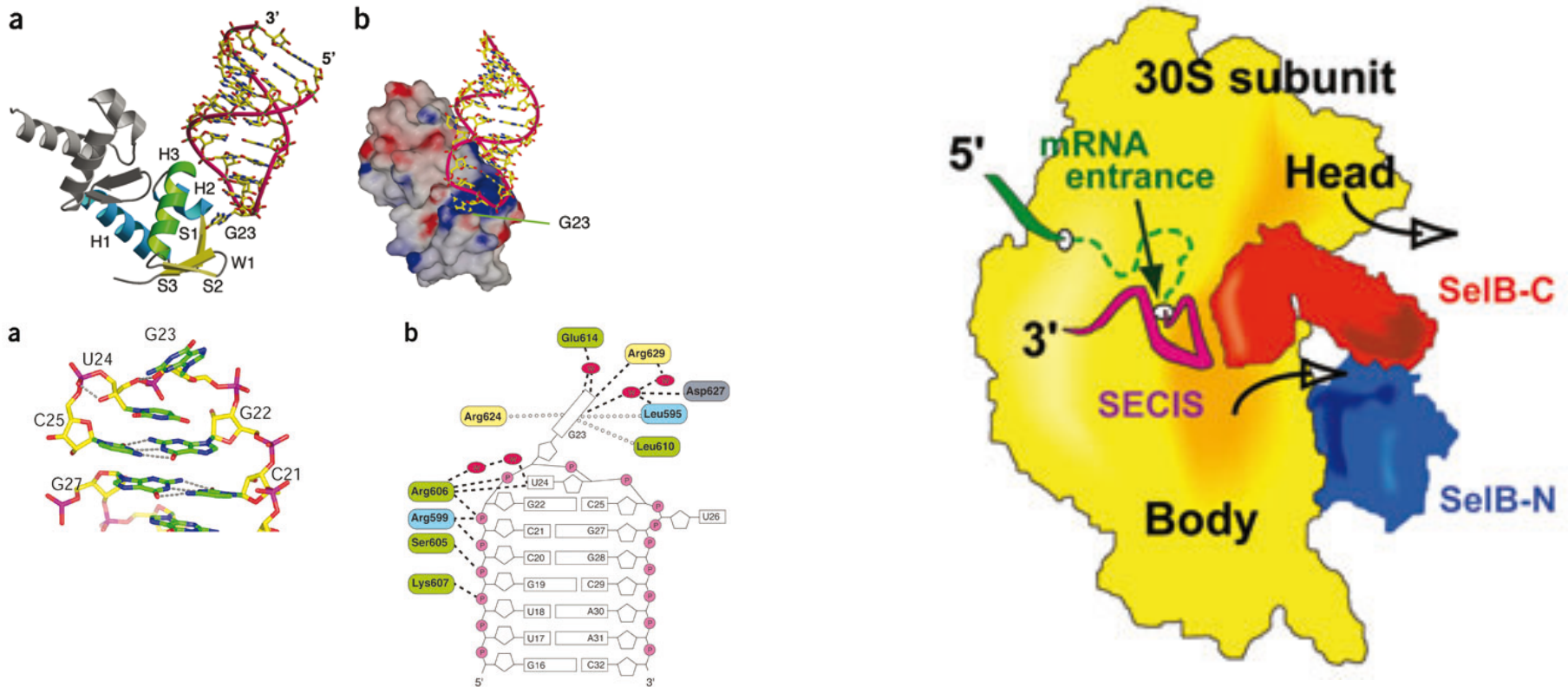
⚠ The addition of Sec to the code repertoire requires a specific elongation factor: **SelB** in Bacteria and Archaea, and **EF-Sec** in eukaryotes

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

