



>AF215832 TCTCTAAGGCGGCCCTCCTTCACCCCAAGGTGGTCAGGACAAGCCCTTGCTTCTACCT



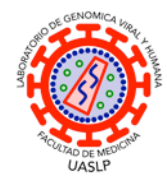
# The Genetic Code

**San Luis Potosi State University (UASLP) Mexico**  
**Molecular Biology Course, Faculty of Medicine post-graduate program**

*Dr. Christian A. García-Sepúlveda*

Viral & Human Genomics BSL-3 Laboratory

Last updated October 07, 2025 v2



# Medical relevance of The Genetic Code

---

Set of rules by which information encoded in DNA/RNA is translated into proteins.

Universal to nearly all living organisms and the foundation for understanding how genes dictate cellular function and organismal traits.

Understanding the genetic code is pivotal in modern medicine.

It allows for the identification of genetic mutations that cause inherited diseases, guiding the development of targeted therapies and precision medicine.

Knowledge of the genetic code facilitates advancements in gene therapy, where faulty genes can be corrected,

In pharmacogenomics, tailors drug treatments to an individual's genetic profile, enhancing efficacy and reducing adverse effects.

Moreover, insights into the genetic code enable the development of novel vaccines and the understanding of cancer genomics.

# Gamow's Diamonds

Georgiy Antonovich Gamow (1904 – 1968).

Ruso-American theoretical physicist and cosmologist.

Early advocate and developer of Big Bang theory.

First mathematical model of the atomic nucleus.

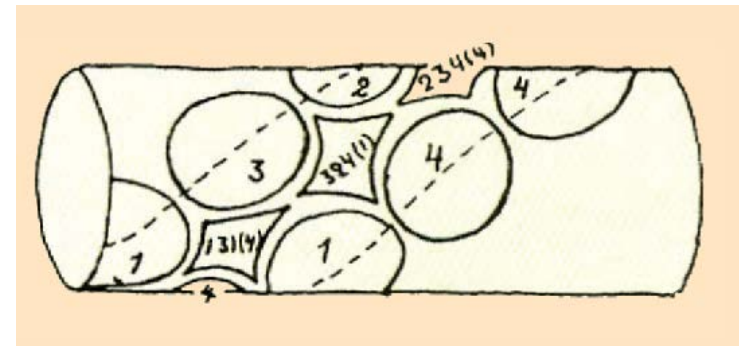
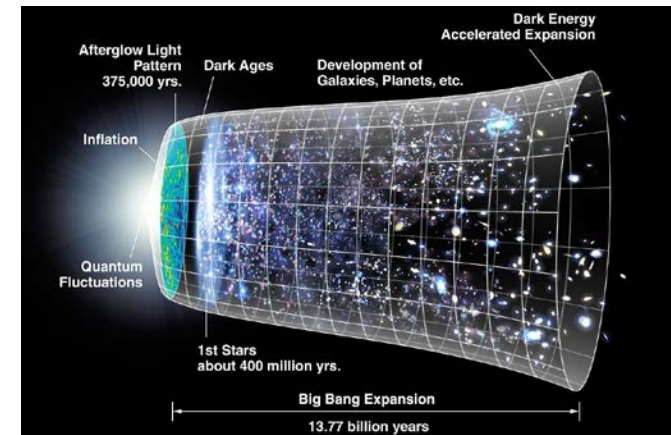
Discovered alpha decay.

Worked on star formation, stellar nucleosynthesis, big bang nucleosynthesis, nucleocosmogenesis and molecular genetics.

Proposed "Gamow's diamonds" for an overlapping, non-degenerate code.

*"The Most Beautiful False Theory in Biochemistry"*

Francis Crick based his work on Gamow's theoretical background.



# Crick-Brenner experiment (codons are triplets)

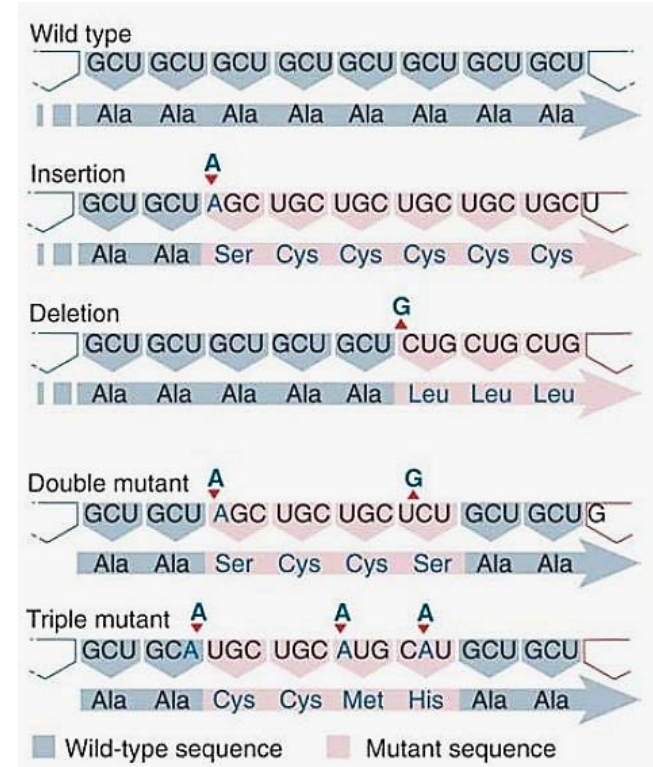
That codons did consist of three DNA bases was first demonstrated in the Crick- Brenner experiment (1961).

The experiment elucidated the nature of gene expression and frameshift mutations.

Proflavin-induced indel mutations (insertions/deletions) of the T4 bacteriophage.

Mutants with 1, 2 or 4 indels did not produce rIIB (frameshifts).

Mutants with 3 indels (or multiples) did produce rIIB (albeit anomalous).



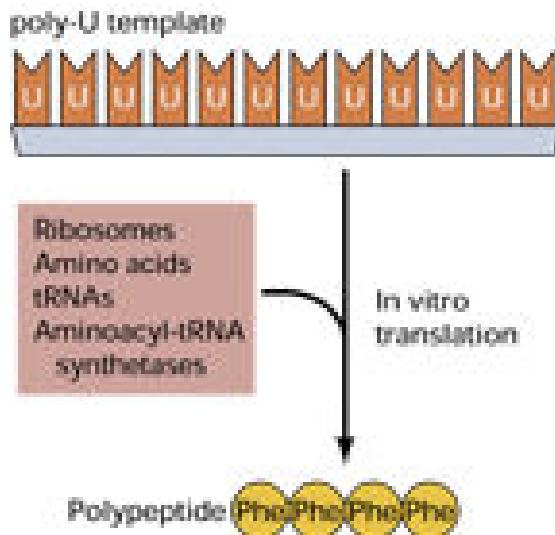
ORF	DNA sequence	Protein sequence
Normal =	ATG TGC TGA CTG ATC GGT	= M <b>C S L I G</b>
-1 =	TGT GCT GAC TGA TCG GT	= C A D S S
-2 =	GTG CTG ACT GAT CGG T	= V L T D R
-3 =	TGC TGA CTG ATC GGT	= <b>C S L I G</b>

# Nirenberg-Mathei experiment (descyphering codons)

First descyphering of a codon in 1961 at the NIH.

Used a cell-free system to translate a poly-uracil RNA sequence to a phenylalanine repeat peptide.

Deduced that UUU = phenylalanine.



	U	C	A	G	
U	UUU } Phe UUC } UUA } Leu UUG }	UCU } Ser UCC } UCA } UCG }	UAU } Tyr UAC } UAA } Stop UAG } Stop	UGU } Cys UGC } UGA } Stop UGG } Trp	U C A G
C	CUU } Leu CUC } CUA } CUG }	CCU } Pro CCC } CCA } CCG }	CAU } His CAC } CAA } Gln CAG }	CGU } Arg CGC } CGA } CGG }	U C A G
A	AUU } Ile AUC } AUA } AUG }	ACU } Thr ACC } ACA } ACG }	AAU } Asn AAC } AAA } Lys AAG }	AGU } Ser AGC } AGA } Arg AGG }	U C A G
G	GUU } Val GUC } GUA } GUG }	GCU } Ala GCC } GCA } GCG }	GAU } Asp GAC } GAA } Glu GAG }	GGU } Gly GGC } GGA } GGG }	U C A G

Extending this work, Nirenberg & Matthaei determined the nucleotide makeup of each codon.

# Nirenberg-Mathei experiment (decyphering codons)

\* Marshall Nirenberg \*  
-see sheets

	C <sup>14</sup> Phe	C <sup>14</sup> Arg	C <sup>14</sup> Asp	C <sup>14</sup> Asn	H <sup>3</sup> Cys	C <sup>14</sup> Glu	C <sup>14</sup> Gln	C <sup>14</sup> Gly	C <sup>14</sup> His	C <sup>14</sup> Ileu	C <sup>14</sup> Leu	C <sup>14</sup> Lys	C <sup>14</sup> Met	C <sup>14</sup> Phe	C <sup>14</sup> Pro	C <sup>14</sup> Ser	C <sup>14</sup> Thr	H <sup>3</sup> Trp	C <sup>14</sup> Tyr	C <sup>14</sup> Val
UUU	-0.11	-0.30	-0.03	-0.05	-	-0.03	-0.07	-0.03	-0.01	0.03	-0.24	-0.11	-0.10	1.29 <sup>a</sup>	0.03	-0.04	-0.26	-0.13	0.05	0.01
UUC	0.01	-0.27	0.03	0.10	-	0.01	0.13	-0.05	0.04	0.01	0.31	0.02	0.01	1.59 <sup>a</sup>	-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	0.02	-0.02	-0.02	0.04	0.01	-0.03	0.05	0.01	0.00
UUG	-0.22	-0.02	-0.07	-0.05	-	0.02	-0.08	0.01	0.05	0.00	0.39	-0.06	0.17	-0.02	-0.07	-0.24	-0.10	0.05	0.01	0.01
UUU	0.01	-0.31	0.05	-0.05	-	0.06	0.19	-0.18	0.00	0.05	-0.11	0.09	-0.17	0.02	-0.01	1.37	-0.09	-	-0.02	0.01
UUC	0.01	-0.32	-0.01	-0.02	-	0.03	-0.16	-0.01	-0.01	0.00	-0.06	0.01	-0.02	-0.19	-0.09	0.13	-0.05	0.03	0.04	0.02
UUA	-0.01	0.08	-0.08	-0.05	-	-	0.20	0.17	-	-0.01	-	0.00	-	0.02	-0.05	0.12	-	-0.03	0.01	-0.06
UUG	-0.20	0.00	0.00	0.07	-	0.01	-0.23	-0.41	0.01	0.02	0.00	0.04	-0.11	0.03	0.06	1.09 <sup>e</sup>	0.04	0.03	0.05	0.03
UUAU	-0.02	-0.41	-0.21	-0.02	-	0.00	-0.11	-0.23	-0.04	0.01	-0.24	-0.09	0.03	-0.37	0.00	-0.02	-0.05	0.01	0.81	0.03
UAC	-0.01	-0.37	-0.02	0.02	-	-0.02	-0.16	-0.13	-0.04	-0.01	-0.21	-0.06	-0.04	-0.23	0.02	0.00	-0.11	0.02	0.56	0.00
UAA	-0.02	-0.41	-0.01	-0.01	-	0.02	-0.30	-0.31	-0.02	0.00	-0.05	0.02	-0.09	-0.38	0.02	0.02	-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	0.00
UUAU	-0.05	-0.03	0.06	0.03	0.93	0.03	-0.16	-0.73	-0.01	0.03	-0.05	-0.10	0.02	-0.01	-0.02	-0.06	-0.09	0.06	0.02	0.14
UUC	0.22	-0.18	0.05	0.12	0.74	0.04	-0.05	-0.56	-0.02	-0.02	-0.01	0.00	-0.15	0.04	0.02	0.02	-	0.04	0.03	0.02
UUA	-0.12	-0.42	0.07	0.14	0.10	0.10	-0.13	-0.30	-0.02	0.02	-0.04	-0.11	0.00	-0.37	0.00	0.00	-0.05	0.03	0.03	0.05
UUG	0.04	-0.36	0.01	0.02	-0.01	-0.01	0.28	-	-0.02	-0.05	-0.01	0.00	0.06	0.06	0.14	0.14	-0.02	0.03	0.03	-0.01

# Canonical genetic code (three positions)

- Canonical Genetic Code
- From the codon perspective
- Interpretation based on codon sequence

	U	C	A	G	
U	UUU Phe UUC UUA Leu UUG	UCU UCC Ser UCA UCG	UAU Tyr UAC UAA Stop UAG Stop	UGU Cys UGC UGA Stop UGG Trp	U C A G
C	CUU Leu CUC CUA CUG	CCU Pro CCC CCA CCG	CAU His CAC CAA Gln CAG	CGU Arg CGC CGA CGG	U C A G
A	AUU Ile AUC AUA AUG	ACU Thr ACC ACA	AAU Asn AAC AAA Lys	AGU Ser AGC AGA Arg AGG	U C A G
G	GUU Val GUC GUA GUG	GCU Ala GCC GCA GCG	GAU Asp GAC GAA Glu GAG	GGU Gly GGC GGA GGG	U C A G

First position (5' end) | Third position (3' end)

# AUG

### Amino acid names:

Ala = alanine	Gln = glutamine	Leu = leucine	Ser = serine
Arg = arginine	Glu = glutamate	Lys = lysine	Thr = threonine
Asn = asparagine	Gly = glycine	Met = methionine	Trp = tryptophan
Asp = aspartate	His = histidine	Phe = phenylalanine	Tyr = Tyrosine
Cys = cysteine	Ile = Isoleucine	Pro = proline	Val = valine

# Alternative genetic codes

Slight variations on the canonical code had been predicted

Alternative codes were discovered in 1979, in human mitochondria.

Many alternative mitochondrial codes now known.

Mycoplasma variants translate **UGA** as tryptophan.

In bacteria and archaea, **GUG** and **UUG** are common start codons.

	U	C	A	G	
First position (5' end)	UUU } Phe UUC } UUA } <b>UUG</b> } Leu	UCU } UCC } Ser UCA } UCG }	UAU } Tyr UAC } UAA } Stop UAG } Stop	UGU } Cys UGC } <b>UGA</b> } Stop UGG } Trp	U C A G
	CUU } CUC } Leu CUA } CUG }	CCU } CCC } Pro CCA } CCG }	CAU } His CAC } CAA } Gln CAG }	CGU } CGC } Arg CGA } CGG }	U C A G
	AUU } AUC } Ile AUA } AUG }	ACU } ACC } Thr ACA } ACG }	AAU } Asn AAC } AAA } Lys AAG }	AGU } AGC } Ser AGA } Arg AGG }	U C A G
	GUU } GUC } Val GUA } <b>GUG</b> }	GCU } GCC } Ala GCA } GCG }	GAU } Asp GAC } GAA } Glu GAG }	GGU } GGC } Gly GGA } GGG }	U C A G

Amino acid names:

Ala = alanine	Gln = glutamine	Leu = leucine	Ser = serine
Arg = arginine	Glu = glutamate	Lys = lysine	Thr = threonine
Asn = asparagine	Gly = glycine	Met = methionine	Trp = tryptophan
Asp = aspartate	His = histidine	Phe = phenylalanine	Tyr = Tyrosine
Cys = cysteine	Ile = Isoleucine	Pro = proline	Val = valine

# Alternative genetic codes

In certain proteins, non-standard amino acids are encoded for by standard stop codons.

**UGA** = Selenocysteine (21<sup>st</sup>)

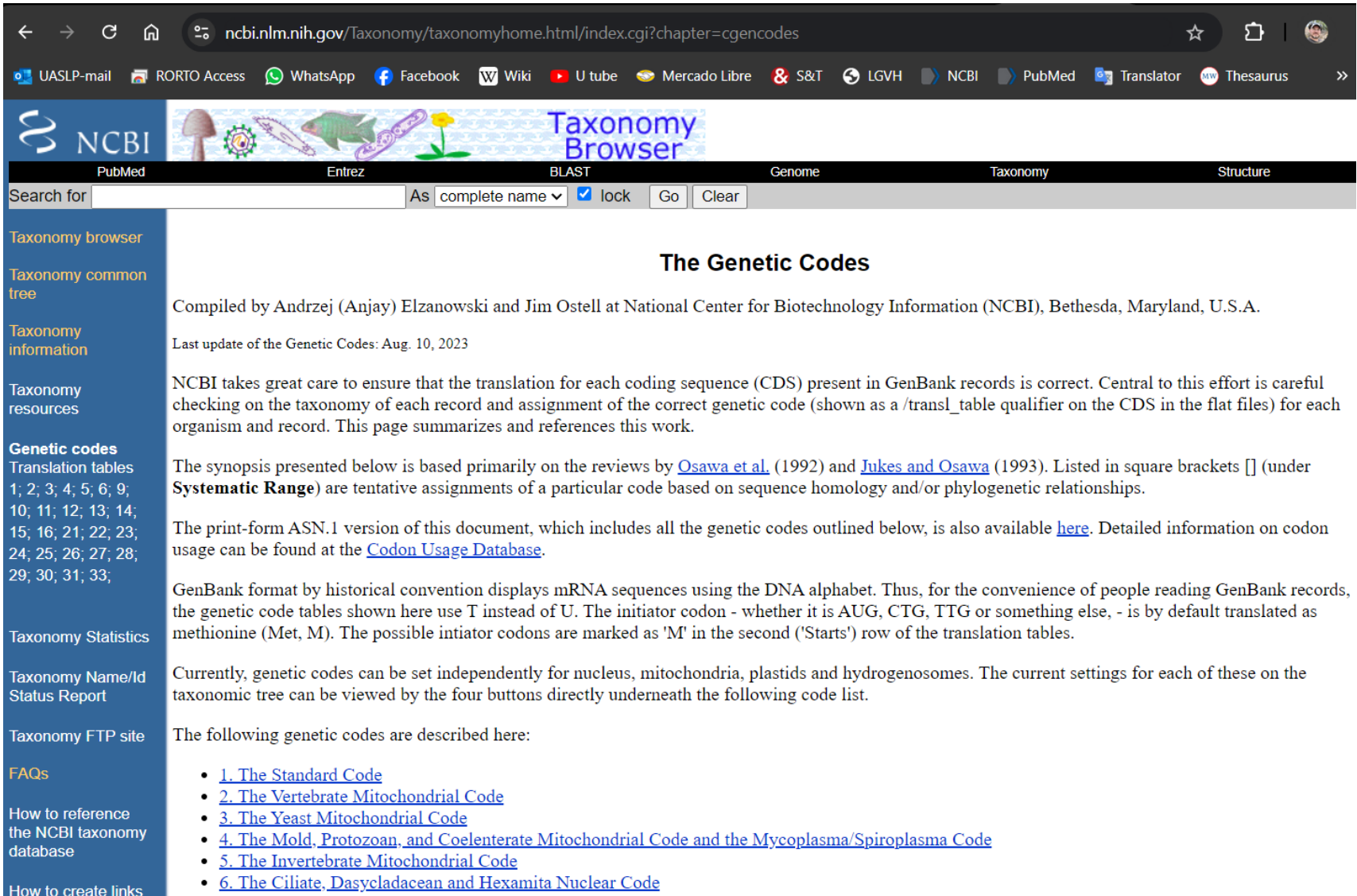
**UAG** = Pyrrolysine (22<sup>nd</sup>)

Depend on associated signal sequences in the mRNA.

	U	C	A	G	
U	UUU } Phe	UCU } Ser	UAU } Tyr	UGU } Cys	U
	UUC } Phe	UCC } Ser	UAC } Tyr	UGC } Cys	C
	UUA } Leu	UCA } Ser	UAA } Stop	UGA } Stop	A
	UUG } Leu	UCG } Ser	UAG } Stop	UGG } Trp	G
C	CUU } Leu	CCU } Pro	CAU } His	CGU } Arg	U
	CUC } Leu	CCC } Pro	CAC } His	CGC } Arg	C
	CUA } Leu	CCA } Pro	CAA } Gln	CGA } Arg	A
	CUG } Leu	CCG } Pro	CAG } Gln	CGG } Arg	G
A	AUU } Ile	ACU } Thr	AAU } Asn	AGU } Ser	U
	AUC } Ile	ACC } Thr	AAC } Asn	AGC } Ser	C
	AUA } Ile	ACA } Thr	AAA } Lys	AGA } Arg	A
	AUG } Met	ACG } Thr	AAG } Lys	AGG } Arg	G
G	GUU } Val	GCU } Ala	GAU } Asp	GGU } Gly	U
	GUC } Val	GCC } Ala	GAC } Asp	GGC } Gly	C
	GUA } Val	GCA } Ala	GAA } Glu	GGA } Gly	A
	GUG } Val	GCG } Ala	GAG } Glu	GGG } Gly	G

Amino acid names:

Ala = alanine	Gln = glutamine	Leu = leucine	Ser = serine
Arg = arginine	Glu = glutamate	Lys = lysine	Thr = threonine
Asn = asparagine	Gly = glycine	Met = methionine	Trp = tryptophan
Asp = aspartate	His = histidine	Phe = phenylalanine	Tyr = Tyrosine
Cys = cysteine	Ile = Isoleucine	Pro = proline	Val = valine



The screenshot shows the NCBI Taxonomy Browser interface. The browser address bar displays `ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html/index.cgi?chapter=cgencodes`. The page title is "The Genetic Codes". The main content area includes the following text:

Compiled by Andrzej (Anjay) Elzanowski and Jim Ostell at National Center for Biotechnology Information (NCBI), Bethesda, Maryland, U.S.A.

Last update of the Genetic Codes: Aug. 10, 2023

NCBI takes great care to ensure that the translation for each coding sequence (CDS) present in GenBank records is correct. Central to this effort is careful checking on the taxonomy of each record and assignment of the correct genetic code (shown as a `/transl_table` qualifier on the CDS in the flat files) for each organism and record. This page summarizes and references this work.

The synopsis presented below is based primarily on the reviews by [Osawa et al.](#) (1992) and [Jukes and Osawa](#) (1993). Listed in square brackets [] (under **Systematic Range**) are tentative assignments of a particular code based on sequence homology and/or phylogenetic relationships.

The print-form ASN.1 version of this document, which includes all the genetic codes outlined below, is also available [here](#). Detailed information on codon usage can be found at the [Codon Usage Database](#).

GenBank format by historical convention displays mRNA sequences using the DNA alphabet. Thus, for the convenience of people reading GenBank records, the genetic code tables shown here use T instead of U. The initiator codon - whether it is AUG, CTG, TTG or something else, - is by default translated as methionine (Met, M). The possible initiator codons are marked as 'M' in the second ('Starts') row of the translation tables.

Currently, genetic codes can be set independently for nucleus, mitochondria, plastids and hydrogenosomes. The current settings for each of these on the taxonomic tree can be viewed by the four buttons directly underneath the following code list.

The following genetic codes are described here:

- 1. [The Standard Code](#)
- 2. [The Vertebrate Mitochondrial Code](#)
- 3. [The Yeast Mitochondrial Code](#)
- 4. [The Mold, Protozoan, and Coelenterate Mitochondrial Code and the Mycoplasma/Spiroplasma Code](#)
- 5. [The Invertebrate Mitochondrial Code](#)
- 6. [The Ciliate, Dasycladacean and Hexamita Nuclear Code](#)

# Alternative genetic codes

The Standard Code

The Vertebrate Mitochondrial Code

The Yeast Mitochondrial Code

The Mold, Protozoan & Coelenterate Mitochondrial Code

The Mycoplasma/Spiroplasma Code

The Invertebrate Mitochondrial Code

The Ciliate, Dasycladacean and Hexamita Nuclear Code

The Echinoderm and Flatworm Mitochondrial Code

The Euplotid Nuclear Code

The Bacterial and Plant Plastid Code

The Alternative Yeast Nuclear Code

The Ascidian Mitochondrial Code

The Alternative Flatworm Mitochondrial Code

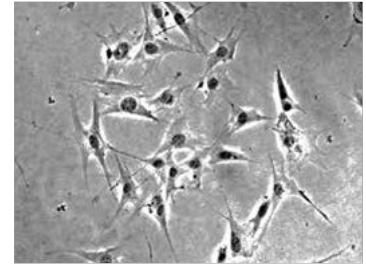
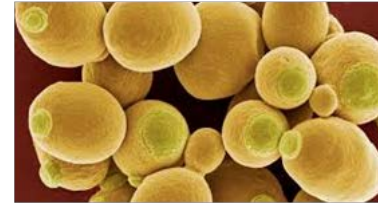
Blepharisma Nuclear Code

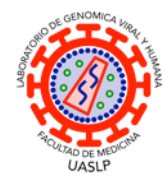
Chlorophycean Mitochondrial Code

Trematode Mitochondrial Code

Scenedesmus Obliquus Mitochondrial Code

Thraustochytrium Mitochondrial Code





# Alternative genetic codes





---

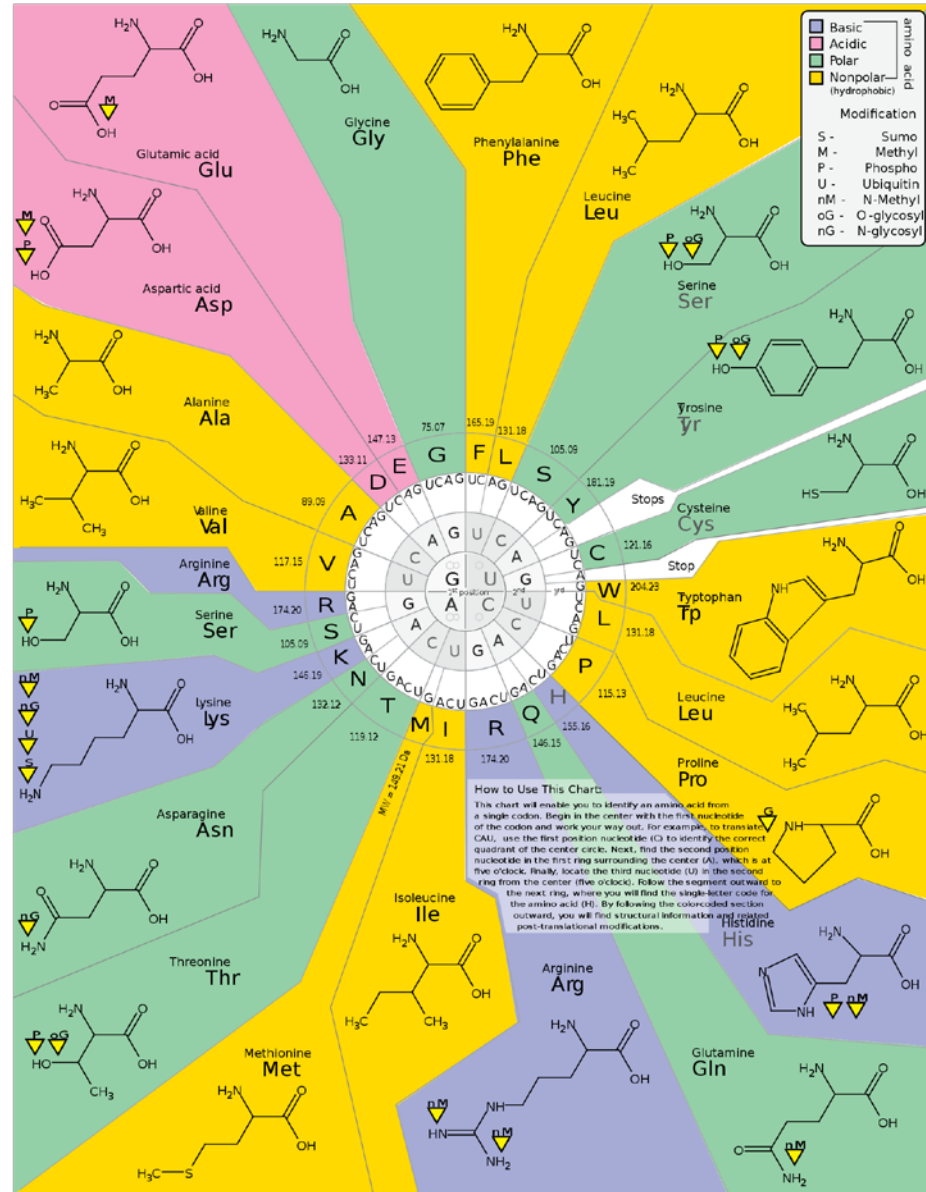
However, all known codes have strong similarities and the coding mechanism is the same for all organisms:

- three-base codons
- tRNA & ribosomes
- read the code in the same direction
- translating the code by codons
- 5´ to 3´
- Amino (NH)-terminus to Carboxy (COOH)-terminus

# Standard genetic code aminoacid properties

Uses colours to identify aminoacids having similar properties.

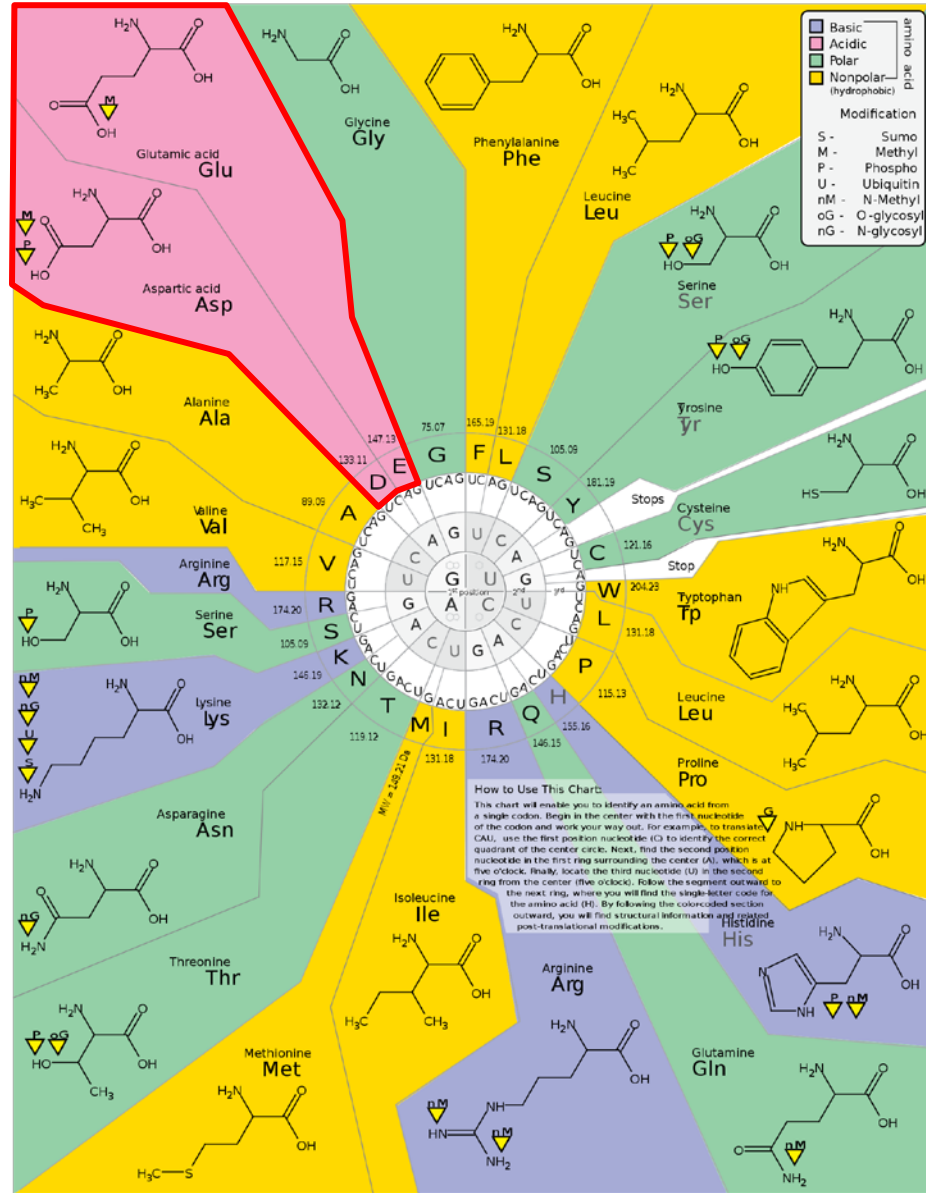
-  Basic
-  Acidic
-  Polar
-  Non-polar



# Acidic amino acids

## Acidic (Asp, Glu):

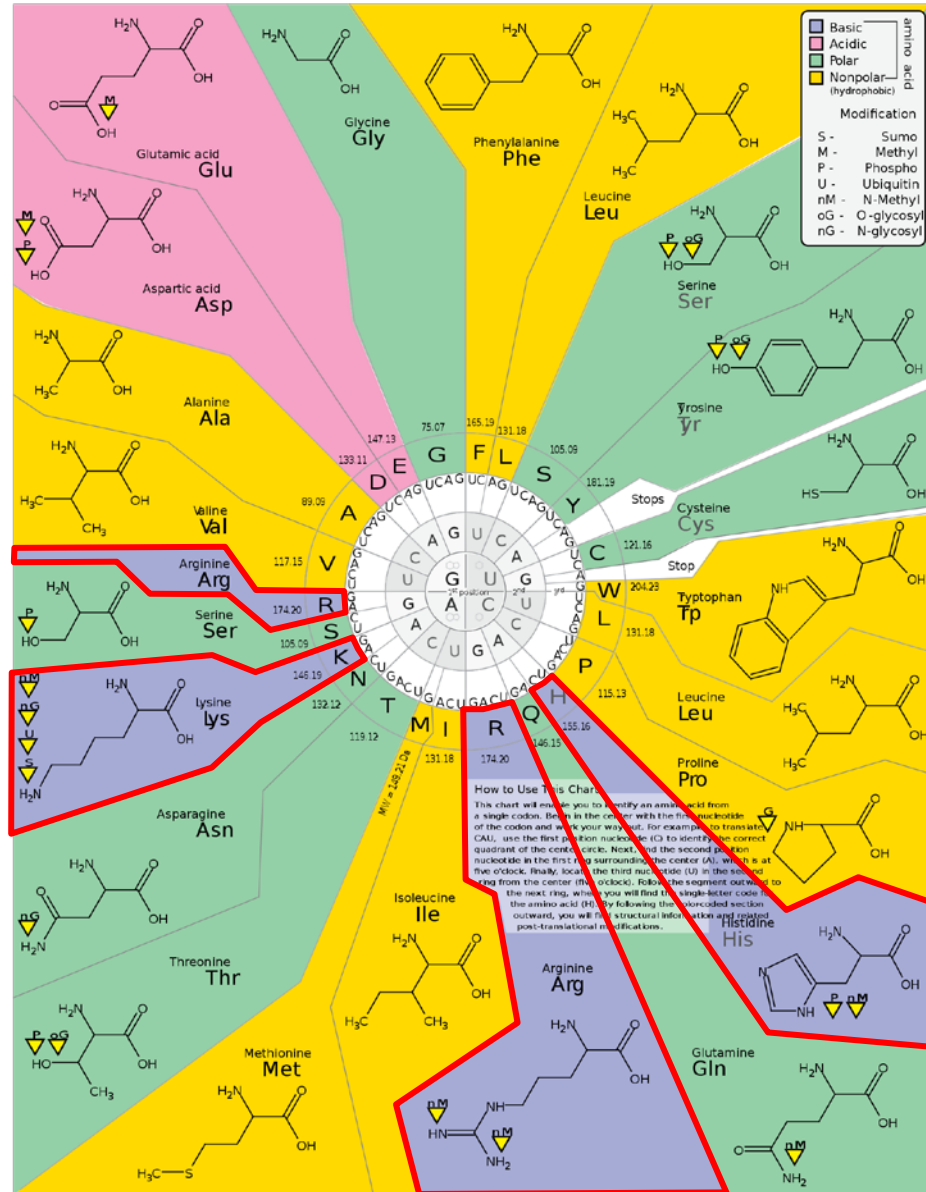
- Participate in enzyme active sites, facilitating catalysis through proton donation.
- Play a crucial role in maintaining protein structure by forming salt bridges with basic amino acids.
- within the lipid bilayer.



# Basic amino acids

## Basic Amino Acids (Lys, Arg, His):

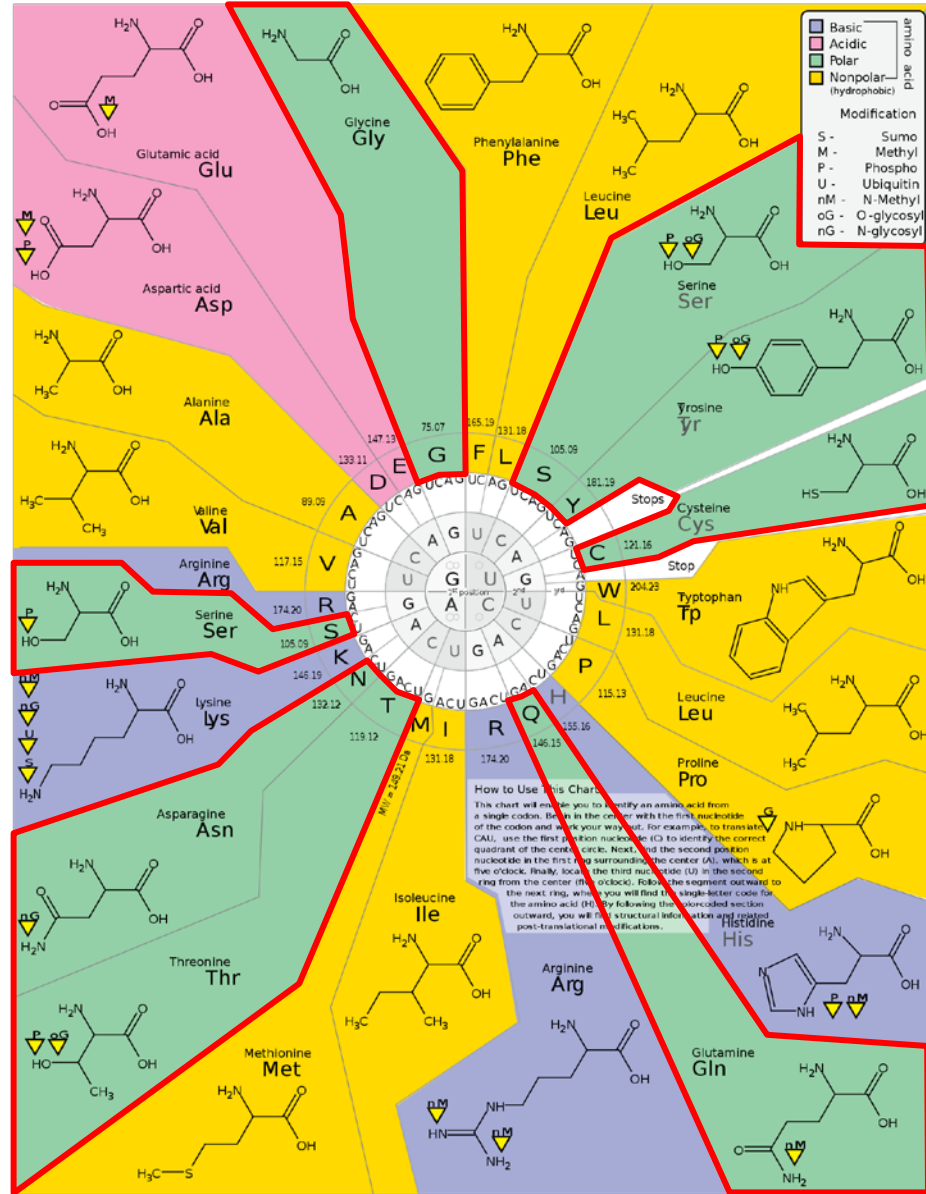
- Often found in DNA-binding proteins, interacting with the negatively charged phosphate backbone of DNA.
- Contribute to the formation of salt bridges, stabilizing protein structures.



# Polar amino acids

## Polar Amino Acids (Ser, Thr, Asn, Gln):

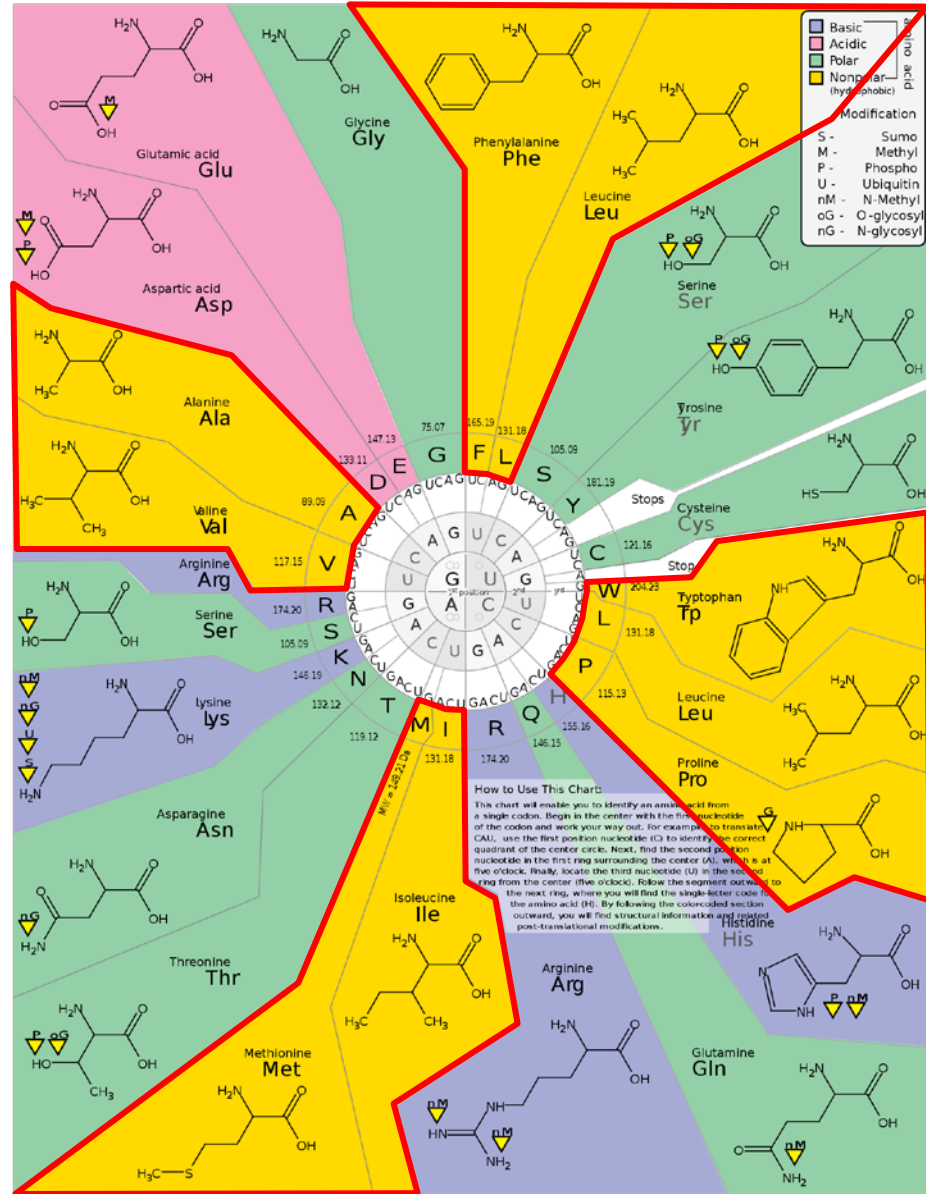
- Frequently involved in hydrogen bonding, contributing to the secondary and tertiary structures of proteins.
- Essential in active sites of enzymes, facilitating chemical reactions by stabilizing transition states.



# Non-polar amino acids

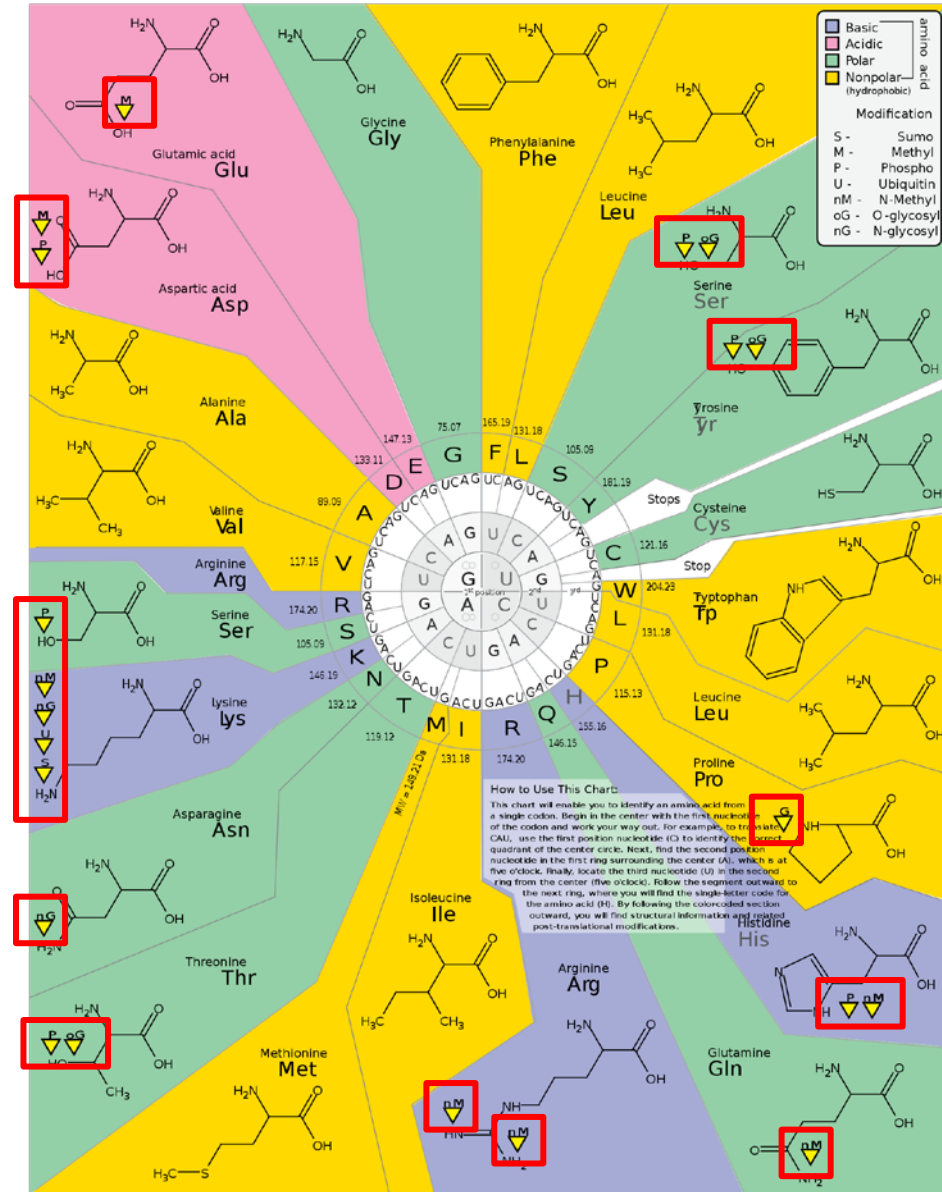
## Non-Polar Amino Acids (Ala, Val, Leu, Ile):

- Predominantly found in the interior of proteins, driving the folding process through hydrophobic interactions.
- Play a key role in forming membrane-spanning regions of transmembrane proteins, anchoring them within the lipid bilayer.



# Post-translational modifications

- S - Sumolation
- M - Methylation
- P - Phosphorylation
- U - Ubiquitination
- nM - N-Methylation
- oG - O-Glycosylation
- nG - N-Glycosylation



# Genetic Code degeneracy/redundancy

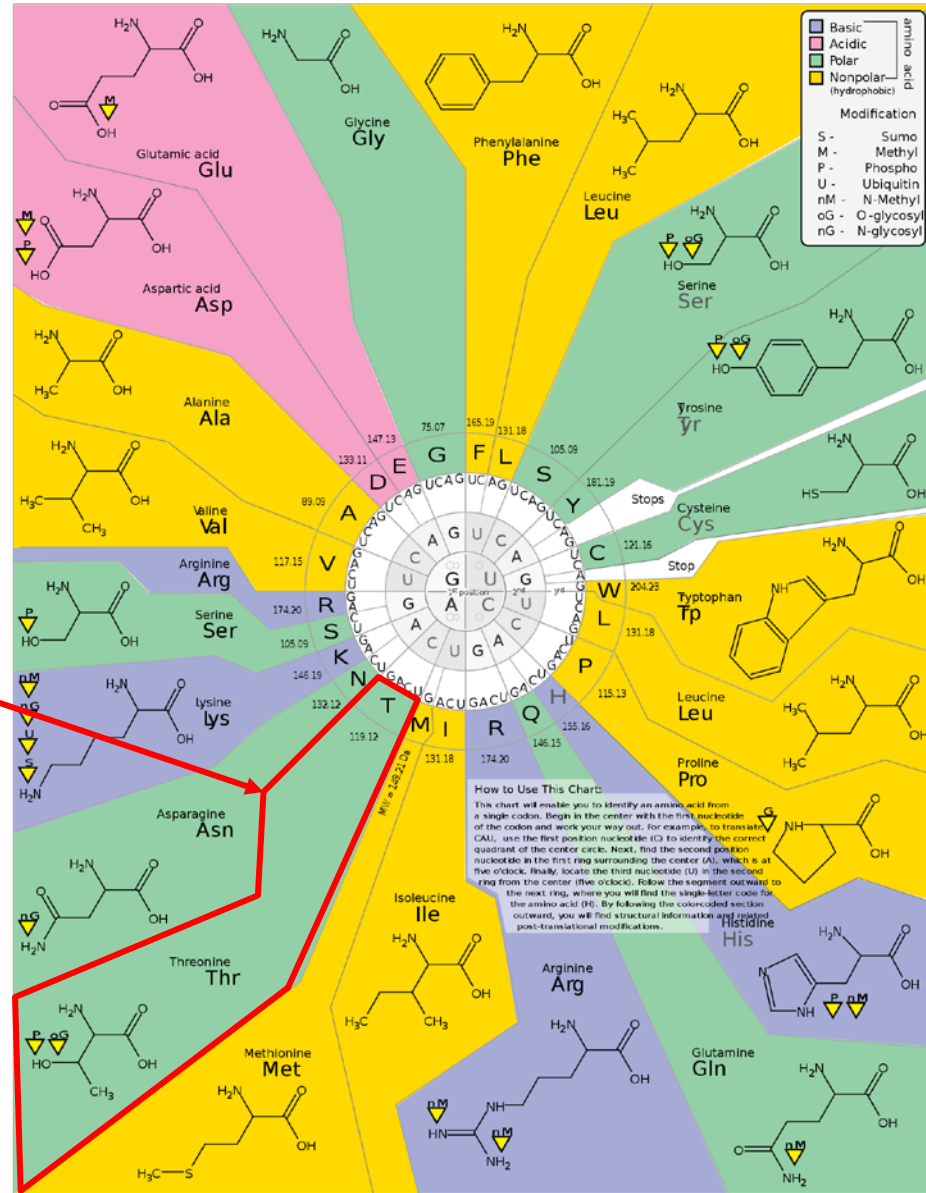
## Degeneracy or redundancy

More than one codon encoding for same amino acid,

ACC  
ACG  
ACU  
ACA

} **Thr**

Which means that the third bases is relatively free to mutate or allowed a wider degree of evolutionary freedom !

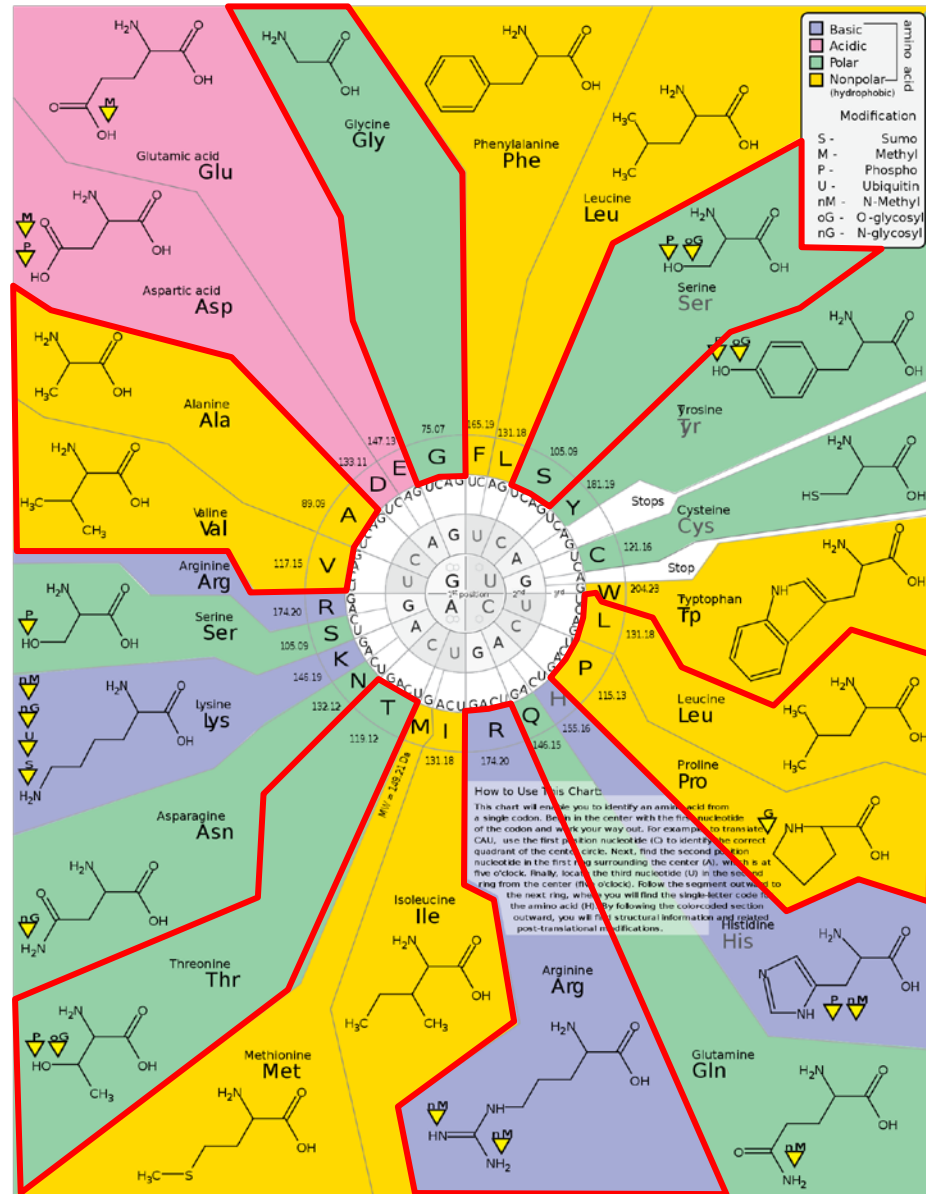


# Genetic Code degeneracy/redundancy

A position of a codon is said to be a **fourfold degenerate** site if any nucleotide at this position specifies the same amino acid.

For example, the third position of the glycine codons (GGA, GGG, GGC, GGU) is a fourfold degenerate site = all nucleotide substitutions at this site are synonymous.

Only the third positions of some codons may be fourfold degenerate.

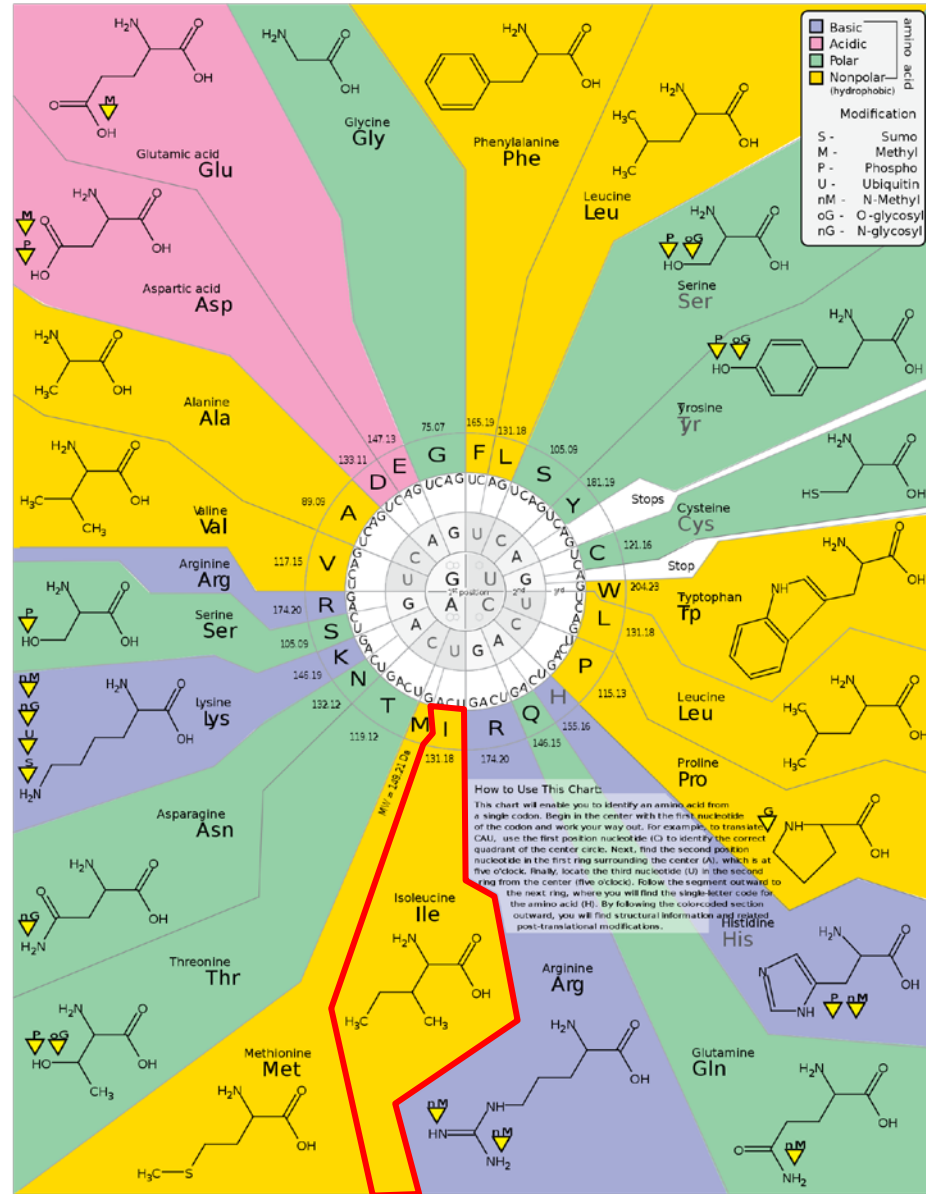


# Genetic Code degeneracy/redundancy

There is only one possible **threefold degenerate** site.

Where changing three of the four nucleotides has no effect on the amino acid, while changing the fourth possible nucleotide results in a NS substitution.

This is the case of Ile codon: AUU, AUC, or AUA all encode isoleucine, but AUG encodes methionine.



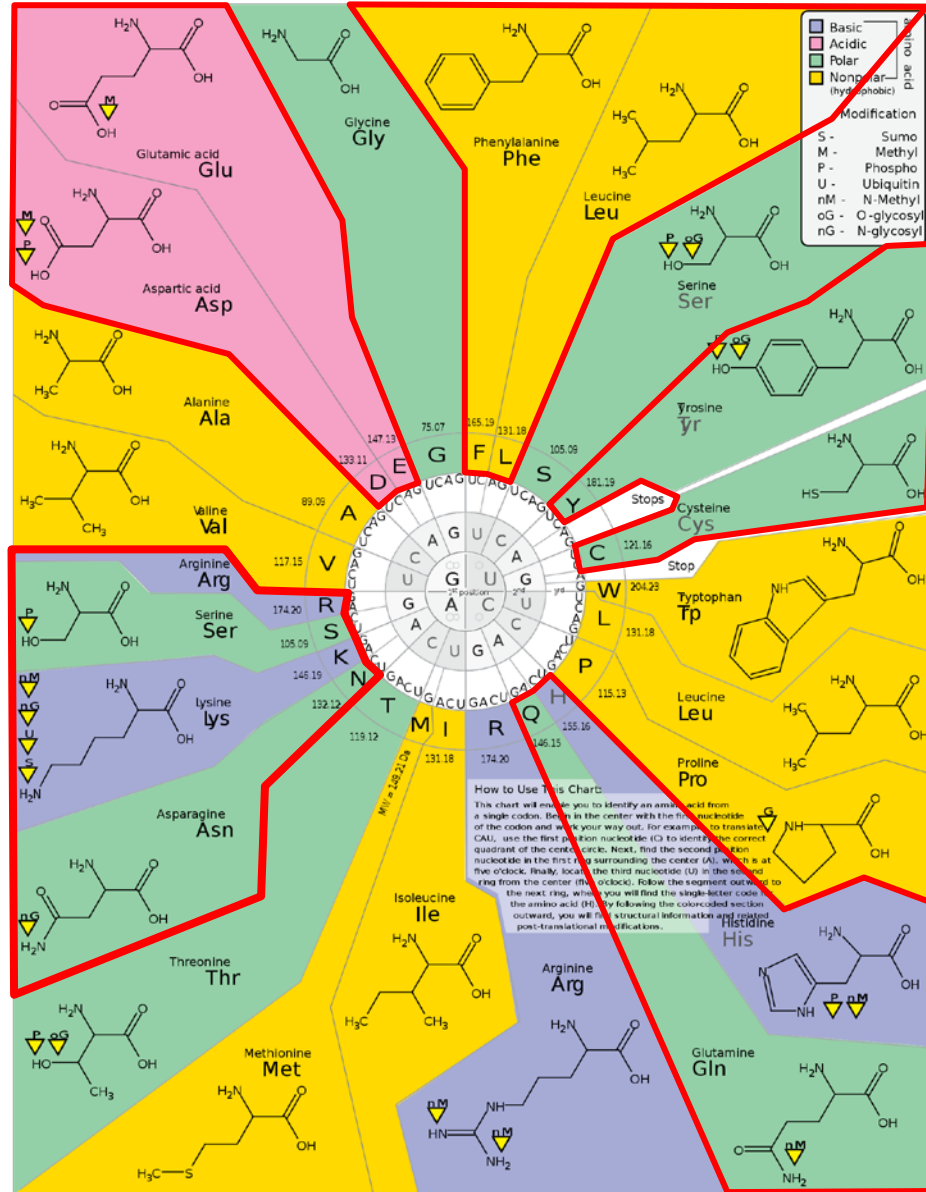
# Genetic Code degeneracy/redundancy

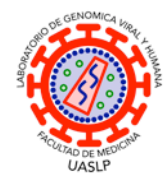
A position of a codon is said to be a **twofold degenerate** site if only two of four possible nucleotides at this position specify the same amino acid.

I.E. glutamic acid codons.

In twofold degenerate sites, the equivalent nucleotides are always either two purines (A/G) or two pyrimidines (C/U) = transitional substitutions.

Transversions are NS





# Genetic Code degeneracy/redundancy

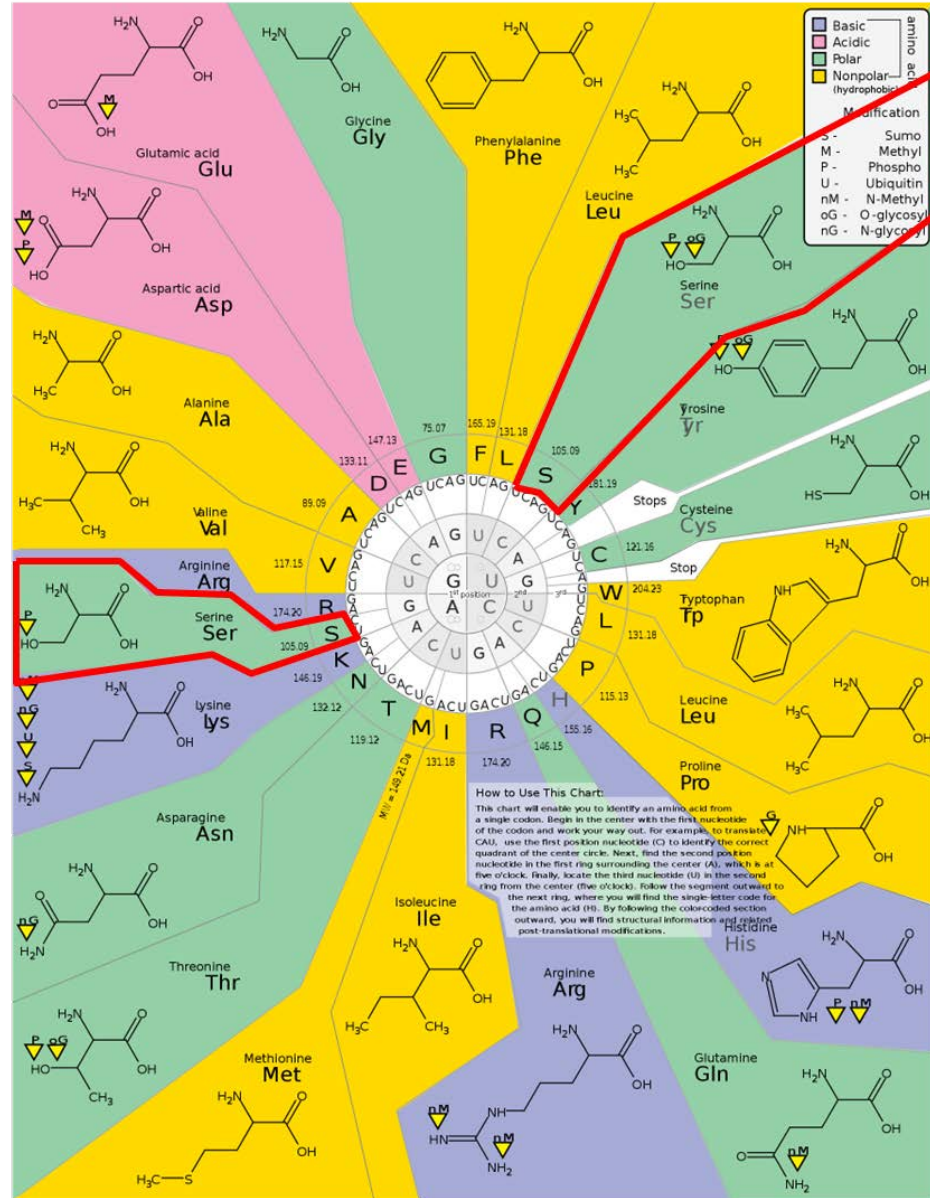
---

- The genetic code has redundancy (degeneracy) but no ambiguity.
- Redundancy: Different codons code for the same amino acid.
  - GAA & GAG both code for Glutamic Acid (**redundant**) but DO NOT code for any other amino acid (**Ambiguity**).
- Codon differences may fall in any position:
  - GAA & GAG = Glutamic Acid (difference in 3<sup>rd</sup> position).
  - UUA, UUG, CUU, CUC, CUA & CUG = Leucine (1<sup>st</sup> & 3<sup>rd</sup>).
  - UCA, UCG, UCC, UCU, AGU & AGC = Serine (1<sup>st</sup>, 2<sup>nd</sup> & 3<sup>rd</sup> ).

# Genetic Code degeneracy/redundancy

Only three amino acids are encoded by six different codons:

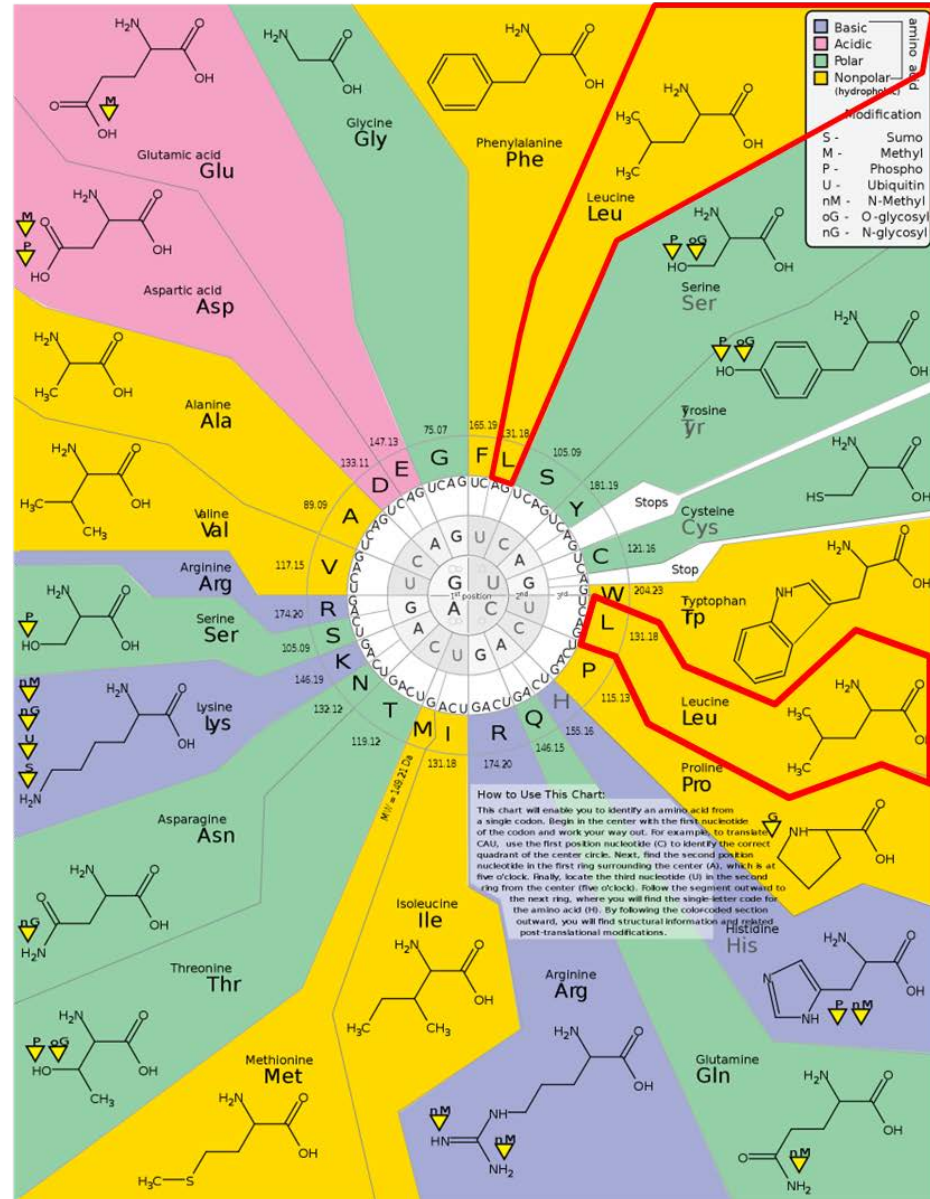
- Serine
- Leucine
- Arginine.



# Genetic Code degeneracy/redundancy

Only three amino acids are encoded by six different codons:

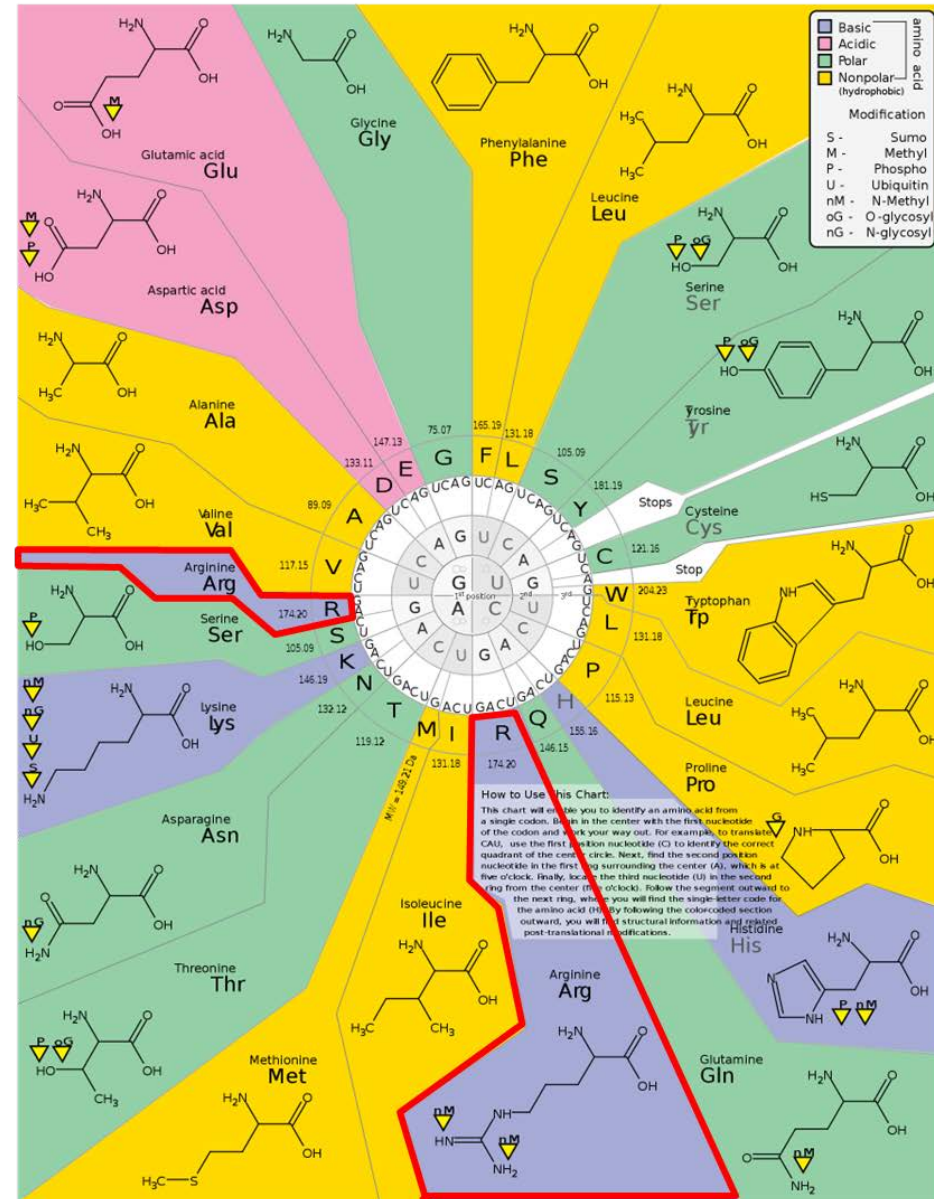
- Serine
- Leucine
- Arginine.



# Genetic Code degeneracy/redundancy

Only three amino acids are encoded by six different codons:

- Serine
- Leucine
- Arginine.

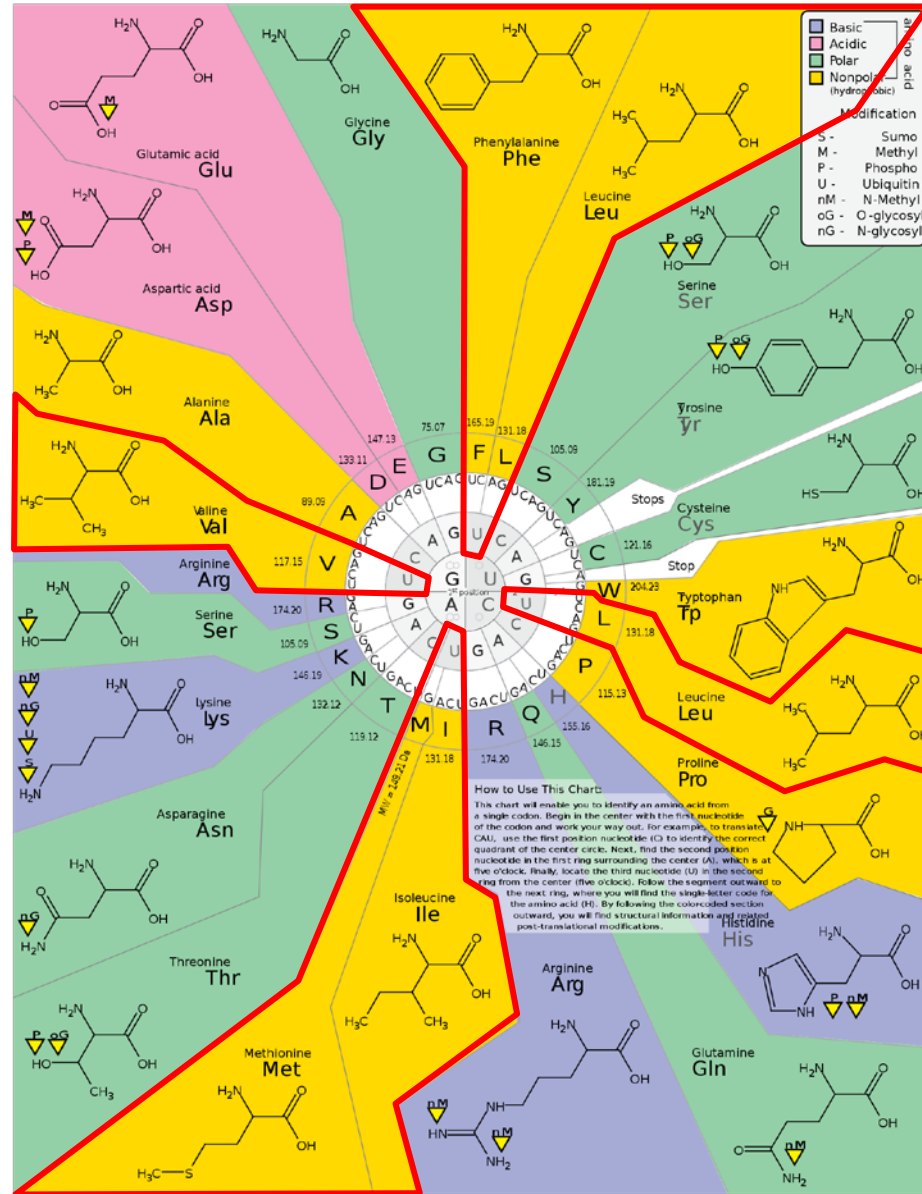


# Genetic Code degeneracy NUN

A practical consequence of redundancy is that some errors in the genetic code only cause a silent mutation.

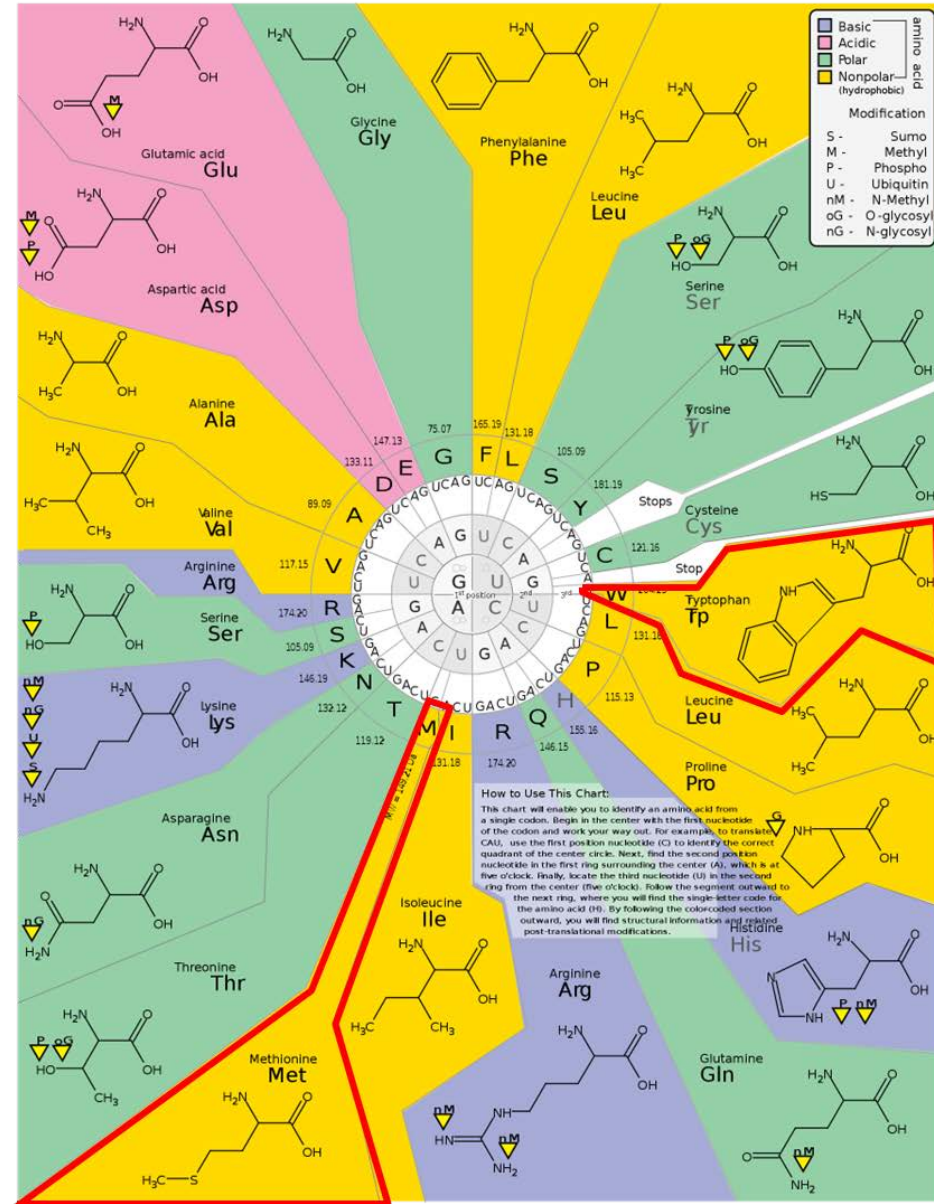
An error that would not affect the protein because the hydrophobicity is maintained by equivalent substitution of amino acids.

For example, a codon of **NUN** (where N = any nucleotide) tends to code for **hydrophobic amino acids**.



# Single codon amino acids

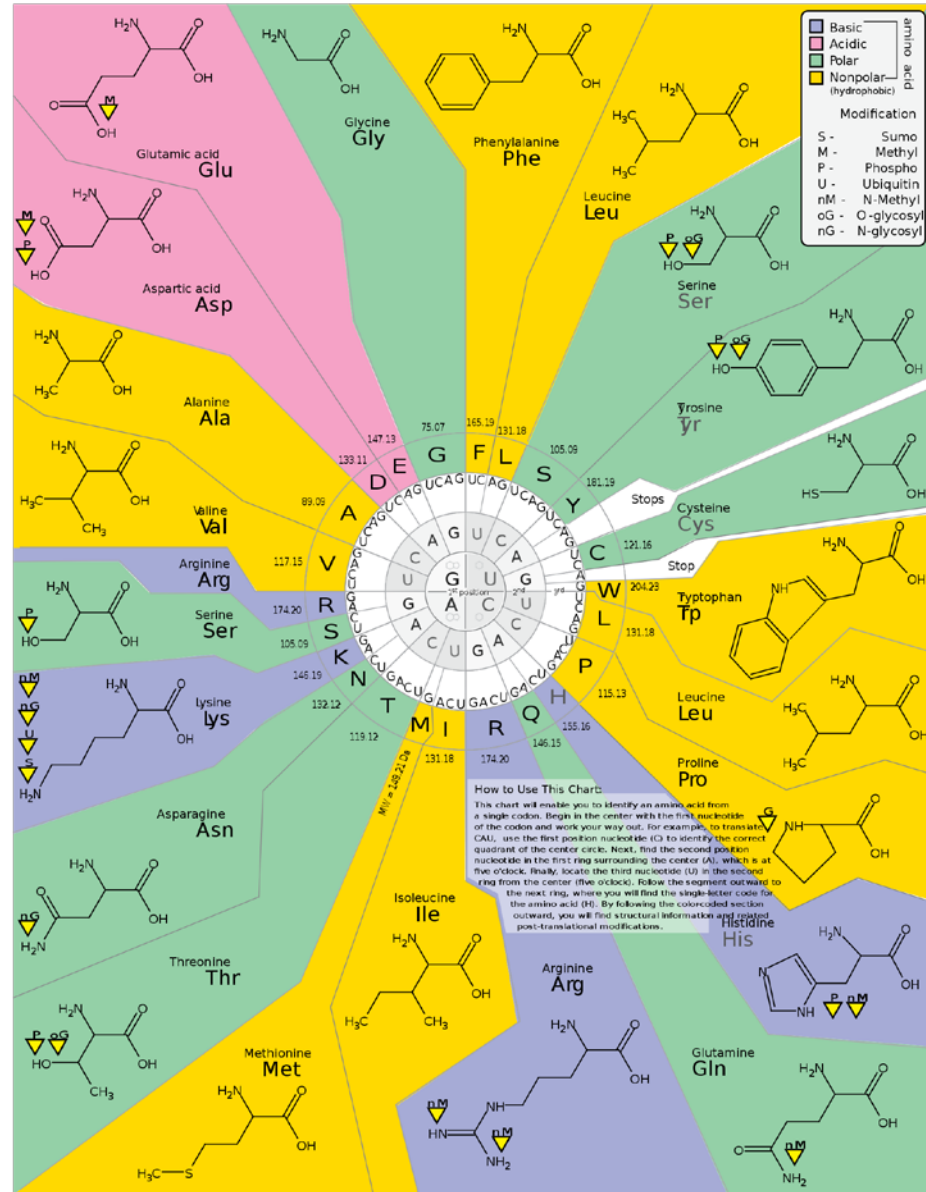
- Only two amino acids are specified by a single codon.
- **Methionine**, specified by the codon AUG, which also specifies the start of translation.
- **Tryptophan**, specified by the codon UGG.



# Non-synonymous conserved physicochemical properties

Also, most second and third site non-synonymous substitutions lead to “conserved” chemical properties.

UCn - Ser }  
 UAn - Tyr } Polar  
 UGU/C - Cys }

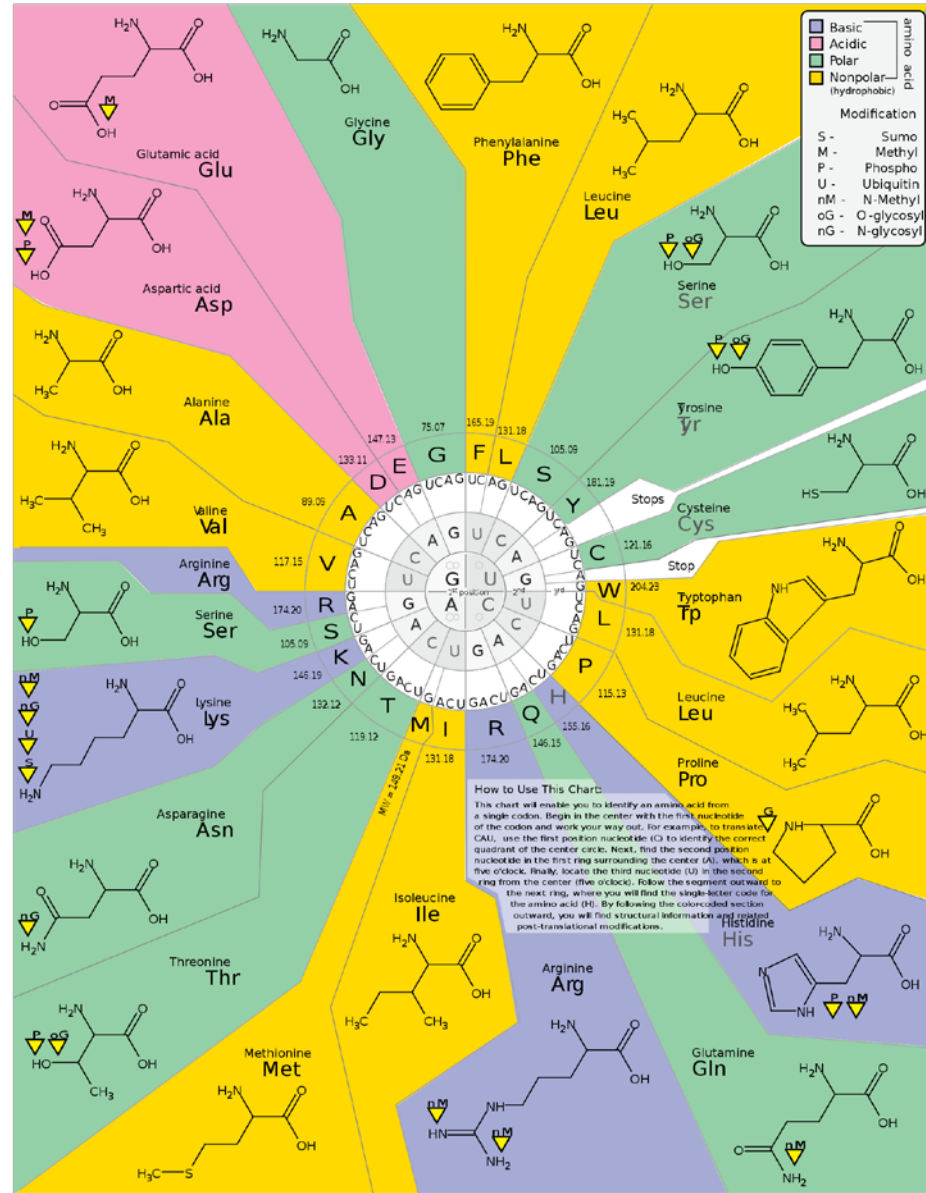


# Standard genetic code

There are  $4^3 = 64$  different codon combinations possible with a triplet codon of three nucleotides.

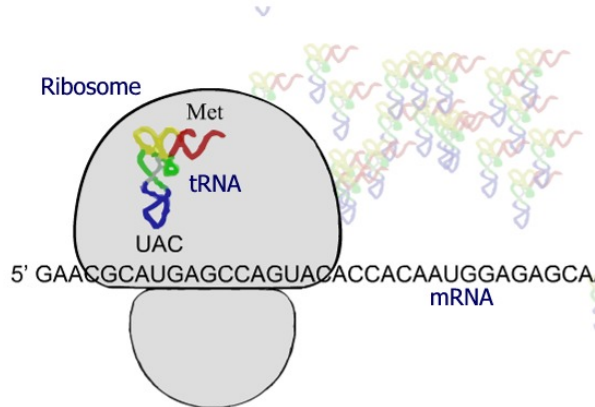
$$4 \times 4 \times 4 = 64$$

All 64 codons of the canonical genetic code are assigned to either amino acids or stop signals during translation.



# Start codon and the initiator tRNA

- Translation starts with a **chain initiation codon** (start codon AUG).
- Unlike stop codons, the **start codon** alone is **not sufficient** to begin the process.
- Nearby **sequences and initiation factors** are also required to start translation.
- The **most common start codon is AUG**, which codes for methionine, so **most amino acid chains start with methionine**.





# Stop codons and termination factors

There are **three known stop codons** which have been given names:

**UAG** is amber

**UGA** is opal (also called umber)

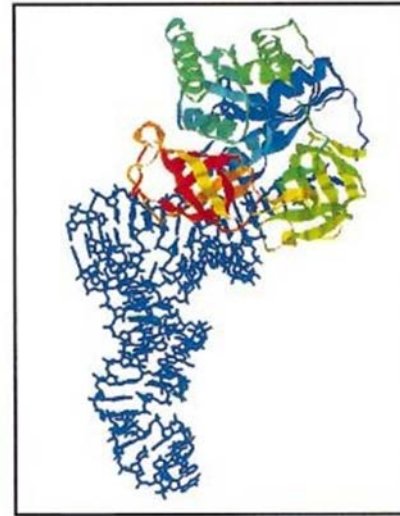
**UAA** is ochre

Also called **termination codons**.

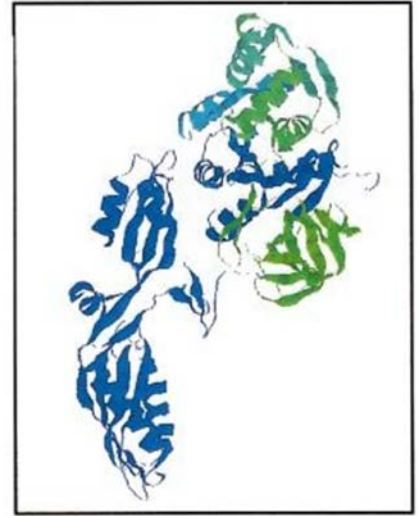
**Do not have tRNA anti-codons**, instead they **bind release factors**.

They **signal release of the nascent polypeptide** from the ribosome due to binding of release factors.

Suffice termination.



tRNA molecule



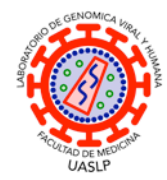
Termination factor



## Laboratorio de Genómica Viral y Humana

Instalaciones de Alta Contención Biológica Nivel de Bioseguridad 3 (BSL-3) CDC-certificadas

Facultad de Medicina UASLP  
San Luis Potosí, México



# Content copyright and license

---

The Viral and Human Genomics Laboratory is committed to promoting the human rights of free access to knowledge and to receiving the benefits of scientific progress and its applications by providing universal access to all the resources and publications it produces. This is in agreement with article 15 of the United Nations International Covenant on Economic, Social and Cultural Rights published on April 30, 2020.

All information included in this document is in the public domain, was compiled by the licensor and is distributed under a Creative Commons Attribution 4.0 International (CC BY 4.0 DEED) license which grants the licensee (you) the right to copy, remix, transform, develop and redistribute the material in any medium or format for any purpose, including commercial purposes provided that:

- 1) Corresponding credit is given to the licensor as “CA García-Sepúlveda, Laboratory of Viral and Human Genomics UASLP”,
- 2) Any changes to the original document are indicated and,
- 3) In no way suggest that the licensor endorses the derivative work.

All rights reserved © 2024 CA García-Sepúlveda, Laboratory of Viral and Human Genomics UASLP

(Last updated: August 23, © 2024.)