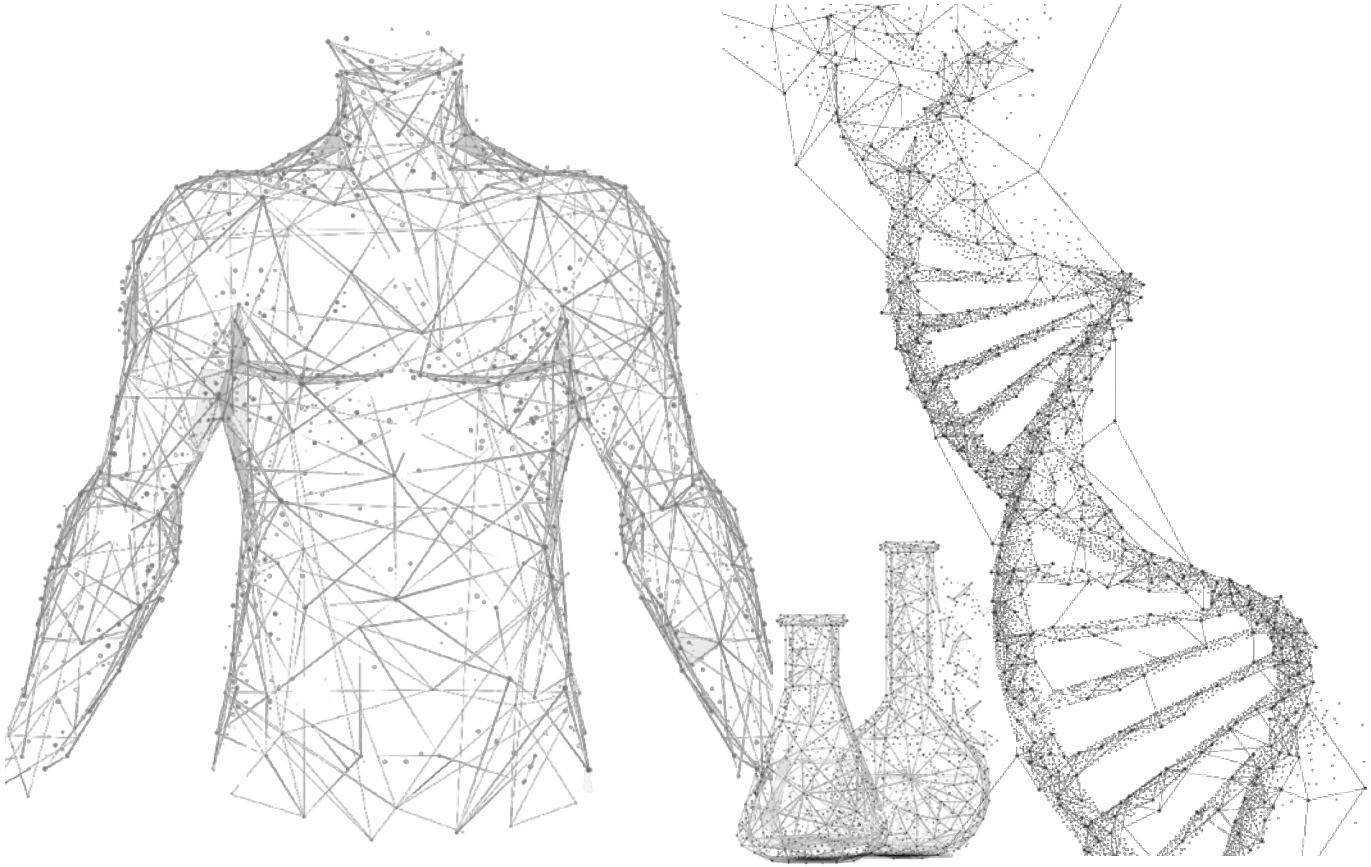


The Genetic Code

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

Dr. Christian A. García-Sepúlveda

Viral & Human Genomics BSL-3 Laboratory

Last updated February 10, 2025 v3

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

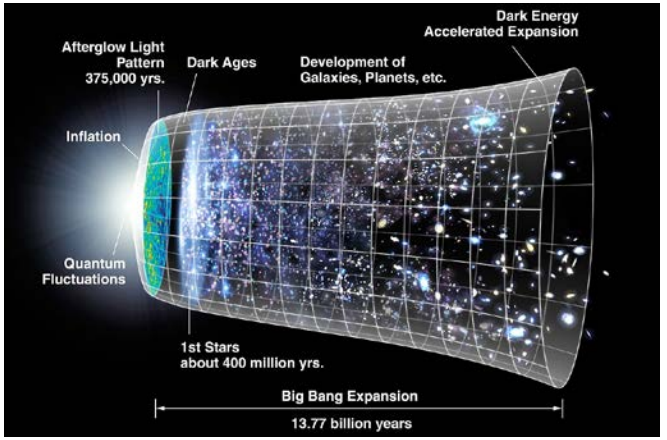
Georgiy Antonovich Gamow (1904 – 1968).

Ruso-American theoretical physicist and cosmologist.

Early advocate and developer of Big Bang theory.

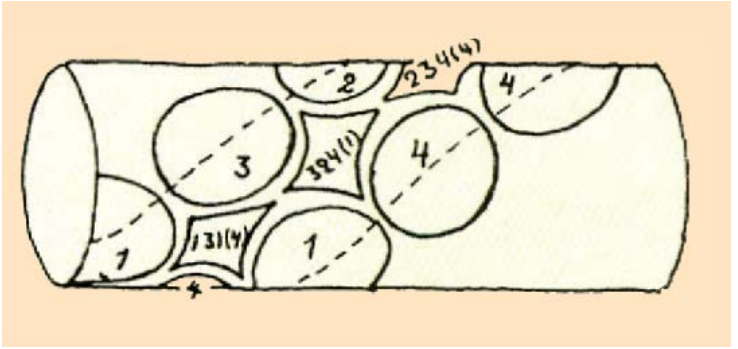
First mathematical model of the atomic nucleus.

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).

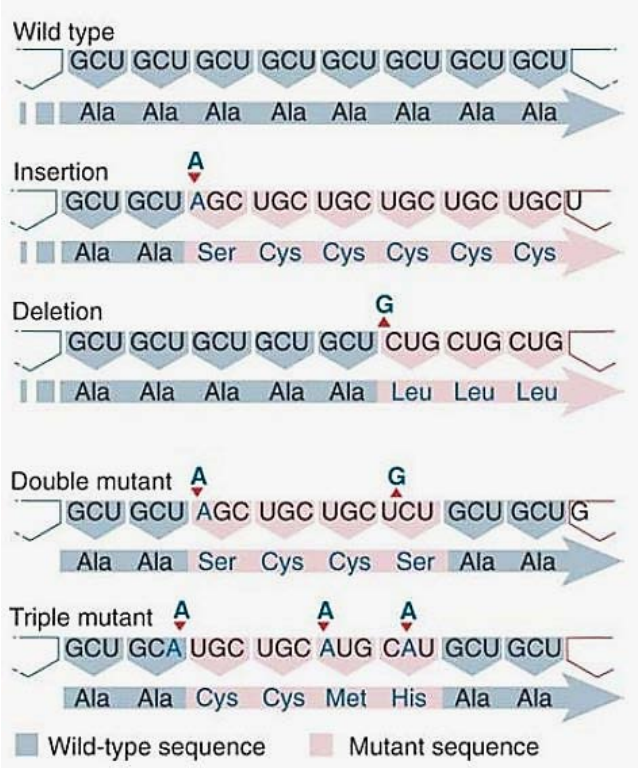
The experiment elucidated the nature of gene expression and frameshift mutations.

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).



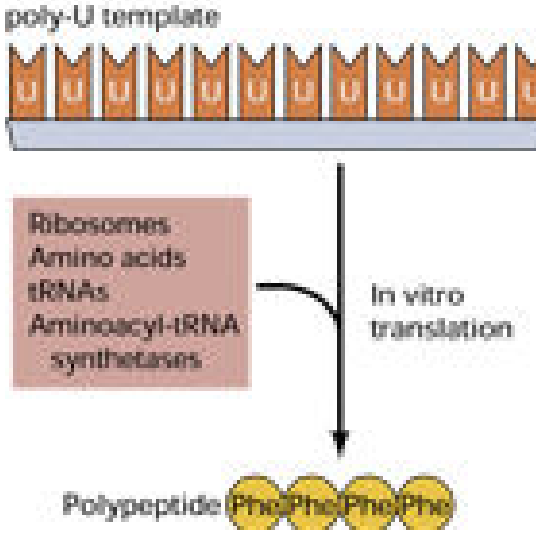
ORF	DNA sequence	Protein sequence
Normal =	ATG TGC TGA CTG ATC GGT	= M C S L I G
-1 =	TGT GCT GAC TGA TCG GT	= C A D S S
-2 =	GTG CTG ACT GAT CGG T	= V L T D R
-3 =	TGC TGA CTG ATC GGT	= C S L I G

Nirenberg-Mathei experiment (descyphering codons)

First descypherement 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.

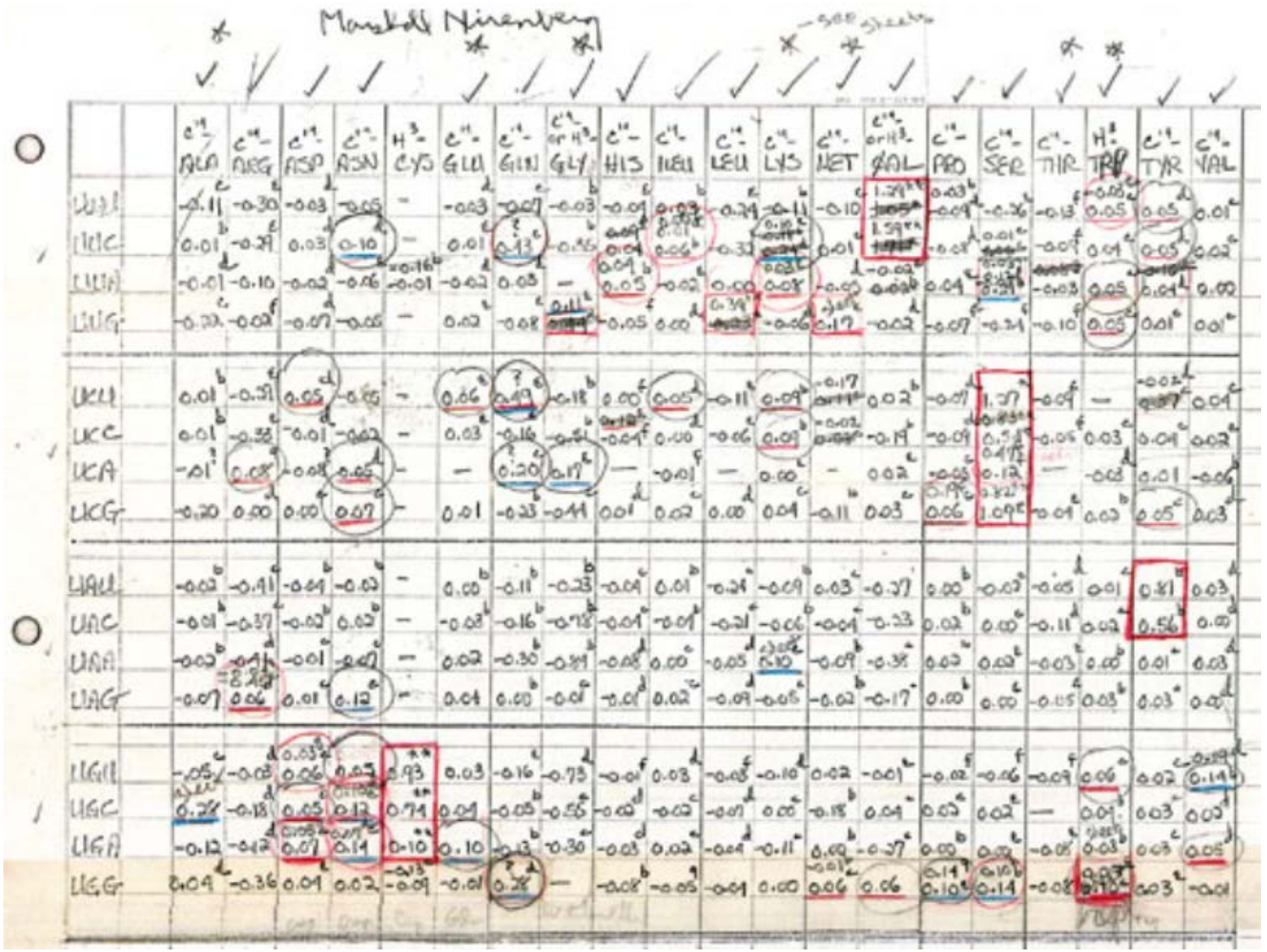
Deduced that UUU = phenylalanine.



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

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

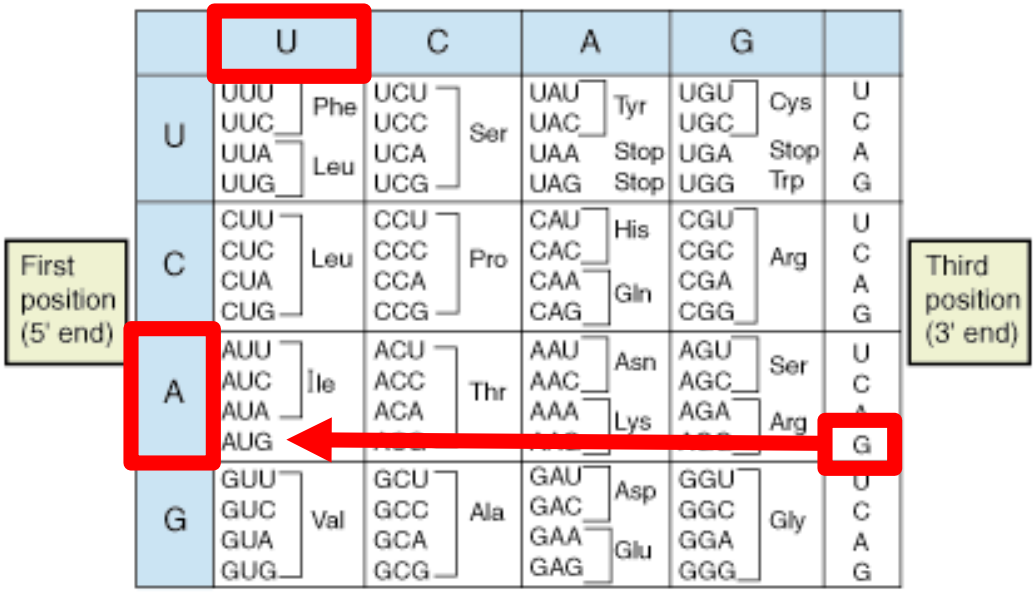
Nirenberg-Mathei experiment (descyphering codons)



Canonical genetic code (three positions)

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

AUG



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

Amino acid names:

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

Alternative genetic codes

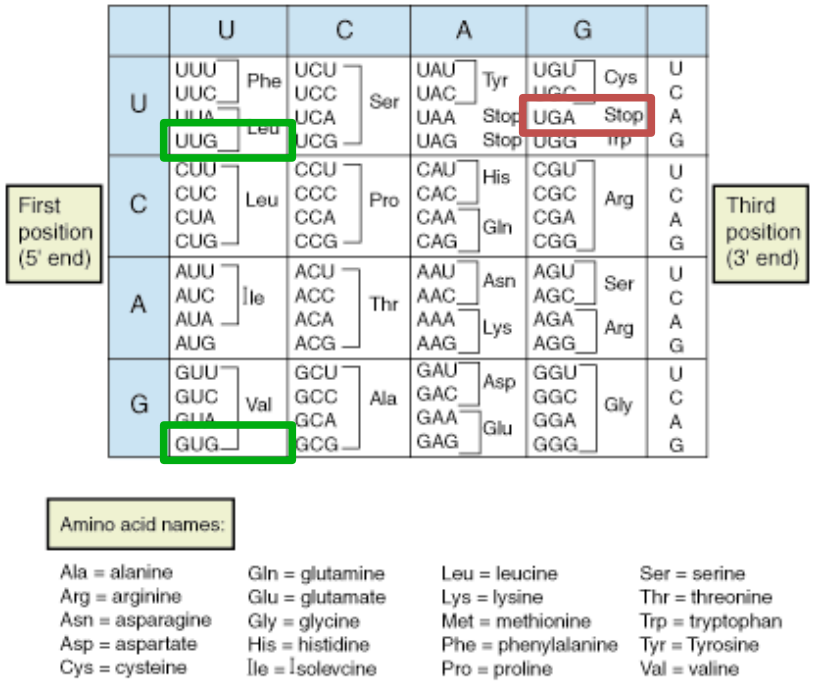
Slight variations on the canonical code had been predicted

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.



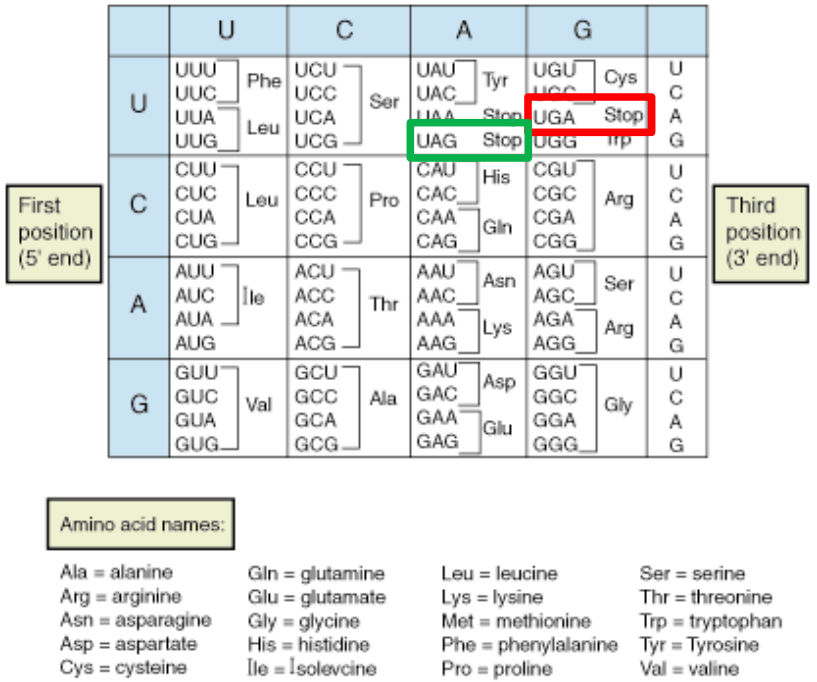
Alternative genetic codes

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

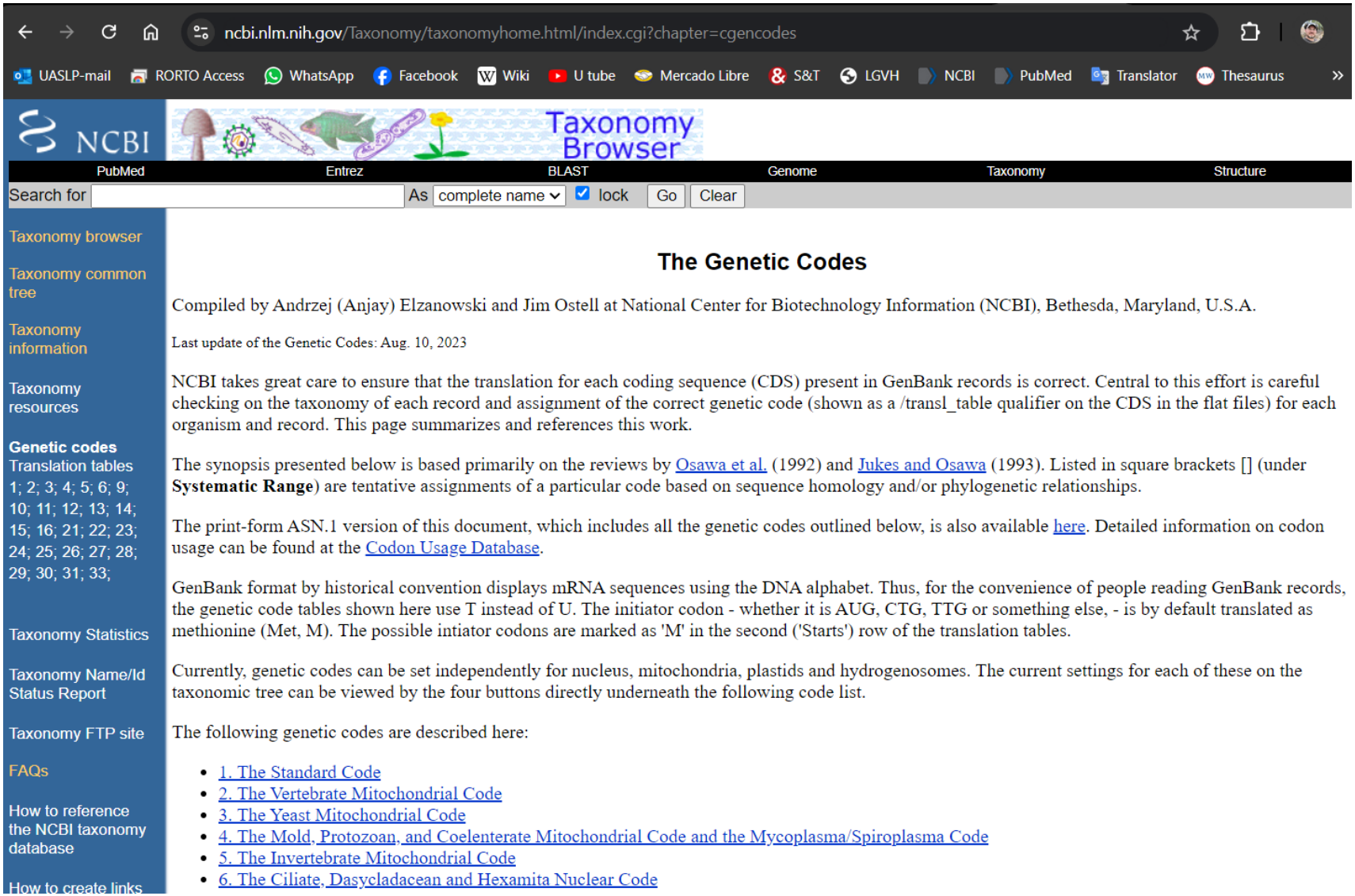
UGA = Selenocysteine (21st)

UAG = Pyrrolysine (22nd)

Depend on associated signal sequences in the mRNA.



Alternative genetic codes



ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html/index.cgi?chapter=cgencodes

UASLP-mail | RORTO Access | WhatsApp | Facebook | Wiki | U tube | Mercado Libre | S&T | LGVH | NCBI | PubMed | Translator | Thesaurus

NCBI Taxonomy Browser

PubMed | Entrez | BLAST | Genome | Taxonomy | Structure

Search for [ ] As [complete name] lock [Go] [Clear]

- Taxonomy browser
- Taxonomy common tree
- Taxonomy information
- Taxonomy resources
- **Genetic codes** Translation tables 1; 2; 3; 4; 5; 6; 9; 10; 11; 12; 13; 14; 15; 16; 21; 22; 23; 24; 25; 26; 27; 28; 29; 30; 31; 33;
- Taxonomy Statistics
- Taxonomy Name/Id Status Report
- Taxonomy FTP site
- FAQs
- How to reference the NCBI taxonomy database
- How to create links

The Genetic Codes

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 intiator codons are marked as 'M' in the second ('Starts') row of the translation tables.

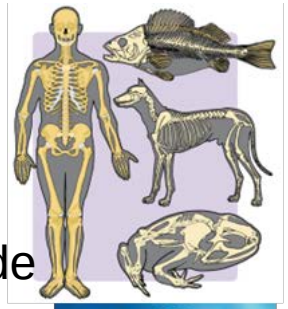
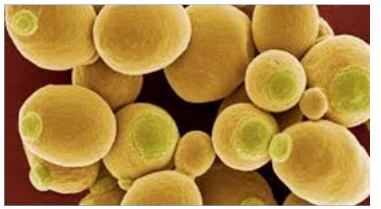
Currently, genetic codes can be set independently for nucleus, mitochondria, plastids and hydrogenosomes. The current settings for each of these on the taxonomic tree can be viewed by the four buttons directly underneath the following code list.

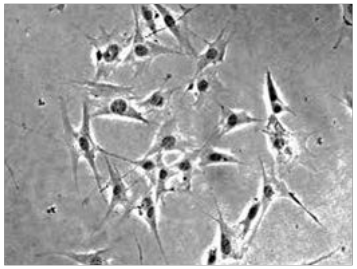
The following genetic codes are described here:

- 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

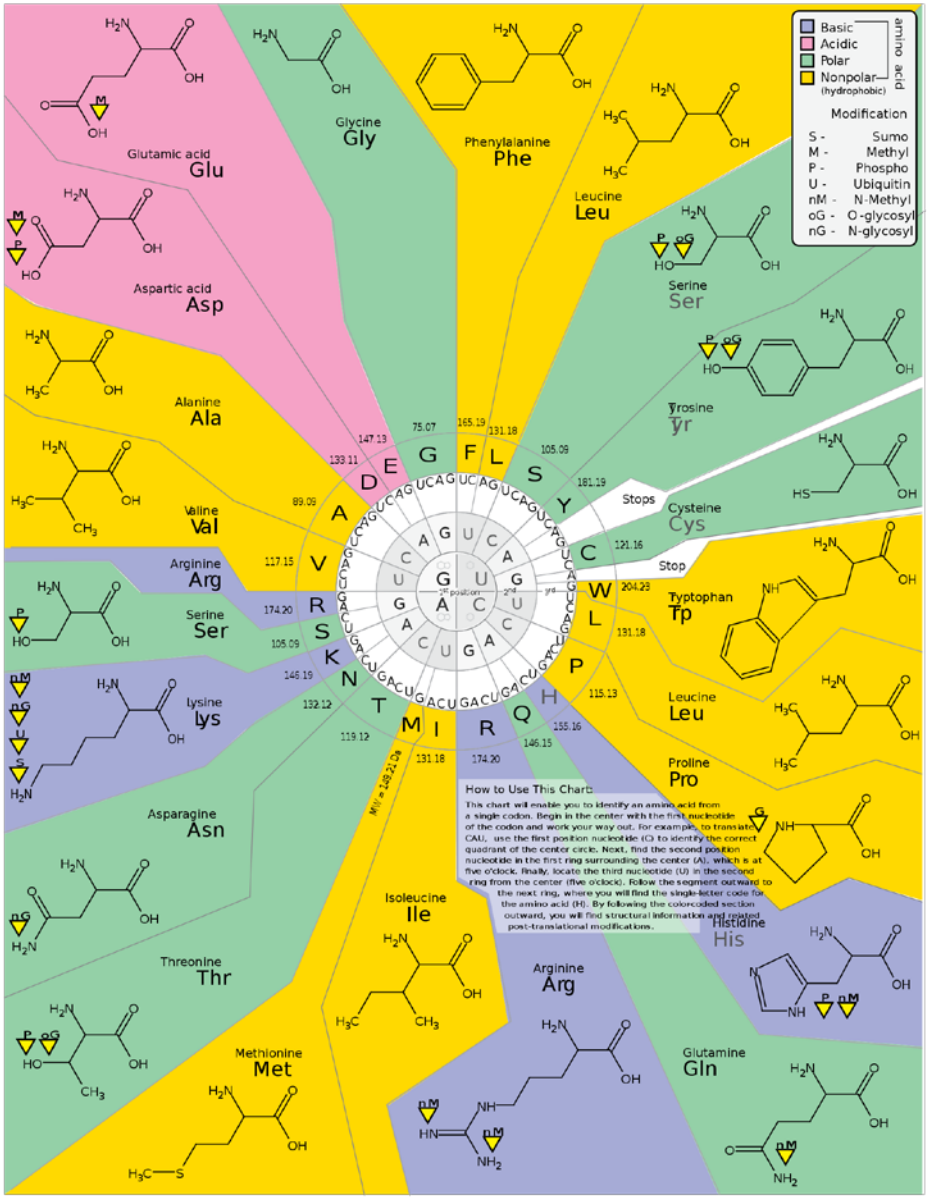
However, all known codes have strong similarities and the coding mechanism is the same for all organisms:

- 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

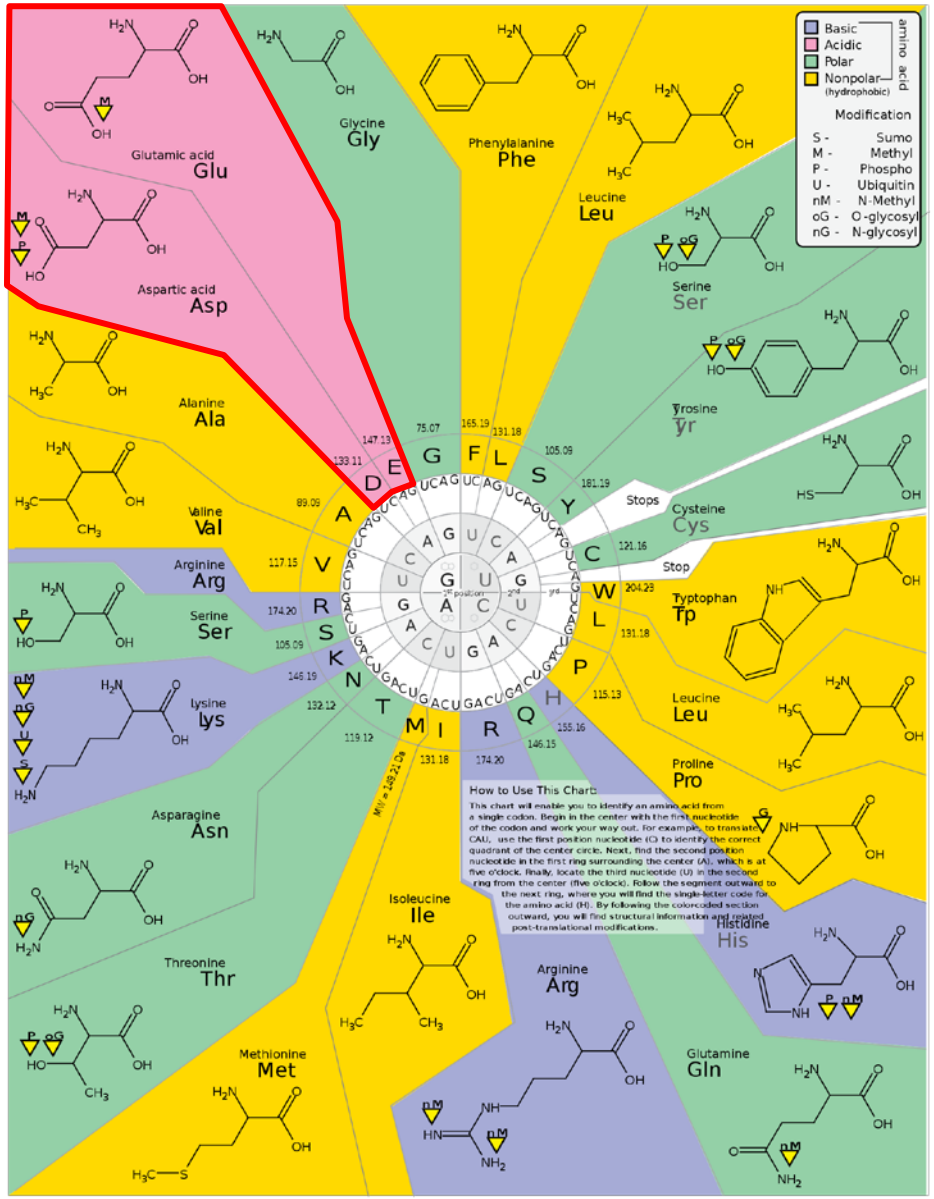


http://en.wikipedia.org/wiki/File:GeneticCode21-version-2.svg

Acidic amino acids

Acidic (Asp, Glu):

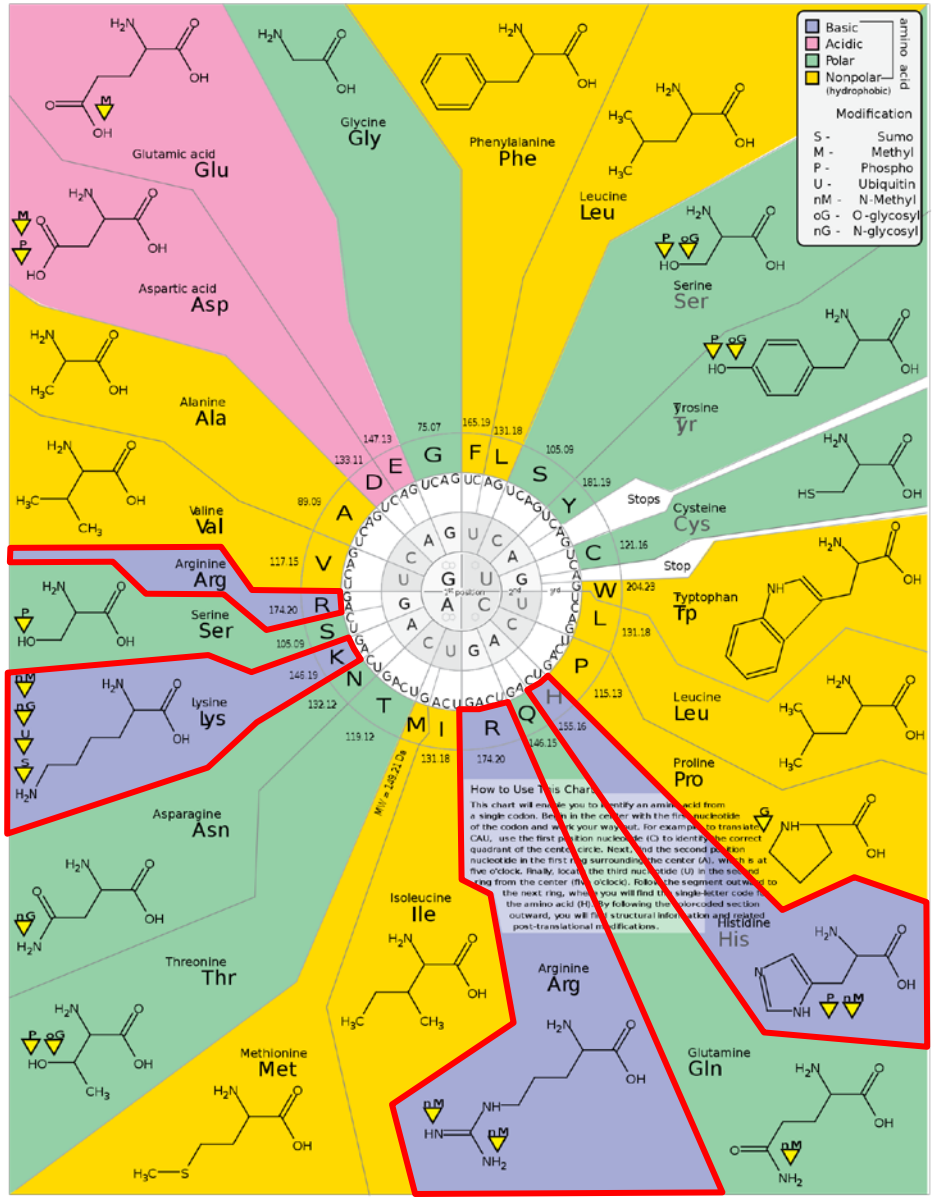
- Participate in enzyme active sites, facilitating catalysis through proton donation.
- Play a crucial role in maintaining protein structure by forming salt bridges with basic amino acids.
- within the lipid bilayer.



Basic amino acids

Basic Amino Acids (Lys, Arg, His):

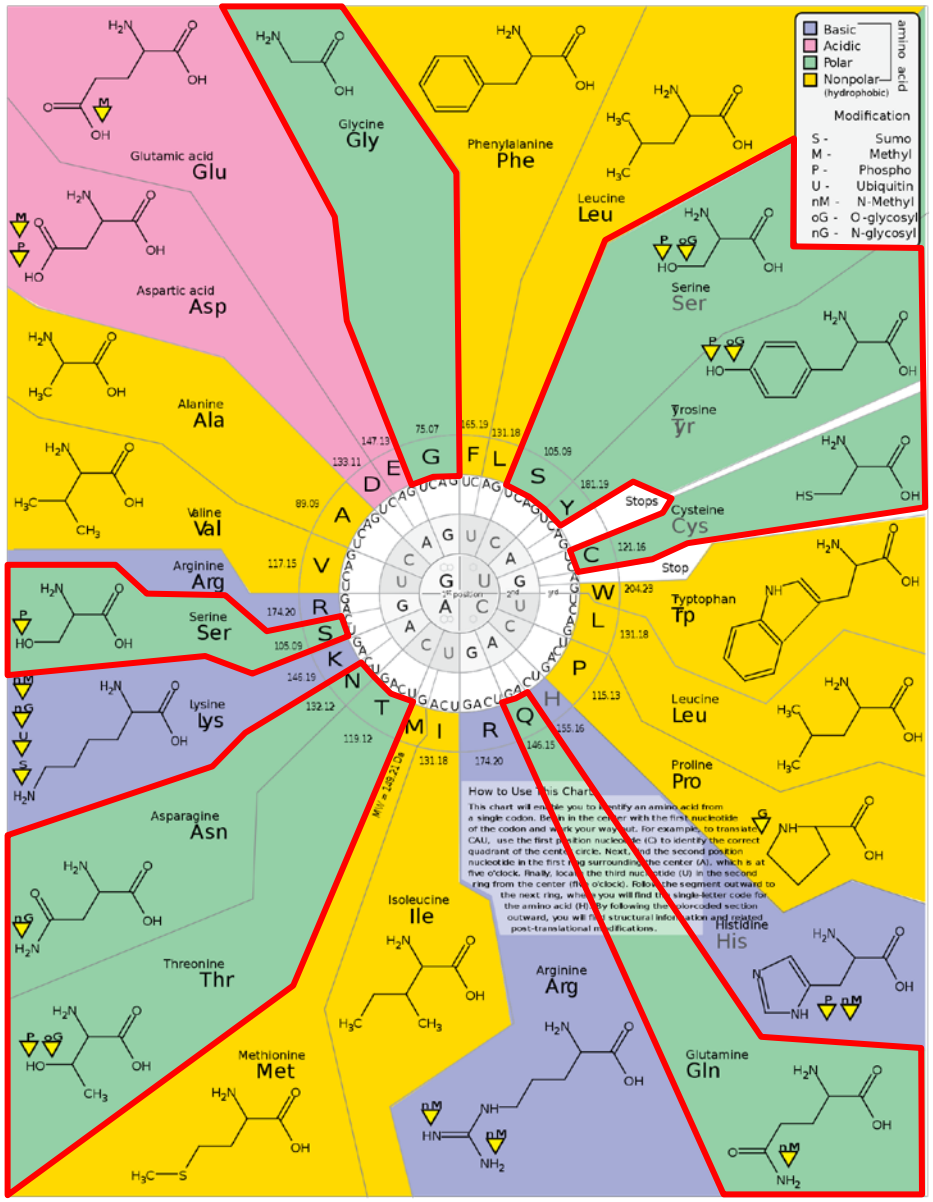
- Often found in DNA-binding proteins, interacting with the negatively charged phosphate backbone of DNA.
- Contribute to the formation of salt bridges, stabilizing protein structures.



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.

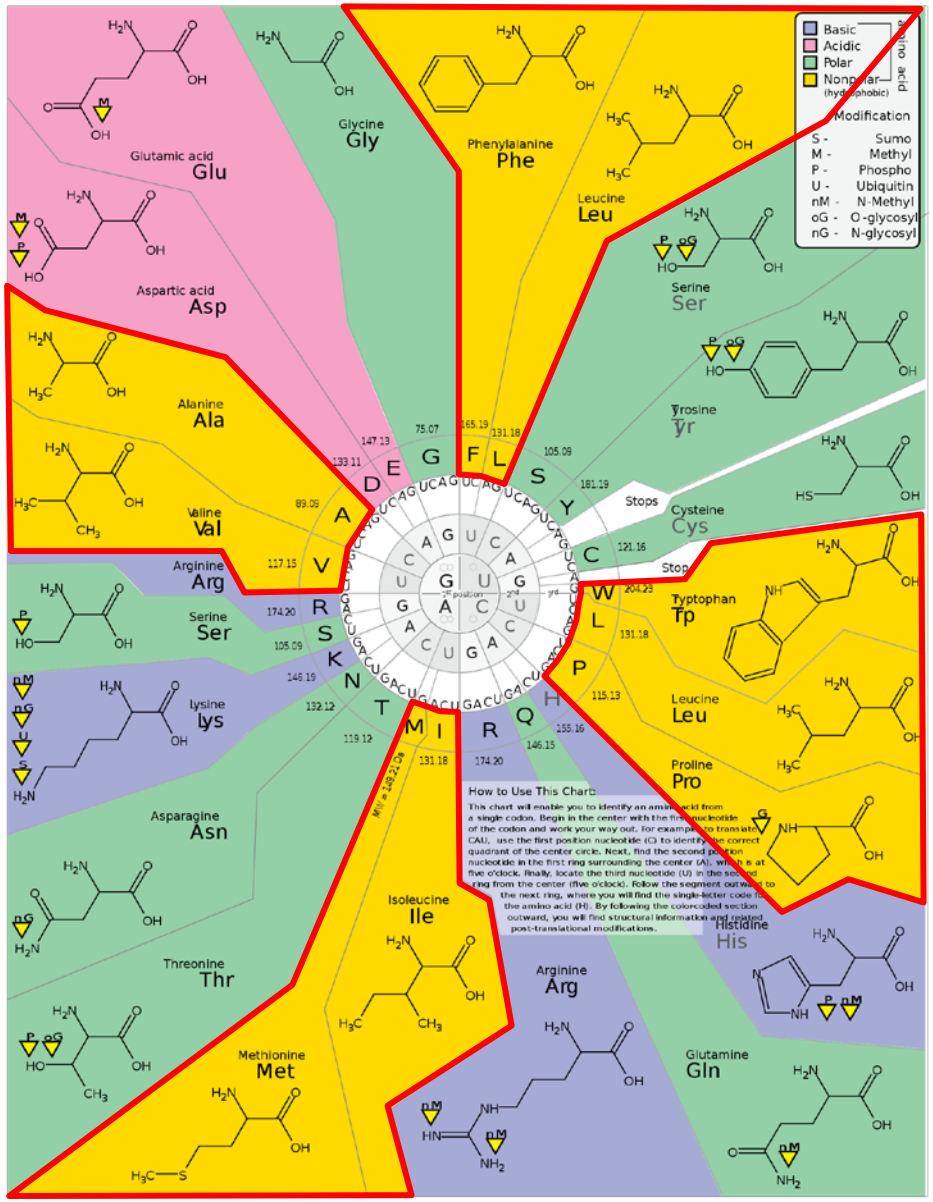


http://en.wikipedia.org/wiki/File:GeneticCode21-version-2.svg

Non-polar amino acids

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

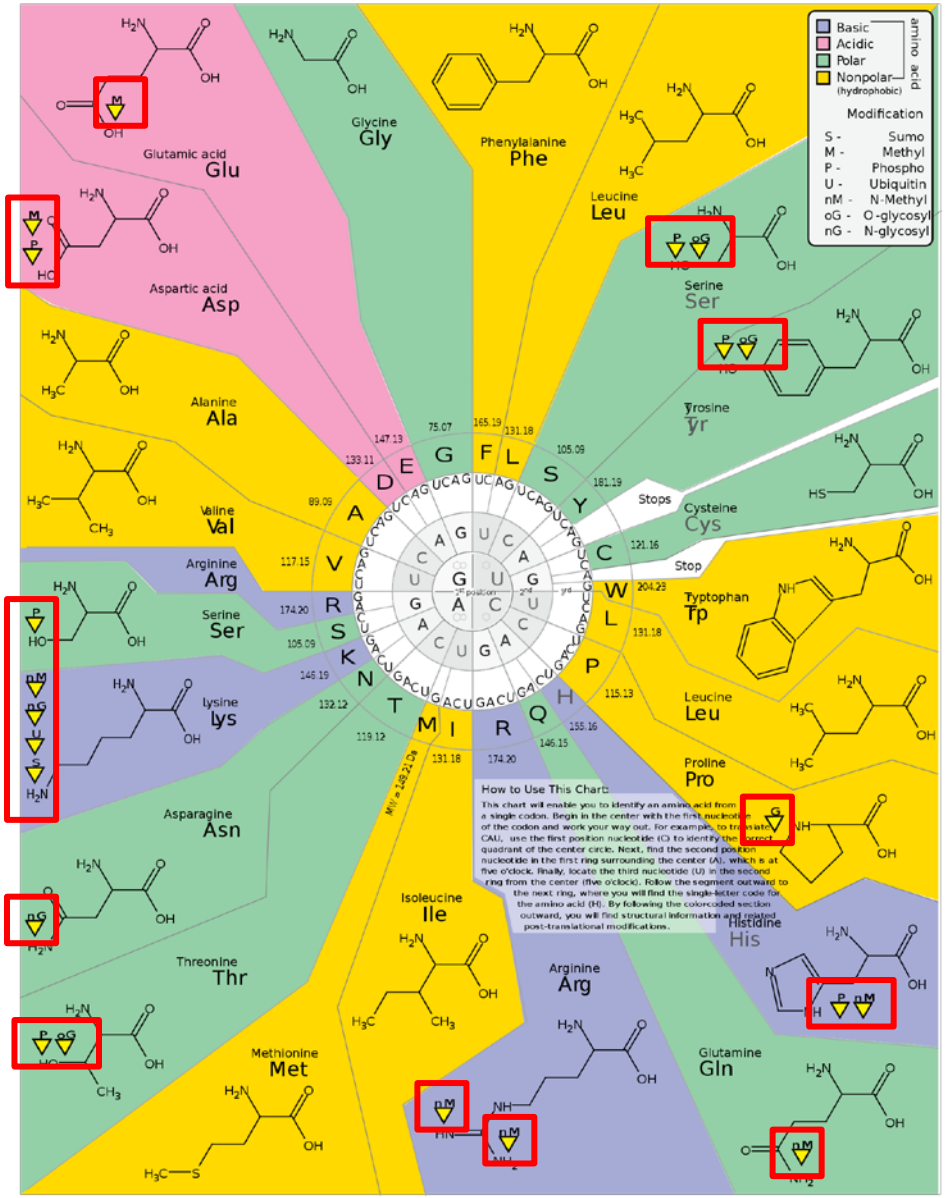
- Predominantly found in the interior of proteins, driving the folding process through hydrophobic interactions.
- Play a key role in forming membrane-spanning regions of transmembrane proteins, anchoring them within the lipid bilayer.



http://en.wikipedia.org/wiki/File:GeneticCode21-version-2.svg

Post-translational modifications

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



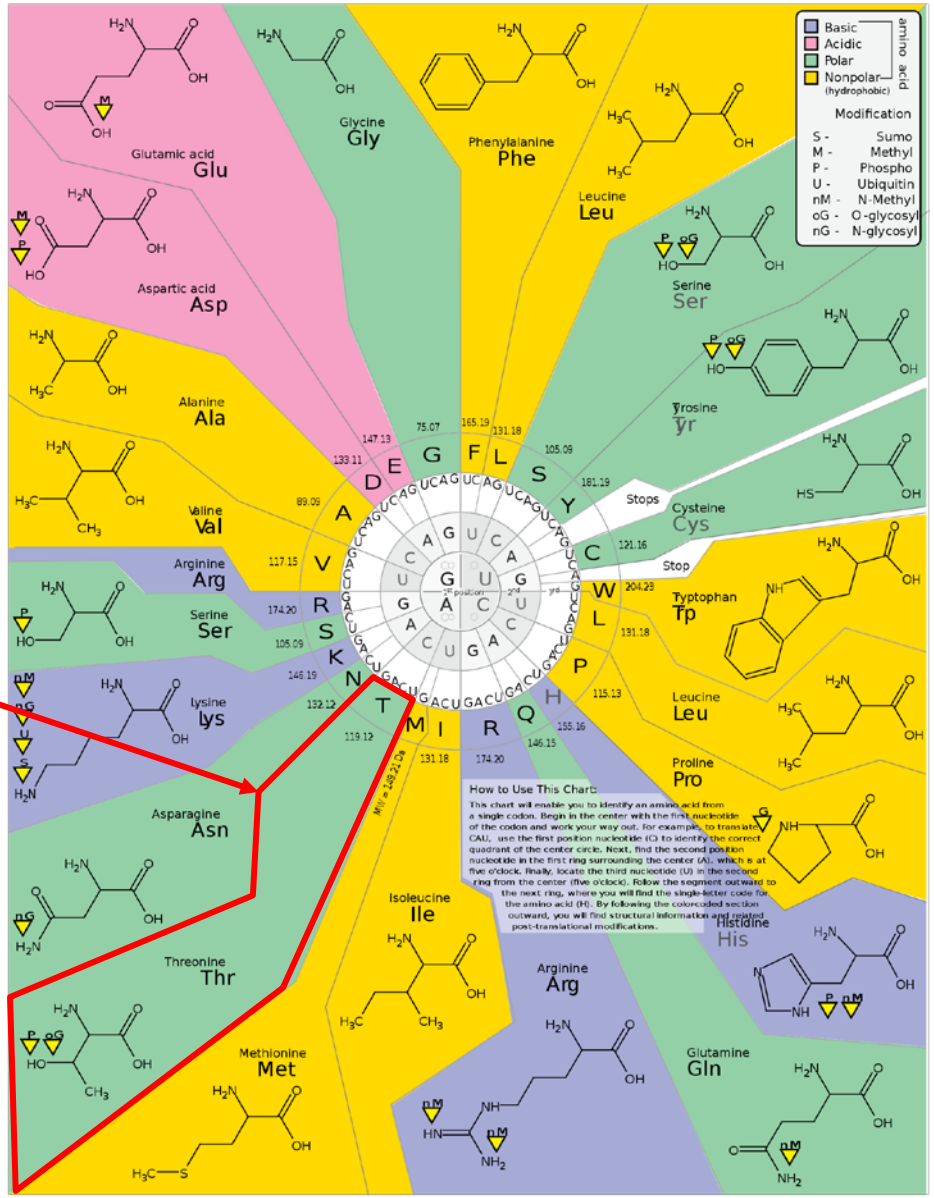
Genetic Code degeneracy/redundancy

Degeneracy or redundancy

More than one codon encoding for same aminoacid,

ACC
ACG
ACU
ACA
} Thr

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

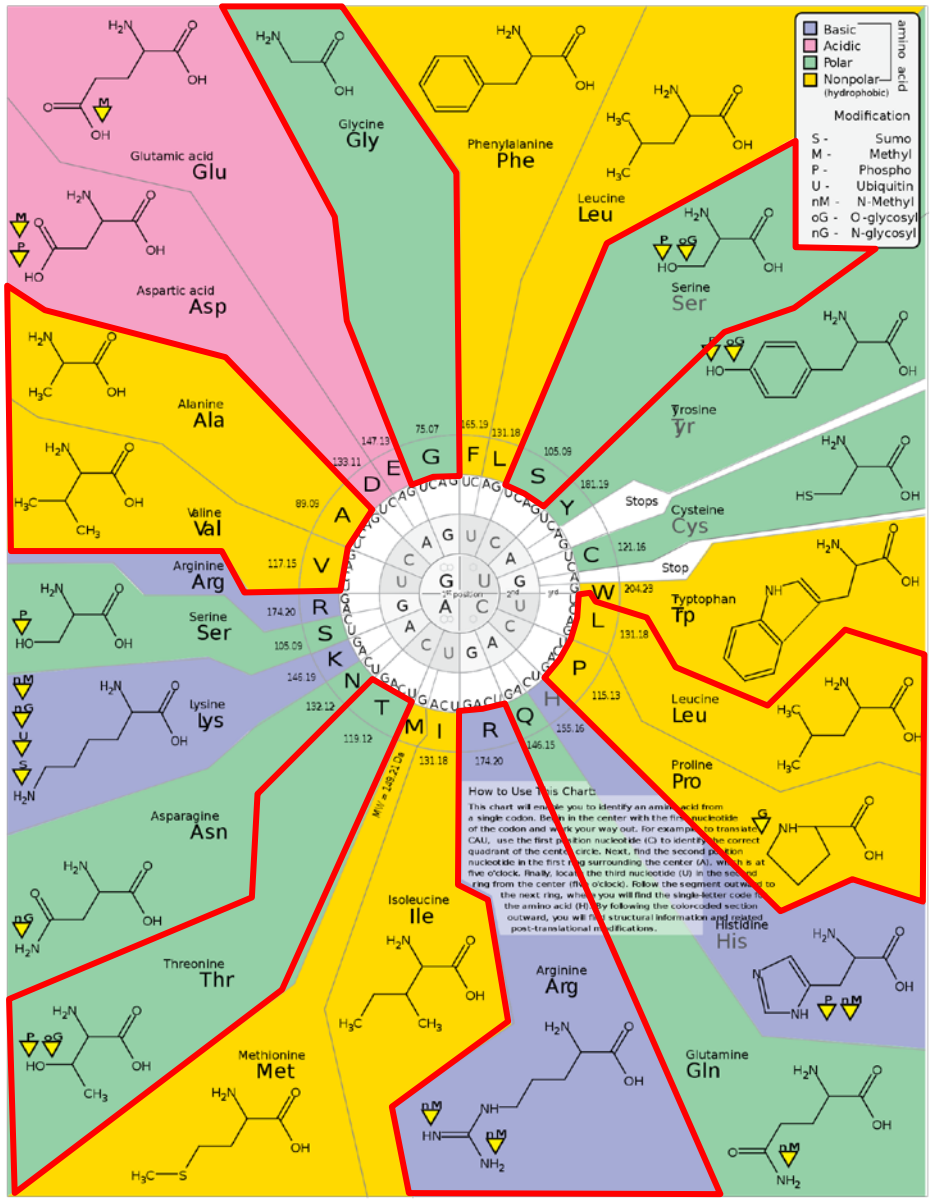


Genetic Code degeneracy/redundancy

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.

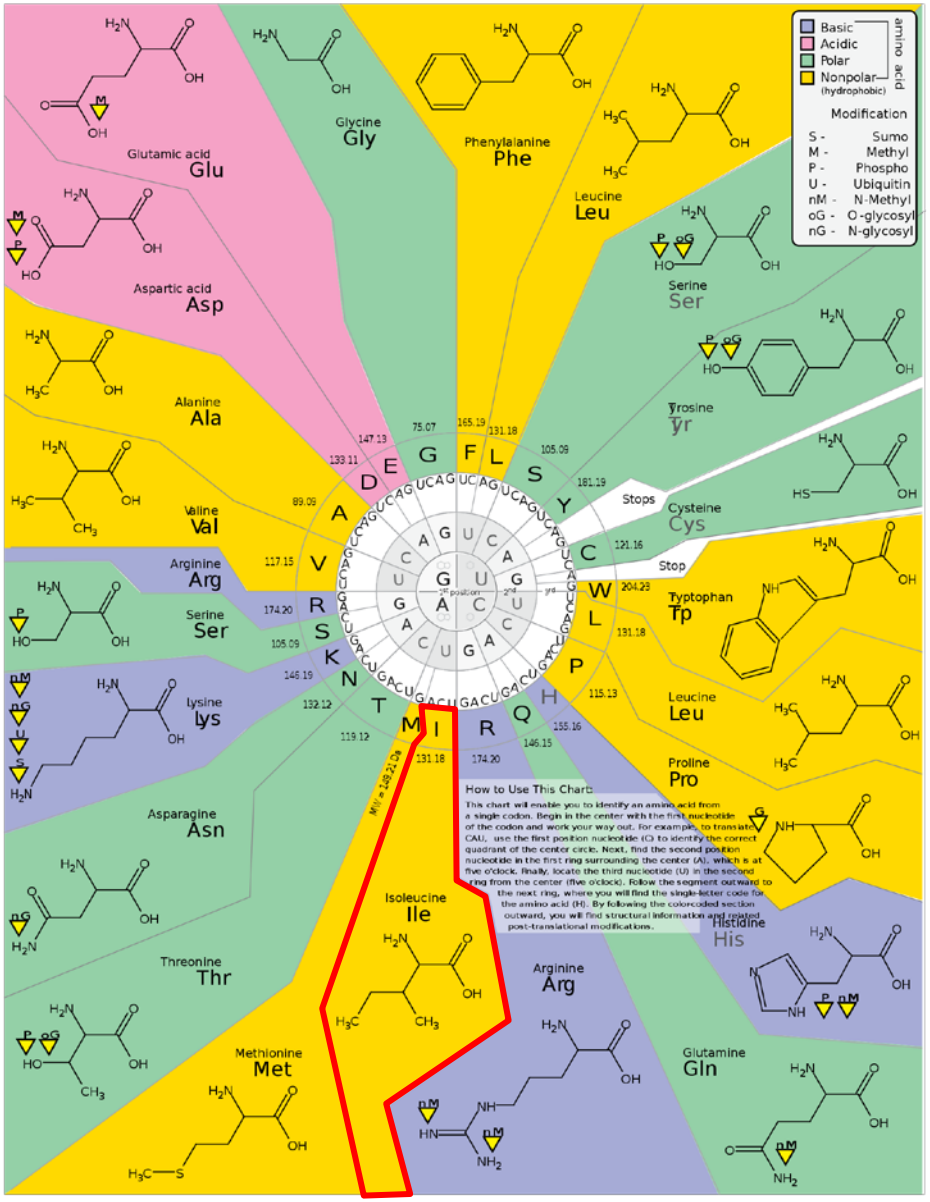


Genetic Code degeneracy/redundancy

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.



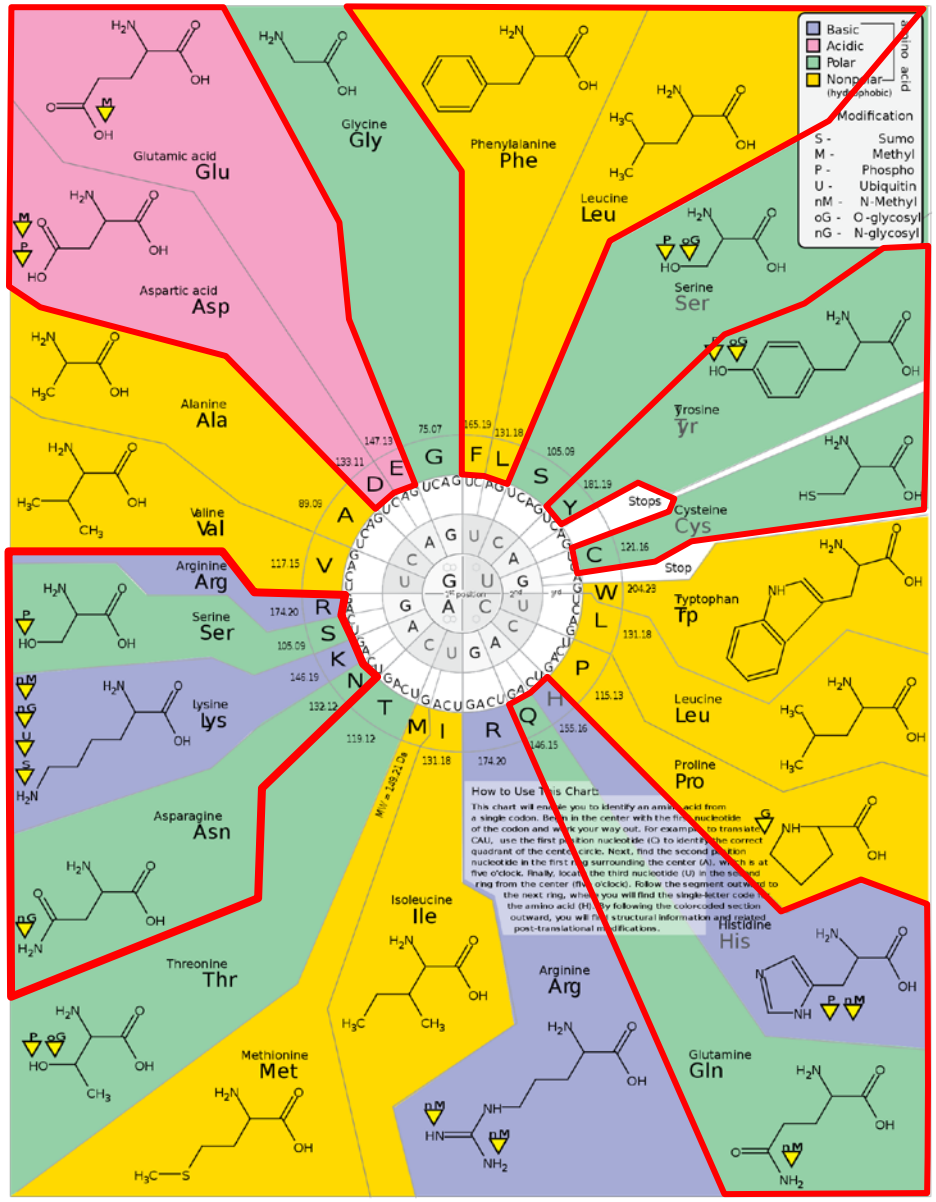
Genetic Code degeneracy/redundancy

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) = transicional substitutions.

Transversions are NS



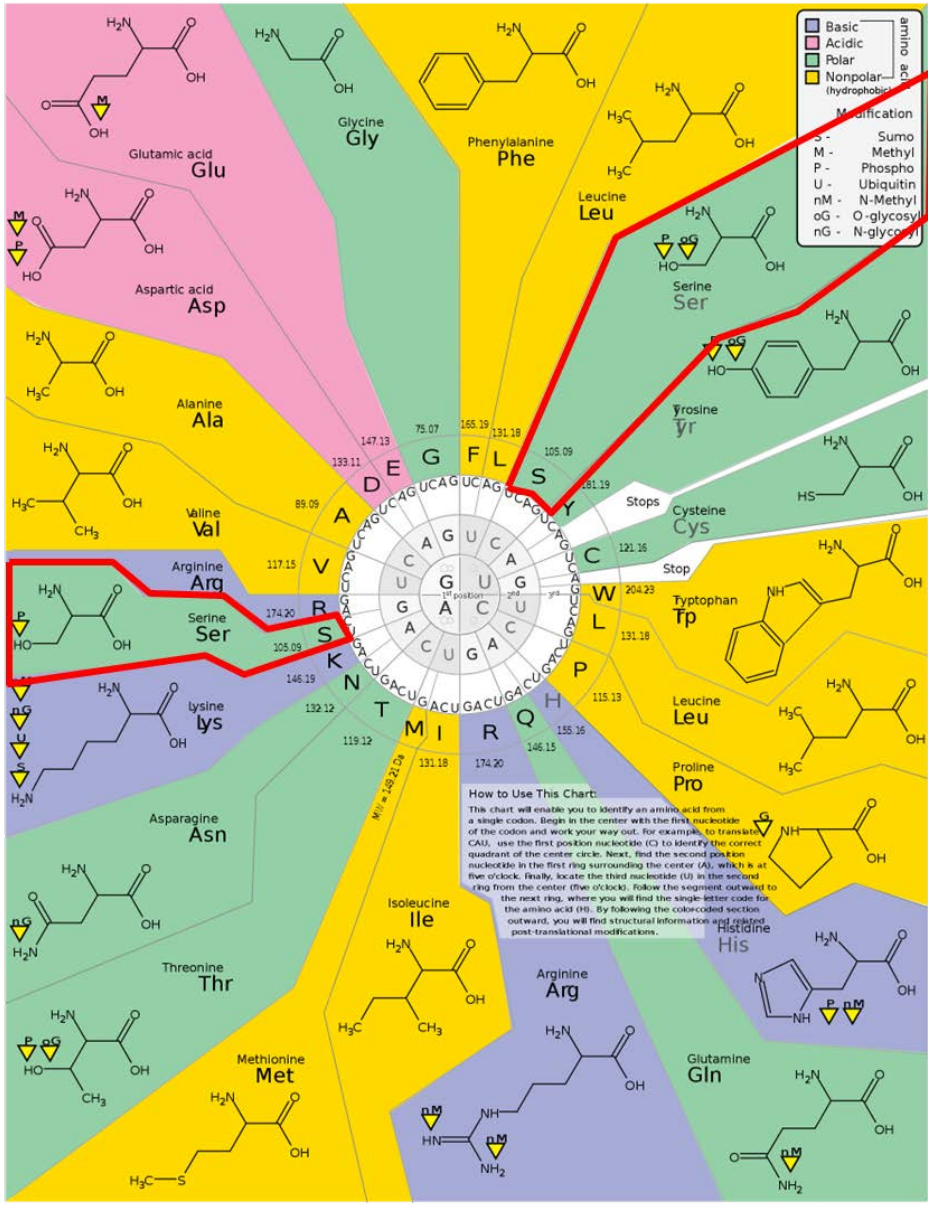
Genetic Code degeneracy/redundancy

- 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 3rd position).
  - UUA, UUG, CUU, CUC, CUA & CUG = Leucine (1st & 3rd).
  - UCA, UCG, UCC, UCU, AGU & AGC = Serine (1st, 2nd & 3rd).

Genetic Code degeneracy/redundancy

Only three amino acids are encoded by six different codons:

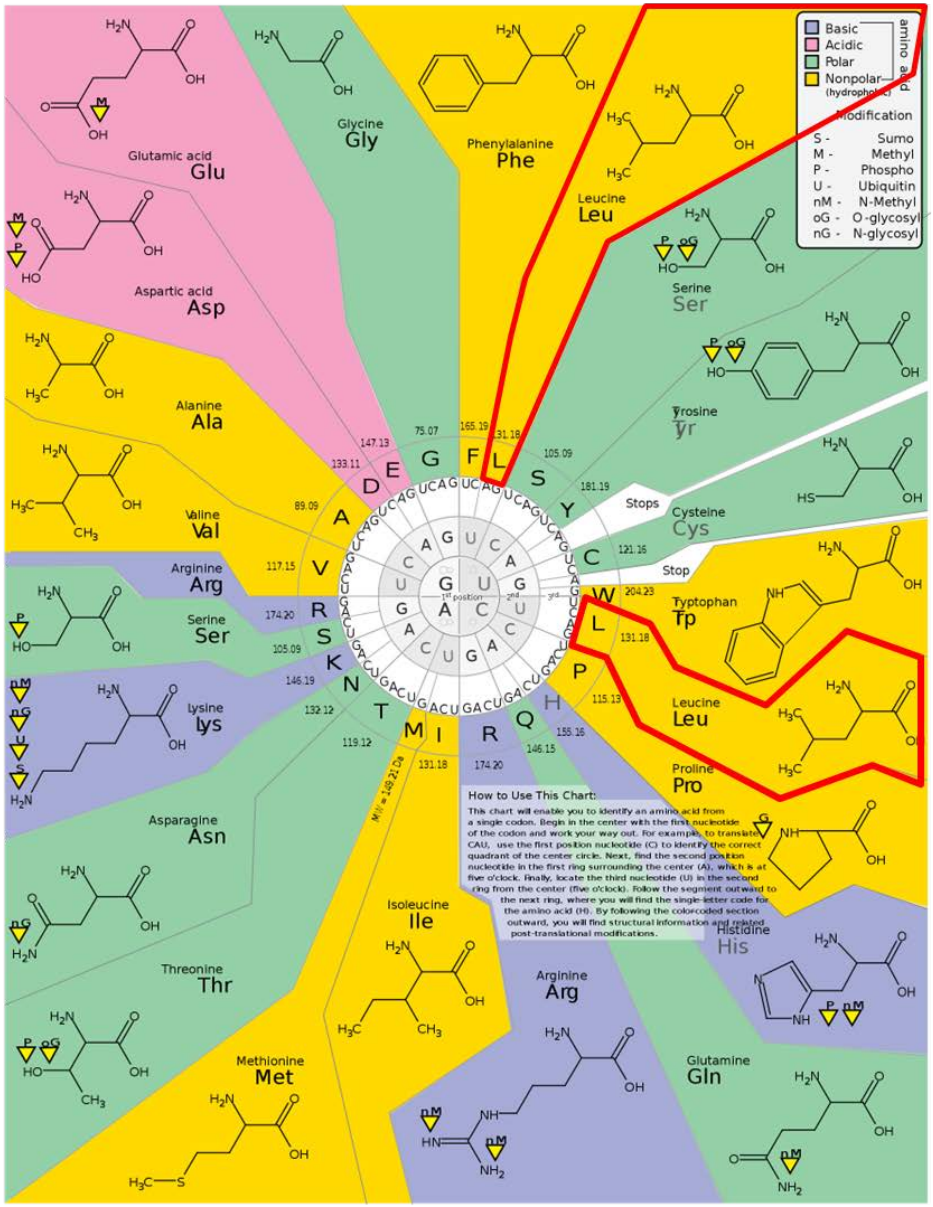
- Serine
- Leucine
- Arginine.



Genetic Code degeneracy/redundancy

Only three amino acids are encoded by six different codons:

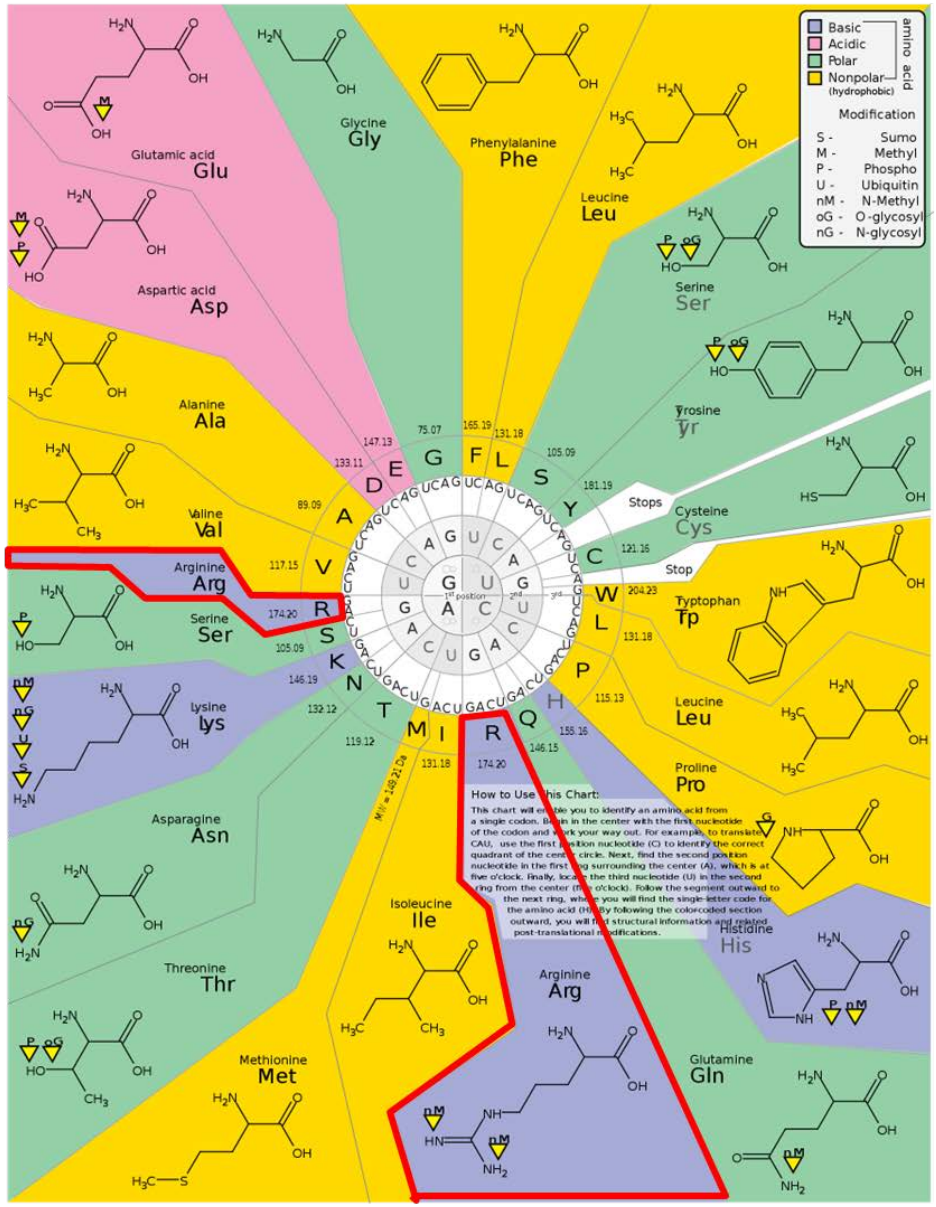
- Serine
- Leucine
- Arginine.



Genetic Code degeneracy/redundancy

Only three amino acids are encoded by six different codons:

- Serine
- Leucine
- Arginine.

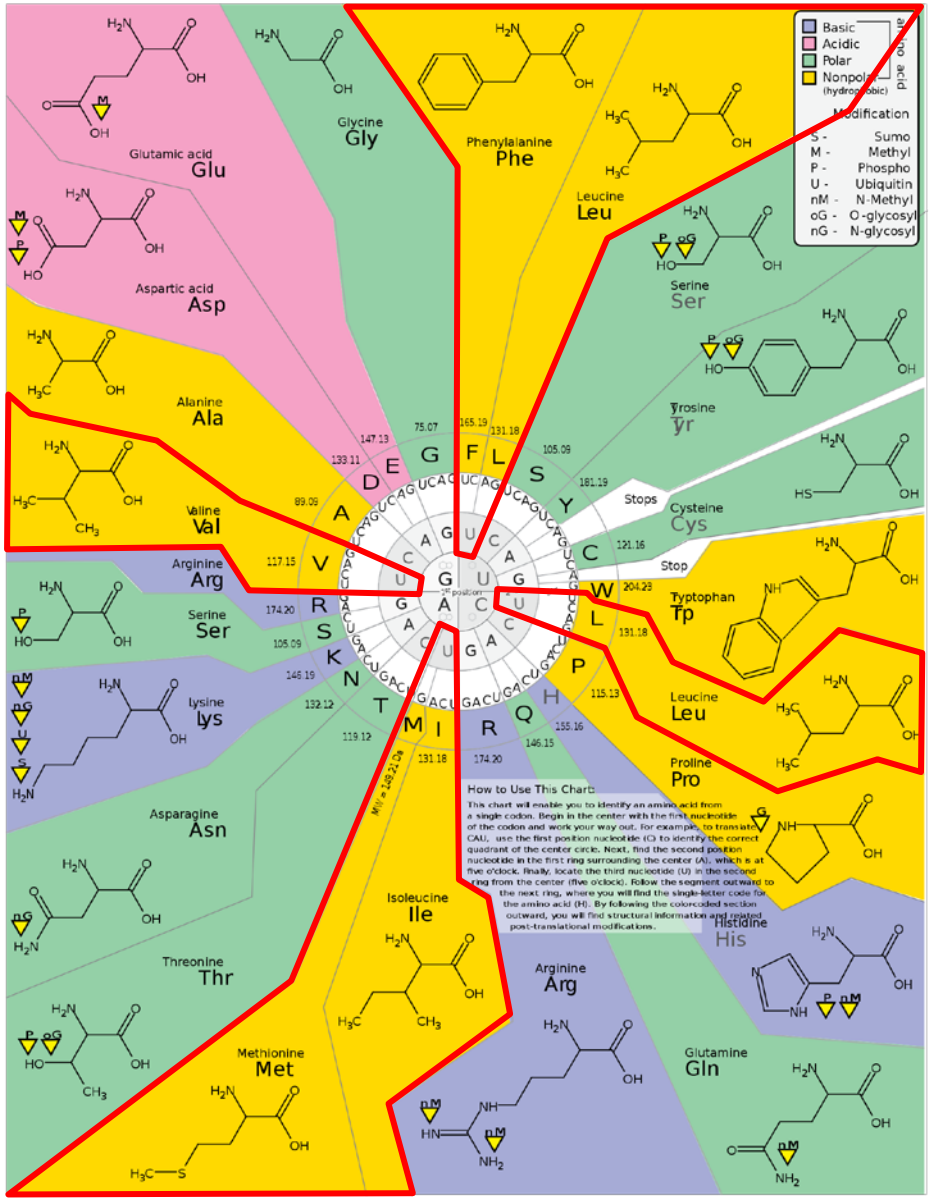


Genetic Code degeneracy NUN

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

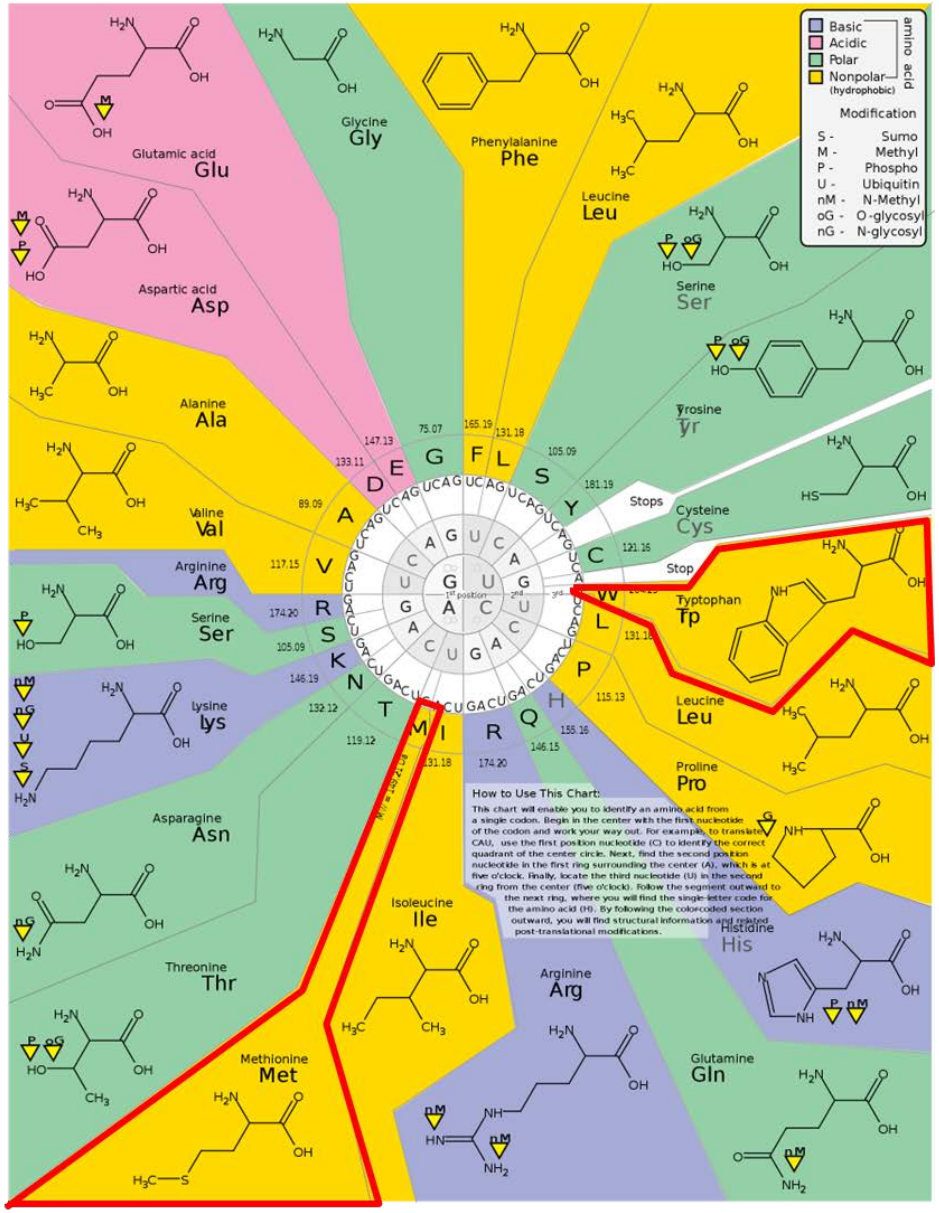
An error that would not affect the protein because the hydrophobicity is maintained by equivalent substitution of amino acids.

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

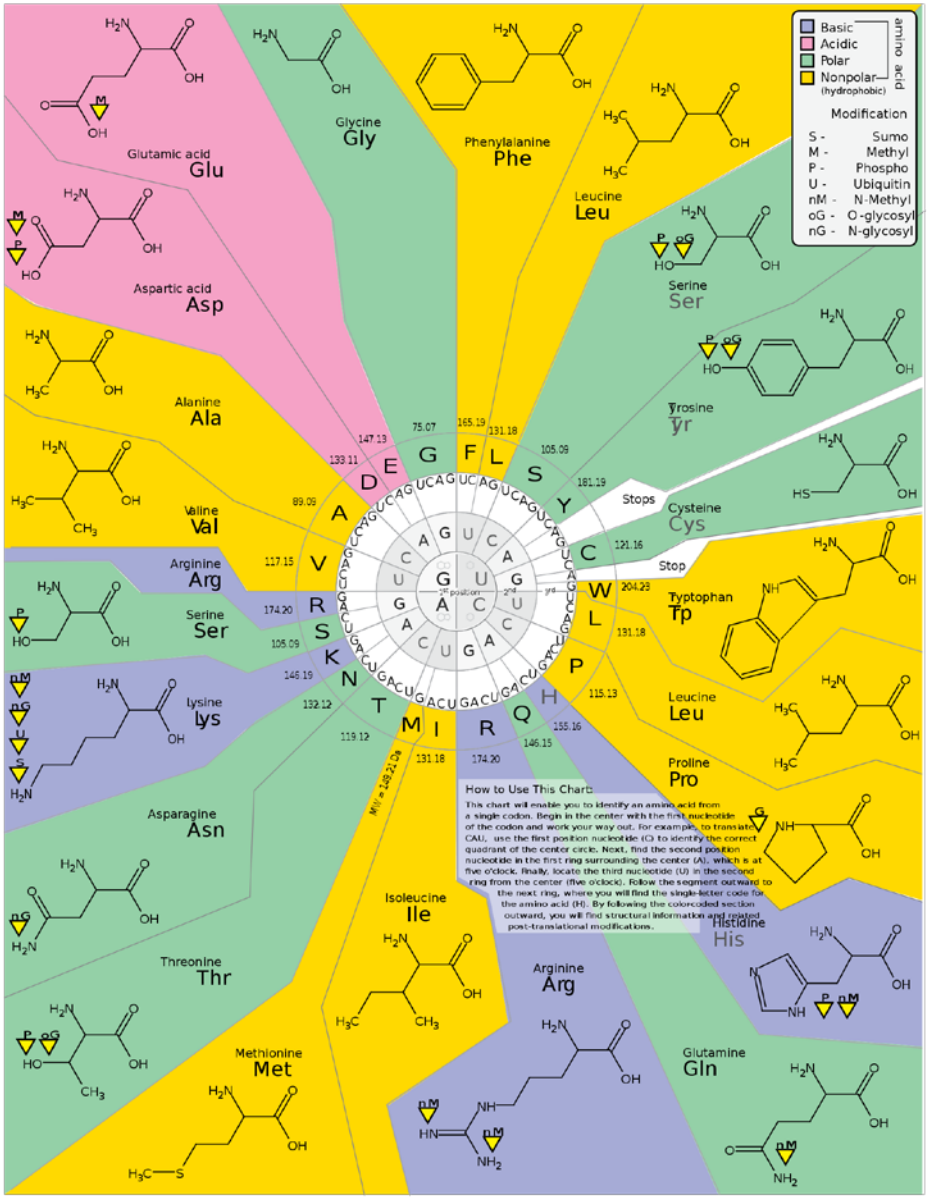
Also, most second and third site non-synonymous substitutions lead to “conserved” chemical properties.

UCn - Ser

UAn - Tyr

UGU/C - Cys

} Polar

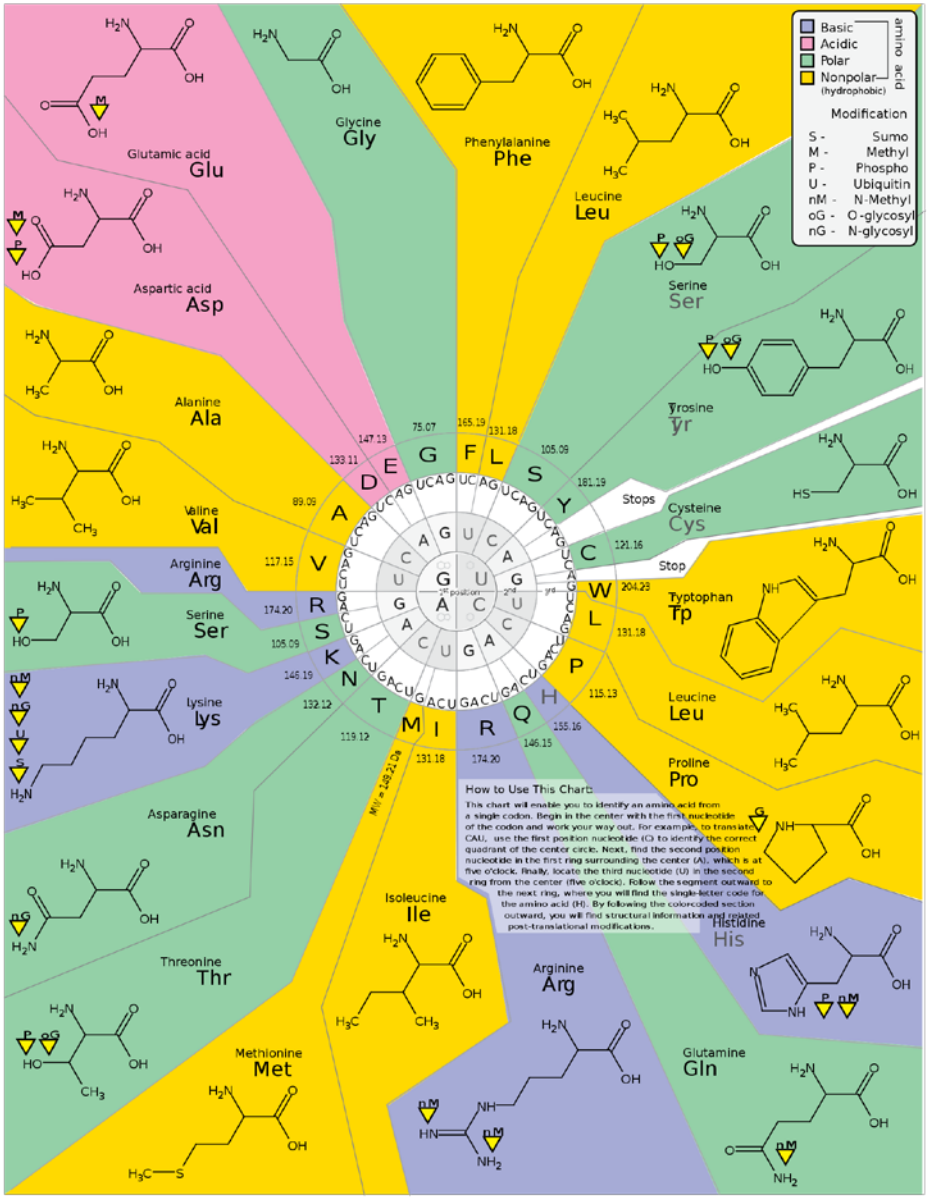


Standard genetic code

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

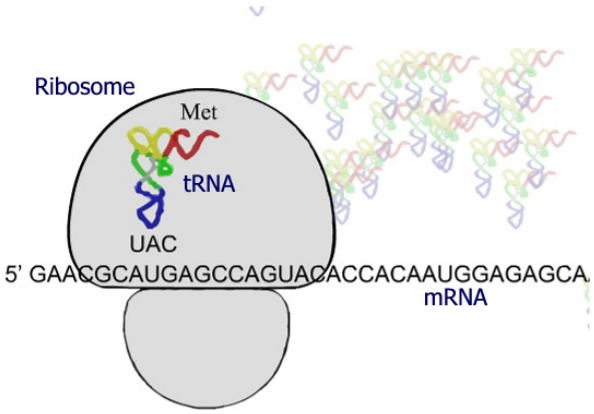
4x4x4 = 64

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.



Start codon and the initiator tRNA

All eukaryote proteins start with methionine residue.

Leader signal peptide is usually cleaved during protein trafficking.

Mature proteins correspond to biologically functional versions.

Maize embryo globulin S allele (7S-like) mRNA, complete cds

Human MHC class I transplantation antigen (hla) gene

Dosidicus gigas isolate VolcanMari-2 NADH dehydrogenase

S.cerevisiae genes HSS1, NPL4 and HSP

GenBank: X72224.1

FASTA Graphics

Go to:

```
LOCUS       X72224                5380 bp    DNA     linear   PLN 12-JUN-2006
DEFINITION  S.cerevisiae genes HSS1, NPL4 and HSP.
ACCESSION   X72224
VERSION     X72224.1
KEYWORDS    .
SOURCE      Saccharomyces cerevisiae (baker's yeast)
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Dikarya; Ascomycota; Saccharomycotina;
            Saccharomycetes; Saccharomycetales; Saccharomycetaceae;
            Saccharomyces.
REFERENCE   1
  AUTHORS   Kurihara,T. and Silver,P.
  TITLE     Suppression of a sec63 mutation identifies a novel component of the
            yeast endoplasmic reticulum translocation apparatus
  JOURNAL   Mol. Biol. Cell 4 (9), 919-930 (1993)
   PUBMED   8257794
REFERENCE   2  (bases 1 to 5380)
  AUTHORS   DeHoratius,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAY-1993) C.M. DeHoratius, Princeton University, Dept
            of Molecular Biology, Lewis Thomas Laboratory, Washington Road,
            Princeton, NJ 08540, USA
FEATURES             Location/Qualifiers
     source          1..5380
                     /organism="Saccharomyces cerevisiae"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4932"
                     /chromosome="II"
     gene            complement(160..780)
                     /gene="HSS1"
     CDS             complement(160..780)
                     /gene="HSS1"
                     /codon_start=1
                     /protein_id="CAA51025.1"
                     /db_xref="GOA:P33754"
                     /db_xref="InterPro:IPR018624"
                     /db_xref="SGD:S000000375"
                     /db_xref="UniProtKB/Swiss-Prot:P33754"
                     /translation="MSEFNETKFSNNGTFFETEEPIVETKSISVYTPLIYVFILVVSL
                     VMFASSYRKKQAKKISEQPSIFDENDAHDLYFQIKEMSENEKIHEKVLKAALLNRGAE
                     SVRRSLKLKELAPQINLLYKNGSIGEDYWKRFETEVKLIELEFKDTLQEAERLQPGWV
```

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)

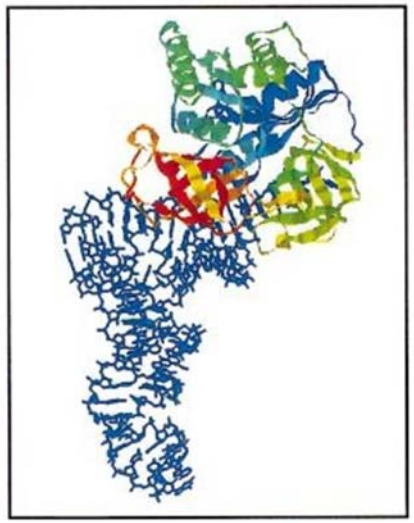
UAA is ochre

Also called termination codons.

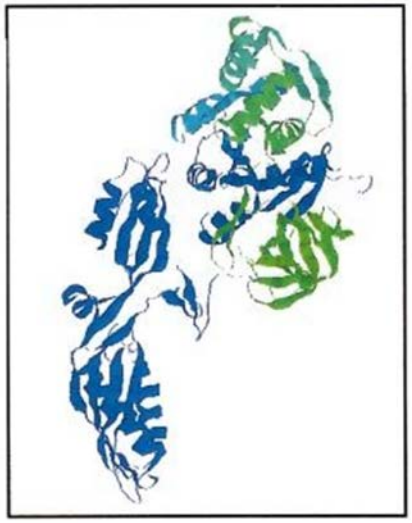
Do not have tRNA anti-codons, instead they bind release factors.

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

www.genomica.uaslp.mx

GenomicaUASLP

YouTube GenomicaUASLP

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.

(Last updated: August 23, © 2024.)