# The Genetic Code

The genetic code consists of 64 triplets of nucleotides. These triplets are called **codons**.With three exceptions, each codon encodes for one of the 20 amino acidsused in the synthesis of proteins. That produces some redundancy in the code: most of the amino acids being encoded by more than one codon.

**Index to this page**



One codon, **AUG** serves two related functions:

- it signals the start of translation
- it codes for the incorporation of the amino acid methionine (Met) into the growing polypeptide chain

The genetic code can be expressed as either RNA codons or DNA codons. RNA codons occur in messenger RNA (**mRNA**) and are the codons that are actually "read" during the synthesis of polypeptides (the process called **translation**). But each mRNA molecule acquires its sequence of nucleotides by **transcription** from the corresponding gene. Because DNA sequencing has become so rapid and because most genes are now being discovered at the level of DNA before they are discovered as mRNA or as a protein product, it is extremely useful to have a table of codons expressed as DNA. So here are both.

Note that for each table, the left-hand column gives the first nucleotide of the codon, the 4 middle columns give the second nucleotide, and the last column gives the third nucleotide.

## The RNA Codons

| | Second nucleotide | | | | |
|---|---|---|---|---|---|
| | **U** | **C** | **A** | **G** | |
| **U** | UUU **Phenylalanine** (Phe ) | UCU **Serine** (Ser) | UAU **Tyrosine** (Tyr) | UGU **Cysteine** (Cys) | **U** |
| | UUC Phe | UCC Ser | UAC Tyr | UGC Cys | **C** |
| | UUA **Leucine** (Leu) | UCA Ser | UAA **STOP** | UGA **STOP** | **A** |
| | UUG Leu | UCG Ser | UAG **STOP** | UGG **Tryptophan** (Trp ) | **G** |
| **C** | CUU **Leucine** (Leu) | CCU **Proline** (Pro) | CAU **Histidine** (His) | CGU **Arginine** (Arg) | **U** |
| | CUC Leu | CCC Pro | CAC His | CGC Arg | **C** |
| | CUA Leu | CCA Pro | CAA **Glutamine** (Gln) | CGA Arg | **A** |

| | | | | | |
|---|---|---|---|---|---|
| | CUG Leu | CCG Pro | CAG Gln | CGG Arg | **G** |
| **A** | AUU **Isoleucine** (Ile) | ACU **Threonine** (Thr ) | AAU **Asparagine** (Asn ) | AGU **Serine** (Ser) | **U** |
| | AUC Ile | ACC Thr | AAC Asn | AGC Ser | **C** |
| | AUA Ile | ACA Thr | AAA **Lysine** (Lys) | AGA **Arginine** (Arg) | **A** |
| | AUG **Methionine** (Met) or **START** | ACG Thr | AAG Lys | AGG Arg | **G** |
| **G** | GUU **Valine** Val | GCU **Alanine** (Ala) | GAU **Aspartic acid** (Asp) | GGU **Glycine** (Gly) | **U** |
| | GUC (Val) | GCC Ala | GAC Asp | GGC Gly | **C** |
| | GUA Val | GCA Ala | GAA **Glutamic acid** (Glu) | GGA Gly | **A** |
| | GUG Val | GCG Ala | GAG Glu | GGG Gly | **G** |

# The DNA Codons

These are the codons as they are read on the sense (5' to 3') strand of DNA. Except that the nucleotide thymidine (**T**) is found in place of uridine (**U**), they read the same as RNA codons. However, mRNA is actually synthesized using the antisense strand of DNA (3' to 5') as the template. [Discussion]

This table could well be called the Rosetta Stone of life.

### The Genetic Code (DNA)

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| TTT | Phe | TCT | Ser | TAT | Tyr | TGT | Cys |
| TTC | Phe | TCC | Ser | TAC | Tyr | TGC | Cys |
| TTA | Leu | TCA | Ser | TAA | STOP | TGA | STOP |
| TTG | Leu | TCG | Ser | TAG | STOP | TGG | Trp |
| CTT | Leu | CCT | Pro | CAT | His | CGT | Arg |
| CTC | Leu | CCC | Pro | CAC | His | CGC | Arg |
| CTA | Leu | CCA | Pro | CAA | Gln | CGA | Arg |
| CTG | Leu | CCG | Pro | CAG | Gln | CGG | Arg |
| ATT | Ile | ACT | Thr | AAT | Asn | AGT | Ser |
| ATC | Ile | ACC | Thr | AAC | Asn | AGC | Ser |
| ATA | Ile | ACA | Thr | AAA | Lys | AGA | Arg |
| ATG | Met* | ACG | Thr | AAG | Lys | AGG | Arg |
| GTT | Val | GCT | Ala | GAT | Asp | GGT | Gly |
| GTC | Val | GCC | Ala | GAC | Asp | GGC | Gly |
| GTA | Val | GCA | Ala | GAA | Glu | GGA | Gly |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| GTG | Val | | GCG | Ala | | GAG | Glu | | GGG | Gly |

*When within gene; at beginning of gene, ATG signals start of translation.

# Codon Bias

All but two of the amino acids (Met and Trp) can be encoded by from 2 to 6 different codons. However, the genome of most organisms reveals that certain codons are preferred over others. In humans, for example, alanine is encoded by GCC four times as often as by GCG. This probably reflects a greater translation efficiency by the translation apparatus (e.g., ribosomes) for certain codons over their synonyms. [More]

# Exceptions to the Code

The genetic code is **almost** universal. The same codons are assigned to the same amino acids and to the same START and STOP signals in the vast majority of genes in animals, plants, and microorganisms. However, some exceptions have been found. Most of these involve assigning one or two of the three STOP codons to an amino acid instead.

### Mitochondrial genes

When mitochondrial mRNA from animals or microorganisms (but not from plants) is placed in a test tube with the cytosolic protein-synthesizing machinery (amino acids, enzymes, tRNAs, ribosomes) it fails to be translated into a protein.

The reason: these mitochondria use UGA to encode tryptophan (Trp) rather than as a chain terminator. When translated by cytosolic machinery, synthesis stops where Trp should have been inserted.

In addition, most

- animal mitochondria use AUA for methionine not isoleucine and
- all vertebrate mitochondria use AGA and AGG as chain terminators.
- Yeast mitochondria assign all codons beginning with CU to threonine instead of leucine (which is still encoded by UUA and UUG as it is in cytosolic mRNA).

Plant mitochondria use the universal code, and this has permitted angiosperms to transfer mitochondrial genes to their nucleus with great ease.

Link to discussion of mitochondrial genes.

### Nuclear genes

Violations of the universal code are far rarer for nuclear genes.

A few unicellular eukaryotes have been found that use one or two (of their three) STOP codons for amino acids instead.

**Nonstandard Amino Acids**

The vast majority of proteins are assembled from the 20 amino acids listed above even though some of these may be chemically altered, e.g. by phosphorylation, at a later time.

However, two cases have been found where an amino acid that is not one of the standard 20 is inserted by a **tRNA** into the growing polypeptide.

- **selenocysteine**. This amino acid is encoded by UGA. UGA is still used as a chain terminator, but the translation machinery is able to discriminate when a UGA codon should be used for selenocysteine rather than STOP. This codon usage has been found in certain Archaea, eubacteria, and animals (humans synthesize 25 different proteins containing selenium).
- **pyrrolysine**. In several species of Archaea and bacteria, this amino acid is encoded by UAG. How the translation machinery knows when it encounters UAG whether to insert a tRNA with pyrrolysine or to stop translation is not yet known.

Welcome&Next Search

---

22 April 2010