

## Genetic Code

The genetic code is the set of rules by which information encoded within genetic material (DNA or mRNA sequences) is translated into proteins by living cells. The 'genetic code' determines a protein's amino acid sequence.

The set of all the codons that specify the 20 amino acids is termed as the genetic code.

A code of 3 nucleotides could code for a maximum of  $4^3 = 64$  amino acids

The Crick, Brenner et al. experiment first demonstrated that codons consist of three DNA bases.

Marshall Nirenberg and Heinrich J. Matthaei were the first to elucidate the nature of a codon in 1961 at the National Institutes of Health.

Har Gobind Khorana identified the rest of the genetic code.

Robert W. Holley determined the structure of transfer RNA (tRNA), the adapter molecule that facilitates the process of translating RNA into protein.

Nirenberg and Philip Leder revealed the triplet nature of the genetic code and deciphered the codons of the standard genetic code.

Leder and Nirenberg were able to determine the sequences of 54 out of 64 codons in their experiments.

In 1968, Khorana, Holley and Nirenberg received the Nobel Prize in Physiology or Medicine for their work.

### What is a Genetic Code?

The genetic code can be defined as the set of certain rules using which the living cells translate the information encoded within genetic material (DNA or mRNA sequences). The ribosomes are responsible to accomplish the process of translation. They link the amino acids in an mRNA-specified (messenger RNA) order using tRNA (transfer RNA) molecules to carry amino acids and to read the mRNA three nucleotides at a time.

## GENETIC CODE TABLE

		SECOND LETTER				
		U	C	A	G	
FIRST LETTER	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 Met	ACU } Thr ACC } ACA } ACG }	AAU } Asn AAC } AAA } Lys AAG }	AGU } Ser AGC } AGA Stop AGG Stop	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

- The genetic code is the set of rules by which a linear sequence of nucleotides specifies the linear sequence of a polypeptide.
- That is, they specify how the nucleotide sequence of an mRNA is translated into the **amino acid** sequence of a polypeptide.
- Thus, the relationship between the nucleotide sequence of the mRNA and the amino acid sequence of the polypeptide is the genetic code.
- The nucleotide sequence is read as triplets called codons.

### PRINCIPLES OF THE GENETIC CODE

- The genetic code consists of 64 different codons, each of which codes for 1 of the 20 amino acids.
- A codon consists of a triplet of nucleotide bases.

### CHARACTERISTICS OF THE GENETIC CODE

The genetic code is endowed with many characteristic properties which have actually been proved by definite experimental evidences.

#### 1. Triplet nature:

- Singlet and doublet codes are not adequate to code for 20 amino acids; therefore, it was pointed out that triplet code is the minimum required.

#### 2. Degeneracy

- The code is degenerate which means that the same amino acid is coded by more than one base triplet.
- Degeneracy does not imply lack of specificity in protein synthesis.
- It merely means that a particular amino acid can be directed to its place in the peptide chain by more than one base triplets.
- For example, the three amino acids arginine, alanine and leucine each have six synonymous codons.
- The code degeneracy is basically of 2 types: partial and complete.
- In partial degeneracy, the first two nucleotides are identical but the third (i.e., 3' base) nucleotide of the degenerate codon differs; for example, CUU and CUC code for leucine.
- Complete degeneracy occurs when any of the 4 bases can take third position and still code for the same amino acid; for example, UCU, UCC, UCA and UCG all code for serine.

### 3. Non-overlapping

- The genetic code is nonoverlapping, i.e., the adjacent codons do not overlap.
- A nonoverlapping code means that the same letter is not used for two different codons. In other words, no single base can take part in the formation of more than one codon.

### 4. Commaless

- The genetic code is commaless (or comma-free). There is no signal to indicate the end of one codon and the beginning of the next.
- There are no intermediary nucleotides (or commas) between the codons.

### 5. Non-ambiguity

- Non-ambiguous code means that there is no ambiguity about a particular codon.
- A particular codon will always code for the same amino acid.
- While the same amino acid can be coded by more than one codon (the code is degenerate), the same codon shall not code for two or more different amino acids (non-ambiguous).

### 6. Universality

- Universality of the code means that the same sequences of 3 bases encode the same amino acids in all life forms from simple microorganisms to complex, multicelled organisms such as human beings.
- Although the code is based on work conducted on the bacterium *Escherichia coli* but it is valid for other organisms.
- The genetic code applies to all modern organisms with only minor exceptions, such as the yeast, mitochondria, and the *Mycoplasma*.

### 7. Polarity

- The genetic code has polarity, that is, the code is always read in a fixed direction, i.e., in the 5' → 3' direction.

### • SPECIAL CODONS

#### A. Chain Initiation Codons

- The triplets AUG and GUG play double roles in *E. coli*.
- When they occur in between the two ends of a cistron (intermediate position), they code for the amino acids methionine and valine, respectively in an intermediate position in the protein molecule.
- But when they occur immediately after a terminator codon, they act as “chain initiation” (C.I.) signals or “starter codons” for the synthesis of a polypeptide chain.

#### B. Chain Termination Codons

- The 3 triplets UAA, UAG, UGA do not code for any amino acid.
- When any one of them occurs immediately before the triplet AUG or GUG, it causes the release of the polypeptide chain from the ribosome.
- They are also called as stop codons.
- They are also called chain termination codons because these codons are used by the cell to signal the natural end of translation of a particular peptidyl chain. However, their inclusion in any mRNA results in the abrupt termination of the message at the point of their location even though the polypeptide chain has not been completed.

#### C. Sense Codons

61 codons, which code for particular amino acids are termed as sense codons.

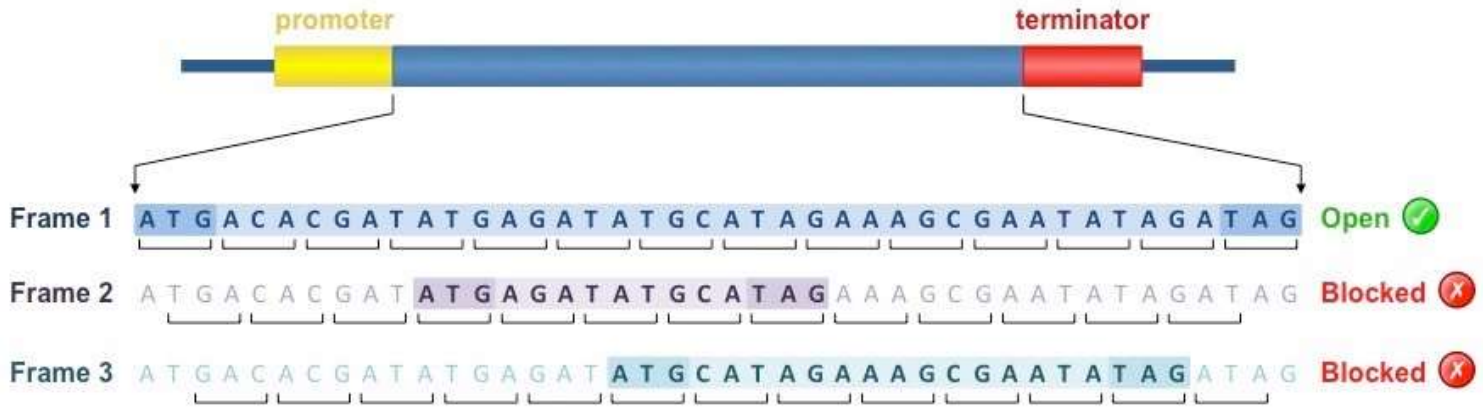
#### D. Non-Sense Codons

- Triplets UAA, UAG, UGA do not code for any amino acid.
- They were originally described as non-sense codons.
- However, the so-called non-sense codons have now been found to be of “special sense”.
- These special-sense codons perform the function of punctuating genetic message like a full stop at the end of a sentence.

### PATTERNS OF THE GENETIC CODE

8. Amino acids with similar structural properties tend to have related codons.
  - Aspartic acid codons (GAU, GAC) are similar to glutamic acid codons (GAA, GAG); the difference being exhibited only in the third base (toward 3' end).
  - Similarly, the codons for the aromatic amino acids phenylalanine (UUU, UUC), tyrosine (UAU, UAC) and tryptophan (UGG) all begin with uracil (U).
9. All codons with U in the second position specify hydrophobic amino acids (Ile, Leu, Met, Phe, Val).
10. All codons with A in the second position specify the charged amino acids, except Arg.
11. All the acidic (Asp, Glu) and basic (Arg, Lys) amino acids have A or G as the second base.

### READING FRAMES AND OPEN READING FRAMES



- The mRNA sequence can be read by the ribosome in three possible reading frames.
- Usually only one reading frame codes for a functional protein since the other two reading frames contain multiple termination codons.
- However, in some bacteriophage, overlapping genes occur which use different reading frames.
- An open reading frame (ORF) is a run of codons that starts with ATG and ends with a termination codon, TGA, TAA or TAG.
- Coding regions of genes contain relatively long ORFs unlike non-coding DNA where ORFs are comparatively short.
- The presence of a long open reading frame in a DNA sequence therefore may indicate the presence of a coding region.
- Computer analysis of the ORF can be used to deduce the sequence of the encoded protein.

#### WOBBLE HYPOTHESIS

Wobble hypothesis: The first two bases on the codon are usually the important ones, the third base may be in some cases one of the four bases and the triplet would still code for the same amino acid as long as the first two bases are the same and have the same sequence. This concept is known as the Wobble hypothesis.