The Genetic Code

The genetic code is the set of rules by which information encoded in genetic material (DNA or RNA sequences) is translated into proteins (amino acid sequences) by living cells. The genetic code consists of the sequence of nitrogen bases—A, C, G, U—in an mRNA chain. The four bases make up the "letters" of the genetic code. The letters are combined in groups of three to form code "words," called codons. Each codon codes for (encodes) one amino acid, except the codons for start or stop signal.

It helps in translating the information contained in a linear sequence of nucleotides into that of the linear sequence of a polypeptide. In other words, they specify how the nucleotide sequence of an mRNA is translated into the amino acid sequence of a polypeptide.

Characteristics of the genetic code

1. **Code is triplet**

    There are 20 Amino acids. A three letter code could make a genetic code for 64 different combinations (4 X 4 X 4)   of genetic code and is sufficient to specify all 20 amino acids. When experiments were performed to crack the genetic code it was found to be a code that was triplet. These three letter codes of nucleotides (AUG, AAA, etc.) are called codons.

2. Code is degenerate

    The occurrence of more than one codon for a single amino acid is referred to as degenerate. A review of genetic code dictionary reveals that most of the amino acids have more than one codon. Out of 20 amino acids, only Met and Trp have one codons each, AUG and UGG . Arginine, Alanine and Leucine each have six codons.
    Explanation for degeneracy of codons is provided by wobble hypothesis

3. **The Code is Non-overlapping**
    A non overlapping code means that the same letter is not used for two different codons. In case of non-overlapping code, mutation in one base will

result in change of only one amino acid where as overlapping codon will result in change of three amino acids which is not the case.

According to Barrel (1976) in φx174 same base can be used for different codons but only at different occasions, in time and space, so the same base can't be used for two different codons during synthesis of same protein. (Has very short nucleic acid. For maximum utilization it changes the reading frame)

4. The Code is Comma Less

A comma less code means that no nucleotide or comma (or punctuation) is present in between two codons. Therefore, code is continuous and comma less and no letter is wasted between two words or codons.

RNA Bases	UUU	[C]UC	GUA	UCC	ACC
Amino Acids	Phe	Leu	Val	Ser	Thr

5. Code has Polarity

Message in mRNA is read in the 5 -3 direction. Thus the polarity of genetic code is from 5 end to 3 end.

Codons	UUG	GUC	UCG	CCA	AGG
Reding left to right	Leu	Val	Ser	Pro	Arg
Reading Right to left	Val	Leu	Ala	Thr	Gly

6. Code is Non-ambiguous

Non-ambiguous code means 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 same codon shall not code for two or more different amino acids. Rarely GUG

act as start codon and codes for Methionine but at intermediate positions it codes for valine.

7. Code is universal

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 multicellular organisms such as human beings.
Although the code is based on work conducted on the bacterium *Escherichia coli* but it is valid for all the other organisms. In 1981, Barrel et. al. observed some minor differences in human mitochondrial genetic code and genetic codes of some primitive organisms like yeast, ciliated protozoa like Tetrahymena and Mycoplasma. (it was perhaps to protect these organisms against viruses that use the universal genetic code).
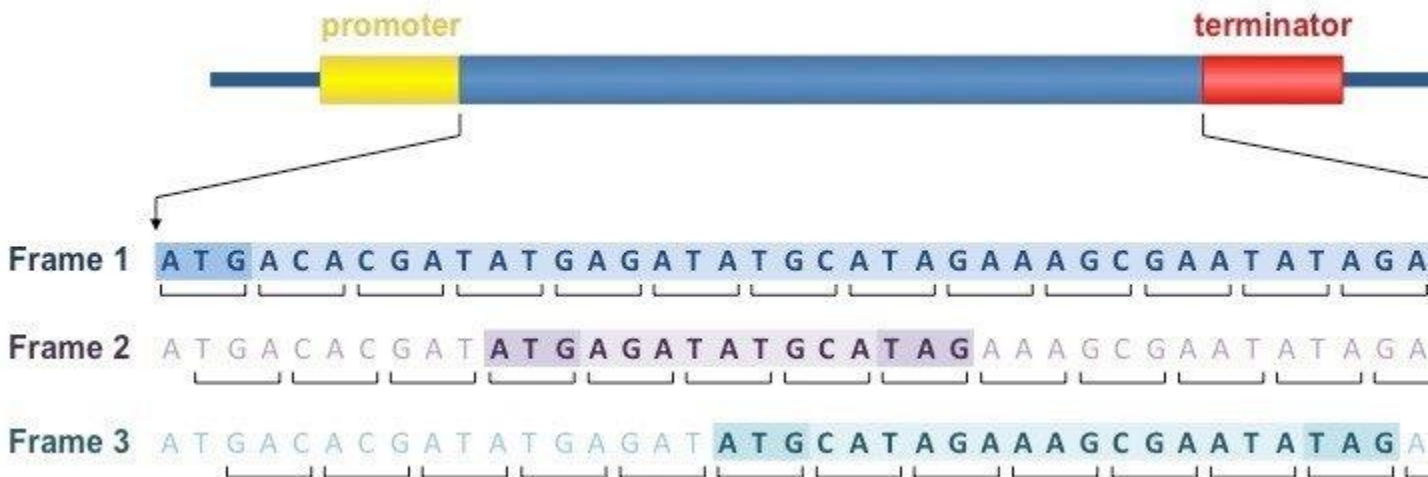
Chain Initiation Codons (Start codon)

The triplet AUG is the start codon or initiating codon, present in the beginning of the mRNA encoding methionine. Rarely GUG, which codes for valine at intermediate positions acts as start codon in some bacteria and encodes methionine at initiating position.

Chain Termination Codons (Stop codon)

The 3 triplets UAA, UAG, UGA do not code for any amino acid. They are called stop codons or chain termination codons because they signal the end of the translation process.

***

READING FRAMES AND OPEN READING FRAMES

promoter

terminator

Frame 1 ATGACACGATATGAGATATGCATAGAAAGCGAATATAGA

Frame 2 ATGACACGAT ATGAGATATGCATAG AAAGCGAATATAGA

Frame 3 ATGACACGATATGAGAT ATGCATAGAAAGCGAATATAGA

- 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.