GENETIC CODE

The number and the sequences of bases in mRNA, specifying an amino acid, is known as codon, while the set of bases in a tRNA, that base pair with a codon of an mRNA is known as anticodon. For example the codon 5' AUG 3' has the anticodon 3' UAC 5'. The set of all the codons, that specify the 20 amino acids, is termed as the genetic code, genetic language or coding dictionary.

The genetic language has only four bases – A, T, G & C. The four bases making up polynucleotide. 20 different amino acids are incorporated during translation. Thus at least 20 different codons must be formed using the four symbols available in the mRNA. Two bases per codon would yield only 4^2 or 16 possible codon – clearly not enough. Three bases per codon yield 4^3 or 64 possible codon – an apparent excess.

Properties of Genetic Code:

Each codon is written as it appears in mRNA & reads in a 5' to 3' direction. The characteristics of genetic code are as follows:

1. The code is a triplet code – each mRNA codon that specifies an amino acids in a polypeptide chain consists of three nucleotides.
2. The code is comma free – the mRNA is read continuously. Three nucleotides (one codon) at a time without skipping any nucleotides of the message.
3. The code is non-overlapping – the mRNA is read in successive groups of three nucleotides. A message of AAGAAGAAG…..in a cell would be read as lysine-lysine-lysine…., which is what the AAG specifies. Theoretically three readings are possible from this message, depending on where the reading is began.
4. The code is almost universal – all organisms share the same genetic language. Thus for example lysine is coded for by AAA or AAG in the mRNA of all organisms, arginine by CGU, CGC, CGA, CGG, AGA & AGG & so on. The code, however is not completely universal. For example the mitochondria of some organisms such as mammals have minor changes in the code.
5. The code is degenerate – with two exceptions – AUG (methionine) & UGG (tryptophan). More than one codon occurs for each amino acid. This multiple coding is called the degeneracy of genetic code. Thus when the first two nucleotides in a codon are identical & the third letter is U or C, the codon often codes for the same

amino acids. For examples UUU & UUC specify phenyl alanine. Similarly CAU & CAC specify histidine. Also when the first two nucleotides in a codon are identical and the third letter is A or G, the same amino acid is often specified. For example, UUA & UUG specify leucine. In some cases with identity in the first two positions, the base in the third position may be U, C, A or G, the same amino acid will be specified. For example CUU, CUC, CUA & CUG, all for leucine. Even though there is degeneracy of codon, that does not mean, that all codons are used equally. For example, UUC is more frequently used codon for phenyl alanine followed by UUU.

6. The code has start and stop signals – in both prokaryotes and eukaryotes, AUG is most commonly used as a start codon for protein synthesis. Although in rare cases GUG may also be used.

   Only 61 of the 64 codons specify amino acids, these codons are called sense codons. The three other codons – UAG (Amber), UAA (Ochre) & UGA (Opal) do not specify an amino acid and no tRNA in normal cell carry the appropriate anticodon. These three codons are the stop or non-sense codons or chain terminating codons.

7. Wobble occurs in the anticodon – since 61 sense codons specify amino acids in the mRNA, a total of 61 tRNA molecules can have the appropriate anticodons. Theoretically though the complete set of 61 sense codons can be read by fewer than 61 distinct tRNA, because of the wobble in the anticodon. The wobble hypothesis proposed by F. Crick shows that the base at the 5' end of the anticodon (complementary to the base of the 3' end of the codon) is not as constrained as the other two bases. This feature allows for less exact base pairing, so that the base at the 5' end of anticodon can potentially pair with one or three different bases at the 3' end of the codon.