The Genetic Code

We know that the genetic information in the nucleic acids is expressed through synthesis of proteins (see Chapter 38-40). We also know that, while proteins have 20 different kinds of essential amino acids, nucleic acids (particularly in living calls) have only four different kinds of bases. (Recently, it has been shown that in artificially synthesized DNA segments, the number of bases can extend to more than four i.e. 6 or 8). It may therefore, be realized that while 20 amino acids are the alphabets of the language of proteins, four bases are the alphabets of the language of nucleic acids. Since it is through mRNA (messenger RNA, see Chapter 38) that the genetic information is passed on to proteins, the problem of genetic code was to prepare a dictionary for translating the language of RNA into the language of proteins. Since one is four alphabet language and the other has 20 alphabets, as shown in Figure 36.1, a singlet code would give only four codons, a doublet code would have 16 codons and triplet code would give us 64 ($4 \times 4 \times 4$) triplets. The 64 triplets would be enough to code for 20 amino acids.

In a 64-codons dictionary, either there should be more than one triplets coding for the same amino acid or else, of the 64 triplets, 44 triplets should mean nothing (nonsense). It has now been proved by experimental evidence that the code is triplet and that all 64 codons carry some meaning, although three of these 64 triplets cannot be translated in terms of amino acids, since they mean stop signals. Whenever any one of these three triplets is encountered on mRNA, it orders that the end of the polypeptide chain has reached and that the chain should now be released, before further synthesis may take place.

A
G
C
T

singlet code

AA	AG	AC	AT
GA	GG	GC	GT
CA	CG	CC	CT
TA	TG	TC	TT

doublet code

AAA	GAA	CAA	TAA
AAG	GAG	CAG	TAG
AAC	GAC	CAC	TAC
AAT	GAT	CAT	TAT
AGA	GGA	CGA	TGA
AGG	GGG	CGG	TGG
AGC	GGC	CGC	TGC
AGT	GGT	CGT	TGT
ACA	GCA	CCA	TCA
ACG	GCG	CCG	TCG
ACC	GCC	CCC	TCC
ACT	GCT	CCT	TCT
ATA	GTA	CTA	TTA
ATG	GTG	CTG	TTG
ATC	GTC	CTC	TTC
ATT	GTT	CTT	TTT

triplet code

Fig. 36.1. A singlet code, a doublet code and a triplet code.

Properties of Genetic Code

Following properties of the genetic code were proved by definite experimental evidence : (i) the code is **triplet**, (ii) the code is **degenerate**, (iii) the code is **non-overlapping**, (iv) the code is **commaless**, (v) the code is **non-ambiguous** and (vi) the code is **universal**. Although it may not be necessary to present the experimental evidences which proved the validity of these properties, it may be useful to explain the meaning of these six properties of the genetic code listed above.

The code is triplet

As earlier outlined, singlet and doublet codes are not enough to code for 20 amino acids; it was pointed out that triplet code is the minimum required. But it could be a quadruplet code or of a higher order. As pointed out above, in a triplet code of 64 codons, there is an excess of 44 codons and, therefore, more than one codons are present for the same amino acid. This excess will be still greater if more than three-letter words are used. In the quadruplet code there will be $4 \times 4 \times 4 \times 4 = 256$ possible words.

The code is degenerate

In a triplet code, as pointed out earlier also, for a particular amino acid more than one words (synonyms) can be used. This phenomenon is described by saying that the code is degenerate. A non-degenerate code would be one where there is one to one relationship between amino acids and the codons, so that 44 codons out of 64, will be useless or nonsense codons. It has been definitely shown that there are no nonsense codons. The codons which were earlier called nonsense codons are also now known to mean stop signals.

The code is non-overlapping

Non-overlapping code means that a base in a mRNA is not used for two different codons. In Figure 36.2 it is shown that an overlapping code can mean coding for four amino acids from six bases. In actual practice six bases code for not more than two amino acids. However, in overlapping genes described in Chapter 35, it is shown that the same base can be used for different codons, but only at different occasions in time and/or space, so that the same base can not be used for two different

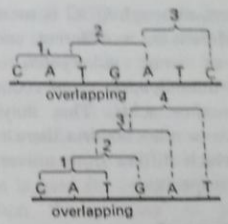

Fig. 36.2. Overlapping of codons due to one letter or two letters.

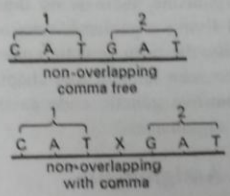

Fig. 36.3. Genetic code, without comma and with comma.

codons during synthesis of the same protein. This is exactly, what we mean by non-overlapping code.

The code is commaless

A commaless code means that no punctuations are needed between any two words. In other words, we can say that after one amino acid is coded, the second amino acid will be automatically coded by the next three letters and that no letters are wasted for telling that one amino acid has been coded and that now second should be coded (Fig. 36.3).

The code is non-ambiguous

Non-ambiguous code means that there is no ambiguity about a particular codon. A particular codon will always code for the same amino acid, wherever it is found. In an ambiguous code, the same codon could have different meanings, or in other words, the same codon could code two or more than two different amino acids. Such is not the case. While the same amino acid can be coded by more than one codons (the code is degenerate), the same codon shall never code for two different amino acids (non ambiguous). However, there is some element of ambiguity when AUG and GUG are considered; both may code for *methionine* as initiating codons, although GUG is meant for *valine* (Fig. 36.4). Moreover, a different code exists in mitochondria of some eukaryotes, so that in cytoplasm and mitochondria same codon may code for different amino acids. This may not mean ambiguity, since in mitochondria there is a separate genetic code which differs from universal code in some essential respects.

The code is universal

Although the code has been worked out by using *in vitro* systems prepared usually from micro-organisms, there is no doubt now that in all kinds of living organisms, micro- or macro-, plants or animals, the same genetic code is used. However, as will be seen later in this chapter, a different and more primitive genetic code exists in mitochondria of some organisms.