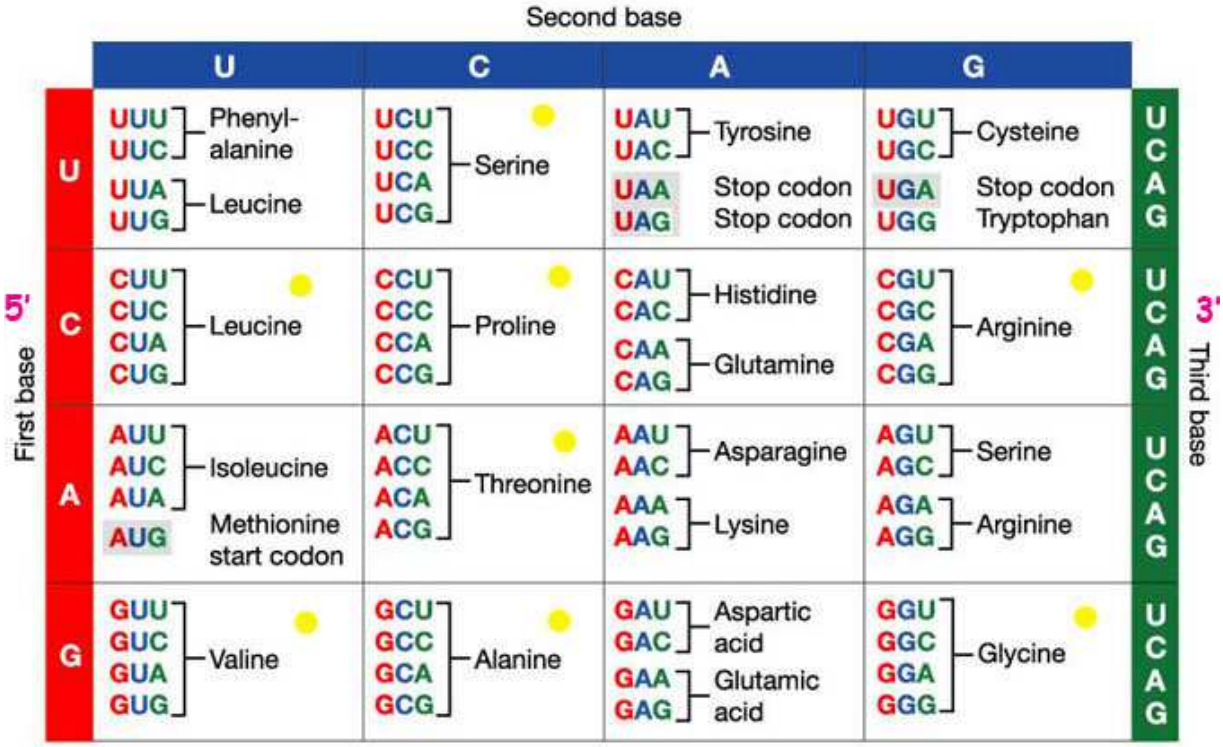
THE GENETIC CODE

Dr.H.B.Mahesha, Yuvaraja's College, University of Mysore, Mysuru.

As DNA is a genetic material, it carries genetic information from cell to cell and from generation to generation. There are only four bases in DNA and twenty amino acids in protein, so some combinations of the bases is needed to specify a particular amino acid. The set of such combinations is called genetic code. A base sequence corresponding to a particular amino acid is a codon.

The surface distinctions among twenty amino acids are neither specific enough nor sufficiently large for pattern of bases to be recognized by means of hydrogen bonding; thus recognition required another intermediate an adopter RNA molecule. This adopter, *i.e.*, tRNA contains a site to which a particular amino acid is attached and a base sequence called anticodon, which hydrogen bonds via complementary base pairing to a given codon.

Figure 1: The genetic code showing the codons and their respective amino acids



Properties of Genetic Code

1. The genetic code is a triplet code:

DNA contains four kinds of nucleotides (of A, T, G and C), and proteins are synthesized from 20 different types of amino acids. A basic problem regarding the genetic code was: how many bases of DNA specify one amino acid? In a singlet code each base or letter would specify one amino acid. Only 4 of the 20 types of amino acids would be coded unambiguously by a singlet code (Table). In a two-letter or doublet code two bases would specify one amino acid. Here 16 (4 x 4) of the 20 amino acids can be specified, but there would be ambiguous determination of a number of amino acids. A triplet or three-letter code was first suggested by the physicist Gamow in 1954. According to the triplet code three letters or bases specify one amino acid. Thus 64 (4 x 4 x 4) distinct triplets of purine and/or pyrimidine bases determine the 20 amino acids. These triplets have been called codons. Since there are 64 codons and only 20 amino acids it is obvious that there are many more codons than there are amino acids, *i.e.*, the code is degenerate. Experimental evidence shows that the code is a triplet one, and that 61 of the 64 codons code for individual amino acids during protein synthesis.

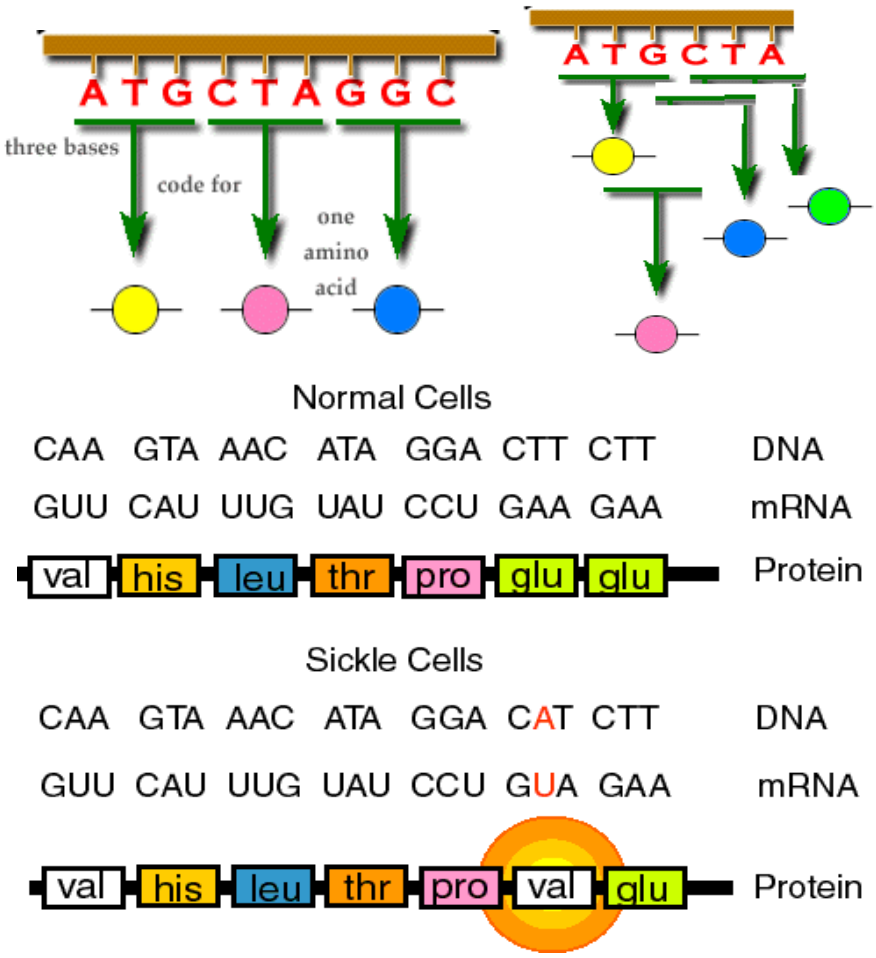
A quadruplet code would have 4x4x4x4 = 256 codons, and would show even more degeneracy than the triplet code.

2. The code is non-overlapping:

Since the DNA molecule is a long chain of nucleotides it could be read either in an overlapping or non-overlapping manner. The genetic code could thus be overlapping or non-overlapping. The reading of the code by these two different ways would yield different results. In the non-overlapping code six nucleotides would code for two amino acids, while in the overlapping code up to four could be coded (Fig.). In the non-overlapping code each letter is read only once while in the over- lapping code it would be read three times, each time as a part of a different word. Mutational changes in one letter would affect only one word-in the non-overlapping code while it would affect three words in the overlapping code.

Studies on normal and sickle cell hemoglobin show that a single mutational change results in the substitution of only one amino acid.

Non Overlapping and Overlapping



Genetic code is Non-Overlapping, if it is overlapping three amino acids Should have been changed in sickle cell anemia

3. The code is comma less:

Is the genetic code read in an uninterrupted manner from one end of the nucleic acid chain to the other? Or are there bases (commas) between successive codons? A code with commas could be represented as follows (the X represents a base acting as a comma).

DDD X CUC X GUA X UCC X ACC------Bases

Phe Leu Val Ser Thr---------Amino acids

A mutation resulting in an addition or deletion of a base would affect only one amino acid of the polypeptide chain. The total genetic message would be only slightly changed.

DUU X –UC X GUA X UCC X ACC ---Bases

Phe Changed Val Ser Thr ----amino Acids

aa

A commaless code would not have the comma bases and can be represented thus:

UUU CUC GUA UCC ACC ----Bases

Phe Leu Val Ser Thr ----Amino acids

In such a code any mutation involving a deletion of a base (-C) would result in a drastic change in the genetic message.

UUU UCG UAU CCA CC ----Bases

Phe Ser Tyr Pro ----Amino acids

The entire series of amino acids following the deletion would change.

All the available evidence indicates that the code is commaless, i.e. there are no demarcating signals between codons. The work of Khorana and his associates cited below gives clear evidence of a commaless code. Long synthetic polynucleotides with specific repeating sequences were used for translation of protein chains. Thus the repeating sequence

CUCUCU contains the codons CUC (for leucine) and UCU (for serine). When this sequence is used for translation of proteins, neither amino acid is incorporated into the protein unless the other IS also present. This result can only be explained by a commaless triplet code where there would have to be alternate translation of CUC and UCU codons.

4. The code is non ambiguous: A particular codon will always code for the same amino acid. In an ambiguous code, the same codon could coded two or more than one codon (i.e. the code is degenerate), the same codon shall not code for 2 or more different amino acids (non ambiguous) except when same codon in the nucleus and mitochondria may code for different amino acids.

5. The code is universal: the genetic code is valid for all organisms ranging from bacteria to man. Therefore it I said to be universal. Marshal et al 1967, demonstrated that the amino acyl tRNAs of E. coli, *Xenopus laevis* (amphibian) and guinea pig use the same genetic code. (Except mitochondria and ciliate protozoa *i.e.*,

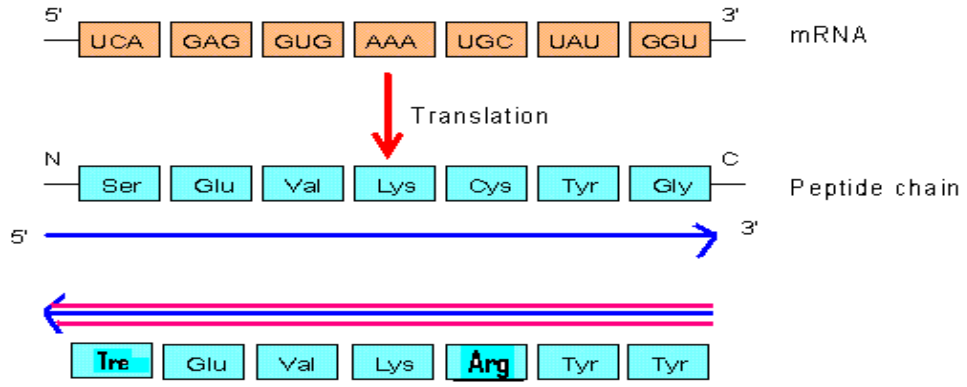
i. in yeast mitochondria UGA codes for tryptophan, although in the nuclear genes UGA is a termination codon.

ii. In ciliate protozoa (*Mycoplasma capricolum*) in this genetic code, codon UAA and UAG specify glutamine instead of stop signals. In future more such cases may be discovered, showing diversity in the genetic code.

6. The code has polarity:

If a gene is to specify the same protein repeatedly it is essential that the code must be read between fixed start and end points. These points are the initiation and the termination codons, respectively. It is also essential that the code must be read in a fixed direction. In other words the code must have polarity. It is obvious that if the code is read in opposite directions it would specify two different proteins, since the codons would. have reversed base sequences. Thus if the message given below is read from left to right the first codon, UCA (first codon from left 5' end), would specify serine.

If read from right to left the codon would become ACU and would specify threonine. It is thus seen that the sequence of amino acids constituting the protein would undergo a drastic change if the code is read in the opposite direction. The available evidence indicates that the message in mRNA is read in the 5'→3' direction. The polypeptide chain is synthesized in the N→C direction, i.e. from the amino (NH_2) terminal to the carboxyl (COOH) terminal



7. Codons and anticodons:

During translation the codons of mRNA pair with complementary anticodons of tRNA. Since mRNA is read in a polar manner in the 5'→3' direction the codons are also written in the 5'→3' direction. Thus the codon AUG is written as 5'AUG3'. The corresponding anticodon on tRNA should therefore be written as 5'CAU3', In such a configuration the first bases of both codon and anticodon would be the ones at the 5' end and third bases at the 3' end.

Base number		1	2	3	
Codon (mRNA)	5'	A	U	G	3'
Anticodon (tRNA)	3'	U	A	C	5'
Base number		3	2	1	

Often, however, the anticodon is written in the 3'→5' direction so as to bring about an easier correlation between the bases of the codon and anticodon. Thus the anticodon for AUG is written as 3' UAC 5' or, more simply, UAC. Here the first letter in the codon is at the 5' end and the first letter of the anticodon at the 3' end.

8. Initiation codons:

The starting amino acid in the synthesis of most protein chains is methionine (eukaryotes) or N-formyl methionine (prokaryotes). MethionyI or N-formyJ methionyl-tRNA specifically binds to initiation sites containing the AUG codon. This codon is therefore called the initiation codon. Less often, GUG also serves as the initiation codon in bacterial protein synthesis. Normally GUG is the codon for valine. In the phage MS2, GUG is the initiation codon for the A protein. GUG has been found to initiate protein synthesis when the normal AUG codon is lost by deletion. However, initiation by GUG is less, efficient since it has a lower affinity for fMet-tRNA. Both AUG and GUG codons show ambiguity III one sense, since each of them codes for two different amino acids. When these two codons are at initiation positions of mRNA they code for N-formyl methionine. In internal positions AUG codes for methionine and GUG for valine.

9. Termination codons:

Three of the 64 codons do not specify any tRNA and were hence called nonsense codons. These codons are VAG (amber), VAA (ochre) and UGA (opal or umber). Since they bring about termination of polypeptide chain synthesis they are also called termination codons. VAG was the first termination codon to be discovered. It was named 'amber' after a graduate student named Bernstein (the German for 'amber') who helped in the discovery of a class of mutations. Apparently to give uniformity the other two termination codons were also named after colors.

Termination codons do not code for any amino acids and hence cause termination and release of polypeptide chains. Apparently no tRNA species has anti codons complementary to the termination codons. There are mRNAs with single termination codons and also mRNAs with two successive termination codons (e.g. MS2 coat protein mRNA). Termination codons are not read by any tRNA molecules but by proteins called release factors. In prokaryotes there are three release factors RF-I, RF-2 and RF-3. RF-I recognizes UAA and VAG, while RF-2 recognizes VAA and UGA. RF-3 stimulates RF-l and RF·2. In eukaryotes a single release factor (RF) recognizes all three termination codons.

10. The code is degenerate:

As mentioned previously, there are 64 possible codons in a triplet code of which 61 have been shown to code amino acids. Since only 20 amino acids take part in protein synthesis it is obvious that there are many more codons than amino acid types. Except for tryptophan and methionine, which have a single codon each, all other amino acids involved in protein synthesis have more than one codon. Phenylalanine, tyrosine, histidine, glutamine, asparagine, lysine, aspartic acid, glutamic acid and cysteine have two codons each. Isoleucine has three codons. Valine, proline, threonine, alanine and glycine have four codons each. Leucine, arginine and serine have six codons each (Table). This variability in the number of codons for different amino acids may at least partially account for the unequal distribution of the different amino acids in protein. In general, the frequency of appearance of amino acids in proteins roughly corresponds to the number of available codons.

Table: Number of codons coding for different amino acids. Amino acids in categories 2-5 are coded by more than one codon. Such codons are called degenerate.

Amino Acids	Number of codons
1.Tryptophan. methionine ..... ……..	1
2.Phenylalanine, tyrosine, histidine, glutamine, asparagine …….	2
lysine, aspartic acid, glutamic acid, cysteine …….....	..2
3.Isoleucine	3
4.Valine, proline, threonine, alanine. glycine	…..4
5. Leucine, arginine, serine	…· .. 6

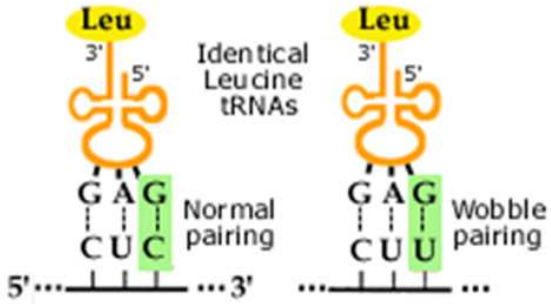
The wobble hypothesis:

The triplet code is a degenerate one with many more codons than the number of amino acid types coded. An explanation for this degeneracy is provided by the 'wobble hypothesis' proposed by Crick (1966). Since there are 61 codons specifying amino acids, the cell should contain 61 different tRNA molecules, each with a different anticodon. Actually, however, the number of tRNA molecule types discovered is much less than 61. This implies that the anticodons of some tRNAs read more than one codon on mRNA.

According to the wobble hypothesis only the first two positions of a triplet codon on mRNA have a precise pairing with the bases of the tRNA anticodon. The pairing of the third position bases of the codon may be ambiguous, and varies according to the nucleotide present in this

position. Thus a single tRNA type is able to recognize two or more codons differing only in the third base. The anticodon UCG of serine tRNA recognizes two codons, AGC and AGU. The bonding between UCG and AGC follows the usual Watson-Crick pairing pattern. In UCG. AGU pairing, however, hydrogen bonding takes place between G and U. This is a departure from the usual Watson-Crick pairing mechanism where G pairs with C and A with U. Such interaction between the third bases is referred to as 'wobble pairing'.

The degeneracy of the code is not random. Mostly the different co dons for a particular amino acid have the same first two letters (leucine, serine and arginine are exceptions). Thus the first two letters of all the four codons for valine are GU and for alanine GC.



When only two codons specify an amino acid the third letters of the codons are either both purines or both pyrimidines: never one purine and one pyrimidine.

It is possible to predict the minimum number of tRNAs required to translate the different codons specifying a particular amino acid. The amino acid leucine is specified by six codons: UUA, UUG, CUU, CUC, CUA and CUG. The first two letters of two codons are UU and of four codons CU. Hence at least two different tRNAs are required, since the first two letters of a codon do not have wobble pairing with the anticodon.

Of the four codons having CU two (CUU and CUC) have pyrimidines as their third bases and two (CUA and CUO) have purines. Hence they cannot be read by the same anticodon, because the purine of the anticodon can only pair with a pyrimidine and vice versa. The CU - codons must therefore be read by at least two different anticodons. Thus at least three co dons are required to read to anticodons for leucine.

Wobble pairing takes place in only certain combinations. Three types have been proposed:

(i) U in the wobble position of the tRNA anticodon can pair with A or G of the mRNA codon, (ii) G can pair with U or C and (iii) I can pair with A, U or C.
