GENETIC CODE, DEGENERACY OF THE GENETIC CODE & WOBBLE HYPOTHESIS

Dr. R. Prasad
Department of Zoology,
Eastern Karbi Anglong College,
Sarihajan

The essence of gene expression lies in the relationship between the nucleotide base sequence of DNA molecule and the linear order of amino acids in protein molecules. This relationship is based on a set of rules known as the genetic code. How information residing in a DNA base sequence could be passed to mRNA through the mechanism of complementary base pairing and how does a base sequence in mRNA use its "message" to guide the synthesis of a protein molecule, which consists of a sequence of amino acids? What is needed, of course, is the knowledge of the appropriate code – the set of rules that determines which nucleotides in mRNA correspond to which amino acids. The cracking of that code, which tells us how DNA can code for proteins, is one of the major landmarks of twentieth-century biology. The awareness of the exact coding relationship between the base sequence of a DNA molecule and the amino acid sequence of a protein arose from the discovery that mutations in DNA can lead to changes in proteins.

Features of genetic code

The genetic code is a triplet code: One of the first questions about the genetic code to be addressed was how many nucleotides are necessary to specify a single amino acid? The basic unit of genetic code – the set of bases that encode a single amino acid is a codon. The investigations made by many early scientists recognized that codons must contain a minimum of three nucleotides. Each nucleotide's position in mRNA can be occupied by one of the four bases; A, G, C, or U. If a codon consisted of a single nucleotide, only four different codons (A,G, C, and U) would be possible, which is not enough to encode 20 different amino acids commonly found in proteins. If codons were made up of two nucleotides each (i.e., GU, AC and so on), there would be 4 x 4 = 16 possible codons – still not enough to encode all 20 amino acids. With three nucleotides per codon, there are 4 x 4 x 4 = 64 possible codons, which is more than enough to specify 20 different amino acids and that would be far less cumbersome than a four nucleotides per codon, 4 x 4 x 4 x 4 = 256 possible codons. Therefore, a triplet code requiring three nucleotides per codon is the most efficient way to encode all 20 amino acids. Using mutations in bacteriophage, Francis Crick and his colleagues in 1961 confirmed that genetic code is indeed a triplet code.

The genetic code is degenerate and unambiguous

One amino acid is encoded by three consecutive nucleotides in mRNA, and each nucleotide can have one of the four possible bases (A, G, C, and U) at each nucleotide position, thus permitting $4^3 = 64$ possible codons. Three of these codons are stop codons (UAA, UAG, and UGA), specifying the end of translation. Thus, 61 codons, called sense codons, encode amino acids. Because there are 61 sense codons and only 20 different amino acids commonly found in proteins, the code contains more information than is needed to specify the amino acids and is said to be degenerate code.

Degenerate is the term that Francis Crick borrowed from quantum physics, where it describes multiple physical states that have equivalent meaning. The degeneracy of the genetic code means that the amino acids may be specified by more than one codon. Tryptophan and methionine are exceptions as they are encoded by a single codon. All other amino acids are encoded by two codons, and some, such as leucine are specified by six different codons (see Fig. 12.25). Codons that specify the same amino acid are said to be synonymous. It is also clear from the Figure 12.25 below that the genetic code is unambiguous: every codon has one and only one meaning. Base pairing between codon and anticodon in which there is nonstandard pairing, usually at the third (3´) position of the codon and is called as **wobble hypothesis.** This nonstandard pairing allows more than one codon to pair with the same anticodon of the tRNA.

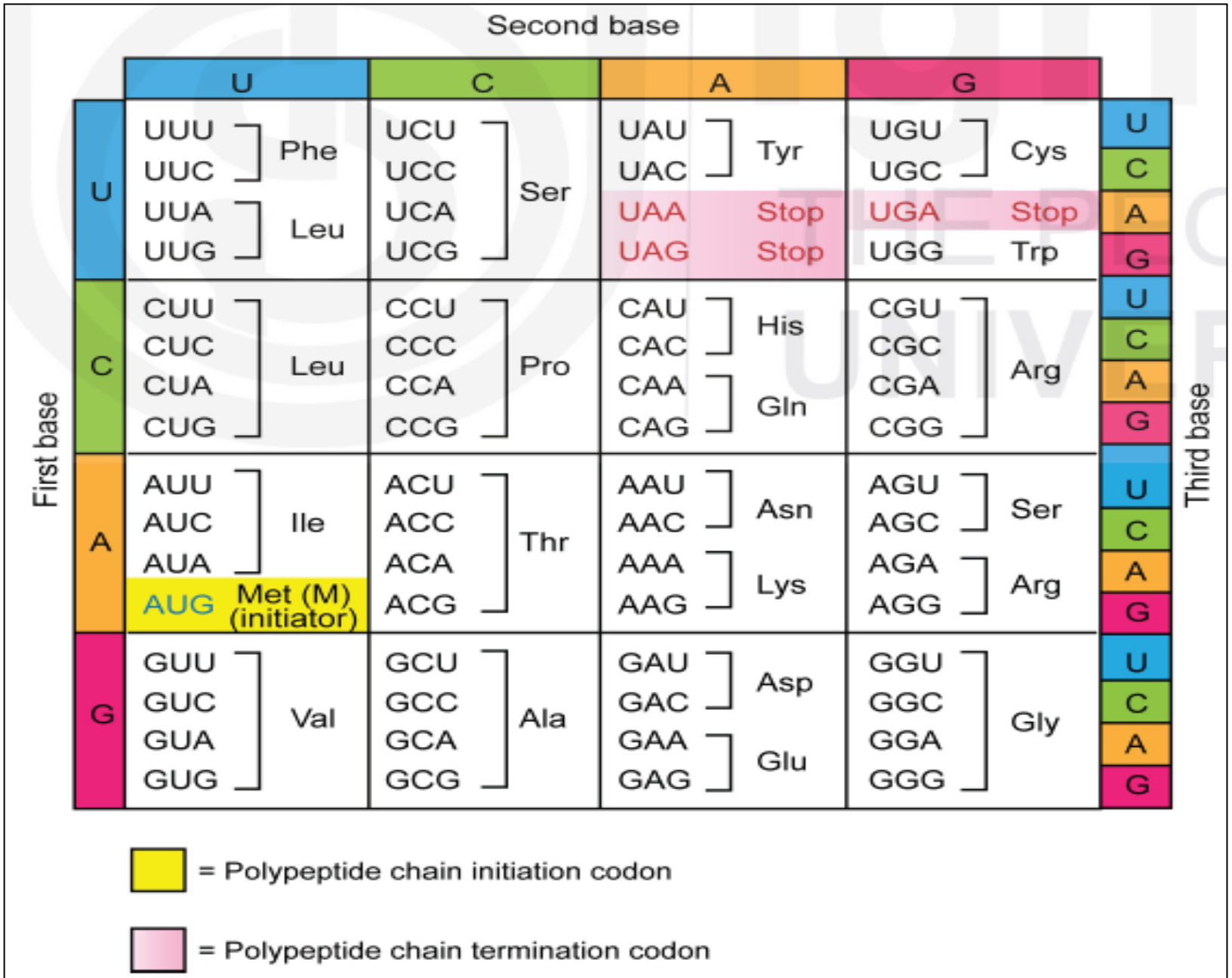

Fig. 12.25: Genetic code consisting of 64 codons. The codons are written in 5′ → 3′ as they appear in mRNA.

Genetic code is nonoverlapping and commaless

The findings of Francis Crick, Sydney Brenner, and their colleagues indicated that the genetic code is generally nonoverlapping. An overlapping code is one in which a single nucleotide may be included in more than one codon frame, as for example:

Nucleotide sequence: A U A C G A G U C

Nonoverlapping code: A U A C G A G U C

Ile Arg Val

Overlapping code: A U A U A C A C G.....

Ile Tyr Thr

Generally, each nucleotide is part of a single codon. However, a few overlapping genes are found in viruses, but codons within the same gene do not overlap, and the genetic code is considered nonoverlapping. The different way of reading the sequence of nucleotides is called a reading frame. The

reading frame is set by the initiation codon, which is the first codon of mRNA to specify an amino acid. After the initiation codon, the other codons are read as successive groups of three nucleotides. No bases are skipped between the codons: so there are no punctuation marks to separate the codons. Therefore, genetic code is **comma less/comma free**. If the DNA duplex is mutated by insertion of single base pair, the mRNA will have an additional nucleotide. This insertion alters the reading frame beyond that point, so that remainder of the mRNA is read incorrectly and all the amino acids are wrong.

The genetic code is nearly universal

The final property of the genetic code is its near universality. Except for a few cases, all organisms studied so far - prokaryotes, eukaryotes, viruses use the same basic genetic code. In other words, the 64 codons almost always stand for the same amino acid suggesting that this coding system was established early in the history of life on earth and has remained largely unchanged over billions of year of evolution. However, several exceptions to the standard genetic code do exist, most notably in mitochondria and few bacteria and other unicellular organisms (see Table 12.5). For example one difference involves the codon UGA, which is stop codon in the standard case but is translated as tryptophan in mammalian and yeast mitochondria.

Table 12.5: Some exceptions to the univarsal genetic code.

Genome	Codon	Universal Code	Altered Code
Bacterial DNA			
Mycoplasma capricolum	UGA	Stop	Trp
Mitochondrial DNA			
Human	UGA	Stop	Trp
Human	AUA	Ile	Met
Human	AGA, AGG	Arg	Stop
Yeast	UGA	Stop	Trp
Trypanosomes	UGA	Stop	Trp
Plants	CGG	Arg	Trp
Nuclear DNA			
Tetrahymena	UAA	Stop	Gln
Paramecium	UAG	Stop	Gln

Wobble Hypothesis

The cells of most organisms possess from about 30 to 50 different tRNAs which serve as adaptor molecules, binding particular amino acids and delivering them to a ribosome, where amino acids are assembled to form polypeptide chain. Since, there are only 20 different amino acids in proteins, thus some amino acids are carried by more than one tRNA. Different tRNAs that accept the same amino acid but have different anticodons are called **isoaccepting tRNAs**. Even though some amino acids have multiple (isoaccepting) tRNAs, there are still more codons than anticodons, because different codons can sometimes pair with the same anticodon through flexibility in base pairing at the third position of the codon. Figure 12.26 shows reveals that many synonymous codons differ only in the third position. For example, alanine is encoded by codons GCU, GCC, GCA, and GCG, all of which begin with GC. When the codon on the mRNA and the anticodon of the tRNA join, the first (5′) base of the codon pairs with the third (3′) base of the anticodon strictly according to Watson and Crick base pair ruling: A with U ; C with G. Next the middle bases of codon and anticodon also strictly follow Watson and Crick ruling. After these bases have hydrogen bonded, the third base pairs weakly, there may be flexibility, or wobble, in their pairing. This is called **wobble hypothesis** developed by Francis Crick which proposed that some nonstranded pairings of bases could take place at the third position of a codon (see Fig.12.26). For example, a G in the anticodon may pair with either C or a U in third position of the codon.

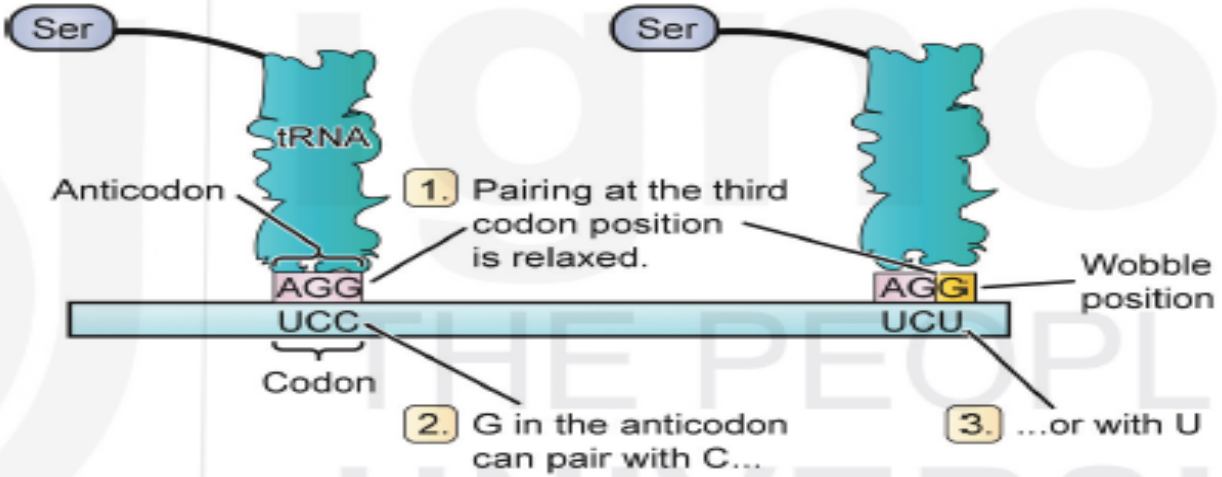

Fig. 12.26: Existence of wobble in pairing of a codon and anticodon.