Subject - M.Sc Botany, (Sem-III)

CC – 11, Topic – Genetic code

Dr. Punam Ranjan

Assistant Professor

Department of Botany

Patna Science College, Patna.

e-mail gayatripunam@gmail.com

The genetic code is a non-overlapping code, with each amino acid plus polypeptide initiation and termination specified by RNA codons composed of three nucleotides.

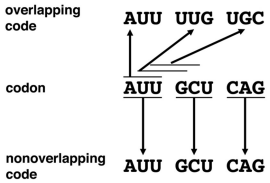
Genes controlled the structure of polypeptides, by focusing on how the sequence of the four different nucleotides in DNA could control the sequence of the 20 amino acids present in proteins. With the discovery of the mRNA intermediates, the question became one of how the sequence of the four bases present in mRNA molecules could specify the amino acid sequence of a polypeptide.

PROPERTIES OF THE GENETIC CODE: AN OVERVIEW

The main features of the genetic code were worked out during the 1960s. Cracking the code was one of the most exciting events in the history of science, with new information reported almost daily. By the mid-1960s, the genetic code was largely solved.

Properties of Genetic code are :-

1. **The genetic code is composed of nucleotide triplets**. Three nucleotides in mRNA specify one amino acid in the polypeptide product; thus, each codon contains three nucleotides.
2. **The genetic code is non-overlapping**. Each nucleotide in mRNA belongs to just one codon except in rare cases where genes overlap and a nucleotide sequence is read in two different reading frames.
3. **The genetic code is comma-free**. There are no commas or other forms of punctuation within the coding regions of mRNA molecules. During translation, the codons are read consecutively.
4. **The genetic code is degenerate**. All but two of the amino acids are specified by more than one codon.
5. **The genetic code is ordered**. Multiple codons for a given amino acid and codons for amino acids with similar chemical properties are closely related, usually differing by a single nucleotide.
6. **The genetic code contains start and stop codons**. Specific codons are used to initiate and to terminate polypeptide chains.
7. **The genetic code is nearly universal**. With minor exceptions, the codons have the same meaning in all living organisms, from viruses to humans.



Genetic code: The coding problem

General approach in understanding the genetic code (experiments)

- Experiments of rII locus of T4 Bacteriophage
- Frameshift mutations using acridine dye
- Tobacco necrosis satellite vius –RNA (1200 nt), Protein of 400 aa
- Findings - Triplet, linear/commaless, non-overlapping

Specific approach in understanding the genetic code (experiments)

- Discovery of ribosome and tRNA
- Experiments on cell free media
- Artificial mRNA and Radiolabelled Amino acids
- Experiments by Zamecnik, Nirenberg, Hogland, Metthaei, Ochoa , H. Khurana etc.

Characteristics of Genetic code

Codons are triplet

20 different amino acids are incorporated into polypeptides during translation. Thus, at least 20 different codons must be formed with the four bases available in mRNA.

This problem is solved in 1961. Francis Crick and colleagues published the first strong evidence in support of a *triplet code* (three nucleotides per codon). Crick and coworkers carried out a genetic analysis of mutations induced at the *r*II locus of bacteriophage T4 by the chemical proflavin. Proflavin is a mutagenic agent that causes single base-pair additions and deletions. Phage T4 *r*II mutants are unable to grow in cells of *E. coli* strain K12, but grow like wild-type phage in cells of *E. coli* strain B. Wild-type T4 grows equally well on either strain. Crick and coworkers isolated proflavin-induced revertants of a proflavin-induced rII mutation. These revertants were shown to result from the occurrence of additional mutations at nearby sites rather than reversion of the original mutation. Second-site mutations that restore the wild-type phenotype in a mutant organism are called suppressor mutations because they cancel, or suppress, the effect(s) of the original mutation.

Crick and colleagues reasoned that if the original mutation was a single base-pair addition or deletion, then the suppressor mutations must be single base-pair deletions or additions, respectively, occurring at a site or sites near the original mutation. If sequential nucleotide triplets in an mRNA specify amino acids, then every nucleotide sequence can be recognized or read during translation in three different ways. For example, the sequence AAAGGGCCCTTT can be read (1)

AAA, GGG, CCC, TTT, (2) A, AAG, GGC, CCT, TT, or (3) AA, AGG, GCC, CTT, T. The reading frame of an mRNA is the series of nucleotide triplets that are read (positioned in the A site of the ribosome) during translation. A single base-pair addition or deletion will alter the reading frame of the gene and mRNA for that portion of the gene distal to the mutation. This effect is illustrated below in Figure 1 (Next page). The suppressor mutations were then isolated as single mutants by screening progeny of backcrosses to wild-type. Like the original mutation, the suppressor mutations were found to produce rII mutant phenotypes. Crick and colleagues next isolated proflavin-induced suppressor mutations of the original suppressor mutations, and so on.

Crick and colleagues then classified all the isolated mutations into two groups, plus (+) and minus (-) (for additions and deletions, although they had no idea which group was which), based on the reasoning that a (+) mutation would suppress a (-) mutation but not another (+) mutation, and vice versa.

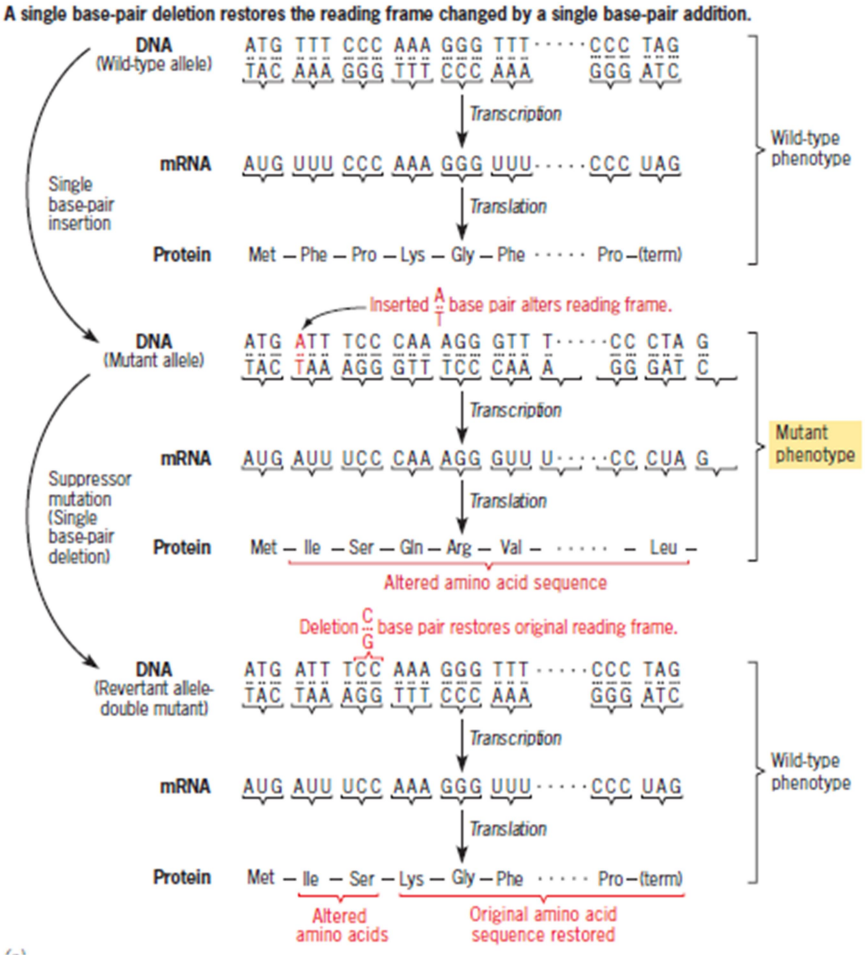


Figure 1- Genetic codes are triplet

Source- Snustad and Simmons, Principles of Genetics, 5th Edition

Experiment of Crick, indicated that the addition of three base pairs or the deletion of three base pairs left the distal portion of the gene with the wild-type reading frame. This result would be expected only if each codon contained three nucleotides.
Evidence from in-vitro translation studies soon supported the results of Crick and colleagues and firmly established the triplet nature of the code. Some of the more important results were:

(1) Trinucleotides were sufficient to stimulate specific binding of aminoacyl-tRNAs to ribosomes. For example, 5 -UUU-3 stimulated the binding of phenylalanyl-tRNAPhe to ribosomes.

(2) Chemically synthesized mRNA molecules that contained repeating dinucleotide sequences directed the synthesis of copolymers (large chainlike molecules composed of two different subunits) with alternating amino acid sequences. For example, when poly(UG)n was used as an artificial mRNA in an in vitro translation system, the repeating copolymer (cys-val)m was synthesized. (The subscripts n and m refer to the number of nucleotides and amino acids in the respective polymers.)

(3) In contrast, mRNAs with repeating trinucleotide sequences directed the synthesis of a mixture of three homopolymers (initiation being at random on such mRNAs in the in vitro systems). For example, poly(UUG)n directed the synthesis of a mixture of polyleucine, polycysteine, and polyvaline.

These results are consistent only with a triplet code, with its three different reading frames. When poly(UUG)n is translated in reading frame 1, UUG, UUG, polyleucine is produced, whereas translation in reading frame 2, UGU, UGU, yields polycysteine, and translation in reading frame 3, GUU, GUU, produces polyvaline. Ultimately, the triplet nature of the code was definitively established by comparing the nucleotide sequences of genes and mRNAs with the amino acid sequences of their polypeptide products.

DECIPHERING THE CODE: Cracking of the genetic code: Solving the Specific problem

The cracking of the genetic code in the 1960s took several years and involved intense competition between many different research laboratories. New information accumulated rapidly but sometimes was inconsistent with earlier data. Indeed, cracking the code proved to be a major challenge.

Deciphering the genetic code includes, obtaining answers to several questions.

(1) Which codons specify each of the 20 amino acids?

(2) How many of the 64 possible triplet codons are utilized?

(3) How is the code punctuated?

(4) Do the codons have the same meaning in viruses, bacteria, plants, and animals?

The answers to these questions were obtained primarily from the results of two types of experiments, both of which were performed with cell-free systems. The first type of experiment involved translating artificial mRNA molecules in vitro and determining which of the 20 amino acids were incorporated into proteins. In the second type of experiment, ribosomes were activated with mini-mRNAs just three nucleotides long. Then, researchers determined which aminoacyl-tRNAs were stimulated to bind to ribosomes activated with each of the trinucleotide messages.

The decade of the 1960s—the era of the cracking of the genetic code—was one of the most exciting times in the history of biology. Deciphering the genetic code was a difficult and laborious task, and progress came in a series of breakthroughs. By combining the results of in vitro translation experiments performed with synthetic mRNAs and **trinucleotide binding assays**, Marshall Nirenberg, Severo Ochoa, H. Ghobind Khorana, Philip Leder, and their colleagues worked out the meaning of all 64 triplet codons. **Nirenberg and Khorana shared the 1968 Nobel Prize** in Physiology or Medicine for their work on the code with **Robert Holley**, who determined the complete nucleotide sequence of the yeast alanine tRNA. Ochoa had already received the 1959 Nobel Prize for his discovery of RNA polymerase.

Below is the list of all 64 codons and their amino acids are given in table, that was the outcome of deciphering of genetic codes. The genetic codes which are degenerate and which are not universal (shows variations) are also given below.

First letter of codon (5′ end)

Second letter of codon

	U		C		A		G	
U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys
	UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys
	UUA	Leu	UCA	Ser	UAA	Stop	UGA	Stop
	UUG	Leu	UCG	Ser	UAG	Stop	UGG	Trp
C	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg
	CUC	Leu	CCC	Pro	CAC	His	CGC	Arg
	CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg
	CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg
A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser
	AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser
	AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg
	AUG	Met	ACG	Thr	AAG	Lys	AGG	Arg
G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly
	GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly
	GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly
	GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly

Degeneracy of Genetic code

Amino acid	Number of codons	Amino acid	Number of codons
Met	1	Tyr	2
Trp	1	Ile	3
Asn	2	Ala	4
Asp	2	Gly	4
Cys	2	Pro	4
Gln	2	Thr	4
Glu	2	Val	4
His	2	Arg	6
Lys	2	Leu	6
Phe	2	Ser	6

Variation in Codons

	Codons*				
	UGA	AUA	AGA AGG	CUN	CGG
Normal code assignment	Stop	Ile	Arg	Leu	Arg
Animals					
Vertebrates	Trp	Met	Stop	+	+
Drosophila	Trp	Met	Ser	+	+
Yeasts					
Saccharomyces cerevisiae	Trp	Met	+	Thr	+
Torulopsis glabrata	Trp	Met	+	Thr	?
Schizosaccharomyces pombe	Trp	+	+	+	+
Filamentous fungi	Trp	+	+	+	+
Trypanosomes	Trp	+	+	+	+
Higher plants	+	+	+	+	Trp
Chlamydomonas reinhardtii	?	+	+	+	?

*N indicates any nucleotide; +, codon has the same meaning as in the normal code; ?, codon not observed in this mitochondrial genome.

Figure Source- Lehninger, Principles of Biochemistry, 6th Edition

Suggested Books and References

1. Principles of Genetics 6th Edition by Snustad and Simmons.
2. Lehninger, Principles of Biochemistry, 6th Edition.