Santosh Murmu

4th semester Honours

C8T:MOLECULAR BIOLOGY,

Unit-4,Central dogma(adaptor hypothesis)Genetic code characteristics, deciphering of code.

Central dogma

- Deoxyribonucleic acid DNA is the genetic material of the cell, carrying information in a coded form from cell to cell and from parent to progeny. When a gene is active or expressed, it is first copied (transcribed) into another nucleic acid, RNA, which, in turn, directs the synthesis of the ultimate gene product, the protein (translation).
- Central dogma suggests the transfer of information from linear sequence of four letter alphabet of the polynucleotide chain into the twenty amino acid language of the polypeptide chain (protein).

- Process of protein synthesis is called translation. Translation of RNA into protein is unidirectional and irreversible. Proteins are polypeptide chains of 20 amino acids. The actual process of protein synthesis involves linking together of amino acids in a specific sequence of polypeptide chain.
- The genetic information exists in coded form called genetic code.

Adaptor hypothesis

- The ***adaptor hypothesis*** is part of a scheme to explain how information encoded in DNA is used to specify the amino acid sequence of proteins. It was formulated by Francis Crick in the mid-1950s, together with the central dogma of molecular biology and the sequence hypothesis.

Explanation : The adaptor hypothesis was framed to explain how information could be extracted from a nucleic acid and used to put together a string of amino acids in a specific sequence, that sequence being determined by the nucleotide sequence of the nucleic acid (DNA or RNA) template. Crick proposed that each amino acid is first attached to its own specific "adaptor"

- piece of nucleic acid (in an enzyme-catalysed reaction). The order of assembly of the amino acids is then determined by a specific recognition between the adaptor and the nucleic acid which is serving as the informational template. In this way the amino acids could be lined up by the template in a specific order. Coupling between adjacent amino acids would then lead to the synthesis of a polypeptide whose sequence is determined by the template nucleic acid.

- **Basis :**Crick's thinking behind this proposal was based on a general consideration of the chemical properties of the two classes of molecule — nucleic acids and proteins. The amino acids are characterised by having a variety of side chains which vary from being hydrophilic to hydrophobic: their individual characters reside in the very different properties these side chains have. By contrast, a nucleic acid is composed of a string of nucleotides whose sequence presents a geometrically defined surface for hydrogen bonding.

- This makes nucleic acids good at recognising each other, but poor at distinguishing the varied side chains of amino acids. It was this apparent lack of any possibility of specific recognition of amino acid side chains by a nucleotide sequence which led Crick to conclude that amino acids would first become attached to a small nucleic acid — the adaptor — and that this, by base-pairing with the template (presumably as occurs between DNA strands in the double helix), would carry the amino acids to be lined up on the template.

Proof:

- That such adaptors do exist was discovered by Mahlon Hoagland and Paul Zamecnik in 1958. These “soluble RNAs” are now called transfer RNAs and mediate the translation of messenger RNAs on ribosomes according to the rules contained in the genetic code. Crick imagined that his adaptors would be small, perhaps 5-10 nucleotides long. In fact, they are much larger, having a more complex role to play in protein synthesis, and are closer to 100 nucleotides in length.

- **Genetic codes** ,the sequence of nucleotides in deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) that determines the amino acid sequence of proteins. Though the linear sequence of nucleotides in DNA contains the information for protein sequences, proteins are not made directly from DNA. Instead, a messenger RNA (mRNA) molecule is synthesized from the DNA and directs the formation of the protein.

Characteristics of genetic code.

- **DNA Genetic Codon is Triplet:**
  - If a genetic codon consisted of two consecutive bases, the number of codons would be $4^2 = 16$. Since the number of amino acids is 20, this is insufficient. Therefore, three is the minimum number of bases needed to code for 20 amino acids $4^3 = 64$. George Gamow in 1964 pointed out that the code would contain at least three consecutive bases.
  - The length of the coding portion of a gene called reading frame depends upon the length of the message to be translated. For example, a sequence of 600 nucleotides will code for a polypeptide having a chain of 200 amino acids. Therefore, the length of mRNA depends upon the length of polypeptide it codes for.

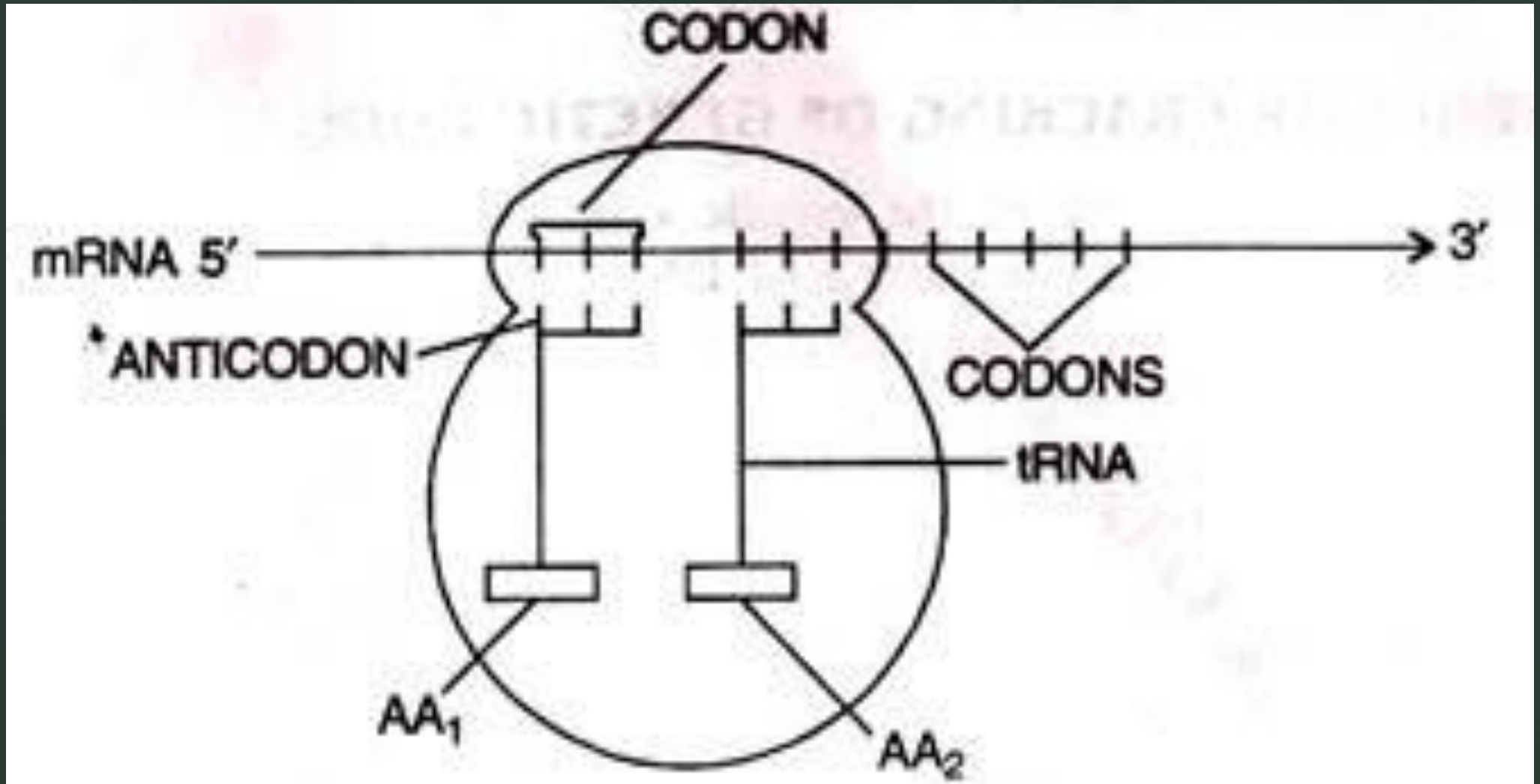


Fig. 11.1. Triplet codons on mRNA dictate the arrangement of amino acids of a polypeptide chain.

- **Genetic codon is a triplet codon:** The consecutive three nucleotides of the coding strand of DNA code for one amino acid.
- **Redundancy of the code:** Out of 64 codons, 61 codons represent amino acids, the remaining three are stop codons. As there are only 20 amino acid, coded by 61 codons, several codons specify the same amino acids. In this way the codons are synonyms
- This phenomenon is called redundancy of the code or degenerate code. Except for methionine and tryptophan all other codons are multiple codons. Each of the three amino acids — leucine, serine and arginine is represented by six different codons

- **Number of codons coding for different amino acids:**

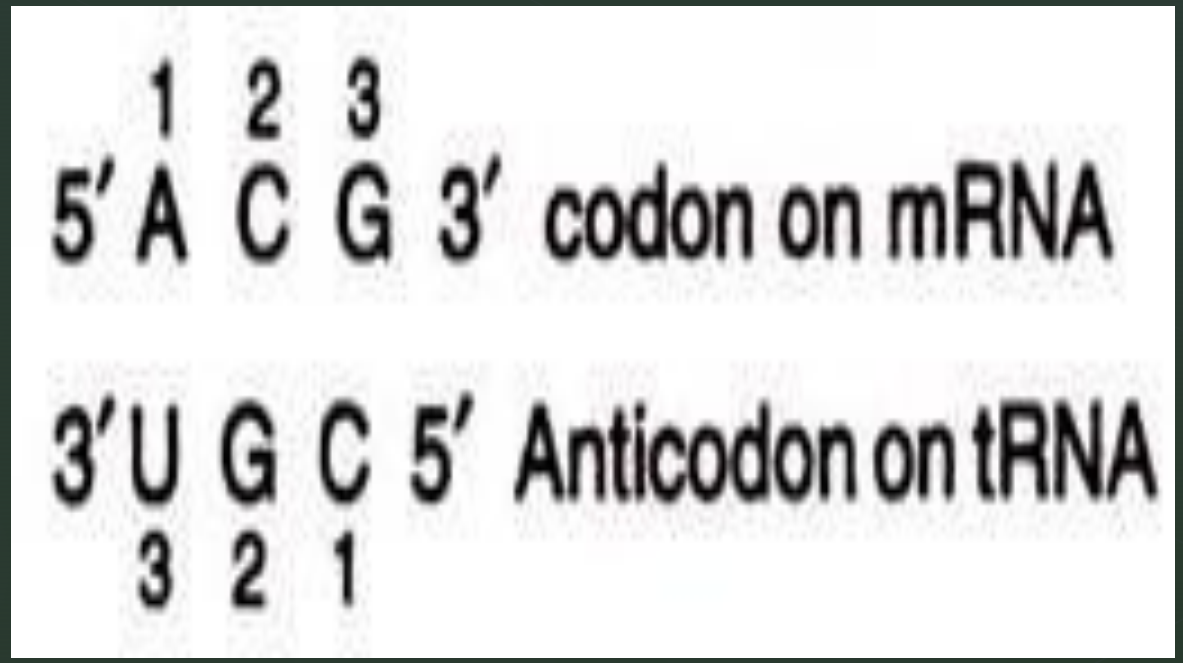
	Amino acids	Number of codons
1.	Methionine, Tryptophan	1
2.	Histidine, Glutamine, Lysine, Aspartic acid, Phenylalanine, tyrosine, asparagine, glutamic acid, cystine.	2
3.	Isoleucine	3
4.	Proline, Valine, Theonine, glycine, Alanine	4
5.	Leucine, serine, arginine	6

- **The codons are non-overlapping:** The same base cannot be a part of the two consecutive codons. They lie adjacent to each other.
- **The codons are comma less:** The three bases on DNA code for one amino acid and next three bases will code for next amino acid and so on. There is no gap or pause between the consecutive triplets.
- **Start codons:** The codons which initiate protein synthesis is called the start codon. The first amino acid of the polypeptide chain is always methionine coded by AUG codon. Therefore AUG is the start codon. Rarely the first amino acid is valine coded by GUG. In prokaryotes AUG codon codes for a modified amino acid, formyl methionine (f-Met).

- **Stop codons:**
  - Protein synthesis stops before UAA codon, UAG codon and UGA codon. This indicates that these three codons are stop codons. They terminate the protein synthesis. The completed polypeptide chain is released. Release factors (RF) enter the 'A' site of the ribosome and trigger hydrolysis of the peptidyl-tRNA occupying the 'P' site resulting in the release of newly synthesized protein.
  - Stop codons are also called non-sense codons. No tRNA can bind to these codons. UAG is called amber codon, UAA is called ochre codon and UGA is called opal codon.
- **Genetic code is universal:**
  - A particular codon codes for the same amino acid in all organisms from prokaryotes to plants and animals including viruses.

- The universality of genetic code provides strong evidence that life on the earth started from a common ancestor. When living forms appeared on the earth, the genetic code was established. It has not changed since then, throughout the evolution of living forms and has been preserved throughout the biological evolution.
- **Co-linearity:**
- The sequence of codons on mRNA and the sequence of corresponding amino acids in polypeptide chain are co-linear
- Translation of mRNA occurs in 5′ → 3′ direction.

- **Codons on mRNA and anticodons n tRNA are written as follows:** First base of the codon pairs with the third base of the anticodon. Codons are written in 5′ 3′ but anticodon sequence is written with a backward arrow.



Deciphering or Cracking of Genetic Code:

- The most important feature of the genetic code is that it is a triplet codon. Three consecutive nucleotides of a single strand of DNA contain the information for coding a specific amino acid. It is known as a triplet codon. Translation takes place in such a way that these nucleotide triplets are read in a successive non-overlapping fashion. The information is first transcribed into messenger RNA, which has a sequence of bases complementary to DNA from which it is copied.
- DNA has four types of bases C, T, G, A while RNA has four complementary bases G, A, C, U. The four base language of DNA is translated into language of 20 amino acids.
- Nirenberg and Mattaei gave the first experimental proof for the triplet codon.They used artificial system made of only uracil nucleotides (Poly U) in a cell free system.

- It resulted in the synthesis of polypeptide chain made up of only one kind of amino acid, phenylalanine. It was concluded that codon for phenylalanine was uridylic acid basis (uracil), UUU.
- Similarly poly C(CCC) codon represented amino acid proline and poly A(AAA) codon represents am: no acid chain of lysine.
- Later Hargobind Khorana confirmed the genetic code to be triplet codon. Using synthetic mRNA have alternating polynucleotides in a cell free system, discovered the chain of alternating amino acid using alternating uracil (U) and guanine (G) triplets which showed the following results.

- Similarly, alternating ACA and CAC triplets produced a chain of following amino acids.

- This also confirmed that each codon is a triplet. The cell free protein-synthesizing system was an extract of E. coli without walls. It contained ribosomes, tRNA, tRNA synthetase enzymes, ATP and radioactive amino acids. Use of artificial trinucleotide templates resulted in determination of base composition of all genetic codons.

- **Universal Genetic Code:**
- **Properties of Genetic Codons:**

64 Types of codons

First Base 5′-end	Second Base				Third Base 3′-end
	U	C	A	G	
U	UUU Phe UUC UUA Leu UUG	UCU UCC UCA Ser UCG	UAU Tyr UAC UAA Stop UAG Stop	UGU Cys UGC UGA Stop UGG Tryp	U C A G
C	CUU CUC CUA Leu CUG	CCU CCC CCA Pro CCG	CAU His CAC CAA CAG Glu	CGU CGC CGA Arg CGG	U C A G
A	AUU AUC Ileu AUA AUG Met	ACU ACC ACA Thr ACG	AAU AAC Asp N AAA AAG Lys	AGU Ser AGC AGA AGG Arg	U C A G
G	GUU GUC GUA Val GUG	GCU GCC GCA Ala GCG	GAU Asp GAC GAA GAG Glu	GGU GGC Gly GGA GGG	U C A G

Phe–Phenylalanine
Cys–Cystaine
Glu–Glutamic acid
Thr–Threonine
Gly–Glycine.

Leu–Leueine
Tryp–Tryptophan
Arg–Arginine
Val–Valine

Ser–Serine
Pro–Proline
Ileu–Isoleucine
Ala–Alanine

Tyr–Tyrosine
His–Histidine
Met–Methionine
Asp–Aspartic acid

Wobble Hypothesis:

- This was put forward by Francis Crick in 1965. According to this, hydrogen bonding between the codon of mRNA and anticodon of tRNA, there is a strict base pairing rule only for first two bases of the codon, while the base pairing involving the third base of codon appears to be less important. This is known as wobble hypothesis.
- The first two bases of each codon are primary determinants of specificity. The third base pairing is not very stable and wobbles. For example, CUU, CUG, CUC, CUA codons, which differ only at the third base represent the same amino acid leucine. The first two bases of the codon form strong base pairs with the corresponding bases of the anticodon but the third base forms weak hydrogen bond.
- At third position even unusual base pairing which does not conform to Watson and Crick base pairing rule can occur. Like adenine, cytosine and Uralic from A-I. C-I and U-I base pairs respectively at the third position, where I is inosine base.
- Several codons meant for the same amino acids are recognized by the same tRNA. In this way a minimum of 32 tRNAs are required to translate 61 codons.