GENETIC CODE

This concept pertains to the relationship between the sequence of amino acids within a polypeptide chain and the sequences of nucleotides found in mRNA or DNA molecules. The discovery of the genetic code is attributed to Nirenberg and Matthaei. Crick, in 1961, proposed that the deletion or addition of one or two bases in DNA could disrupt its functionality.

Nirenberg and Matthaei further elaborated that a single codon could potentially specify four amino acids ($4^1 = 4$), while a double codon could specify 16 amino acids ($4^2 = 16$), which is inadequate for encoding the essential 20 amino acids. However, they argued that a triplet codon could specify 64 amino acids ($4^3 = 64$), providing sufficient coverage for the full set of 20 amino acids.

George Gamow introduced the concept of the triplet codon and also coined the term "genetic code."

Features of Genetic Code

(i) Triplet Codon

The genetic code consists of triplet codons, each comprising three consecutive nitrogen bases. A codon represents a sequence of three nucleotides that dictates the incorporation of a specific amino acid into a polypeptide chain.

(ii) Start Signal or Initiation Codon

Typically, the start signals or initiation codon is AUG, which codes for the amino acid methionine. However, in prokaryotes, alternatives such as GUG and UUG (as noted by Lewin in 2000) may also serve as initiation codons. Regardless of the specific codon used, they all specify the amino acid methionine. It's worth noting that GUG and UUG code for different amino acids within the polypeptide chain, with GUG representing valine and UUG representing leucine.

(iii) Stop Signal or Termination Codon

The termination of a polypeptide chain is indicated by three termination codons: UAA (ochre), UAG (amber), and UGA (opal). These codons do not correspond to any amino acid and are therefore termed nonsense codons. Instead, they signify the end of translation and the completion of protein synthesis.

Table Assignment of mRNA codons to Amino Acids

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

(iv) Non ambiguous codon

Typically, each codon specifically designates only one particular amino acid and no other.

(v) Non overlapping code

Each nitrogen base within a DNA or mRNA molecule contributes to the composition of only one codon and does not overlap with adjacent codons.

(vi) Universal code

The genetic code is consistent across all organisms, ranging from viruses to humans. Regardless of the organism, a given codon specifies the same amino acid.

(vii) Comma less

The genetic code is devoid of pauses, allowing for continuous reading. Any addition or deletion of a nucleotide would alter the entire sequence of the genetic code.

(viii) Collinearity

The sequence of codons within DNA or mRNA molecules corresponds directly to the sequence of amino acids within a polypeptide chain.

(ix) Related codons

Codons associated with amino acids possessing similar properties exhibit similarity. For instance, aromatic amino acids like tryptophan (UGG), phenylalanine (UUC, UUU), and tyrosine (UAC, UAU) share related codons.

(x) Degeneracy of codons

While there are 64 possible triplet codons, only 20 amino acids need to be encoded. Consequently, certain amino acids are specified by more than one codon. Tryptophan (UGG) and methionine (AUG) are exceptions, each being specified by a single codon. However, the remaining amino acids can be encoded by 2 to 6 different codons, leading to the classification of these additional codons as degenerate codons.

Wobble hypothesis (crick, 1966)

Degenerate codons display a distinctive pattern where the initial two nitrogen bases remain identical, whereas the third base varies. It's intriguing that this variation in the third base doesn't alter the coding specificity. What's fascinating is that the 5’ end base of the t-RNA anticodon possesses a remarkable "wobble" ability, allowing it to form pairs with non-complementary bases of the m-RNA. For instance, codons like CCA, CCC, CCG, and CCU all correspond to the amino acid proline.

Mutations and Genetic Code

The intricate relationship between genes and DNA is most effectively elucidated through mutation studies. These investigations allow for a clearer comprehension of the effects caused by significant deletions and rearrangements within specific DNA segments. Such alterations can lead to the loss or acquisition of a gene and consequently impact its function. An archetypal instance of gene mutation, or more specifically a point mutation, is exemplified by the alteration of a single base pair within the gene responsible for the beta globin chain. This minute change results in the substitution of the amino acid residue glutamate with valine, leading to the manifestation of sickle cell anemia. Moreover, the insertion or deletion of one or two bases can disrupt the reading frame starting from the point of insertion or deletion. Conversely, the insertion or deletion of three or multiples thereof bases can either add or remove one or multiple codons, subsequently altering one or several amino acids while leaving

the reading frame unchanged from that juncture onwards. Such mutations are categorized as frame shift mutations, which fundamentally underscore the genetic evidence supporting the codon's triplet nature.

tRNA - the Adapter Molecule

Francis Crick proposed the existence of transfer RNA (tRNA), which was initially referred to as soluble RNA (SRNA) before the formulation of the genetic code. These molecules constitute approximately 15% of the total cellular RNA.

Crick hypothesized the presence of an adapter molecule capable of both deciphering the genetic code and binding to specific amino acids. This intermediary molecule bridges the triplet code of mRNA and the amino acid sequence of a polypeptide chain. Most tRNAs share a nearly identical basic structure, with over 60 distinct types identified. Kim and Klug suggested that the three-dimensional structure of tRNA resembles an inverted L-shape, which has since been confirmed.

The secondary structure of tRNA is often likened to a cloverleaf. All tRNA molecules possess a guanine residue at their 5' terminal end, while an unpaired -CCA sequence is present at their 3' end, where amino acids are attached. Each tRNA molecule is specific to a particular amino acid. For initiation, a specific tRNA known as initiator tRNA is required. Notably, there are no tRNAs for stop codons.

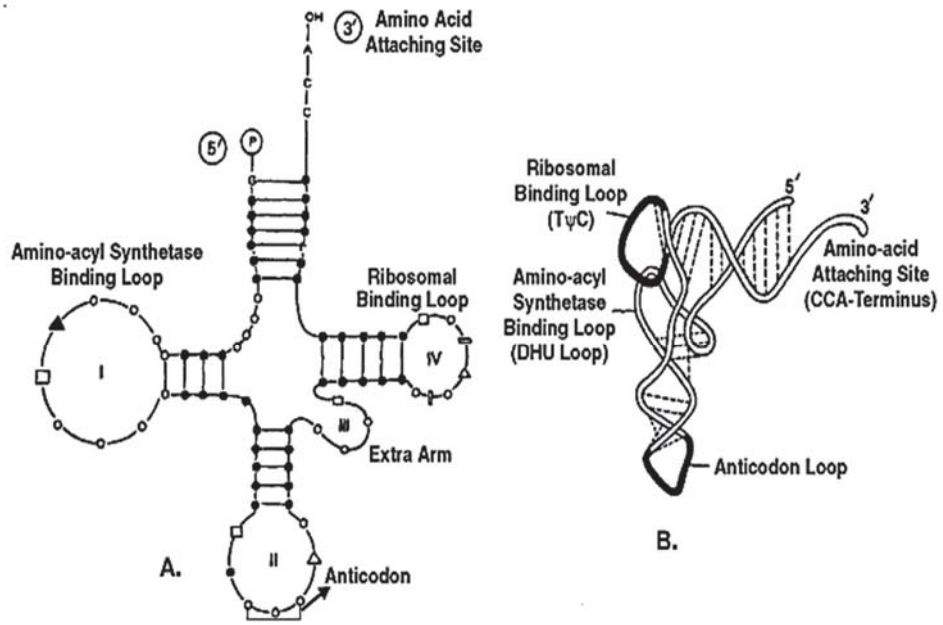

Fig. : Structure of tRNA:

A. Clover leaf model to show basic plan of tRNA secondary structure or 2D structure
B. Three-Dimensional Structure showing inverted L-shaped configuration

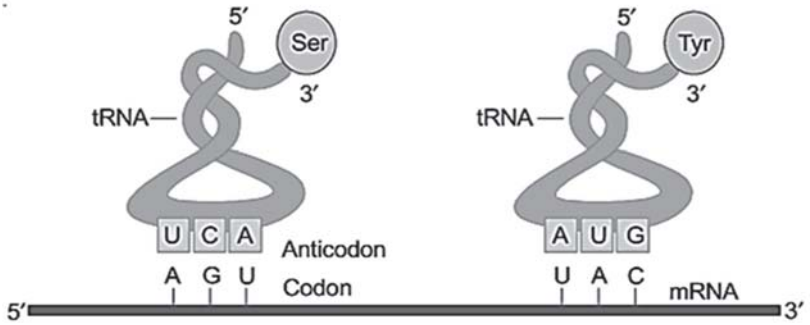

Fig. : tRNA – the adapter molecule

Transfer RNA (tRNA) consists of three distinct loops:

(i) The Aminoacyl synthetase binding loop, also known as the DHU loop (dihydroxyacridine loop), is situated as the first loop from the 5' end.

(ii) The Ribosomal binding loop, which contains 7 unpaired bases, is positioned as the first loop from the 3' end. This loop is also referred to as the TYC loop.

(iii) The Anticodon loop, also featuring 7 unpaired bases, is crucial for tRNA's function. Among these 7 bases within the anticodon loop, 3 bases serve as the anticodon, specifically recognizing and pairing with a particular triplet codon found on the mRNA molecule during translation.

Example: Suppose we have the coding strand sequence of a transcription unit as provided below:

5' - C G T A T C G A T C G G T T A C G A – 3'

To find the complementary strand sequence in the 3' to 5' direction, we follow base pairing rules where adenine (A) pairs with thymine (T) and cytosine (C) pairs with guanine (G).

Solution: The complementary strand sequence is as follows:

3'-G C A T A G C T A G C C A A T G C T - 5'