(28S, 18S, and 5.8S), whereas the RNA polymerase III is responsible for transcription of **tRNA**, **5srRNA**, and **snRNAs** (**small nuclear RNAs**). The RNA polymerase II transcribes precursor of mRNA, the **heterogeneous nuclear RNA** (**hnRNA**).

(ii) The second complexity is that the primary transcripts contain both the exons and the introns and are non-functional. Hence, it is subjected to a process called **splicing** where the introns are removed and exons are joined in a defined order. hnRNA undergoes additional processing called as capping and tailing. In **capping** an unusual nucleotide (methyl guanosine triphosphate) is added to the 5′-end of hnRNA. In **tailing**, adenylate residues (200-300) are added at 3′-end in a template independent manner. It is the fully processed hnRNA, now called mRNA, that is transported out of the nucleus for translation (Figure 6.11).

The significance of such complexities is now beginning to be understood. The split-gene arrangements represent probably an ancient feature of the genome. The presence of introns is reminiscent of antiquity, and the process of splicing represents the dominance of **RNA-world**. In recent times, the understanding of RNA and RNA-dependent processes in the living system have assumed more importance.

6.6 Genetic Code

During replication and transcription a nucleic acid was copied to form another nucleic acid. Hence, these processes are easy to conceptualise on the basis of complementarity. The process of translation requires transfer of genetic information from a polymer of nucleotides to synthesise a polymer of amino acids. Neither does any complementarity exist between nucleotides and amino acids, nor could any be drawn theoretically. There existed ample evidences, though, to support the notion that change in nucleic acids (genetic material) were responsible for change in amino acids in proteins. This led to the proposition of a genetic code that could direct the sequence of amino acids during synthesis of proteins.

If determining the biochemical nature of genetic material and the structure of DNA was very exciting, the proposition and deciphering of genetic code were most challenging. In a very true sense, it required involvement of scientists from several disciplines – physicists, organic chemists, biochemists and geneticists. It was George Gamow, a physicist, who argued that since there are only 4 bases and if they have to code for 20 amino acids, the code should constitute a combination of bases. He suggested that in order to code for all the 20 amino acids, the code should be made up of three nucleotides. This was a very bold proposition, because a permutation combination of 4^3 (4 × 4 × 4) would generate 64 codons; generating many more codons than required.

Providing proof that the codon was a triplet, was a more daunting task. The chemical method developed by Har Gobind Khorana was

instrumental in synthesising RNA molecules with defined combinations of bases (homopolymers and copolymers). Marshall Nirenberg's cell-free system for protein synthesis finally helped the code to be deciphered. Severo Ochoa enzyme (polynucleotide phosphorylase) was also helpful in polymerising RNA with defined sequences in a template independent manner (enzymatic synthesis of RNA). Finally a checker-board for genetic code was prepared which is given in Table 6.1.

Table 6.1: The Codons for the Various Amino Acids

First position	Second position U	Second position C	Second position A	Second position G	Third position
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	UGA Stop	A
	UUG Leu	UCG Ser	UAG Stop	UGG Trp	G
C	CUU Leu	CCU Pro	CAU His	CGU Arg	U
	CUC Leu	CCC Pro	CAC His	CGC 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	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

The salient features of genetic code are as follows:

(i) The codon is triplet. 61 codons code for amino acids and 3 codons do not code for any amino acids, hence they function as stop codons.

(ii) Some amino acids are coded by more than one codon, hence the code is **degenerate**.

(iii) The codon is read in mRNA in a contiguous fashion. There are no punctuations.

(iv) The code is nearly **universal**: for example, from bacteria to human UUU would code for Phenylalanine (phe). Some exceptions to this rule have been found in mitochondrial codons, and in some protozoans.

(v) AUG has dual functions. It codes for Methionine (met) , and it also act as **initiator** codon.

(vi) UAA, UAG, UGA are stop terminator codons.

If following is the sequence of nucleotides in mRNA, predict the sequence of amino acid coded by it (take help of the checkerboard):

-AUG UUU UUC UUC UUU UUU UUC-

Now try the opposite. Following is the sequence of amino acids coded by an mRNA. Predict the nucleotide sequence in the RNA:

Met-Phe-Phe-Phe-Phe-Phe-Phe

Do you face any difficulty in predicting the opposite?

Can you now correlate which two properties of genetic code you have learnt?

6.6.1 Mutations and Genetic Code

The relationships between genes and DNA are best understood by mutation studies. You have studied about mutation and its effect in Chapter 5. Effects of large deletions and rearrangements in a segment of DNA are easy to comprehend. It may result in loss or gain of a gene and so a function. The effect of point mutations will be explained here. A classical example of point mutation is a change of single base pair in the gene for beta globin chain that results in the change of amino acid residue glutamate to valine. It results into a diseased condition called as **sickle cell anemia**. Effect of point mutations that inserts or deletes a base in structural gene can be better understood by following simple example.

Consider a statement that is made up of the following words each having three letters like genetic code.

RAM HAS RED CAP

If we insert a letter B in between HAS and RED and rearrange the statement, it would read as follows:

RAM HAS BRE DCA P

Similarly, if we now insert two letters at the same place, say BI'. Now it would read,

RAM HAS BIR EDC AP

Now we insert three letters together, say BIG, the statement would read

RAM HAS BIG RED CAP

The same exercise can be repeated, by deleting the letters R, E and D, one by one and rearranging the statement to make a triplet word.

RAM HAS EDC AP

RAM HAS DCA P

RAM HAS CAP

The conclusion from the above exercise is very obvious. Insertion or deletion of one or two bases changes the reading frame from the point of insertion or deletion. However, such mutations are referred to as

frameshift insertion or **deletion mutations**. Insertion or deletion of three or its multiple bases insert or delete in one or multiple codon hence one or multiple amino acids, and reading frame remains unaltered from that point onwards.

6.6.2 tRNA– the Adapter Molecule

From the very beginning of the proposition of code, it was clear to Francis Crick that there has to be a mechanism to read the code and also to link it to the amino acids, because amino acids have no structural specialities to read the code uniquely. He postulated the presence of an adapter molecule that would on one hand read the code and on other hand would bind to specific amino acids. The tRNA, then called sRNA (soluble RNA), was known before the genetic code was postulated. However, its role as an adapter molecule was assigned much later.

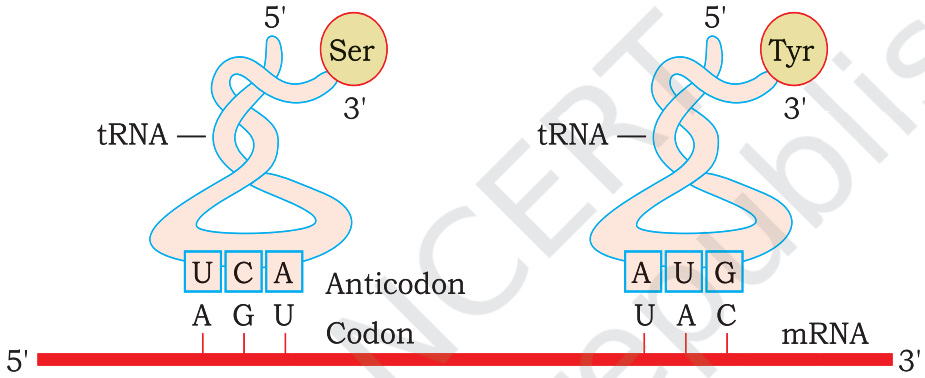

Figure 6.12 tRNA - the adapter molecule

tRNA has an **anticodon loop** that has bases complementary to the code, and it also has an **amino acid acceptor end** to which it binds to amino acids. tRNAs are specific for each amino acid (Figure 6.12). For initiation, there is another specific tRNA that is referred to as **initiator tRNA**. There are no tRNAs for stop codons. In figure 6.12, the secondary structure of tRNA has been depicted that looks like a clover-leaf. In actual structure, the tRNA is a compact molecule which looks like inverted L.

6.7 Translation

Translation refers to the process of polymerisation of amino acids to form a polypeptide (Figure 6.13). The order and sequence of amino acids are defined by the sequence of bases in the mRNA. The amino acids are joined by a bond which is known as a peptide bond. Formation of a peptide bond requires energy. Therefore, in the first phase itself amino acids are activated in the presence of ATP and linked to their cognate tRNA–a process commonly called as **charging of tRNA** or **aminoacylation of tRNA** to be more specific. If two such charged tRNAs are brought close enough, the formation of peptide bond between them

would be favoured energetically. The presence of a catalyst would enhance the rate of peptide bond formation.

The cellular factory responsible for synthesising proteins is the ribosome. The ribosome consists of structural RNAs and about 80 different proteins. In its inactive state, it exists as two subunits; a large subunit and a small subunit. When the small subunit encounters an mRNA, the process of translation of the mRNA to protein begins. There are two sites in the large subunit, for subsequent amino acids to bind to and thus, be close enough to each other for the formation of a peptide bond. The ribosome also acts as a catalyst (23S rRNA in bacteria is the enzyme- ribozyme) for the formation of peptide bond.

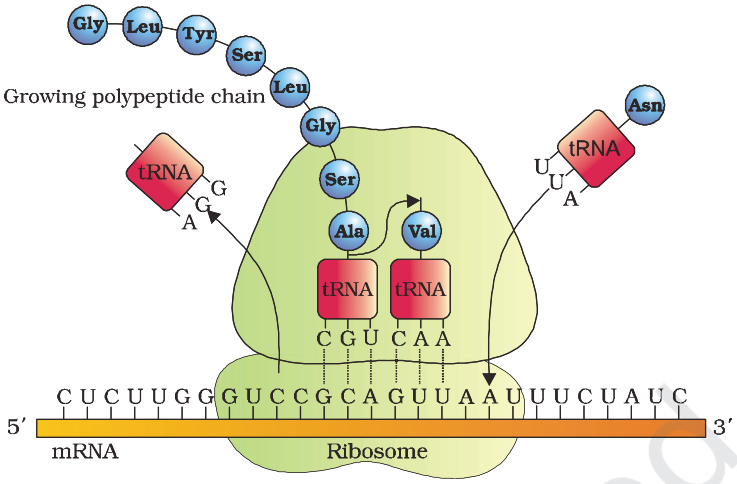

Figure 6.13 Translation

A translational unit in mRNA is the sequence of RNA that is flanked by the start codon (AUG) and the stop codon and codes for a polypeptide. An mRNA also has some additional sequences that are not translated and are referred as **untranslated regions** (**UTR**). The UTRs are present at both 5′-end (before start codon) and at 3′-end (after stop codon). They are required for efficient translation process.

For initiation, the ribosome binds to the mRNA at the start codon (AUG) that is recognised only by the initiator tRNA. The ribosome proceeds to the elongation phase of protein synthesis. During this stage, complexes composed of an amino acid linked to tRNA, sequentially bind to the appropriate codon in mRNA by forming complementary base pairs with the tRNA anticodon. The ribosome moves from codon to codon along the mRNA. Amino acids are added one by one, translated into Polypeptide sequences dictated by DNA and represented by mRNA. At the end, a **release factor** binds to the stop codon, terminating translation and releasing the complete polypeptide from the ribosome.

6.8 Regulation of Gene Expression

Regulation of gene expression refers to a very broad term that may occur at various levels. Considering that gene expression results in the formation of a polypeptide, it can be regulated at several levels. In eukaryotes, the regulation could be exerted at

(i) transcriptional level (formation of primary transcript),
(ii) processing level (regulation of splicing),
(iii) transport of mRNA from nucleus to the cytoplasm,
(iv) translational level.

The genes in a cell are expressed to perform a particular function or a set of functions. For example, if an enzyme called beta-galactosidase is synthesised by *E. coli*, it is used to catalyse the hydrolysis of a disaccharide, lactose into galactose and glucose; the bacteria use them as a source of energy. Hence, if the bacteria do not have lactose around them to be utilised for energy source, they would no longer require the synthesis of the enzyme beta-galactosidase. Therefore, in simple terms, it is the metabolic, physiological or environmental conditions that regulate the expression of genes. The development and differentiation of embryo into adult organisms are also a result of the coordinated regulation of expression of several sets of genes.

In prokaryotes, control of the rate of transcriptional initiation is the predominant site for control of gene expression. In a transcription unit, the activity of RNA polymerase at a given promoter is in turn regulated by interaction with accessory proteins, which affect its ability to recognise start sites. These regulatory proteins can act both positively (activators) and negatively (repressors). The accessibility of promoter regions of prokaryotic DNA is in many cases regulated by the interaction of proteins with sequences termed **operators**. The operator region is adjacent to the promoter elements in most operons and in most cases the sequences of the operator bind a repressor protein. Each operon has its specific operator and specific repressor. For example, *lac* operator is present only in the *lac* operon and it interacts specifically with *lac* repressor only.

6.8.1 The *Lac* operon

The elucidation of the *lac* operon was also a result of a close association between a geneticist, Francois Jacob and a biochemist, Jacque Monod. They were the first to elucidate a transcriptionally regulated system. In *lac* operon (here *lac* refers to lactose), a polycistronic structural gene is regulated by a common promoter and regulatory genes. Such arrangement is very common in bacteria and is referred to as **operon**. To name few such examples, *lac* operon, *trp* operon, *ara* operon, *his* operon, *val* operon, etc.

The *lac* operon consists of one regulatory gene (the *i* gene – here the term *i* does not refer to inducer, rather it is derived from the word inhibitor) and three structural genes (*z*, *y*, and *a*). The *i* gene codes for the repressor of the *lac* operon. The *z* gene codes for beta-galactosidase (β-gal), which is primarily responsible for the hydrolysis of the disaccharide, lactose into its monomeric units, galactose and glucose. The *y* gene codes for permease, which increases permeability of the cell to β-galactosides. The *a* gene encodes a transacetylase. Hence, all the three gene products in *lac* operon are required for metabolism of lactose. In most other operons as well, the genes present in the operon are needed together to function in the same or related metabolic pathway (Figure 6.14).