and human cells. Matthew Meselson and Franklin Stahl performed the following experiment in 1958:

(i) They grew *E. coli* in a medium containing $^{15}NH_4Cl$ (^{15}N is the heavy isotope of nitrogen) as the only nitrogen source for many generations. The result was that ^{15}N was incorporated into newly synthesised DNA (as well as other nitrogen containing compounds). This heavy DNA molecule could be distinguished from the normal DNA by centrifugation in a cesium chloride (CsCl) density gradient (Please note that ^{15}N is not a radioactive isotope, and it can be separated from ^{14}N only based on densities).

(ii) Then they transferred the cells into a medium with normal $^{14}NH_4Cl$ and took samples at various definite time intervals as the cells multiplied, and extracted the DNA that remained as double-stranded helices. The various samples were separated independently on CsCl gradients to measure the densities of DNA (Figure 6.7).

Can you recall what centrifugal force is, and think why a molecule with higher mass/density would sediment faster?

The results are shown in Figure 6.7.

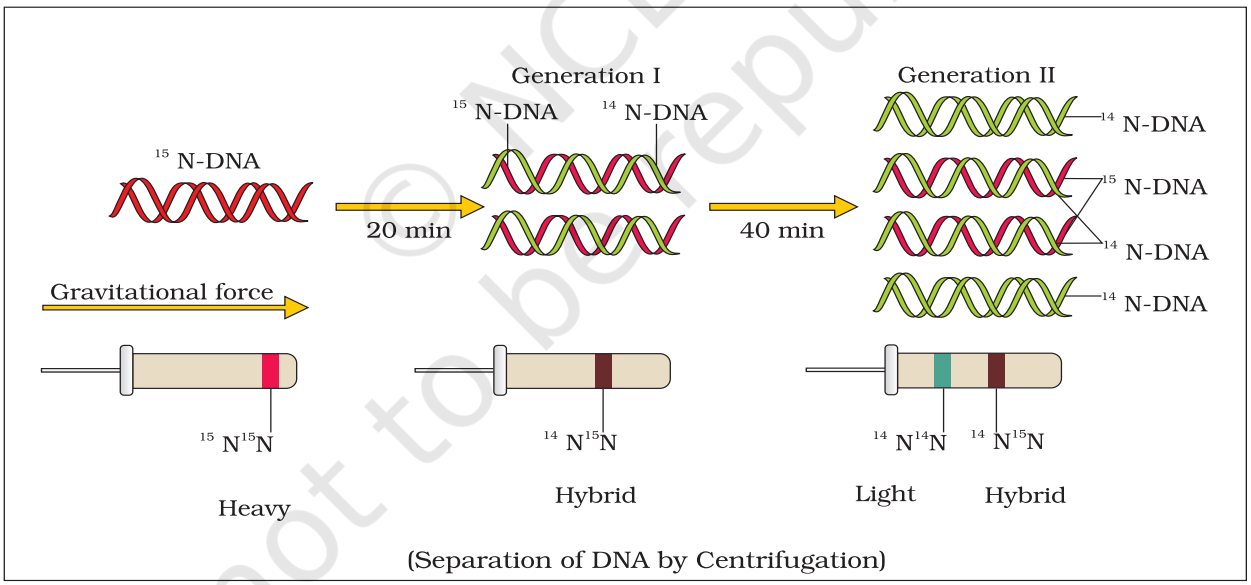

Figure 6.7 Meselson and Stahl's Experiment

(iii) Thus, the DNA that was extracted from the culture one generation after the transfer from ^{15}N to ^{14}N medium [that is after 20 minutes; *E. coli* divides in 20 minutes] had a hybrid or intermediate density. DNA extracted from the culture after another generation [that is after 40 minutes, II generation] was

composed of equal amounts of this hybrid DNA and of 'light' DNA.

If E. coli was allowed to grow for 80 minutes then what would be the proportions of light and hybrid densities DNA molecule?

Very similar experiments involving use of radioactive thymidine to detect distribution of newly synthesised DNA in the chromosomes was performed on *Vicia faba* (faba beans) by Taylor and colleagues in 1958. The experiments proved that the DNA in chromosomes also replicate semiconservatively.

6.4.2 The Machinery and the Enzymes

In living cells, such as *E. coli*, the process of replication requires a set of catalysts (enzymes). The main enzyme is referred to as DNA-dependent **DNA polymerase**, since it uses a DNA template to catalyse the polymerisation of deoxynucleotides. These enzymes are highly efficient enzymes as they have to catalyse polymerisation of a large number of nucleotides in a very short time. *E. coli* that has only 4.6×10^6 bp (compare it with human whose diploid content is 6.6×10^9 bp), completes the process of replication within 18 minutes; that means the average rate of polymerisation has to be approximately 2000 bp per second. Not only do these polymerases have to be fast, but they also have to catalyse the reaction with high degree of accuracy. Any mistake during replication would result into mutations. Furthermore, energetically replication is a very expensive process. Deoxyribonucleoside triphosphates serve dual purposes. In addition to acting as substrates, they provide energy for polymerisation reaction (the two terminal phosphates in a deoxynucleoside triphosphates are high-energy phosphates, same as in case of ATP).

In addition to DNA-dependent DNA polymerases, many additional enzymes are required to complete the process of replication with high degree of accuracy. For long DNA molecules, since the two strands of DNA cannot be separated in its entire length (due to very high energy requirement), the replication occur within a small opening of the DNA helix, referred to as **replication fork**. The DNA-dependent DNA polymerases catalyse polymerisation only in one direction, that is 5′→3′. This creates some additional complications at the replicating fork. Consequently, on one strand (the template with polarity 3′→5′), the replication is **continuous**, while on the other (the template with polarity 5′→3′), it is **discontinuous**. The discontinuously synthesised fragments are later joined by the enzyme **DNA ligase** (Figure 6.8).

The DNA polymerases on their own cannot initiate the process of replication. Also the replication does not initiate randomly at any place in DNA. There is a definite region in *E. coli* DNA where the replication originates. Such regions are termed as **origin of replication**. It is

because of the requirement of the origin of replication that a piece of DNA if needed to be propagated during recombinant DNA procedures, requires a vector. The vectors provide the origin of replication.

Further, not every detail of replication is understood well. In eukaryotes, the replication of DNA takes place at S-phase of the cell-cycle. The replication of DNA and cell division cycle should be highly coordinated. A failure in cell division after DNA replication results into polyploidy(a chromosomal anomaly). You will learn the detailed nature of origin and the processes occurring at this site, in higher classes.

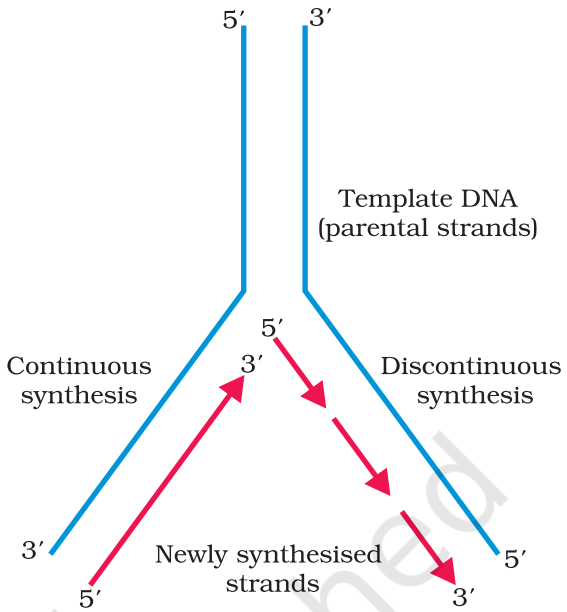

Figure 6.8 Replicating Fork

6.5 Transcription

The process of copying genetic information from one strand of the DNA into RNA is termed as **transcription**. Here also, the principle of complementarity governs the process of transcription, except the adenosine complements now forms base pair with uracil instead of thymine. However, unlike in the process of replication, which once set in, the total DNA of an organism gets duplicated, in transcription only a segment of DNA and only one of the strands is copied into RNA. This necessitates defining the boundaries that would demarcate the region and the strand of DNA that would be transcribed.

Why both the strands are not copied during transcription has the simple answer. First, if both strands act as a template, they would code for RNA molecule with different sequences (Remember complementarity does not mean identical), and in turn, if they code for proteins, the sequence of amino acids in the proteins would be different. Hence, one segment of the DNA would be coding for two different proteins, and this would complicate the genetic information transfer machinery. Second, the two RNA molecules if produced simultaneously would be complementary to each other, hence would form a double stranded RNA. This would prevent RNA from being translated into protein and the exercise of transcription would become a futile one.

6.5.1 Transcription Unit

A transcription unit in DNA is defined primarily by the three regions in the DNA:

(i) A Promoter
(ii) The Structural gene
(iii) A Terminator

There is a convention in defining the two strands of the DNA in the structural gene of a transcription unit. Since the two strands have opposite polarity and the **DNA-dependent RNA polymerase** also catalyse the polymerisation in only one direction, that is, 5′→3′, the strand that has the polarity 3′→5′ acts as a template, and is also referred to as **template strand**. The other strand which has the polarity (5′→3′) and the sequence same as RNA (except thymine at the place of uracil), is displaced during transcription. Strangely, this strand (which does not code for anything) is referred to as **coding strand**. All the reference point while defining a transcription unit is made with coding strand. To explain the point, a hypothetical sequence from a transcription unit is represented below:

3′-ATGCATGCATGCATGCATGCATGC-5′ Template Strand

5′-TACGTACGTACGTACGTACGTACG-3′ Coding Strand

Can you now write the sequence of RNA transcribed from the above DNA?

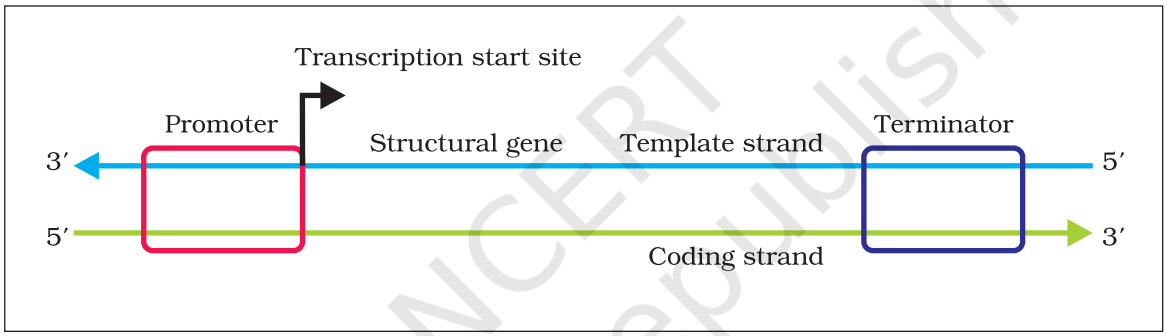

Figure 6.9 Schematic structure of a transcription unit

The **promoter** and **terminator** flank the **structural gene** in a transcription unit. The promoter is said to be located towards 5′-end (upstream) of the structural gene (the reference is made with respect to the polarity of coding strand). It is a DNA sequence that provides binding site for RNA polymerase, and it is the presence of a promoter in a transcription unit that also defines the template and coding strands. By switching its position with terminator, the definition of coding and template strands could be reversed. The terminator is located towards 3′-end (downstream) of the coding strand and it usually defines the end of the process of transcription (Figure 6.9). There are additional regulatory sequences that may be present further upstream or downstream to the promoter. Some of the properties of these sequences shall be discussed while dealing with regulation of gene expression.

6.5.2 Transcription Unit and the Gene

A gene is defined as the functional unit of inheritance. Though there is no ambiguity that the genes are located on the DNA, it is difficult to literally

define a gene in terms of DNA sequence. The DNA sequence coding for tRNA or rRNA molecule also define a gene. However by defining a **cistron** as a segment of DNA coding for a polypeptide, the structural gene in a transcription unit could be said as **monocistronic** (mostly in eukaryotes) or **polycistronic** (mostly in bacteria or prokaryotes). In eukaryotes, the monocistronic structural genes have interrupted coding sequences – the genes in eukaryotes are split. The coding sequences or expressed sequences are defined as **exons**. Exons are said to be those sequence that appear in mature or processed RNA. The exons are interrupted by **introns**. Introns or intervening sequences do not appear in mature or processed RNA. The split-gene arrangement further complicates the definition of a gene in terms of a DNA segment.

Inheritance of a character is also affected by promoter and regulatory sequences of a structural gene. Hence, sometime the regulatory sequences are loosely defined as regulatory genes, even though these sequences do not code for any RNA or protein.

6.5.3 Types of RNA and the process of Transcription

In bacteria, there are three major types of RNAs: mRNA (messenger RNA), tRNA (transfer RNA), and rRNA (ribosomal RNA). All three RNAs are needed to synthesise a protein in a cell. The mRNA provides the template, tRNA brings aminoacids and reads the genetic code, and rRNAs play structural and catalytic role during translation. There is single DNA-dependent RNA polymerase that catalyses transcription of all types of RNA in bacteria. RNA polymerase binds to promoter and initiates transcription (**Initiation**). It uses nucleoside triphosphates as substrate

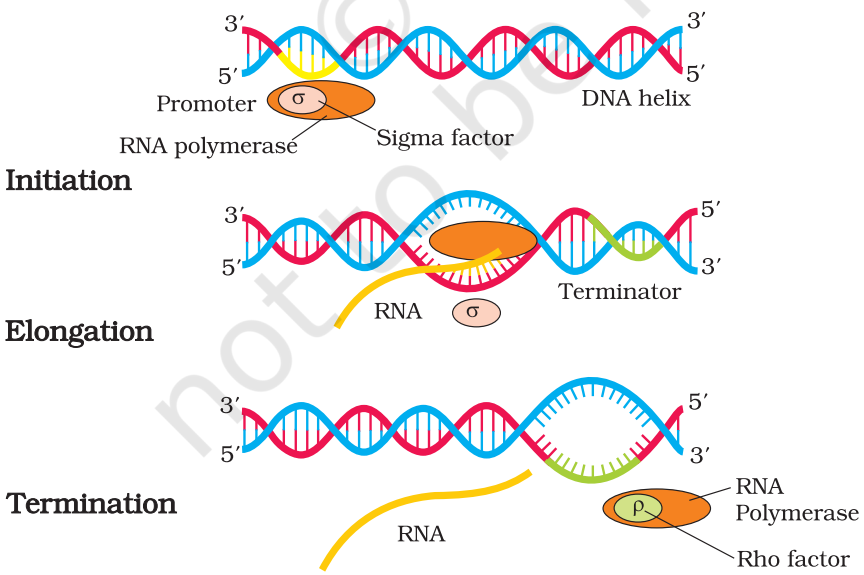

Figure 6.10 Process of Transcription in Bacteria

and polymerises in a template depended fashion following the rule of complementarity. It somehow also facilitates opening of the helix and continues elongation. Only a short stretch of RNA remains bound to the enzyme. Once the polymerases reaches the terminator region, the nascent RNA falls off, so also the RNA polymerase. This results in **termination** of transcription.

An intriguing question is that how is the RNA polymerases able to catalyse all the three steps, which are initiation, elongation and termination. The RNA polymerase is only capable of catalysing the process of elongation. It associates transiently with **initiation-factor** (σ) and **termination-factor** (ρ) to initiate and terminate the transcription, respectively. Association with these factors alter the specificity of the RNA polymerase to either initiate or terminate (Figure 6.10).

In bacteria, since the mRNA does not require any processing to become active, and also since transcription and translation take place in the same compartment (there is no separation of cytosol and nucleus in bacteria), many times the translation can begin much before the mRNA is fully transcribed. Consequently, the transcription and translation can be coupled in bacteria.

In eukaryotes, there are two additional complexities –

(i) There are at least three RNA polymerases in the nucleus (in addition to the RNA polymerase found in the organelles). There is a clear cut division of labour. The RNA polymerase I transcribes **rRNAs**

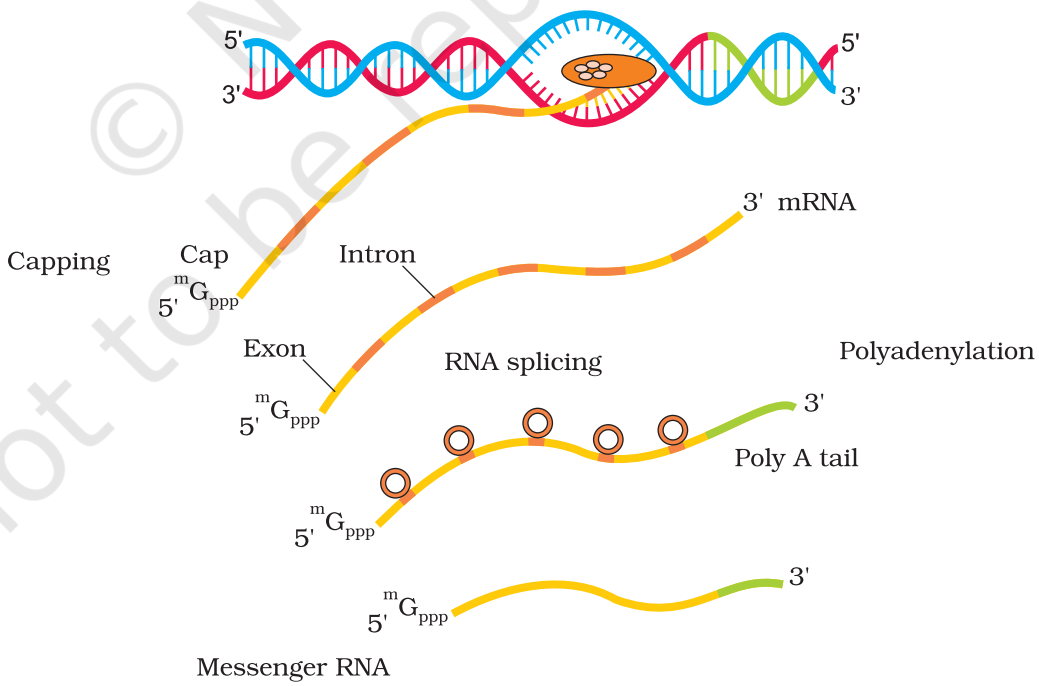

Figure 6.11 Process of Transcription in Eukaryotes

(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.

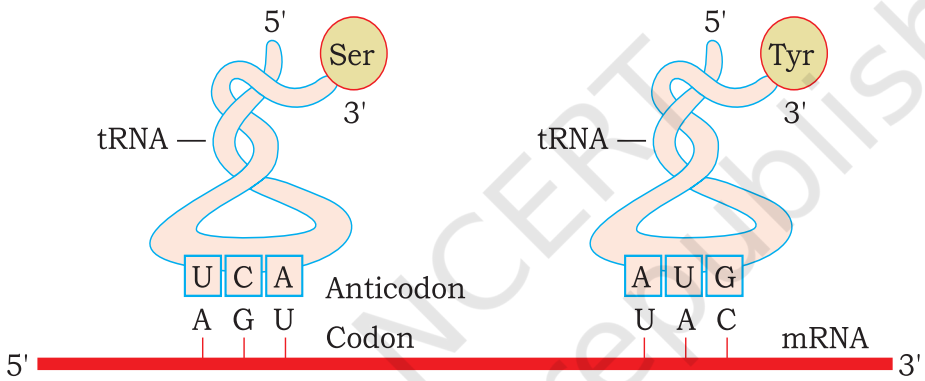

Figure 6.12 tRNA - the adapter molecule

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