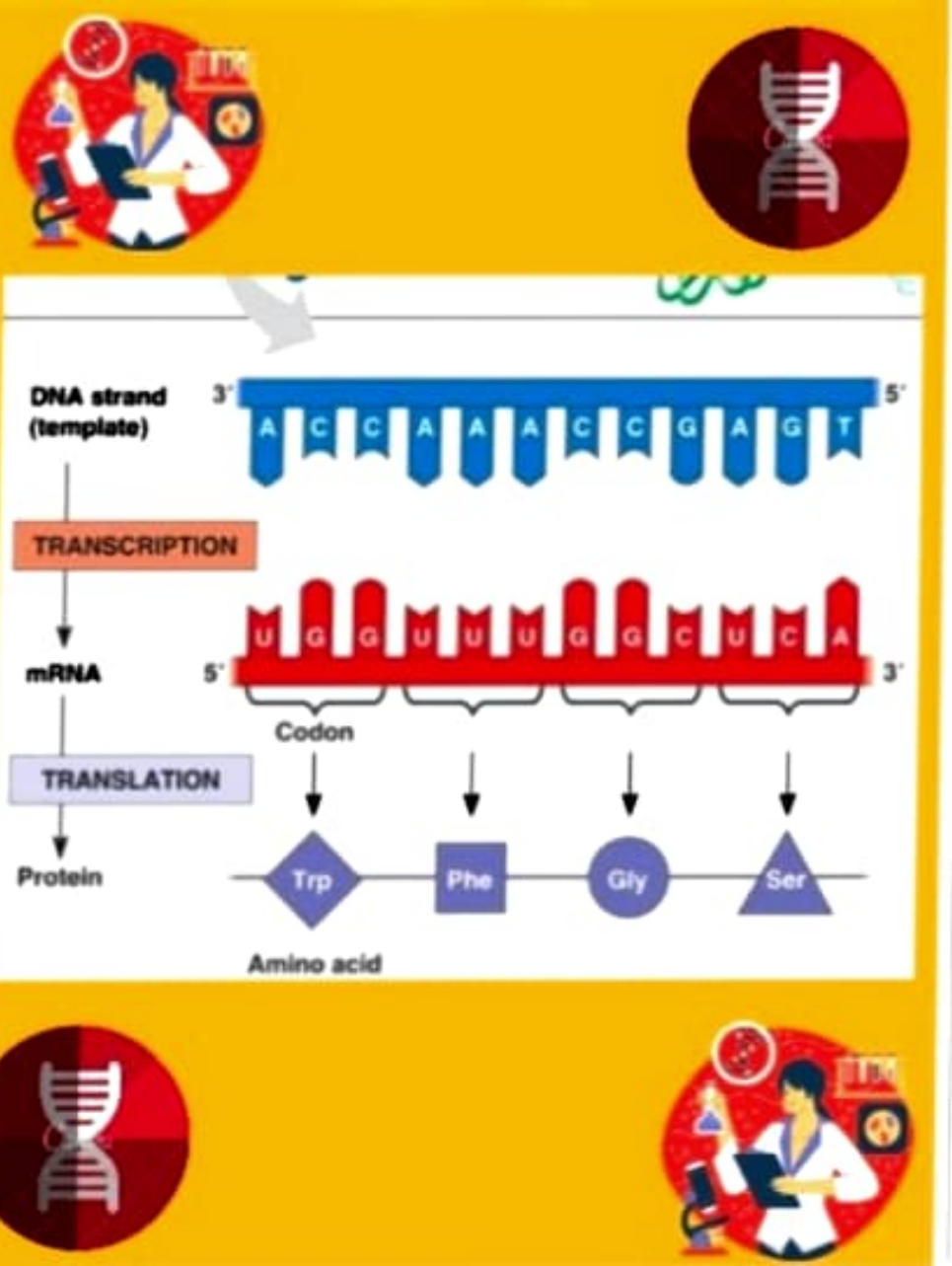
BIOLOGY



KRISHNAGAR ACADEMY

Molecular Basis Of Inheritance

Chapter - 6

CLASS - 12

TRANSLATION (PROTEIN SYNTHESIS)

Translation involves a change in language from nucleotide order in an mRNA molecule to amino acid sequence in a polypeptide. In translation, the order of nucleotides is read as triplet codons and determines the order of amino acids in the growing polypeptide chain. The process involves the following steps:

1. Activation of amino acids
2. Attachment of activated amino acids with tRNAs (Amino acetylation of tRNA)
3. Formation of polypeptide chain (Initiation, elongation and termination).
4. Modification of released polypeptide chain

Flow of Genetic Information (Central Dogma and Central Dogma Reverse)

DNA is the repository or master copy of genetic information. The single-stranded mRNA serves as an intermediate between DNA and polypeptide chain. This is called **one way flow of information** or **central dogma.** This concept was proposed by **Crick** in (1958). The genetic information flows in the following direction:

$$\text{DNA} \xrightarrow{\text{Transcription}} \text{mRNA} \xrightarrow{\text{Translation}} \text{Protein}$$

In RNA tumour viruses where RNA is hereditary material, DNA is synthesised on RNA template with the help of enzyme **RNA-dependent DNA polymerase** or **reverse transcriptase**. These viruses are called **retroviruses**. These include HIV. From this DNA, mRNA is transcribed which participates in protein synthesis. This is called **central dogma reverse**. The process of DNA synthesis on RNA is called **reverse transcription** or **inverse flow of information**. This was described by **D. Baltimore (1970)**. The flow of information can be represented as follows:

$$\text{RNA} \xrightarrow{\text{Reverse transcription}} \text{DNA} \xrightarrow{\text{Transcription}} \text{mRNA} \xrightarrow{\text{Translation}} \text{Protein}$$

Raw Material for Protein Synthesis

Following machinery and substances are required for protein synthesis:

1. **Ribosomes** as protein factories act as site of translation.
2. **Amino acids** as raw material
3. **mRNA** as messenger, carries coded information from cistron/cistrons.
4. **tRNA** as adaptor for transport of amino acids
5. **Enzymes:** (a) Amino acid activating enzymes (*amino acyl*-tRNA *synthetase*)
   (b) Peptide polymerase system
6. **Adenosine triphosphate** (ATP) as an energy source for activation of amino acids
7. **Guanosine triphosphate** (GTP) for synthesis of peptide bond
8. **Soluble protein** initiation and transfer factors
9. Various **inorganic cations** such as Mg^{2+}, K^{+} and NH^{+}

Mechanism of Translation

The process of translation or assembly of amino acids is completed in the following steps:

1. Activation of Amino Acids

In the cytoplasm amino acids occur in inactive form. Before joining to tRNA, they need to be activated. The enzyme *aminoacyl-synthetase* (or *aminoacyl-tRNA synthetase*) catalyses the union of amino acid and ATP to form a complex aminoacyl adenosine monophosphate or aminoacyl adenylate (AMP ~ AA) - enzyme complex.

1. $\underset{\text{(Amino acid)}}{\text{AA} + \text{ATP}} \xrightarrow[Mg^{2+}]{\text{Enzyme}} \underset{\text{(Aminoacyl adenylate enzyme complex)}}{\text{AA} - \text{AMP} - \text{Enzyme}} + \underset{\text{(Pyrophosphate)}}{\text{P} \sim \text{P}}$

2. $\text{P} \sim \text{P} + \text{H}_2\text{O} \xrightarrow{\textit{Pyrophosphatase}} 2\text{Pi} + \text{Energy}$

2. Attachment of Activated Amino Acid with tRNA or Charging of tRNA

The enzyme-bound activated amino acids (aminoacyl-adenylates) attach to the 3′ end of their specific tRNA molecules and form aminoacyl-tRNA. The reaction is catalysed by the same enzyme **aminoacyl synthetase** which catalyses the activation of amino acid. The aminoacyl-tRNA complexes are taken to the ribosomes for protein synthesis.

$$\text{AA} \sim \text{AMP} - \text{Enzyme} + \text{tRNA} \xrightarrow{\textit{Pyrophosphatase}} \underset{\text{(charged tRNA)}}{\text{AA} - \text{tRNA}} + \underset{\text{(Enzyme)}}{\text{E}} + \text{AMP}$$

A specific amino acid attaches to a specific ***aminoacyl tRNA synthetase*** and also to a specific tRNA molecule. There must be minumum 20 different enzymes and 20 different types of tRNA molecules in each cell so that each tRNA is charged with a specific amino acid.

tRNA charged with its cognate amino acid serves as an adaptor molecule for decoding the information of mRNA. Hence, tRNA is also called adaptor RNA. The tRNA with attached amino acid is said to be acylated or charged. The tRNA molecule without an amino acid is uncharged tRNA while with an incorrect amino acid, is mischarged tRNA. There are no tRNA for termination codons. Hence, synthesis of polypeptide chain terminates at a codon for which there is no tRNA.

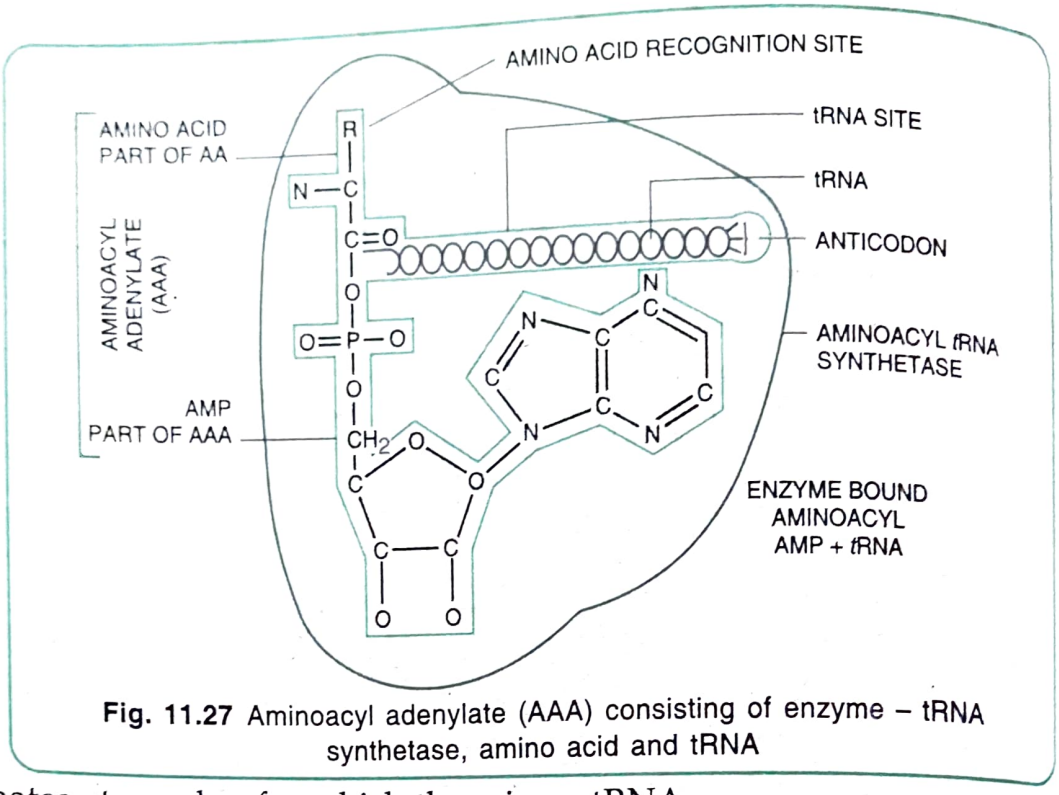


Fig. 11.27 Aminoacyl adenylate (AAA) consisting of enzyme – tRNA synthetase, amino acid and tRNA

3. Formation of Polypeptide Chain

Initiation of Polypeptide Chain in Prokaryotes (Formation of Initiation Complex)

Three initiation factors (IFs) called IF1, IF2 and IF3 bind to the small subunit of ribosome with GTP attached to IF2.

Initiator tRNA with its amino acid and mRNA binds to the small ribosomal subunit. In *E. coli* and other bacteria, the initiator tRNA is called formylatable tRNA and is respresented as $tRNA^{f\text{-}Met}$. It carries N-formylmethionine (f-Met). In prokaryotes, each polypeptide chain starts with N-formylmethionine amino acid. IF2 with its GTP helps identify the location of N-formylmethionin $tRNA^{f\text{-}Met}$ in the P-site of 30S subunit of ribosome. It is the only aminoacyl tRNA that can bind to the P site of small ribosomal subunit.

The $tRNA^{f\text{-}Met}$ which brings formylated methionine to the initiation codon is different from $tRNA^{Met}$ which brings methionine in the polypeptide chain when AUG occupies a median position on mRNA. The formylation of methionine blocks the formation of peptide bond at the amino end of the first amino acid. Consequently the peptide bond is formed between the carboxyl group of first amino acid and amino group of the second amino acid.

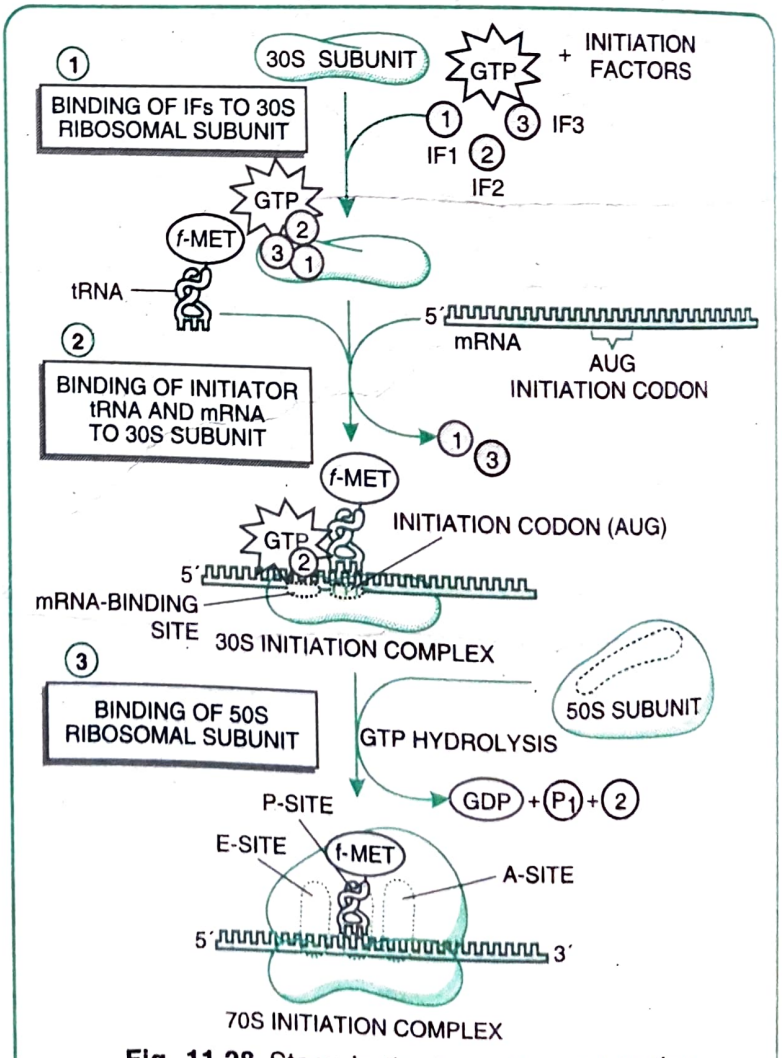


Fig. 11.28 Steps in the formation of initiation complex during initiation of polypeptide formation

The initiation factors, GTP, mRNA and 30S subunit all collectively constitute 30S preinitiation complex.

1. Anticodon of $tRNA^{f\text{-}Met}$ base pairs with the initiation codon, AUG, the first codon on the 5′ end of mRNA. AUG is brought at correct position when mRNA binds to 30S ribosomal subunit by its ribosome binding site. This completes the formation of 30S preinitiation complex.
2. **30S preinitiation complex** formed this way joins with 50S ribosomal subunit, generating **70S complex**. The energy needed for the binding of 50S subunit is provided by the hydrolysis of GTP. Mg^{2+} ions are required during this process.

In 70S initiation complex, the *f*-Met-$tRNA^{f\text{-}Met}$ occupies P-site on 50S subunit of ribosome and the anticodon on *f*-Met-tRNA pairs with the initiation codon AUG in RNA. When initiation codon is occupied by tRNA carrying formylated methionine, the second codon of mRNA lies close to A-site. The second aminoacyl tRNA with anticodon corresponding to the second codon binds with it and occupies the A-site of ribosome.

The ribosomal binding site on mRNA is also called leader nucleotide sequence or Shine-Dalgarno Sequence (after its discoverer). It consists of 3–9 purine nucleiotides (AGGAGG), located slightly upstream of initiation codon. This purine sequence base pairs with a pyrimidine rich complementary sequence on 3′ end of 16S rRNA of 30S subunit of ribosome. The 3′ end of 16S rRNA was earlier called mRNA binding site.

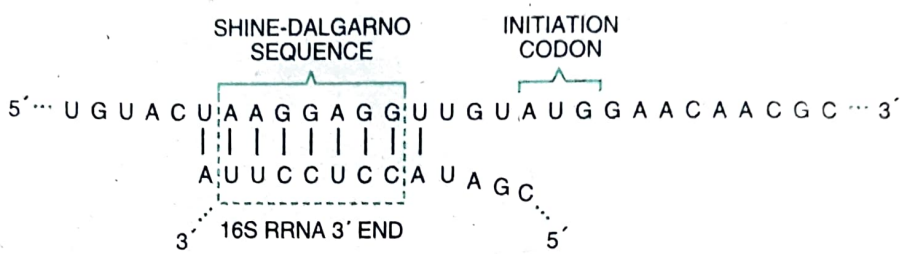


Diagram showing base pairing between Shine-Dalgarno Sequence (leader sequence) in mRNA and complementary region near 3′ terminus of 16S tRNA

4. Elongation of Polypeptide Chain

After the formation of 70S mRNA-*f*-Met-$tRNA^{f\text{-}Met}$ complex, the elongation of polypeptide chain begins by the regular addition of amino acids in following steps:

1. **Binding of AA-tRNA at Site-A of Larger Subunit of Ribosome:** The large subunit of each ribosome has two slots, called **P-site (peptidyl or donor site)** and **A-site (aminoacyl or acceptor site)**. The incoming aminoacyl-tRNA complex (AA-tRNA) attaches to the acceptor site and base-pairs with mRNA codon present in the ribosome's A-site.
2. **Formation of Peptide Bond:** The enzyme **peptidyl synthetase** or **peptidyl transferase** catalyses the formation of peptide bond between the carboxyl group (–COOH) of first amino acid and the amino group ($–NH_2$) of second amino acid present on the P- and A-sites of ribosome respectively. In bacteria, this role is played by 23S rRNA of the ribosome. It acts as ribozyme and catalyses the formation of peptide bond. This causes:
   (a) Transfer of amino acid from P-site to A-site
   (b) Formation of 2-amino acid chain (monopeptide) on A-site
   (c) Release of tRNA from P-site
3. **Translocation:** The enzyme **translocase** shifts ribosome on mRNA by one codon, so that the tRNA at A site carrying the dipeptide, moves to P site. This process is called **translocation**. GTP gets hydrolysed to GDP and provides energy for translocation. The A-site now reaches third codon of mRNA and is free to receive third amino acid corresponding to the third codon.

Thus during elongation of polypeptide chain, ribosome shifts along mRNA in 5′ → 3′ direction so that the next codon on mRNA is available at the A-site. Steps in the elongation are assisted by a number of proteins called **elongation factors (EFs)** and energy is derived from GTP.

INITIATION

An mRNA molecule binds to the small ribosomal subunit at an mRNA binding site. An initiator tRNA molecule carrying methionine then binds at the start codon "AUG". The large ribosomal subunit then binds to the small one.

The initiator tRNA is in the P site. The next codon signals another tRNA to bind. It occupies the a site. A peptide bond is formed between the amino acids in the P and A sites.

ELONGATION

The ribosome translocates three bases along the mRNA, moving the tRNA in the P site to the e site, freeing it and allowing a tRNA with the appropriate anticodon to bind to the next codon and occupy the vacant a site.

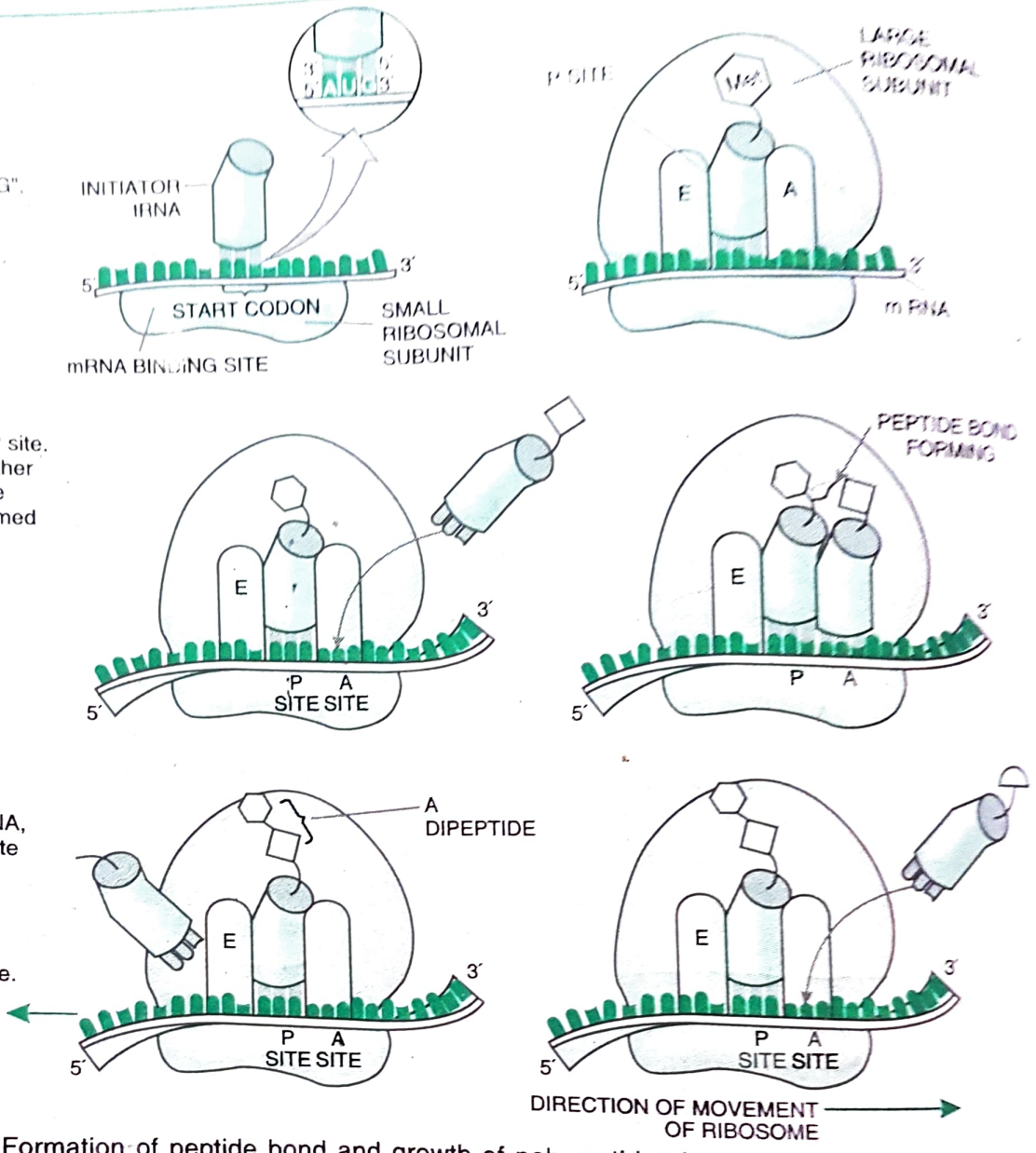


Fig. 11.29 Formation of peptide bond and growth of polypeptide chain on a ribosome

In prokaryotes, the elongation factors are called **EF-Tu, EF-Ts** and **EF-G**.

A growing polypeptide chain is called **nascent polypeptide**. It remains attached to its ribosome till released or terminated. Thus, each ribosome synthesises only one polypeptide chain at a time. After releasing its chain a ribosome dissociates into its two subunits and can start synthesis of a new peptide chain.

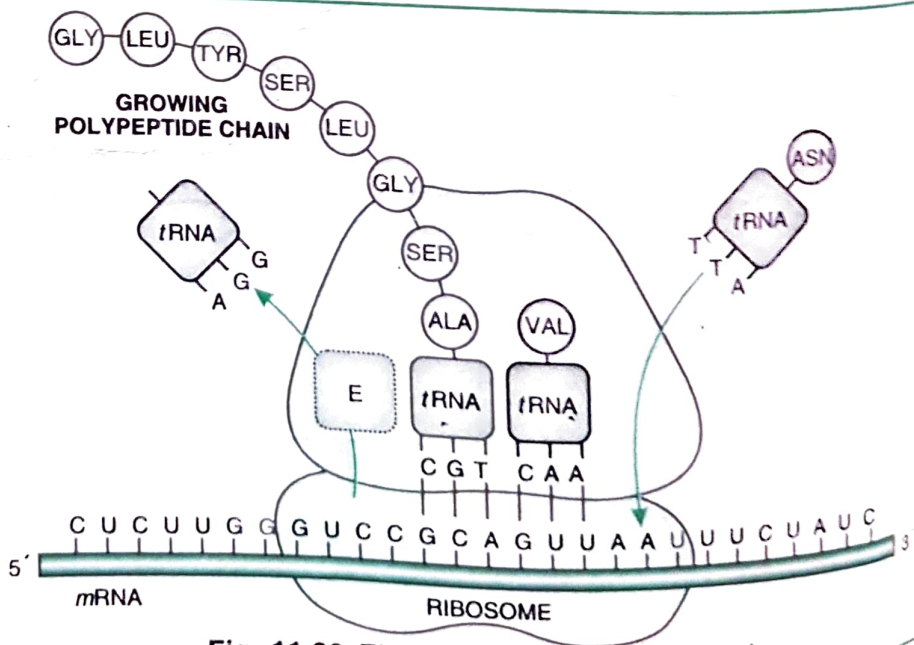


Fig. 11.30 Elongation of polypeptide chain

The whole process of arrival of charged tRNA-amino acid complex, peptide bond formation and translocation takes less than 0.1 seconds and is repeated with the addition of every amino acid. As the ribosome moves over mRNA, all the codons of mRNA come to lie on the A-site of ribosome one by one and each shift adds one amino acid to the peptide chain and the peptide chain grows. With every amino acid incorporated in the peptide chain one ATP and 2 GTP molecules are used.

Termination and Release of Polypeptide Chain

Synthesis or elongation of polypeptide chain is terminated when A-site of ribosome reaches termination codon. There are three termination codons — **UGA**, **UAG** and **UAA**. These are also called nonsense codons. Any one of them is present at the end of each cistron. There are no tRNA molecules to recognise these termination codons. A **releasing factor** (RF) is needed for the separation of polypeptide chain from the terminal tRNA. Prokaryotes have three releasing factors. These are **RF_1**, **RF_2** and **RF_3**. RF_1 is specific for UAG and UAA and RF_2 for UGA. RF_3 stimulates factors RF_1 and RF_2. The free ribosome now dissociates into two subunits with the help of dissociation factor DF or RF_3. The same mRNA can be used repeatedly for the synthesis of multiple copies of polypeptide.

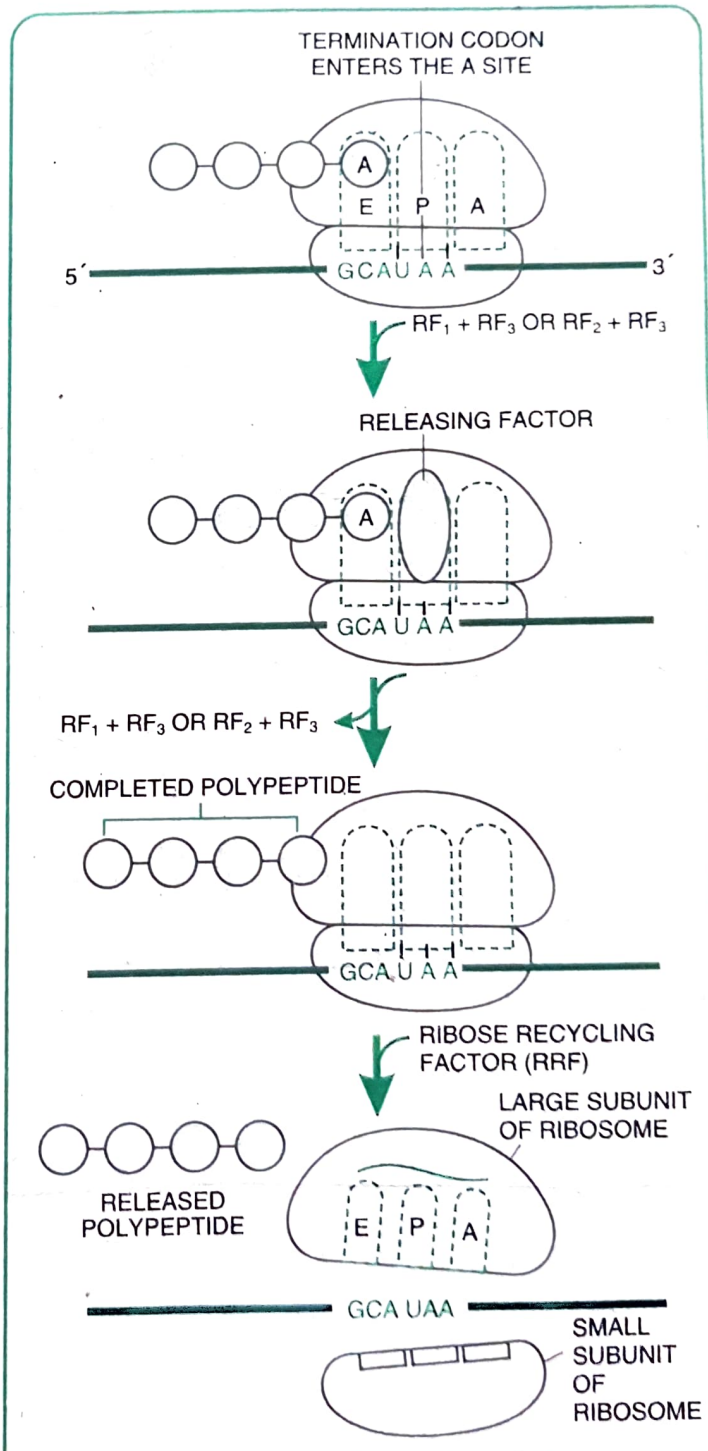


Fig. 11.31 Termination of translation and release of polypeptide chain

5. Modification of Released Polypeptide Chain or Protein Folding

The newly released polypeptide molecule undergoes following modifications:

1. The **formyl group** from the first amino acid of the polypeptide chain, the formylated methionine, is removed by the enzyme **deformylase**. Enzyme **exopeptidase** removes some amino acids either from N-terminal or C-terminal or from both.
2. The unfolded primary structure of polypeptide changes to secondary or tertiary structure by folding, so that it assumes a three dimensional structure. Special proteins called **molecular chaperones** control the correct folding.
3. The physiological activity of some proteins that act as enzymes are altered by **phosphorylation** and **hydroxylation** of some amino acids.
4. Carbohydrate molecules are attached to certain proteins to produce glycoproteins.

TABLE 11.13 Differences between Translation in Prokaryotes and Eukaryotes

Translation in Prokaryotes	Translation in Eukaryotes
1. mRNA is polycistronic.	1. mRNA is monocistronic.
2. Ribosomes are 70S type.	2. Ribosomes are 80S type.
3. First amino acid for the initiation of synthesis is formylated methionine (*f*-Met).	3. First amino acid for initiation of synthesis is methionine (met).
4. Three initiation factors IF1, IF2 and IF3 are needed for initiation.	4. About 10 initiation factors are used for initiation. These are eIF1, eIF2, eIF3, eIf4A, eIF4B, eIF4C, eIF4D, eIF5 and eIF6.
5. Elongation factors are EF-Tu and EF-Ts.	5. Elongation factor is eEF1.
6. Two releasing factors RF_1 and RF_2, are found. RF_1 for UAG and UAA termination codons and RF_2 for UGA codon.	6. Only one releasing factor-RF_1 is found for all three termination codons.
7. Transcription coupled translation occurs in prokaryotes.	7. Translation occurs only after transcription of mRNA is completed and it comes out of the nucleus into the cytoplasm.
8. Prokaryotic mRNA is functional even when it is being transcribed. It does not undergo maturation.	8. Eukaryotic mRNA, when transcribed from DNA, is heterogenous RNA and undergoes splicing and maturation to become functional.
9. mRNA has short lifespan, say only a few seconds	9. Eukaryotic mRNA has a comparatively longer lifespan (say a few hours to a few days).
10. Translation is speedy, about 20 amino acids are added per second.	10. Translation is slow, only one amino acid is added per second.

Translocation of Protein

1. In case of free ribosomes, terminated polypeptide molecules are released into the cytoplasm. They are employed in the synthesis of new cytoplasm, components of different cell organelles or are used as enzymes. Some of them are translocated to mitochondria, chloroplast and nucleus, or to the Golgi body. Inside Golgi, these may undergo glycosylation forming glycoproteins and may form cell secretions.
2. In case, ribosomes are attached to ER membranes, the protein molecules are released in the lumen of ER, but some become integral part of membranes. In ER lumen, they generally reach Golgi apparatus and form hydrolytic enzymes which may be packed in vesicles for lysosome formation.

Inhibition of Protein Synthesis

Some antibiotics inhibit protein synthesis in bacteria. The effect of various antibiotics is listed below:

- **Tetracycline:** Inhibits binding of aminoacyl tRNA to ribosome.
- **Streptomycin:** Inhibits initiation of translation.
- **Neomycin:** Inhibits interaction between tRNA and mRNA.
- **Chloramphenicol:** Inhibits action of enzyme *peptidyl transferase* and formation of peptide bonds.
- **Erythromycin:** Inhibits translocation of ribosome along mRNA.
- **Puromycin:** Binds to C–terminus of growing polypeptide causing premature termination of chain from ribosome in both eukaryotes and prokaryotes.
- **Actinomycin:** Inhibits action of RNA polymerase and synthesis of RNA.

REGULATION OF GENE EXPRESSION (REGULATION OF PROTEIN SYNTHESIS)

DNA has genetic information coded for all the proteins which may be required at different times by a cell during cell cycle. But all the genes do not function at the same time or all the times or at the same rate. It means flow of information from DNA to protein is regulated by some mechanism, i.e., genes are switched on and switched off as per requirement. In all cell types, some genes synthesise enzymes all the time. They are called **constitutive genes.** For other genes transcription of mRNA is initiated only on demand and they are switched off when need is fulfilled. This is described as **regulation of gene action** or **genetic control of protein synthesis.** Gene regulation is essential to achieve maximum cellular economy and to avoid chaos of activities.

The control on gene expression can be illustrated by the following example. In bacterium *Escherichia coli*, enzyme ***β-galactosidase*** is needed to catalyse hydrolysis of disaccharide, lactose into glucose and galactose. Therefore, this enzyme is synthesised only when lactose is present in the culture medium. When there is no lactose in the medium, ***β-galactosidase*** is not synthesised. When lactose is added to the medium, the gene associated with the synthesis of β-galactoside is switched on and when lactose in the medium is exhausted, the concerned gene is switched off.

Therefore, the metabolic, physiological or environmental conditions regulate gene expression. The development and differentiation of embryo in eukaryotes results by coordinated regulation of expression of several sets of genes at different times and at different levels. The regulation of gene expression involves two aspects:

1. Quantitative difference in the amounts of proteins/enzymes of different genes.
2. The amount of protein/enzyme produced by a gene at different times or in different tissues.

Lac Operon System

Jacob and Monod (1961) while studying catabolism of lactose in *E. coli* explained that three enzymes, namely, *β-galactosidase, lac permease* and *transacetylase* are needed for lactose catabolism. The cistrons (DNA segments) for these enzymes are represented by lac-Z, lac-Y and lac-A. These are called structural genes and are located in a linear sequence in the bacterial chromosome. The action of these genes is controlled in a coordinated fashion by regulatory genes which along with structure genes form an operon. The operon model for lactose catabolism is called lac operon. The lac operon consists of two kind of genes:

1. Operator Gene (O): It lies just before first structural gene and overlaps the promoter sequence. It controls transcription of mRNA from structural genes. It is under the control of repressor molecule, produced by the regulator or repressor gene.

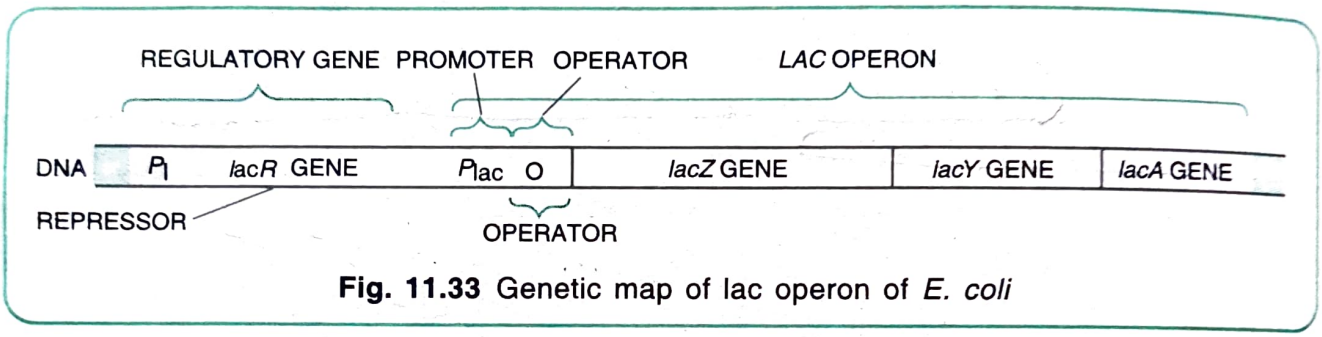


Fig. 11.33 Genetic map of lac operon of *E. coli*

2. Promoter Gene: It lies immediately adjacent to operator gene. *RNA polymerase* binds at this site. Transcription of an operon begins at promoter, the site where *RNA polymerase* binds for the transcription of mRNA from structural genes. It controls the rate of mRNA synthesis.

3. Regulator Gene: It lies outside the operon and produces a repressor substance. This repressor binds to operator gene and supresses the transcription of structural genes.

Functioning of Lac Operon: Beckwith (1967), Epstein and Beckwith (1968) and Martin (1969) have described the operation of lac operon in *E. coli* in the following ways:

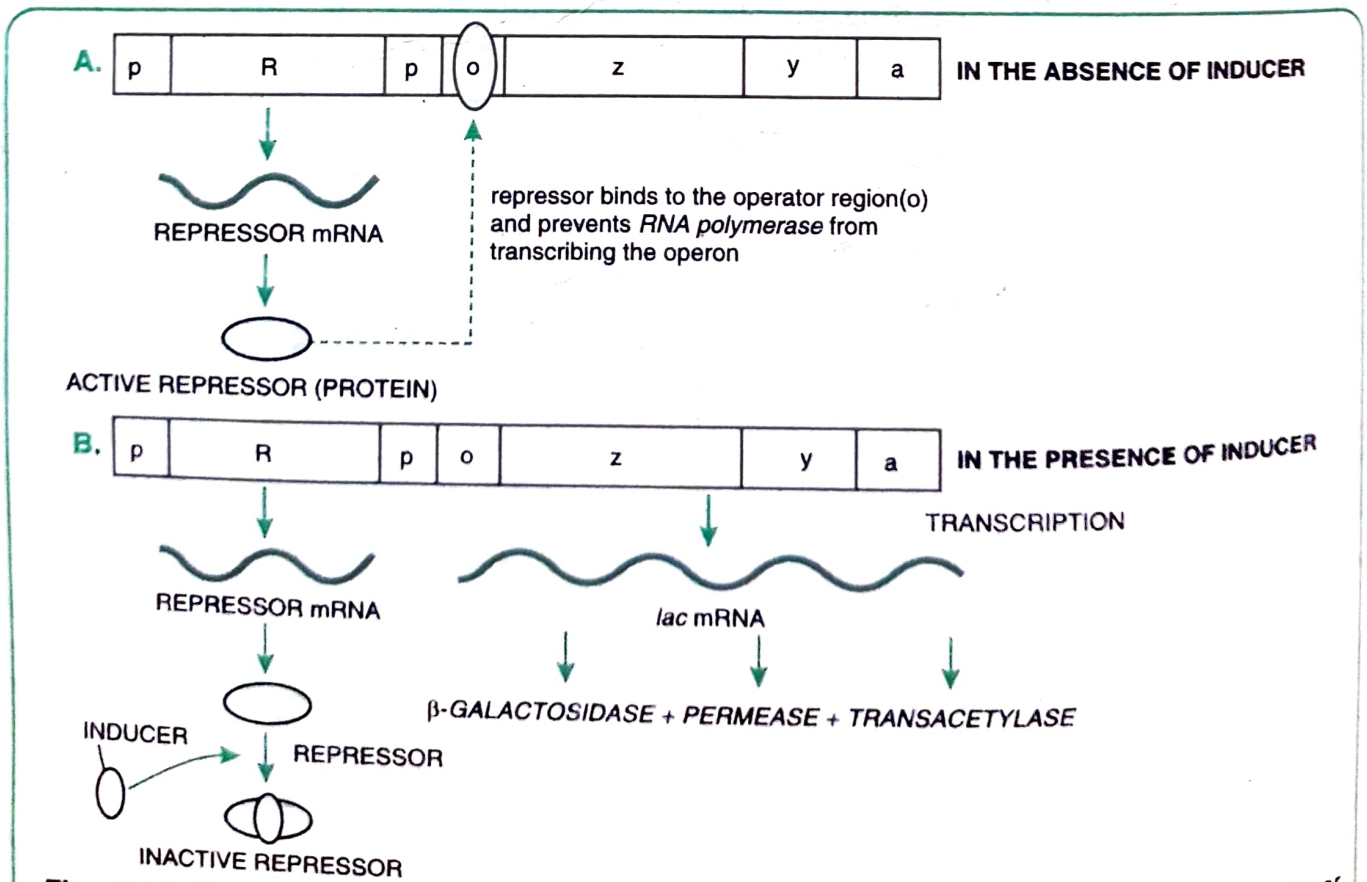


Fig. 11.34 Functioning of lac operon: A–In the absence of inducer; B–In the presence of inducer

1. In the **absence of inducer lactose,** the regulator gene-R produces a repressor protein which binds to operator site and prevents transcription of structural genes.
2. When **inducer lactose is introduced** in the medium, it binds to the **repressor substance.** The repressor fails to bind to the operator. The operator then induces the *RNA polymerase* to bind to the promoter and to transcribe mRNA by the structure genes and enzymes are produced.

mportance of Gene Regulation

1. Gene regulation is essential for growth, division, differentiation and morphogenesis.
2. Gene regulation is an economic device on the part of cell so as to synthesise only those proteins, enzymes or hormones which are needed at a particular time and in required amount only.
3. The genes for a multistep reaction can be switched on or off simultaneously.

THE HUMAN GENOME PROJECT (HGP)

The Human Genome Project, first proposed in 1986, was one of the most ambitious projects ever undertaken in the biological sciences. The goal was nothing less than the complete characterization of the genetic make-up of humans. Though the project was started in the U.S.A., it became a worldwide research effort and 17 other countries were involved directly or indirectly with the project, and these countries were Germany, France, Japan, Australia, Brazil, Canada, Denmark, Israel, Italy, Korea, Mexico, The Netherlands, Russia and Sweden. This international project involved some 100 laboratories. Powerful computers were used to store and share the enormous amount of information derived from the analyses of human DNA.

GOALS OF HUMAN GENOME PROJECT

- Identify all the genes (approx. 20,000 to 25,000) in human DNA.
- Determine the sequences of the three billion chemical base pairs that make up human DNA.
- Store this information in databases.
- Improve tools for data analysis.
- Transfer related technologies to other sectors, such as industries.
- Address the **ethical, legal, and social issues** (ELSI) that may arise from the project.

Methodologies. The methods involved two major approaches. One approach focused on identifying all the genes that are expressed as RNA [referred as **Expressed Sequence Tags** (ESTs)]. The other approach is **sequence annotation.** Here, sequencing the whole set of genome is taken that contained all the coding and non-coding sequence, and later assigning different regions in the sequence with functions. For sequencing, the total DNA from a cell is isolated and converted into random fragments of relatively smaller sizes and cloned in suitable host using specialised vectors. The cloning resulted into amplification of each piece of DNA fragment so that it subsequently could be sequenced with ease. Bacteria and yeast were two commonly used hosts and the vectors were called as **BAC** (bacterial artificial chromosomes) and **YAC** (yeast artificial chromosomes).

The fragments were sequenced using automated DNA sequences*. The sequences were then arranged based on some overlapping regions present in them, using specialised computer based programs. These sequences were subsequently annotated and were assigned to each chromosome.

The genetic and physical maps on the genome was assigned using information on polymorphism of restriction endonuclease recognition sites, and some repetitive DNA sequences (known as **microsatellites**).

Salient features of human genome. A genome is the complete collection of an organism's genetic material. Some main features of human genome are as follows.

(1) Human genome contains 3164.7 million nucleotide bases.
(2) An average gene consists of 3,000 bases; the largest known human gene being dystrophin with 2.4 million bases.
(3) Total number of genes is approx. 30,000.
(4) Almost 99.9 per cent nucleotide bases are exactly the same in all people.
(5) Functions for over 50 per cent of the discovered genes are unknown.
(6) Less than 2 per cent of the genome codes for proteins.
(7) Repeated sequences make up very large portion of the human genome.
(8) Possibly, repetitive sequences have no direct coding functions, but they shed light on chromosome structure, dynamics and evolution.
(9) Chromosome I has 2968 genes whereas chromosome 'Y' has only 231 genes.
(10) Single-base DNA differences **(SNPs - single nucleotide polymorphism,** 'snips') occur in about 1.4 million locations.
(11) Identification of 'snips' is helpful in finding chromosomal locations for disease - associated sequences and tracing human history.

The complete sequencing of the human chromosome 22, followed by sequencing of chromosome 21 was done in May 2000. The final draft was compiled from a total of 22 billion bases covering the total human DNA sequence almost seven folds by the entry of a private company Celera Genomics, Inc., into the genome project arena founded by **Craig Venter** (the pioneer of the shotgun sequencing technique). On 3rd September, 2007 **J. Craig Venter** published the whole structure of his diploid genome (named as 'HUREF-Human Reference') that created a history which was considered as the biologist's equivalence to landing on moon. Similarly, the genome of Watson was studied by HGSC.

Applications of Human Genome Project. By mapping of all of the human chromosomes, it becomes possible to examine a person's DNA and identify genetic abnormalities. This is extremely useful in diagnosing diseases and providing genetic counselling to those considering having children. This kind of information would also create possibilities for new gene therapies. Once it is known where an abnormal gene is located and how it differs in base sequence from the normal DNA sequence, steps could be taken to correct the abnormality. Many disorders of the genes, which are known as **SNP's-single nucleotide polymorphisms, Indels—Inversions and deletions, CNV-Copy Number Variation and Inversions,** are now known by human genome sequencing technique.

Besides providing clues to understanding human biology, learning about non-human organisms DNA sequences can lead to an understanding of their natural capabilities that can be applied toward solving challenges in health care, agriculture, energy production, environmental remediation.

of political power. Insurance companies may refuse to insure people at 'genetic risk'; this would save these companies the expense of future medical bills incurred by 'less-than-perfect' people. Another fear is of the attempts being made to "breed out" certain genes and people from the human population in order to create a 'perfect race.'

Besides humans, many non-human organisms, such as bacteria, yeast, *Caenorhabditis elegans* (a free-living non-pathogenic nematode), *Drosophila* (fruit fly), plants (rice and *Arabidopsis*, etc.) have also been sequenced.

DNA FINGERPRINTING

DNA fingerprinting is a very quick way to compare the DNA sequences of any two individuals. The DNA of a person is unique in much the same way as fingerprints are. There are 23 pairs of human chromosomes, with 1.5 million pairs of genes. As you are well aware that genes are segments of DNA which differ in the sequence of their nucleotides. Not all segments of DNA code for proteins; some DNA segments have a regulatory function, while others are intervening sequences (introns), and still others are repeated DNA sequences. For DNA fingerprinting **short repetitive nucleotide sequences** which are specific for a person are important. These nucleotide sequences are known as **variable number tandem repeats** (VNTR).

DNA fingerprinting involves identifying differences in some specific regions in DNA sequence called as **repetitive DNA,** because in these sequences, a small stretch of DNA is repeated many times. These repetitive DNA are separated from bulk genomic DNA as different peaks during density gradient centrifugation. The bulk DNA forms a major peak and the other small peaks are referred to a **satellite DNA**. Depending on base composition (A : T rich or G : C rich), length of segment and number of repetitive units, the satellite DNA is classified into many sub-categories, such as micro satellites, mini-satellites, etc. These sequences show high degree of polymorphism and form the basis of DNA fingerprinting.

Mechanism of DNA fingerprinting. As every cell contains DNA, extremely small amounts of blood, semen, hair bulb, or any other cells from the body of a person show the same degree of polymorphism, they become very useful identification tool in forensic applications. Further, as the polymorphisms are inheritable from parents to children, DNA fingerprinting is the basis of paternity testing, in case of disputes. The technique of DNA fingerprinting was first developed by **Alec Jeffreys** (recipient of the world's oldest science prize-Royal Society's **Copley Medal** in August, 2014). The process entails :

1. **Extraction of DNA from the sample cells.**
2. **Amplification.** Making many copies of DNA is called **amplification**. It is done by a technique called **polymerase chain reaction**.
3. **Restriction digestion.** DNA is then cut into sections by using restriction endonucleases.
4. **Separation of DNA sequences.** In this process, DNA fragments are separated by using gel electrophoresis.
5. **Southern blotting** (named after its inventor **E. M. Southern**). In this process, separated DNA sequences are transferred from gel onto a nylon membrane.
6. **Hybridisation using labelled VNTR probe.** In this process, radioactive DNA probes are attached to specific portions of the DNA fragments. Thereafter, any DNA not attached to the probes is washed off.

7. **Detection of hybridised DNA fragments by autoradiography.** The remaining DNA is then exposed to X-ray film. The radioactive probes on the DNA are allowed to expose the film which, when developed reveals a unique pattern of dark and light bands *(Fig. 38)*. These bands give a characteristic pattern for an individual DNA. It differs from individual to individual in a population except in the case of monozygotic (identical) twins. The sensitivity of the technique can be increased by use of polymerase chain reaction. Consequently, DNA from a single cell is enough to perform DNA fingerprinting analysis. Currently, many different probes are used to generate DNA fingerprints.

Applications of DNA fingerprinting. DNA fingerprinting can be used to identify individuals who may have carried out crimes such as rape, settling paternity disputes, determining relationships for immigration purposes, detecting inherited diseases and monitoring bone marrow transplants. Besides, the fingerprinting is useful in determining population and genetic diversities. In figure 38, the dark DNA bands show that both children share some of the maternal and paternal DNA.

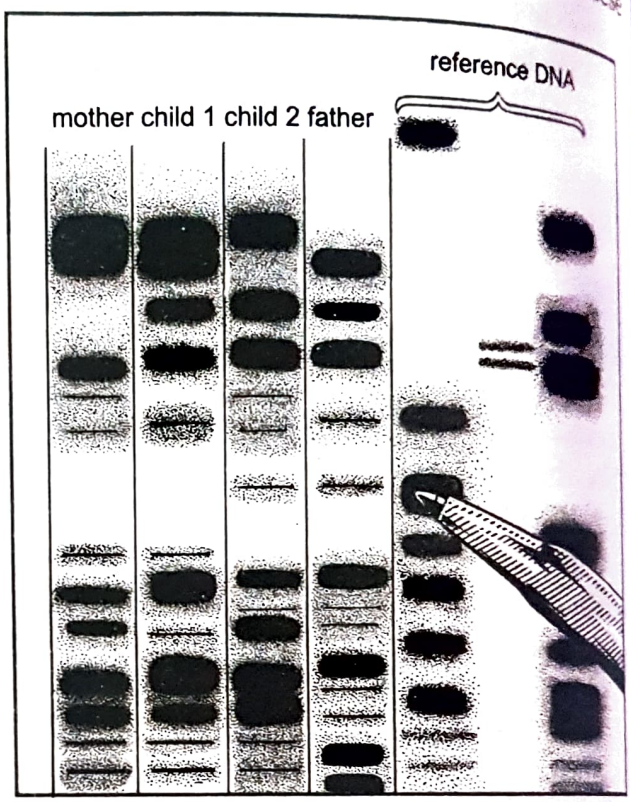


Fig. 38. DNA fingerprinting; note that the figure proves family relationships.

RICE GENOME PROJECT

Rice (*Oryza sativa*) is one of the most important crops in the world. Rice, wheat and maize together account for about half of the world's food production and rice itself is the principal food of half of the world's population. Rice is the obvious choice for the first whole genome sequencing of a cereal crop. The rice genome is well mapped and well characterized, and it is the smallest of the major cereal crop genomes at an estimated 400 to 430 Mb. The rice genome is the third largest public genome project undertaken to date, behind the human and mouse genomes.

The International Rice Genome Sequencing Project (IRGSP) began in September, 1997 at a workshop held in Singapore. Scientists from many nations attended the workshop and agreed to an international collaboration to sequence the rice genome. As a result, representatives from Japan, Korea, China, the United Kingdom and the United States met six months later in Tsukuba to establish the guidelines. The IRGSP has evolved to include 11 nations, *viz.*, Japan, Korea, United States, China, India, Taiwan, France, Brazil, Thailand, Canada and United Kingdom.

SEQUENCING, FINISHING AND ANNOTATION

The Rice Genome Research Programme (Japan) used a shotgun approach to sequence PAC or BAC clones. With this procedure, individual PAC/BAC clones (100 to 200 kb) from a sequence ready contig are shattered by sonication or nebulization, and the fragments are subcloned to produce a shotgun library with an average insert size of 1 to 3 kb. Clones from the shotgun library are then sequenced at random to provide the desired degree of 'coverage' of the total sequence. After sequencing, software such as PHRED and PHRAD is used to order the subclone sequences and reassemble the entire BAC sequence.

CONCLUSION

Sequencing of the rice genome is a monumental task. To date, ~ 3.5% of the genome has been completed (15 of 430 Mb) and another 3 to 5% is in progress. The data released provides valuable information on genome structure and organization. Repetitive DNA is estimated to constitute at least 50% of the rice genome. Complete sequencing of the rice genome will provide valuable information on the effect of repetitive elements on genome organization. It is estimated that the rice genome may contain 37,544 genes. Knowledge of the complete genetic code of rice will help breeders develop strains of the crop with specific characteristics (e.g., stress, tolerance, disease resistance of high yield) much quicker than through traditional methods, which may require years of crossing to achieve the desired property. Functional analysis of rice genomics would be a challenge for rice scientists. India, being a major rice producer and consumer, needs to take a lead in the functional genomics of rice.

SHORT/LONG ANSWER TYPE QUESTIONS

1. What is the general idea behind the following expression?
   DNA ⟶ RNA ⟶ Protein.
2. In which direction 5'-3' or 3'-5' are the new strands of DNA formed during replication?
3. In 1928, the bacteriologist Frederick Griffith conducted transformation experiments. How did these experiments contribute to the history of molecular genetics?
4. A segment of DNA, GCCAGGGGATG was translated into the oligopeptide arginine-serine- proline-tyrosine.
   (a) What was the base sequence in the mRNA transcribed from the DNA segment?
   (b) What are the codons for the 4 amino acids?
   (c) If the first adenine in the DNA segment gets substituted by guanine, what will be the :
      1. mRNA transcript?
      2. sequence of amino acids in the new oligopeptide?
      3. anticodons on tRNA for the amino acids?
5. A hypothetical mRNA, AUG CGC CUA AAG AGG codes for five amino acids. What will happen if you delete the first 'C'? Will five amino acids still be coded for? Give reasons.
6. The genetic code is said to be a triplet code. What does this mean and why is it important?
7. What properties of the bacteriophages made them ideal material for solving the controversy that which is the genetic material, protein or DNA?
8. How did the Meselson-Stahl experiment confirm that DNA replication is semi-conservative?
9. How are errors that occur during DNA replication repaired?
10. How do a code, codon, and anticodon differ?
11. A tRNA molecule has four active sites. Where are these sites and what does each do?
12. What is the difference between the P-site and A-site of a ribosome?
13. Why are there 64 different codons for 20 different amino acids?
14. What are the evidences that indicate DNA is the primary genetic material?
15. What is a codon? How many codons serve as stop (nonsense) codons and how many as initiating codons?
16. In genetics, a reference is made to an abbreviated expression 'AUG'. Write any three points of scientific information embodied in this combination of three letters.
17. Write about the transfer RNAs as follows :
   (a) Any two characteristics in their structure.
   (b) Part of the cell where they are located.
   (c) The technical name for the set of bases that are complementary to the triplet code on the mRNA.
18. Write the transcribed mRNA of a DNA strand with the base sequence GAT CAT ACT. What is the name and the specific role of the last codon of the transcribed mRNA is this case?
19. The base sequence on one of the strands of DNA is TAC TAG GAT.
   (i) Give the base sequence of its complementary strand.
   (ii) What is the distance maintained between the two consecutive paired bases in the DNA molecule?
   (iii) Who contributed the base complementarity rule ?
   (iv) Categorise these bases using their full names.
   (v) What holds these base pairs together ?

Q. 20. Suppose during transcription of DNA code AAA, a mistake occurs due to which UUG code of mRNA is formed. Due to this what change in picking the types of amino acid would occur during synthesis of protein ?

Q. 21. How is the wrong base removed before proceeding to add new bases in the 5'-3' direction during DNA replication?

Q. 22. A U G G A C C U G A U A U U U U G A is the base sequence in a strand of mRNA.
   (i) Write the base sequence of the DNA strand from which it has been transcribed.
   (ii) Upon translation, how many amino acids will the resulting peptide have?

Q. 23. A mRNA strand has a series of codons out of which three are mentioned below : (i) AUG, (ii) UUU, and (iii) UAG.
   (a) What will these codons be translated into?
   (b) What are the DNA codons that would have transcribed these RNA codons?

Q. 24. A tRNA is charged with the amino acid phenylalanine.
   (a) At what end of tRNA is this amino acid attached?
   (b) What is the mRNA codon that coded for phenylalanine?
   (c) What is its anticodon?
   (d) Name the enzyme responsible for this attachment.

Q. 25. The DNA is contained within the nucleus of a eukaryotic cell, the proteins for which it codes are needed within the cytoplasm. Explain the roles of the following in the translation of the genetic code into an active enzyme in the cytoplasm :
   DNA, mRNA, tRNA, rRNA, polysomes.

Q. 26. What is the central dogma in molecular biology? Name the two processes represented in it. How has it been modified with the discovery of retroviruses?

Q. 27. Give an account of Hershey-Chase experiment proving that DNA and not the protein coat of the virus is the infecting agent. How is transformation different from transduction?

Q. 28. Give a brief account of the different steps involved in the translation of mRNA into a polypeptide in prokaryotes.

Q. 29. (a) Explain the experiment performed by Griffith on *Streptococcus pneumoniae*. What did he conclude from this experiment?
   (b) Name the three scientists who followed up Griffith's experiments.
   (c) What did they conclude and how?

Q. 30. What are introns and exons ? What ensures a linear arrangement of amino acids although the genes are discontinuous ?

Q. 31. In *E.coli*, three enzymes β-galactosidase, permease and transacetylase began to be produced as soon as lactose was added. Explain why the enzymes were not forming in the absence of lactose.

Q. 32. What is an operon ? How does an excess of tryptophan cause a 'switching-off' of the tryptophan operon ?

Q. 33. How is the functioning of *lac* operon different from that of *try* operon ?

Q. 34. All the cells in a multicellular organism have the same genetic constitution, yet they function differently. How would you account for this ?

Q. 35. Explain the following terms in connection with the concept of *lac* operon in *E.coli* : (a) structural genes, (b) operon, (c) cistron, (d) polycistron, and (e) repressor.

Q. 36. What is 'reverse transcription' ? Mention the term used for the group of viruses which show this kind of transcription.

Q. 37. Distinguish between structural gene, regulatory gene and operator gene.

Q. 38. What acts as an inducer in *lac* operon ? How does it switch on the operon ?

Q. 39. Explain how structural genes of *lac* operon start transcription on addition of lactose.

Q. 40. Differentiate between induction and repression. When and in what systems do the phenomena of induction and feedback repression operate in a cell system ?

Q. 41. What is reverse transcription ? Explain how single stranded RNA of viruses give rise to double stranded DNA.

Q. 42. What is 'feedback inhibition' ? Compare it with the mechanism of regulation based on operon model.

Q. 43. Mention the differences between the mechanisms of the regulation of gene activity in *lac* operon and *try* operon in *E. coli.*

Q. 44. How has the sequencing of human genome opened new windows for treatment of various genetic disorders?

Q. 45. A low level of expression of *lac* operon occurs at all the time. Can you explain the logic behind this phenomenon?

Ans. In the complete absence of expression of *lac* operon, permease will not be synthesized which is essential for transport of lactose from medium into the cells. And if lactose cannot be transported into the cell then it cannot act as inducers hence cannot relieve the *lac* operon from its repressed state.

BOARD EXAMINATION QUESTIONS

Q. 1. What is genetic code? Mention the essential qualities for a universal genetic code. ***(2015)***

Q. 2. How did Hershey and Chase prove that DNA is the genetic material? ***(2014)***

Q. 3. Describe the experiment performed by Griffith. What conclusions did he infer from his observations? ***(2014)***

Q. 4. Explain the mechanism of transcription in a prokaryotic cell. ***(2017)***

Q. 5. Describe Avery, MacLeod and and McCarty's experiment. State its significance. ***(2018)***

Q. 6. Describe the structure of a nucleosome, with the help of a well-labelled diagram. ***(2019)***