An important characteristic of the polynucleotide strand is its direction, or polarity. At one end of the strand, a phosphate group is attached to the 5'-carbon atom of the sugar in the nucleotide. This end of the strand is therefore referred to as the 5' end. The other end of the strand, referred to as the 3' end, has an OH group attached to the 3'-carbon atom of the sugar.

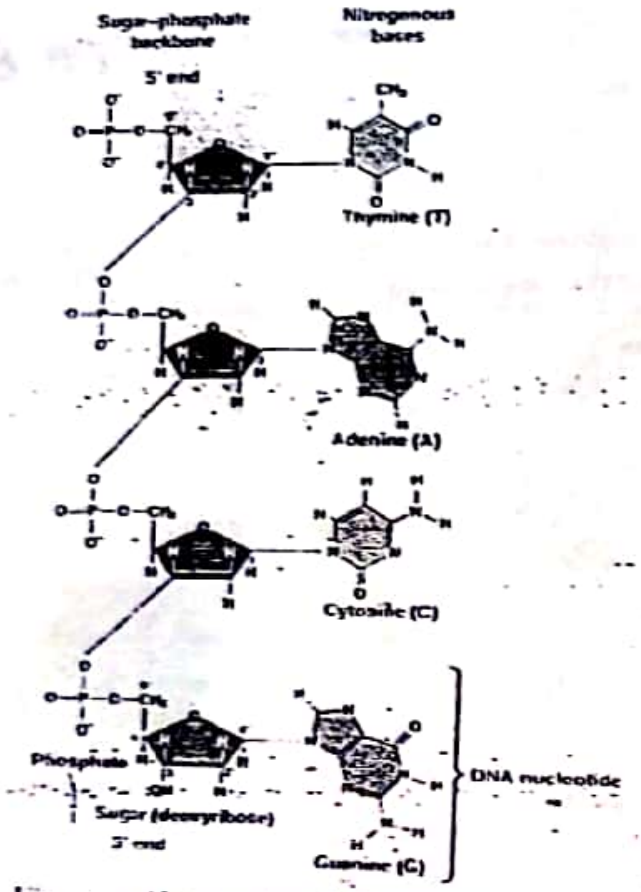


Figure 4: Nucleotide chain in DNA

Secondary structure of DNA

DNA is not a single molecule, but a pair of molecules joined complementarily by hydrogen bonds. The pair of polynucleotide chains describes the secondary structure of DNA.

A fundamental characteristic of DNA's secondary structure is that it consists of two polynucleotide strands wound around each other, forming a double helix. The sugar–phosphate linkages are on the outside of the helix and the bases are stacked in the interior of the molecule.

The details of this double helical structure were provided by James Watson and Francis Crick in 1953, when they published a paper in the British journal *Nature*. (J. D. Watson and F. H. C. Crick, *Molecular Structure of Nucleic Acids: A Structure for Deoxynucleic Acids*. Nature 171 (1953): 738). The double stranded helical structure of DNA is also called the Watson and Crick Model.

The double helical structure of DNA is based on the following experimental findings of other scientists.

1. The base equivalence rule of Erwin Chargaff: Chargaff and co-workers discovered that in all organisms the total amount of adenine is always equal to the amount of thymine (A = T), and the amount of guanine is always equal to the amount of cytosine (G = C). In other words: A+G = T+C (Total purines = Total pyrimidines).

2. X-ray diffraction data of DNA from Maurice Wilkins and Rosalind Frank which showed that DNA is a right handed helical molecule.

3. Tautomeric data on the nitrogenous base from Jerry Donohue at the Cold Spring Harbour Lab showed that *in vivo* Adenine and Cytosine stay mostly in their Amino tautomeric form and Thymine and Guanine remain mostly in their Keto tautomeric form. This is an important determinant of base pairing.

The double helical DNA structural model of Watson and Crick says that:

1. The two polynucleotide chains in the double helix associate by *hydrogen bonding between the complementary nitrogenous bases*. In its *in vivo* tautomeric form, G can hydrogen bond specifically only with C, while A can bond specifically only with T.

2. Two hydrogen bonds stabilize the base pairing between A and T, while three hydrogen bonds stabilize the base pairing between G and C. Therefore, a GC base pair is difficult to separate.
3. The two polynucleotide chains to run in opposite directions (antiparallel).
4. The bases lie on the inside. They are flat structures, lying in complementary pairs perpendicular to the axis of the helix.
5. Adjacent bases in a single strand are separated by 3.4 Å. The helical structure repeats every 34 Å, so there are 10 bases (= 34 Å per repeat/3.4 Å per base) per turn of helix. *However, a number of recent studies indicate that in vivo there are 10.4 nucleotides per complete turns, at least in eukaryotic cells.*
6. Each base pair is rotated ~36° around the axis of the helix relative to the next base pair. So ~10 base pairs make a complete turn of 360°.
7. The twisting of the two strands around one another forms a double helix with a minor groove (~12 Å across) and a major groove (~22 Å across)
8. The diameter of the helix is 20 Å.
9. Apart from Hydrogen bonds between complementary bases, the second force that holds the two DNA strands together is the interaction between the stacked base pairs. These stacking interactions contribute to the stability of the DNA molecule and do not require that any particular base follow another

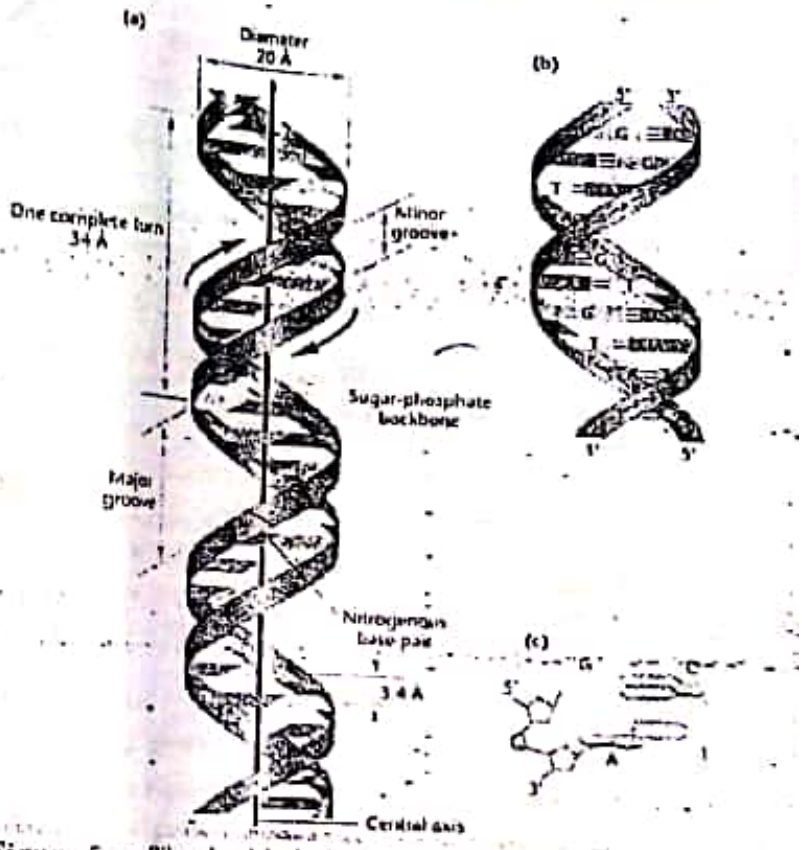
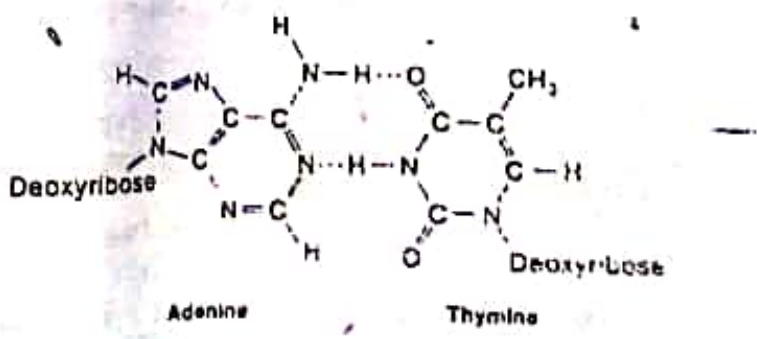
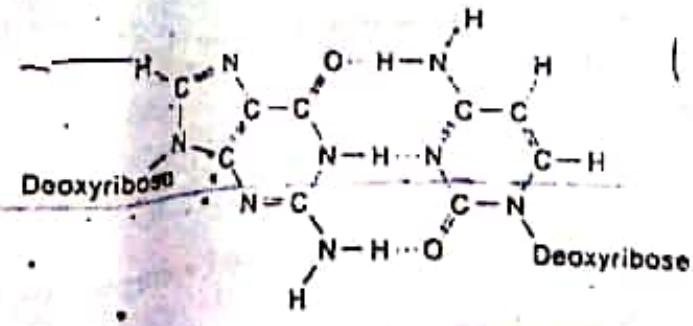


Figure 5: a. The double helical model of DNA. B. Base pairing C. Planar nature of base pairing





Tertiary structure of DNA

In all organisms DNA is complexed with structural proteins to give rise to a condense and stable structure of a chromosome. A chromosome is thus a nucleoprotein complex that is the

1. initiation, in which the transcription apparatus assembles on the promoter and begins the synthesis of RNA;
2. elongation, in which RNA polymerase moves along the DNA, unwinding it and adding new nucleotides, one at a time, to the 3′ end of the growing RNA strand; and
3. termination, the recognition of the end of the transcription unit and the separation of the RNA molecule from the DNA template.

Prokaryotic RNA Polymerase

There is only one type of RNA polymerase in prokaryotes, which catalyzes the synthesis of all classes of bacterial RNA: mRNA, tRNA, and rRNA. Bacterial RNA polymerase (subunit composition shown in Fig. 2) is a large, multimeric enzyme of 480 kD

Molecular Weight

Bacterial RNA polymerases are composed of five subunits of polypeptide chains (Fig. 2) that make up the core enzyme: two copies of a subunit called alpha (α) and single copies of subunits beta (β), beta prime (β′), and omega (ω). The core enzyme catalyzes the elongation of the RNA molecule by the addition of RNA nucleotides.

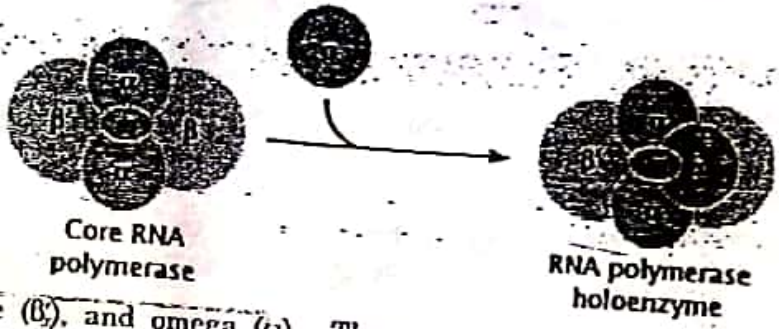


Figure 2: subunit composition of bacterial RNA polymerase

The sigma (σ) factor is not a part of the core enzyme but it controls the binding of RNA polymerase to the promoter. It is a part of the holoenzyme. After sigma has associated with the core enzyme (forming a holoenzyme), RNA polymerase binds stably only to the promoter region and initiates transcription at the proper start site. Sigma is required only for promoter binding and initiation. When a few RNA nucleotides have been joined together, sigma usually detaches from the core enzyme.

Thus, the bacterial RNA Polymerase structure has:

1. α – Two sub units: Involved in making the pentameric core, comprising of $\alpha_2 \beta \beta' \omega$.
2. β – Contains the rNTP binding site
3. β′ – Contains the template DNA binding site
4. ω – Stabilises the catalytic center for polymerization.
5. σ – Contains the promoter-sequence-binding site. The role of this sub unit is in recognizing and binding to the promoter site. Once the nascent transcript reaches the stage of 8-9 nucleotides, σ sub-unit leaves the polymerase complex. However, this sub-unit is essential in ensuring that transcription starts at specific sites. If σ sub-unit is not present, transcription initiates randomly and cannot proceed beyond 20 nucleotides.

cription in prokaryotes

ption in *Bacteria* can be conveniently divided into three stages:

The amino acid is bound to the 3'-terminal A residue with a high energy bond. This bond energy is later used to harness the process of peptide bond formation during the chain elongation.

Protein synthesis

Protein synthesis can be divided into three stages:

1. Initiation
2. Elongation, and
3. Termination.

At initiation the start codon for the protein is recognized: normally it is AUG which codes for methionine, but very occasionally it may be GUG (Valine). A complex is formed between the mRNA, the ribosome and the initiating tRNA.

During elongation, amino acids are added sequentially to the growing peptide chain in accordance with the codon sequence in the mRNA.

At termination, the end of the polypeptide chain is indicated by any of the three termination codons and the complex of mRNA, polypeptide, tRNA, and ribosomes breaks apart.

Each stage requires a number of different protein molecules, termed *factors* to ensure the correct order of events, which are different between prokaryotic and eukaryotic translation. The energy for protein synthesis is provided by the hydrolysis of GTP.

Prokaryotic Translation

Initiation

The first stage of protein synthesis is the binding of the small (30S) ribosomal subunit to the mRNA so that the first AUG codon is positioned in the P site. The correct positioning is achieved by base-pairing between a short purine rich sequence (consensus sequence: 5'-GGAGG-3'), called the *Shine-Dalgarno Sequence*, which is located in the mRNA, 8-13 nucleotides before the start site of translation, and a complementary sequence on the 16S RNA (consensus sequence 3'-CCUCC-5') on the small

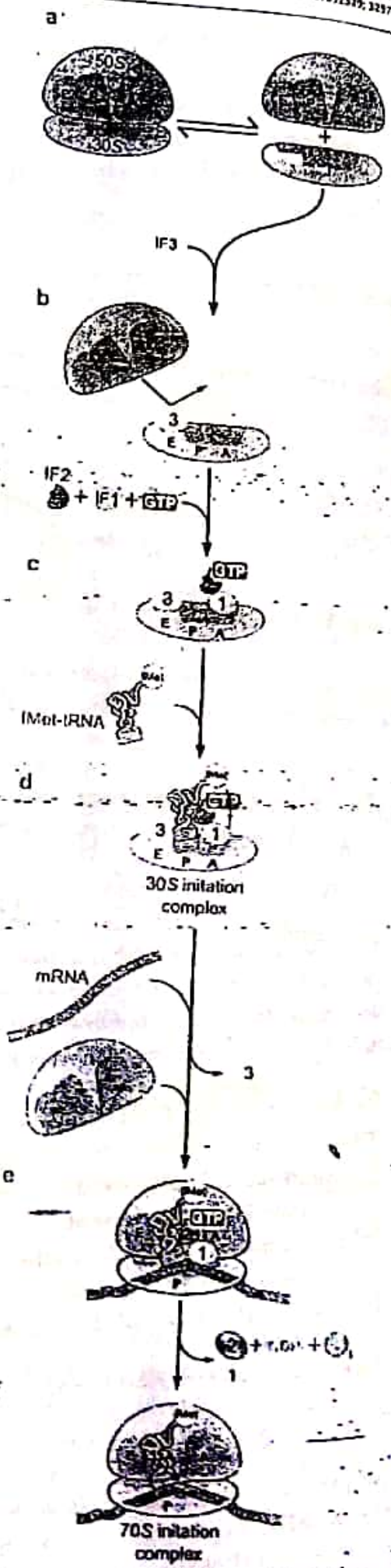


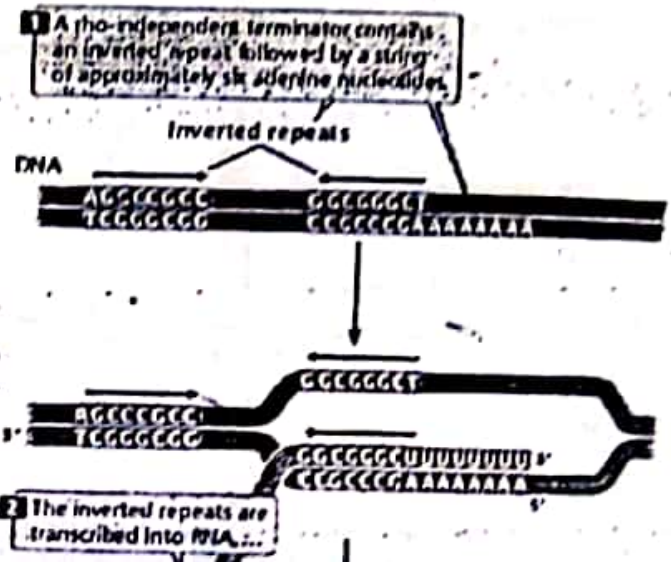
Figure 2: translation initiation in prokaryotes

been added to the transcript, the sigma subunit leaves the RNAP complex. Once this transition occurs, RNAP physically moves away from the promoter, transcribing down the transcription unit, leading to the next phase of the process, elongation.

As elongation proceeds, the polymerase holoenzyme unwinds and rewinds the DNA duplex. At any given moment the transcription bubble is not more than 20 nucleotides. During elongation, the RNAP core carries out the following steps cyclically.

1. DNA Unwinding
2. Nucleotide addition
3. Proof reading
4. Template-Transcript Separation
5. Reannealing of the separated DNA strands.

During elongation on an average, 40 nucleotides are synthesized per second. Elongation process is assisted by elongation factors GRE and Nus. These factors also help in proof reading.



Termination of transcription

Transcription terminates at well-defined

Factor Binding Centre. This ribosomal domain interacts with all those translation factors which bind to GTP. The activation of the Factor Binding Centre stimulates the GTPase activity of IF2. Thus, GTP is hydrolysed into GDP. GTP hydrolysis changes the conformation of IF2, allowing it to leave the complex.

Just after IF2 release, the IF1 also leaves the complex.

Thus with the complete assembly of the ribosome, the 70S initiation complex is ready and Translation initiation is said to have accomplished.

Elongation

Arrival of an appropriate charged tRNA at the A site begins elongation. Elongation is a cyclic process where one cycle is repeated every time when a new amino acid is added to the growing peptide chain. The elongation cycle has three components.

1. Charged tRNA binding at A site
2. Peptide bond formation between the newly arrived amino acid and the last added amino acid in the peptide chain
3. Translocation of the ribosome

The new arriving charged tRNA comes in a ternary complex of charged tRNA + EF-Tu + GTP. It binds to the A site by standard or Wobble codon-anticodon interaction. After the tRNA binds the A site, GDP and EF-Tu are released. The process of correct tRNA and mRNA base pairing is called tRNA accommodation. It is further enhanced by weak hydrogen bonds between rRNA and tRNA.

The activated form of EF-Tu and GTP are regenerated for the next elongation cycle by a process known as EF-Tu/EF-Ts Exchange Cycle.

The peptide bond formation is catalyzed by an inbuilt catalytic centre within the large ribosomal sub-unit - known as **Peptidyl Transferase** - a part of 23S rRNA in the ribosome's large sub-unit. So, this is actually a Ribozyme. This enzyme catalyzes the peptide bond formation between the C end of the AA bound to the tRNA at P site and the N end of the newly arrived AA at the A site. In effect this enzyme transfers the peptide chain formed so far to the newly entered amino acid.

Translocation (Fig 12) is fundamentally a process pushing the tRNAs already present within the ribosome by assistance of a protein factor designated as EF-G and energy from the hydrolysis of GTP. EF-G works by entering the A site of the ribosome.

During translation, the nascent polypeptide moves through a large, water-filled tunnel (approximately 10 nm x 1.5 nm) in the large subunit of the ribosome. The walls of this tunnel are made primarily of 23S rRNA. This structure provides a unique "low affinity" coating through which a polypeptide chain can easily slide.

The elongation cycle is operated repeatedly until the complex encounters a termination codon.

Termination

subunit of the ribosome. This mechanism, which is unique to prokaryotes, allows translation to start in the middle of a mRNA sequence as bacterial mRNAs frequently contain a number of genes which are translated independently.

The events of initiation proceed in the following manner.

1. Initiation factor 3 [IF3] binds to the small subunit of the ribosome and prevents it from binding to the large subunit of the ribosome. Due to this separation, the small ribosomal subunit can bind to mRNA and charged initiator tRNA (which carries *n-formyl methionine*). IF3 keeps the ribosomal subunits apart when protein synthesis is not going on.

2. Initiation factors 1 and 2 [IF1 and IF2] then bind to the small subunit of the ribosome. IF1 binds to the A site of the Ribosome and prevents the entry of the charged initiator tRNA at the A site. For successful translation initiation it is needed that the charged initiator tRNA enters the P site.

3. IF2 binds with GTP. Its role is in guiding the correct localization of the charged initiator tRNA.

Since the small sub-unit binds to the mRNA via the Shine Dalgarno sequence places the AUG initiation codon at the P site, hence the first tRNA, the initiator tRNA, can enter this site and its anticodon can base-pair with the AUG. The initiator tRNA is distinct from the tRNA which inserts methionine in response to AUG in internal sites in the polypeptide in two respects.

1. It carries N-formylmethionine (fMet) instead of methionine (Met).
2. It can directly enter the P site of the ribosome when the large subunit is absent.

Once the fMet-tRNA^{fMet} has bound to the AUG, the large (50S) ribosomal subunit joins to form the complete 70S ribosome. The small subunit of the ribosome aligned with the mRNA and charged initiator tRNA is called the 30S pre initiation complex.

With the joining of the larger subunit of the ribosome, first of all IF3 is released. Rest of the two initiation factors, IF1 and IF2 are released subsequently. When the large subunit binds, it activates a domain called

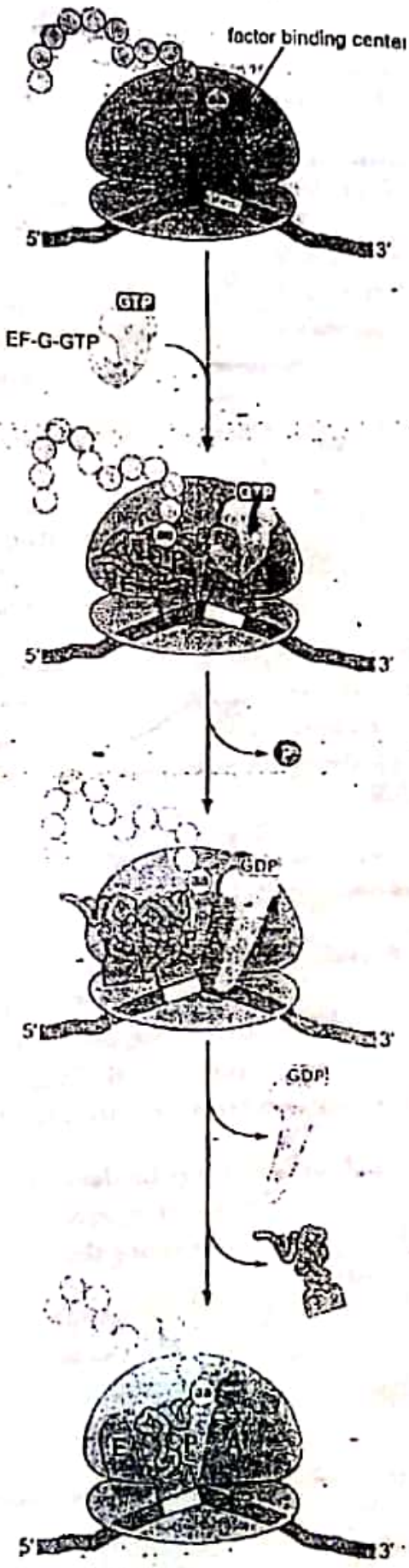


Figure 3: Elongation and Translocation

(*Please refer to your class lecture notes on Gene Structure, for a detailed & diagrammatic account of Operon structure*). The organization of a typical operon is illustrated in Figure 3.

The *lac* Operon of *E. coli*

In 1961, Francois Jacob and Jacques Monod described the "operon model" for the genetic control of lactose metabolism in *E. coli*, Operon is the basic unit of transcriptional control in bacteria.

Lactose is one of the major carbohydrates found in milk; it can be metabolized by *E. coli* bacteria that reside in the mammalian gut. Lactose does not easily diffuse across the *E. coli* cell membrane and must be actively transported into the cell by the enzyme permease (Fig. 4, Step 1).

To utilize lactose as an energy source, *E. coli* must first break it into glucose and galactose, a reaction catalyzed by the enzyme β-galactosidase (Fig. 4, Step 2). This enzyme can also convert lactose into allolactose (Fig. 4, Step 3 and 4), a compound that plays an important role in regulating lactose metabolism. A third enzyme, thiogalactoside transacetylase, also is produced by the *lac* operon.

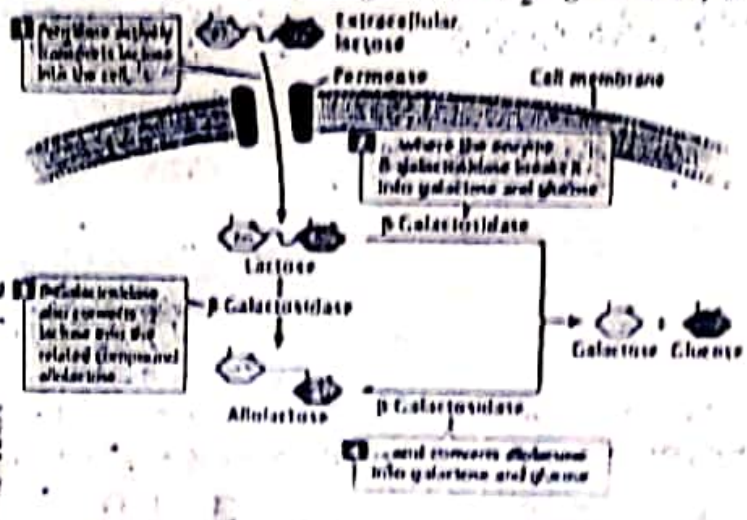


Figure 4: Lactose utilization by *E. coli*

The *lac* operon is an example of a negative inducible operon *but it also has a Positive Control* (as we discuss later). The enzymes β-galactosidase, permease, and transacetylase are encoded by adjacent structural genes in the *lac* operon of *E. coli*. β-galactosidase is encoded by the *lacZ* gene, permease by the *lacY* gene, and transacetylase by the *lacA* gene (Fig. 5). When lactose is absent from the medium in which *E. coli* these genes are not transcribed.

If lactose is added to the medium and glucose is absent, the rate of synthesis of all three enzymes simultaneously increases about a thousand fold. This boost in enzyme synthesis results from the transcription of *lacZ*, *lacY*, and *lacA* and exemplifies coordinate induction, the simultaneous synthesis of several enzymes stimulated by a specific molecule, the inducer.

The structure and operation of *lac* operon are described through the step wise diagrams of Figure 5.

Steps in the operation of the *lac* Operon

1. The *lacZ*, *lacY*, and *lacA* genes have a common promoter (*lacP*) and are transcribed together. Upstream of the promoter is a regulator gene *lacI*, which has its own promotor (*P*).
2. The *lacI* gene is transcribed into a short mRNA that is translated into a repressor. Each repressor consists of four identical polypeptides.

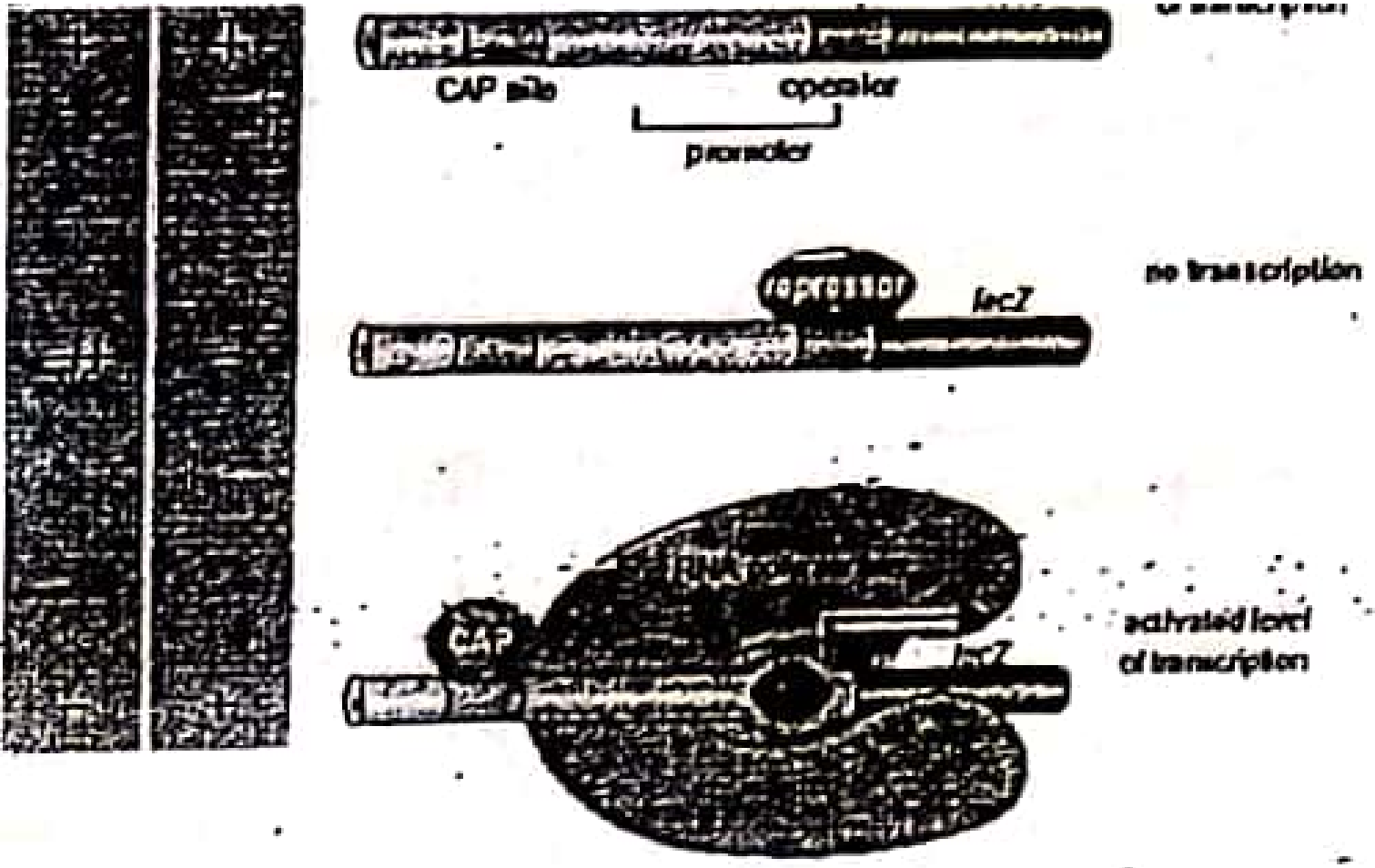


Figure 6: Sugar utilization pattern by *E. coli*

The end of the protein-coding message is signaled by the presence of one of three stop codons (UAA, UAG, or UGA). These are not recognized by a tRNA and do not specify an amino acid. Proteins known as **release factors** bind to any ribosome with a stop codon positioned in the A site, *forcing the peptidyl transferase in the ribosome to catalyze the addition of a water molecule instead of an amino acid to the peptidyl-tRNA*. This reaction releases the carboxyl end of the growing polypeptide chain from its attachment to a tRNA molecule, and since only this attachment holds the growing polypeptide to the ribosome, the completed protein chain is immediately released into the cytoplasm.

There are three types of release factors in prokaryotes falling into two categories.

a. **Class I release factors**, including RF 1 and RF 2, which actually release the newly synthesized peptide chain from the ribosome

b. **Class II release factor**, that is RF 3, a GTP hydrolyzing protein that enables the release of a class I release factor once it has performed its task in the ribosome.

First, any one of the two release factors, RF1 or RF2 binds the A site and then the peptide chain formed so far is released from the ribosome. Later with assistance of another release factor RF3, the class I release factor comes out of the ribosome.

Release factors are an example of molecular mimicry, whereby one type of macromolecule resembles the shape of a chemically unrelated molecule. In this case, the three-dimensional structure of release factors (made entirely of protein) resembles the shape and charge distribution of a tRNA molecule. This shape and charge mimicry helps them enter the A-site on the ribosome and cause translation termination.

Soon afterwards, the ribosome also disintegrates into its sub-units with help of Ribosome Recycling Factor, GTP and Elongation Factor G. The ribosome then releases the mRNA and separates into the large and small subunits. The disintegrated state of ribosome is then stabilized by IF3.

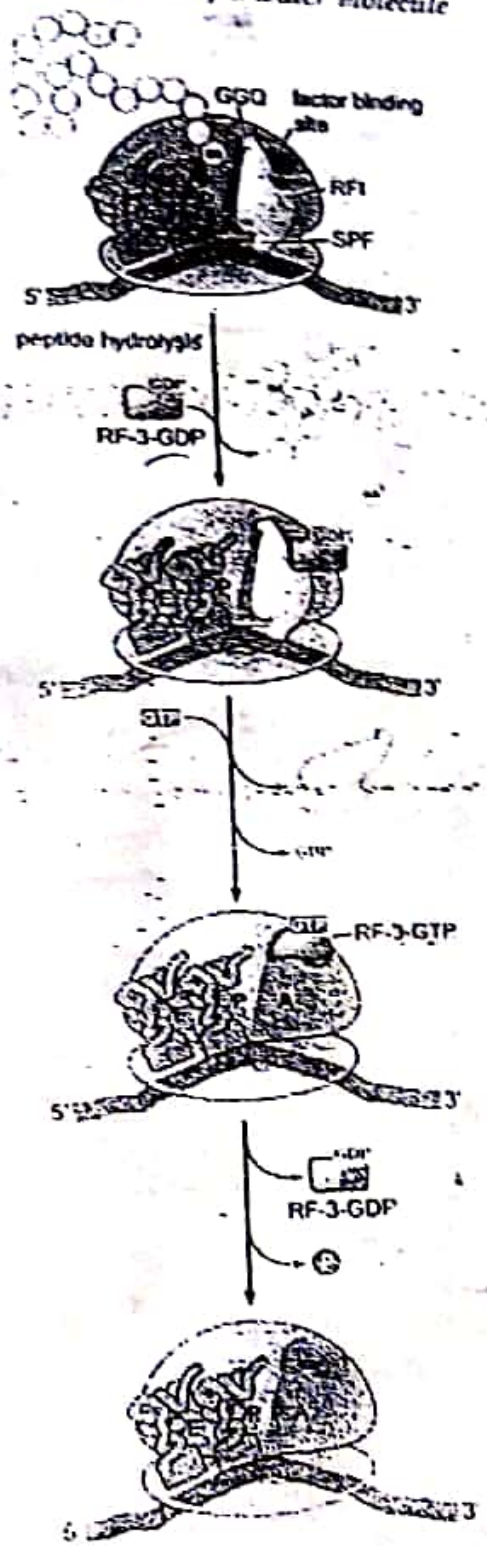


Figure 4: Termination

Eukaryotic Translation

In a eukaryotic cell two types of translation proceed.,

1. Cytoplasmic translation, which is very different from the prokaryotic translation, especially in the process of initiation.

Many bacteria possess multiple types of sigma. *E. coli*, for example, possesses sigma 28 (σ^{28}), sigma 32 (σ^{32}), sigma 54 (σ^{54}), and sigma 70 (σ^{70}), named on the basis of their molecular weights. Each type of sigma initiates the binding of RNA polymerase to a particular set of promoters. For example, σ^{32} binds to promoters of genes that protect against environmental stress, σ^{54} binds to promoters of genes used during nitrogen starvation, etc.

Initiation and elongation of RNA chains

Initiation comprises all the steps necessary to begin RNA synthesis, including

1. promoter recognition,
2. formation of the transcription bubble,
3. creation of the first bonds between rNTPs, and
4. escape of the transcription apparatus from the promoter.

Transcription initiation requires that the transcription apparatus recognize and bind to the promoter.

Promoters in prokaryotes

More than 100 promoter sequences in prokaryotes have been characterized. They differ from each other in some respects, but certain fundamental structures are well conserved. They are (as shown in Fig. 3):

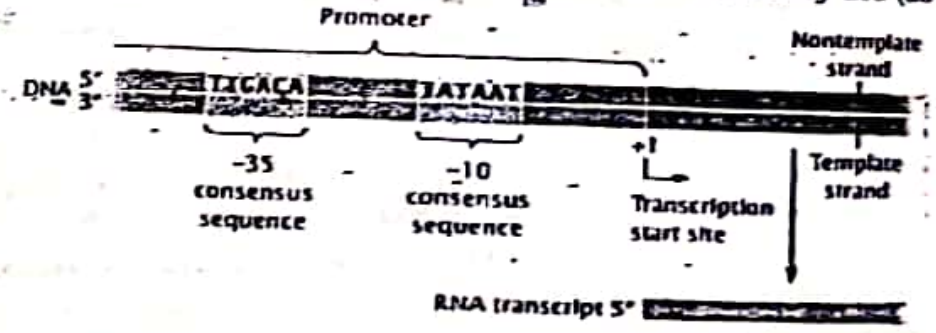


Figure 3: general design of prokaryotic gene promoters

- **-10 sequence**: occurs at about 10 nucleotides before the point of transcription initiation. It contains a consensus sequence of TATAAT. It is also called Pribnow box (after its inventor R. Pribnow). The -10 sequence, being rich in AT is the site of DNA duplex unwinding.
- **-35 sequence**: occurs at about 35 nucleotides before the point of transcription initiation. It contains a consensus sequence of TTGACA. It is also called the recognition sequence because this is where sigma factor binds.

The distance between -10 and -35 sequences is highly conserved in *E.coli*. It is never less than 15 nt and more than 20 nt in length.

Some bacterial promoters contain a third consensus sequence - upstream element - that contains a number of A–T pairs and is found at about –40 to -60. The alpha subunit of the RNA polymerase interacts directly with this upstream element, greatly enhancing the rate of transcription in those bacterial promoters that possess it.

Initiation & Elongation

The process of RNA synthesis in bacteria involves first the binding of the RNA polymerase (RNAP) holoenzyme molecule to the template at the promoter site to form a preinitiation complex, or PIC. Binding is followed by a conformational change of the RNAP, and the first nucleotide then associates with the initiation site on the β subunit of the enzyme. In both prokaryotes and eukaryotes, a purine ribonucleotide is usually the first to be polymerized into the RNA molecule. In the presence of the appropriate nucleotide, RNAP catalyzes the formation of a phosphodiester bond, and the nascent chain is now attached to the polymerization site on the β subunit of RNAP. (See Figure 4.)

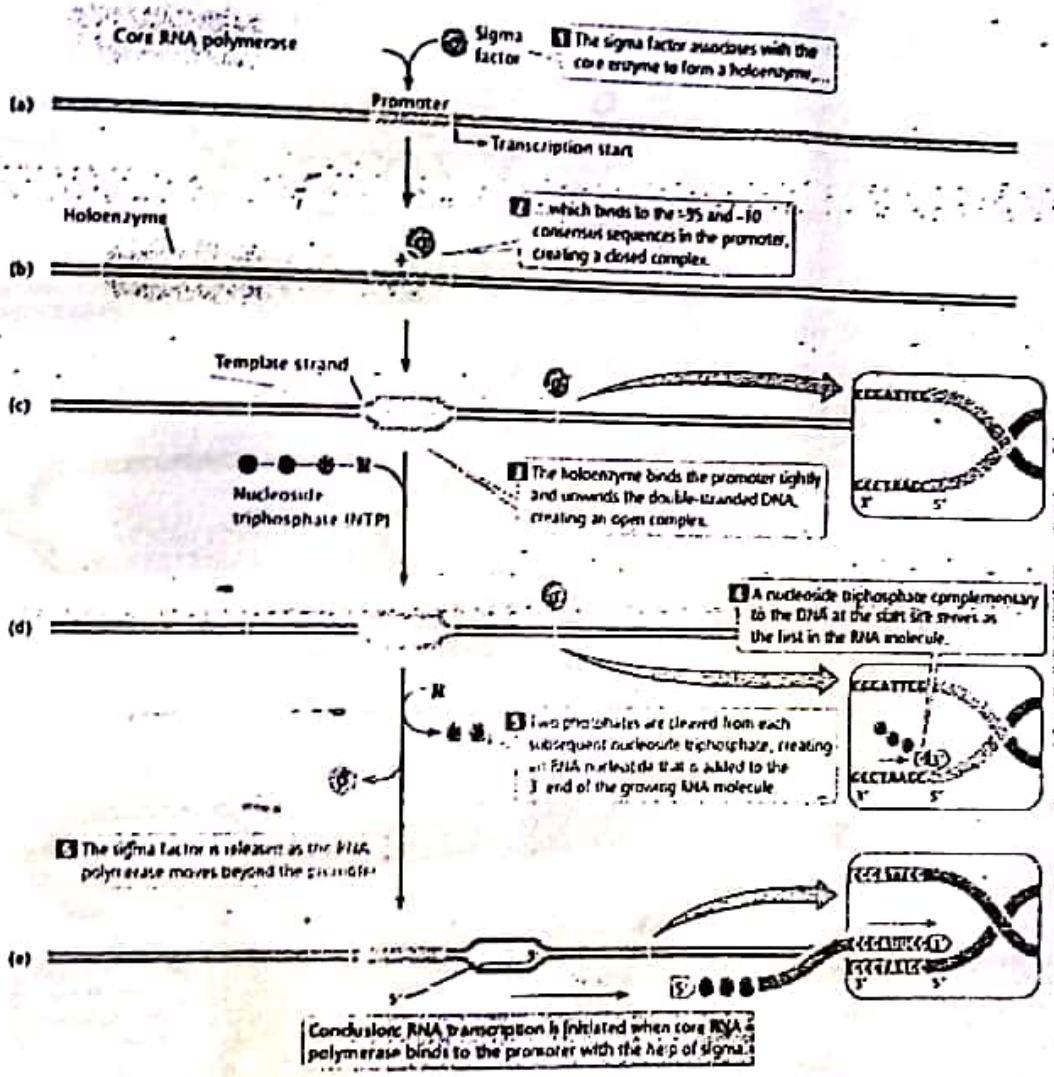


Figure 4: Initiation of prokaryotic transcription

Initiation of formation of the RNA molecule at its 5' end then follows, while elongation of the RNA molecule from the 5' to its 3' end continues antiparallel to its template. The enzyme polymerizes the ribonucleotides in a specific sequence dictated by the template strand and interpreted by Watson-Crick base-pairing rules. Pyrophosphate is released in the polymerization reaction. This pyrophosphate (PP_i) is rapidly degraded to 2 mol of inorganic phosphate (P_i) by ubiquitous pyrophosphatases, thereby providing energy for irreversibility on the overall synthetic reaction. As with eukaryotes, 5' triphosphate of this first nucleotide is maintained in prokaryotic mRNA.

Often in the course of initiation, RNA polymerase repeatedly generates and releases short transcripts, from 2 to 6 nucleotides in length, while still bound to the promoter. This stage is called Abortive Initiation Stage. After several abortive attempts, the polymerase synthesizes an RNA molecule from 10 to 20 nucleotides in length, which allows it to transition to the elongation stage.

After 10–20 nucleotides have been polymerized, RNAP undergoes a second conformational change leading to promoter clearance. A little before this stage i.e. after 8-9 nucleotides have

A. **Purine:** Each purine consists of a six-sided ring attached to a five-sided ring. Both DNA and RNA contain two purines, adenine and guanine (A and G), which differ in the positions of their double bonds and in the groups attached to the six-sided ring.

B. **Pyrimidine:** Each pyrimidine consists of a six-sided ring only. Three pyrimidines are found in nucleic acids: cytosine (C), thymine (T), and uracil (U). Cytosine is present in both DNA and RNA; however, thymine is restricted to DNA, and uracil is found only in RNA. The three pyrimidines differ in the groups or atoms attached to the carbon atoms of the ring and in the number of double bonds in the ring.

In a nucleotide, the nitrogenous base always forms a covalent bond with the 1'-carbon atom of the sugar. A deoxyribose (or ribose) sugar and a base together are termed as a nucleoside.

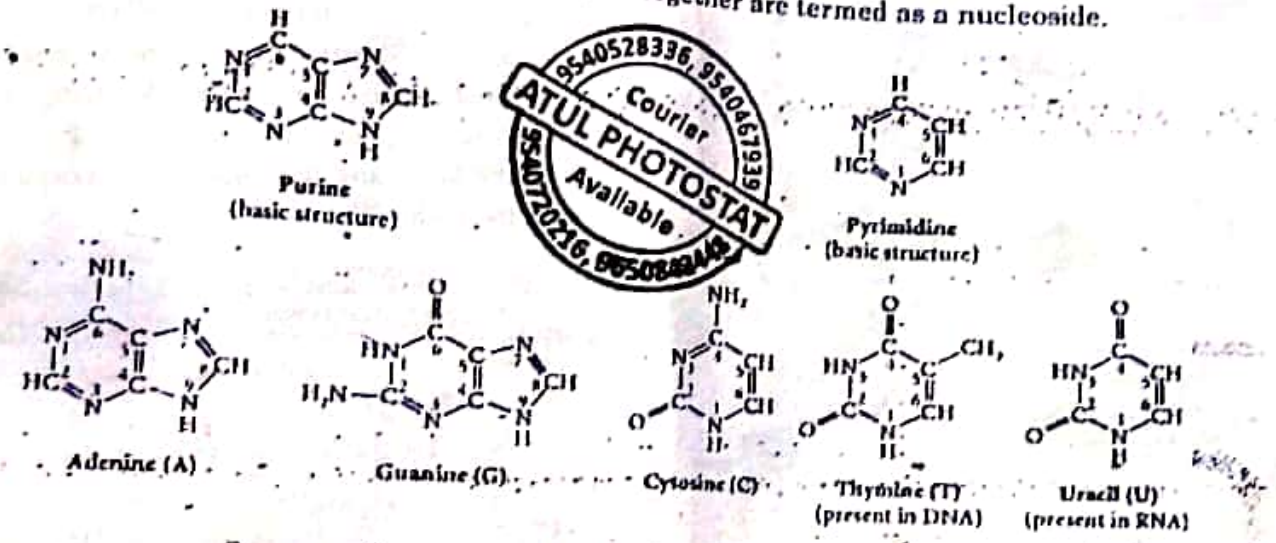


Figure 2: The nitrogenous bases present in nucleic acids.

The DNA nucleotides are properly known as deoxyribonucleotides or deoxyribonucleoside 5'-monophosphates. Because there are four types of bases, there are four different kinds of DNA nucleotides.

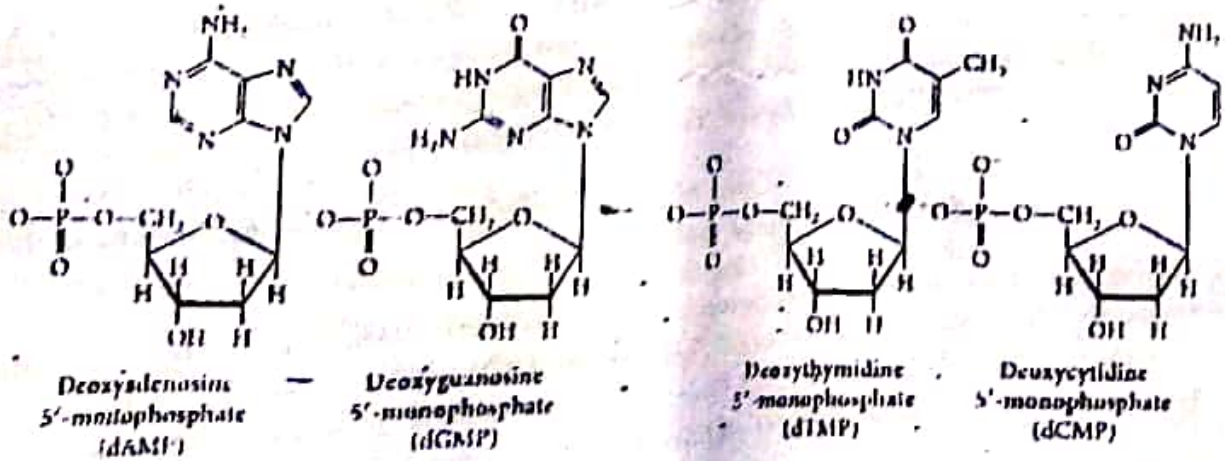


Figure 3: The four types of nucleotides found in DNA

Polynucleotide strands arise when many nucleotides are connected by covalent bonds called phosphodiester linkages, which join the 5'-phosphate group of one nucleotide to the 3'-carbon atom of the next nucleotide. The backbone of the polynucleotide strand is composed of alternating sugars and phosphates.

Deoxyribonucleic acid (DNA) is a linear polymer of deoxyribonucleotides, which is found mostly in the form of a double helix and forms the universal genetic material in all life forms except for RNA viruses, Viroids and Virusoids. It is rarely in a single stranded form (as in some viruses like φX174).

Apart from role as genetic material, the DNA sequences are also involved in various other functions like regulation, structural stability and biocatalysis as DNAzymes.

The Structure of DNA

DNA has a multilevel structural organization, which can be grouped as:

1. Primary structure of DNA
2. Secondary structure of DNA
3. Tertiary structure of DNA

Primary Structure of DNA

Primary structure of DNA is a polymeric-chain-of-deoxyribonucleotides. This chain is always linear in case of eukaryotic cells and many viruses, while in bacterial cells and in some viruses it is a covalently closed (circular) molecule. The repeating units of a DNA chain are nucleotides, each comprising three parts:

1. a pentose sugar- that is Deoxyribose,
2. a phosphate, and
3. a nitrogenous base.

1. **The sugars of nucleic acids:** It is called pentose sugars and has five carbon atoms. The sugars of DNA and RNA are slightly different in structure. RNA's ribose sugar has a hydroxyl group attached to the 2'-carbon atom whereas DNA's sugar, called deoxyribose, has a hydrogen atom at this position. This difference gives rise to the names ribonucleic acid (RNA) and deoxyribonucleic acid (DNA).

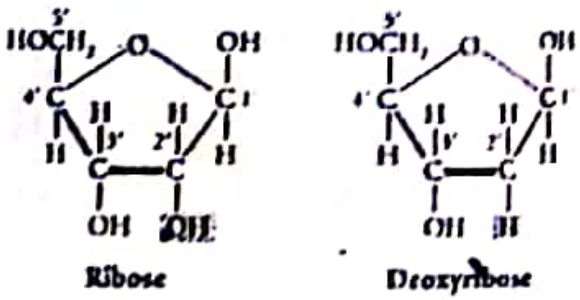


Figure 1: The sugars of nucleic acids

2. **The phosphate group** consists of a phosphorus atom bonded to four oxygen atoms. Phosphate groups are found in every nucleotide and frequently carry a negative charge, which makes DNA acidic. The phosphate group is always bonded to the 5'-carbon atom of the sugar in a nucleotide.

3. **Nitrogen-containing bases** may be of two types: a purine or a pyrimidine.