## THE CENTRAL DOGMA OF PROTEIN SYNTHESIS

Proteins constitute the major part by dry weight of an active growing cell. They are widely distributed in living matter. All enzymes are proteins. Proteins are built up from about 20 amino acids which constitute the basic building blocks. In proteins the amino acids are linked up by peptide bonds to from long chains called *polypeptides*. The sequence of amino acids has a bearing on the properties of a protein, and is characteristic for a particular protein.

The basic mechanism of protein synthesis is that DNA makes RNA, which in turn makes protein. The central dogma of protein synthesis is expressed as follows:

## TEMINISM

In 1958 Crick proposed the central dogma of molecular biology :

$$\text{DNA} \xrightarrow{\text{Transcription}} \text{RNA} \xrightarrow{\text{Translation}} \text{PROTEIN}$$

However, in 1963 Temin reported that in some cancer viruses a reverse flow of genetic information takes place. Certain RNA tumour viruses synthesize DNA, which in turn codes for protein. Temin's hypothesis, put forward in 1964, is that in RNA tumour viruses RNA replicates through a DNA intermediate. This was confirmed by the discovery of an enzyme which makes DNA from a RNA template. It was also found that if the RNA of the virus was destroyed by the enzyme RNase, then DNA synthesis was prevented. This suggests that RNA is acting as a template for DNA synthesis.

Further experiments in other laboratories confirmed the existence of DNA polymerase activity in other RNA tumour viruses. It was also shown that the new DNA made was viral RNA specific. RNA dependent DNA–polymerase has been found in the lymphocytes of leukemia patients, providing some hope for future leukemia control. A modified version of the central dogma has been suggested by Crick in 1970. DNA undergoes replication to form DNA and transcription to form RNA, which then takes part in protein formation. RNA can also undergo replication to form RNA and transcription to form DNA, which then takes part in protein formation. In the diagram (Fig. 12.28) the solid arrows indicate the general transformation of information. The dotted lines show special transfers that do not occur in most cells.

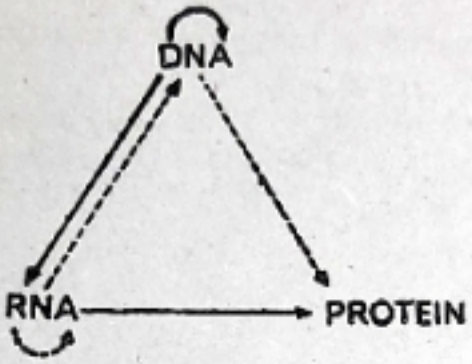



Fig. 12.28. RNA-dependent DNA synthesis.

## 1.2 Mechanism of RNA Transcription in Prokaryotes

The mechanism of RNA synthesis was worked out in late 1950s by the American investigators, **Jerard Horowitz, Samuel B. Weiss** and **Audery Stevens** in *in vitro* experiments. The existence of mRNA was first shown by **Spiegelman** and coworkers and ***RNA polymerase*** activity was discovered by **Weiss** in 1960.

### Material Required

RNA transcription requires:

1. The enzyme ***RNA polymerase*** or ***DNA dependent RNA polymerase***
2. DNA template strand the transcription unit
3. All the four types of ribonucleoside triphosphates (ATP, CTP, GTP and UTP)
4. Divalent metal ions $Mg^{2+}$ or $Mn^{2+}$ as a cofactor

No primer is needed for RNA synthesis

### RNA Polymerase in Prokaryotes

Bacterial ***RNA polymerase*** is a complex **holoenzyme** that catalyses transcription of RNA. In prokaryotes, single ***RNA polymerase*** **(RNAP)** enzyme carries out synthesis of all the three types of RNAs. But in eukaryotes three distinct RNA polymerase enzymes, namely ***RNA polymerase I, RNA polymerase II*** **and** ***RNA polymerase III*** catalyse synthesis of rRNA, mRNA and tRNA respectively. Each bacterial cell of *E. coli* has about 1000 to 2000 RNA polymerase molecules. Mitochondria and chloroplasts have their own RNA polymerases, which are similar to bacterial polymerase enzymes.

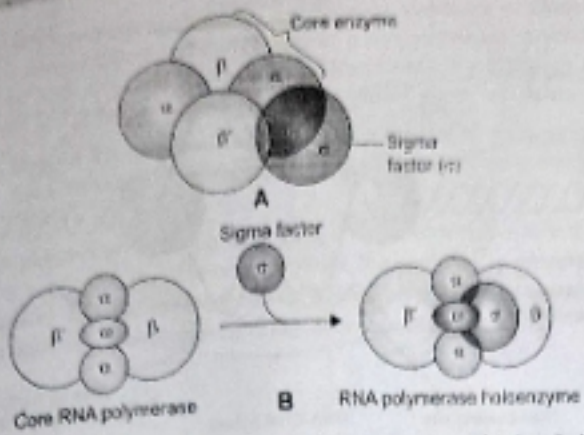



FIGURE 11.1
**RNA polymerase enzyme** of bacterium *E. coli* showing association of its 5 polypeptides $\alpha\alpha\beta\beta'$ and $\sigma$

In prokaryotes single RNA polymerase catalyses synthesis of all the three types of RNAs. In *E. coli*, the holoenzyme ***RNA polymerase*** is a large protein with a molecular weight of 4,80,000. It is a holoenzyme and consists of **core enzyme** and **sigma subunit**.

## 1. Core enzyme

Core enzyme ($\alpha_2\beta\beta'$) is formed of four tightly associated protein chains namely $\beta\beta'\alpha\alpha$ with molecular weights of 1,60,000, 1,50,000, 90,000, and 4,000 respectively. It binds non-specifically to DNA template and migrates downstream till its sigma factor recognises the promoter region.

1. **The α Subunits:** The core enzyme has two molecules of α-polypeptide. This is coded by gene *rpoA*. The α subunits are needed for the assembly of core enzyme; interact with some regulatory proteins, and probably help in the recognition of promoter.
2. **The β Subunit:** Only one copy of β subunit is present in one molecule of core enzyme of RNA polymerase. It is encoded by gene *rpoB*. The β subunit binds with the incoming nucleotide to be added to RNA chain and helps in the formation of first phosphodiester bond.
3. **The β' Subunit:** It is encoded by *rpoC* gene and binds with the template strand or antisense strand of DNA. Therefore, both β and β' are catalytic subunits of core enzyme.
4. **Omega (ω) subunit:** It is coded by gene *rpoZ* and has a molecular mass 10,105. It helps in enzyme assembly but is not required in enzyme activity.

**TABLE 11.1: Subunits of RNA polymerase, their molecular mass and their functions**

| Subunit | Number of subunits | Molecular mass | Functions |
|---|---|---|---|
| 1. Alpha (α) | 2 | 36,511 | Needed in enzyme assembly.<br>Interacts with some regulatory proteins |
| 2. Beta (β) | 1 | 150,616 | Forms a pincer,<br>Is site of rifamycin action |
| 3. Beta' (β') | 1 | 155,159 | Forms a pincer.<br>Provides a conserved motif (–NADFDGD) essential for catalys |
| 4. Omega (ω) | 1 | 10,105 | Helps in enzyme assembly |
| 5. Sigma (σ) | 1 | 70,263 | Directs the enzyme to promoters |

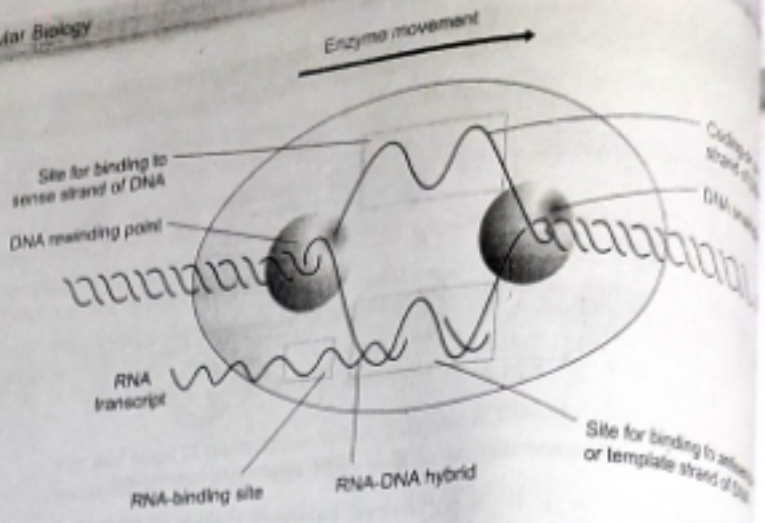



FIGURE 11.2
Active sites on bacterial core enzyme of RNA polymerase

### 2. Sigma subunit (σ)

Sigma subunit is formed of a single polypeptide chain, loosely attached to the enzyme. It is encoded by the gene *rpoD*. It recognises the start signal on molecule and directs the core enzyme of ***RNA polymerase*** to bind to the promoter region upstream of initiation codon. The σ subunit recognises two special sequences in the promoter region of the coding strand (i.e. nontemplate strand DNA. These are **–10 sequence** and **–35 sequence.**

Therefore, ability of holoenzyme ***RNA polymerase*** to recognise promoter sequence is due to σ subunit, which also facilitates opening or melting of DNA helix. The subunit separates from core enzyme once about 10 nucleotides are joined in the RNA transcription.

The core enzyme without sigma factor binds nonspecifically to any region of DNA and can transcribe DNA there, because it fails to recognise specific promoter on DNA template. It is also unable to discriminate between the two strands of DNA helix to be used as a template. **Sigma factor** recognises specific transcription initiation sites where RNA synthesis can begin. These sites are called **promoter sites**. There are different sigma factors, one each for every promoter site.

### 3. Functions of RNA Polymerase

1. ***RNA polymerase*** unwinds about 15 bases of DNA around the initiation to form an open promoter-DNA complex and provides single-strand DNA to act as template for transcription.

## Extension: Functional Sites on Core Enzyme of RNA Polymerase

The core enzyme has following four functional sites:

- DNA unwinding site unwinds DNA duplex and separates two nucleotide strands as the enzyme moves along DNA duplex to be transcribed.
- Site binding to antisense strand or template strand of DNA molecule.
- Site binding to the sense strand or coding strand of DNA molecule.
- DNA rewinding site binds to site where separated sense and antisense strands of DNA start rewinding after the transcription of that segment of DNA is complete.

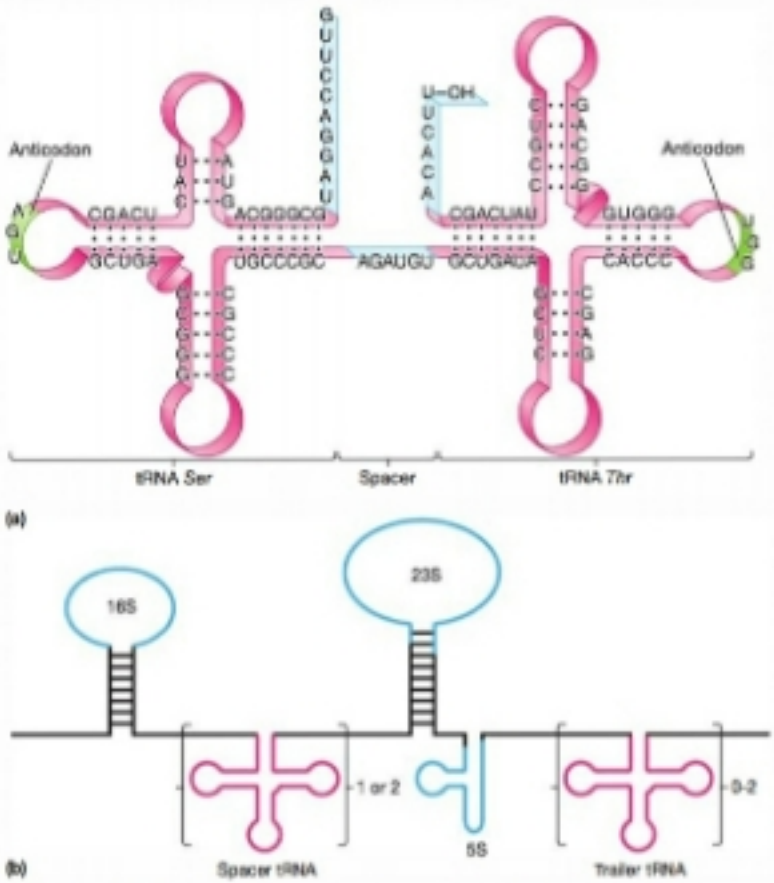



**Figure 13.21 tRNA and rRNA Genes.** (a) A tRNA precursor from *E. coli* that contains two tRNA molecules. The spacer and extra nucleotides at both ends are removed during processing. (b) The *E. coli* rRNA gene codes for a large transcription product that is cleaved into three rRNAs and one to three tRNAs. The 16S, 23S, and 5S rRNA segments are represented by blue lines, and tRNA sequences are placed in brackets. The seven copies of this gene vary in the number and kind of tRNA sequences.

## 13.5 Transcription in Bacteria

After reading this section, you should be able to:

- Illustrate the organization of bacterial genes in a typical operon
- Describe the structure of a typical bacterial RNA polymerase holoenzyme
- Outline the events that occur during the three phases of transcription
- Discuss the role of bacterial promoters and sigma factors in transcription initiation
- Distinguish factor-independent termination of transcription from rho-dependent termination of transcription

Synthesis of RNA under the direction of DNA is called transcription, and the RNA product has a sequence complementary to the DNA template directing its synthesis. Although adenine directs the incorporation of thymine during DNA replication, it usually codes for uracil during RNA synthesis. Transcription generates three major kinds of RNA. **Transfer RNA (tRNA)** carries amino acids during protein synthesis, and **ribosomal RNA (rRNA)** molecules are components of ribosomes. **Messenger RNA (mRNA)** bears the message for protein synthesis. Bacterial genes encoding proteins involved in a related process (e.g., encoding enzymes for synthesis of an amino acid) are often located close to each other and are transcribed from a single promoter. Such a transcriptional unit is termed an **operon**. Transcription of an operon yields an mRNA consisting of a

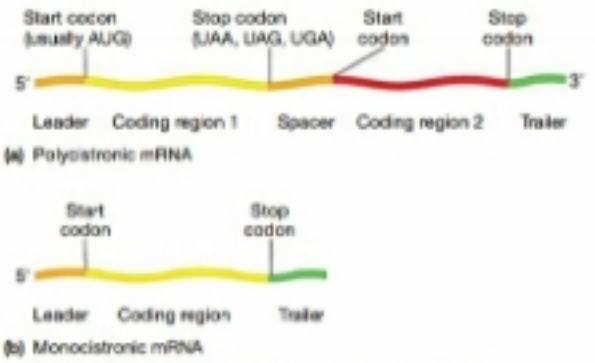



**Figure 13.22 Polycistronic and Monocistronic mRNAs.** Polycistronic mRNAs are commonly observed in bacteria and archaea. Eukaryotic genes usually give rise to monocistronic mRNAs.

leader followed by one coding region, which is separated by a space from the second coding region, and so on, with the final sequence of nucleotides being the trailer. Such mRNAs are said to be **polycistronic mRNAs** (figure 13.22*a*). Each coding region in the polycistronic mRNA is defined by a start and stop codon. Thus each coding region is translated separately to give rise to a single polypeptide. Many archaeal mRNAs are also polycistronic. However, polycistronic mRNAs are rare in eukaryotes. Instead, their mRNAs are usually **monocistronic mRNAs** (figure 13.22*b*), containing information of a single gene. ⏭ *Transcription (section 15.3)*

## Bacterial RNA Polymerases

RNA is synthesized by enzymes called **RNA polymerases.** In bacteria, a single RNA polymerase transcribes all genes. Most bacterial RNA polymerases contain five types of polypeptide chains: α, β, β′, ω, and σ (figure 13.23). The **RNA polymerase core enzyme** is composed of five polypeptides (two α subunits, β, β′, and ω) and catalyzes RNA synthesis. Interestingly, the subunits that form the core enzyme are conserved in archaeal and eukaryal RNA polymerases. However, their RNA polymerases consist of additional proteins and are larger and more complex than the bacterial enzyme (*see table 15.3*). The **sigma factor** (σ) has no catalytic activity but helps the core enzyme recognize the promoter. When sigma is bound to the core enzyme, the six-subunit complex is termed **RNA polymerase holoenzyme.** Only holoenzyme can begin transcription, but the core enzyme completes RNA synthesis once it has been initiated.

## Stages of Transcription

Transcription involves three separate processes: initiation, elongation, and termination, which together are often referred to as the transcription cycle (figure 13.24). Sigma factor is critical to

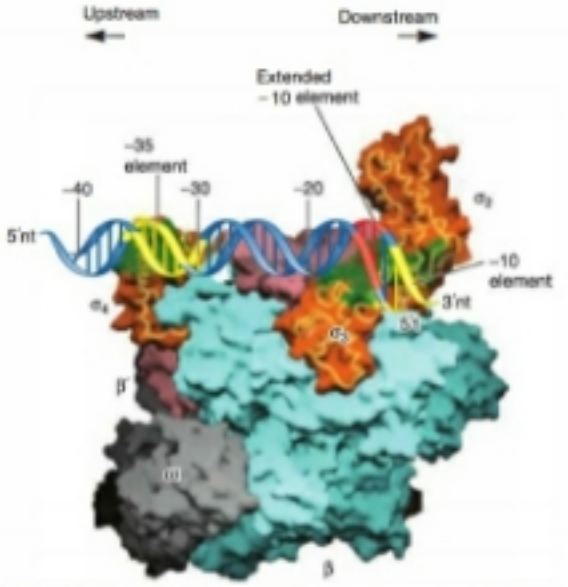



**Figure 13.23 RNA Polymerase Structure.** The holoenzyme-DNA complex of the bacterium *Thermus aquaticus*. Protein surfaces that contact the DNA are in green and are located on the σ factor. The −10 and −35 elements in the promoter are in yellow. The internal active site is covered by the β subunit in this view.

the initiation process. It positions the RNA polymerase core enzyme at the promoter. Many bacterial promoters have two characteristic features: a sequence of six bases (often TTGACA) about 35 base pairs before (upstream) the transcription starting point and a TATAAT sequence called the **Pribnow box,** usually about 10 base pairs upstream of the transcriptional start site (figure 13.25; also figure 13.20). These regions are called the −35 and −10 sites, respectively, because these are their distances in nucleotides upstream of the first nucleotide to be transcribed (i.e., the +1 site). Sigma factor recognizes the −10 and −35 sequences, and directs the RNA polymerase core enzyme to them. ● *Stages of Transcription*

At this point in our discussion, it is worth noting that bacterial cells produce more than one type of sigma factor. Each sigma factor preferentially directs the RNA polymerase to a distinct set of promoters. For instance, in *E. coli*, most genes have promoters recognized by a sigma factor called $\sigma^{70}$. This sigma factor recognizes promoters having the −10 and −35 sequences shown in figure 13.25 and table 13.2. These sequences are the **consensus sequences** for $\sigma^{70}$-recognized promoters. Promoters recognized by other sigma factors have different consensus sequences. The use of different sigma factors to initiate transcription is a common bacterial regulatory mechanism, as we describe in chapter 14. Our focus here is on transcription of genes recognized by $\sigma^{70}$.

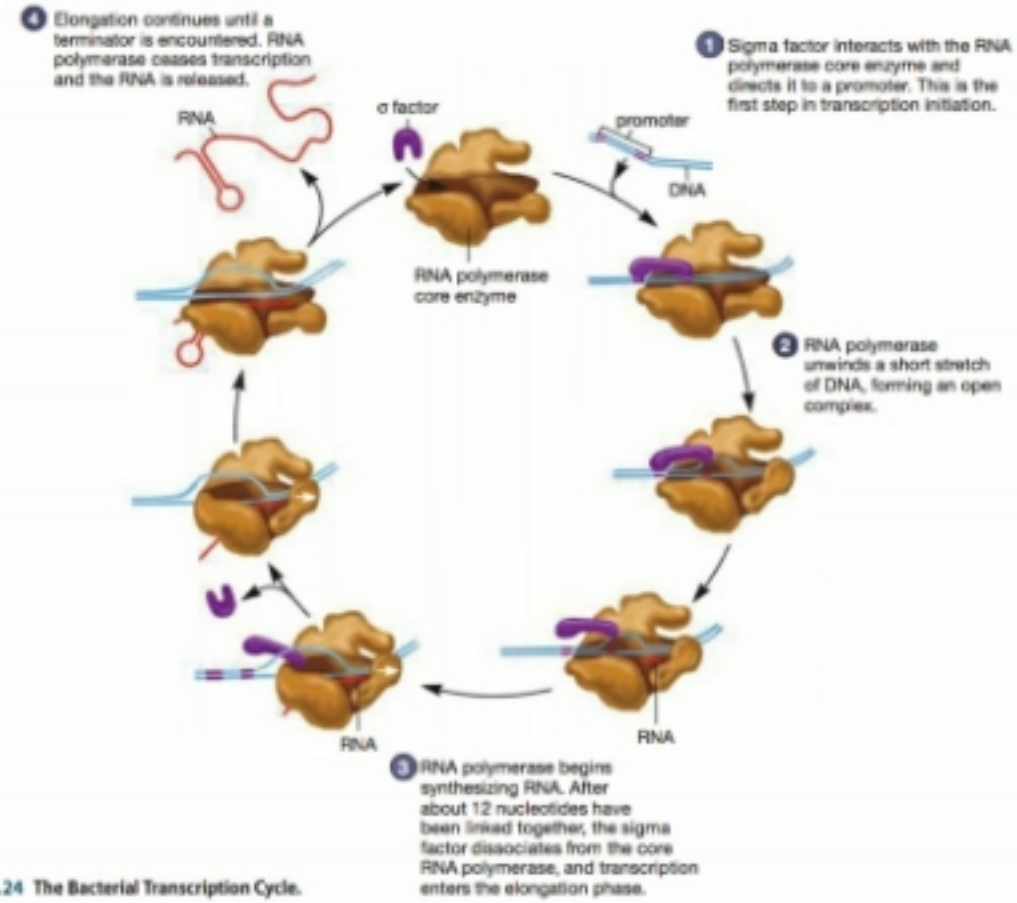



**Figure 13.24 The Bacterial Transcription Cycle.**

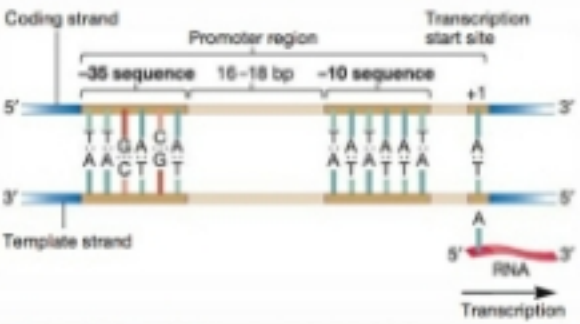



**Figure 13.25 A $\sigma^{70}$ Promoter.** Many bacterial promoters are recognized by the sigma factor $\sigma^{70}$. Its promoters have a characteristic set of nucleotides centered at about −10 and −35. Shown are the consensus sequences for these two important promoter sites. Bacterial cells produce additional sigma factors that recognize different promoter sequences (table 13.2).

**MICRO INQUIRY** *Are the −35 and −10 regions considered "upstream" or "downstream" of the +1 nucleotide?*

Once bound to the promoter, RNA polymerase unwinds the DNA (figure 13.26). The −10 site is rich in adenines and thymines, making it easier to break the hydrogen bonds that keep the DNA double stranded; when the DNA is unwound at this region, it is called an open complex. A region of unwound DNA equivalent to about 16 to 20 base pairs becomes the "transcription bubble," which moves with the RNA polymerase as it synthesizes mRNA from the template DNA strand during elongation (figure 13.27). Within the transcription bubble, a temporary RNA:DNA hybrid is formed. As RNA polymerase holoenzyme progresses along the DNA template, the sigma factor dissociates from the other subunits and can help another RNA polymerase core enzyme initiate transcription (figure 13.24).

The reaction catalyzed by RNA polymerase is quite similar to that catalyzed by DNA polymerase (figure 13.10). ATP, GTP, CTP, and UTP are used to produce RNA complementary to the DNA template, and pyrophosphate is produced as ribonucleoside monophosphates are incorporated into the growing RNA chain. Pyrophosphate is hydrolyzed to fuel the process. RNA synthesis also proceeds in a 5′ to 3′ direction with new nucleotides being added to the 3′ end of the growing chain, making the RNA

**Table 13.2** *E. coli* Sigma Factors and the Sequences They Recognize

| Sigma Factor | Consensus Promoter Sequences[1] | | Genes Transcribed from Promoter |
|---|---|---|---|
| $\sigma^{70}$ | TTGACAT | TATAAT | Most genes |
| $\sigma^{54}$ | CTGGNA[2] | TTGCA | Genes for nitrogen metabolism |
| $\sigma^{38}$ | TTGACA | TCTATACTT | Genes for stationary phase and stress responses |
| $\sigma^{32}$ | TCTCNCCCTTGAA | CCCCATNTA | Genes for heat-shock response |
| $\sigma^{28}$ | CTAA | CCGATAT | Genes for chemotaxis and motility |

1 With the exception of the $\sigma^{54}$ promoters, all consensus sequences are located at −35 and −10, respectively. The $\sigma^{54}$ consensus sequences are located at −24 and −12.
2 N indicates any nucleotide.

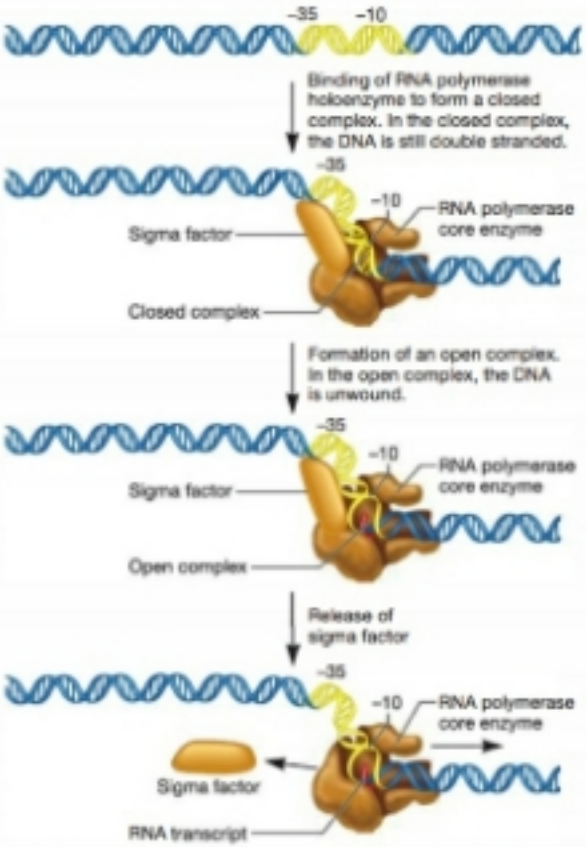
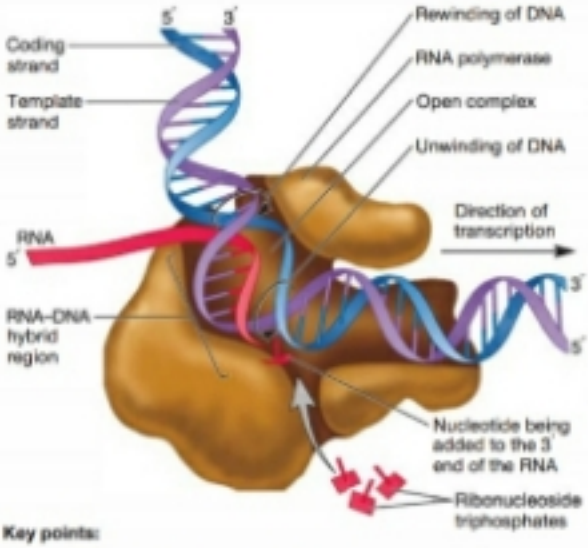



**Figure 13.26 Initiation of Transcription in Bacteria.** The sigma factor of the RNA polymerase holoenzyme is responsible for positioning the core enzyme properly at the promoter. Sigma factor recognizes two regions in the promoter, one centered at −35 and the other centered at −10. Once positioned properly, the DNA at the −10 region unwinds to form an open complex. The sigma factor dissociates from the core enzyme after transcription is initiated.

complementary and antiparallel to the template DNA. As elongation of the mRNA continues, single-stranded mRNA is released, and the two strands of DNA behind the transcription bubble resume their double helical structure. As shown in figure 13.24, RNA polymerase is a remarkable enzyme capable of several activities, including unwinding the DNA, moving along the template, and synthesizing RNA.




**Key points:**

- RNA polymerase slides along the DNA, creating an open complex as it moves.
- The template strand is used to make a complementary copy of RNA as an RNA–DNA hybrid.
- The RNA is synthesized in a 5′ to 3′ direction using ribonucleoside triphosphates as precursors. Pyrophosphate is released (not shown).
- The complementarity rule is the same as the A:T/G:C rule except that U is substituted for T in the RNA.

**Figure 13.27 The "Transcription Bubble."**

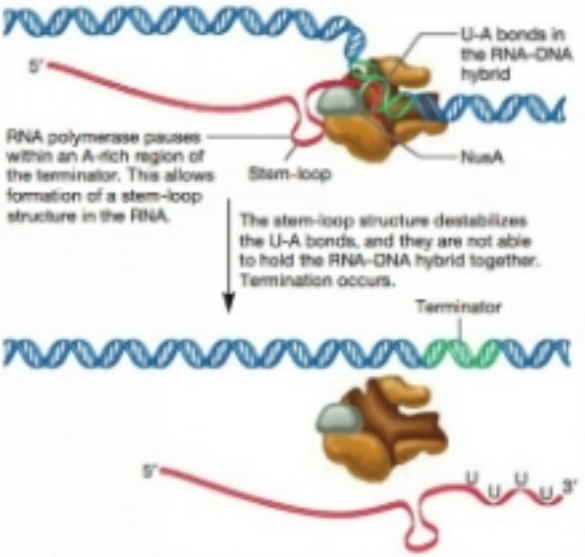



**Figure 13.28 Factor-Independent Termination of Transcription.** This type of terminator contains an inverted repeat and an A-rich sequence downstream from the repeat. A protein called NusA stimulates termination.

Termination of transcription occurs when the core RNA polymerase dissociates from the template DNA. This is brought about by the terminator. There are two kinds of terminators. The first type causes factor-independent termination (figure 13.28). This terminator consists of an inverted repeat followed by an A-rich nucleotide sequence. RNA polymerase transcribes the inverted repeat, but it pauses within the A-rich region. This allows the inverted repeat to fold back on itself, forming a hairpin-shaped stem-loop structure. The A-U base pairs holding the DNA and RNA together in the transcription bubble are too weak to hold the RNA:DNA duplex together and RNA polymerase falls off.

The second kind of terminator is termed factor-dependent terminator because it requires the aid of a protein. The best-studied termination factor is rho factor (ρ). Rho factor can be involved in transcription termination of all types of genes, but its action is best studied for protein-coding genes. Current models propose that rho binds to mRNA at a site called *rut* for *rho*-utilization site. For rho to bind, *rut* must be free of ribosomes, as shown in figure 13.29. Rho uses energy supplied by ATP hydrolysis to move along the mRNA, as it tries to catch up with RNA polymerase. However, rho's rate of movement is slower than that of RNA polymerase. Thus rho can only catch up with RNA polymerase if the polymerase pauses at a rho-dependent pause site. If this occurs, rho catches up with RNA polymerase and causes RNA polymerase to dissociate from DNA. How rho does this is not completely clear. However, it is known that rho factor has hybrid RNA:DNA helicase activity. This activity may cause unwinding of the mRNA-DNA complex. *mRNA Synthesis*

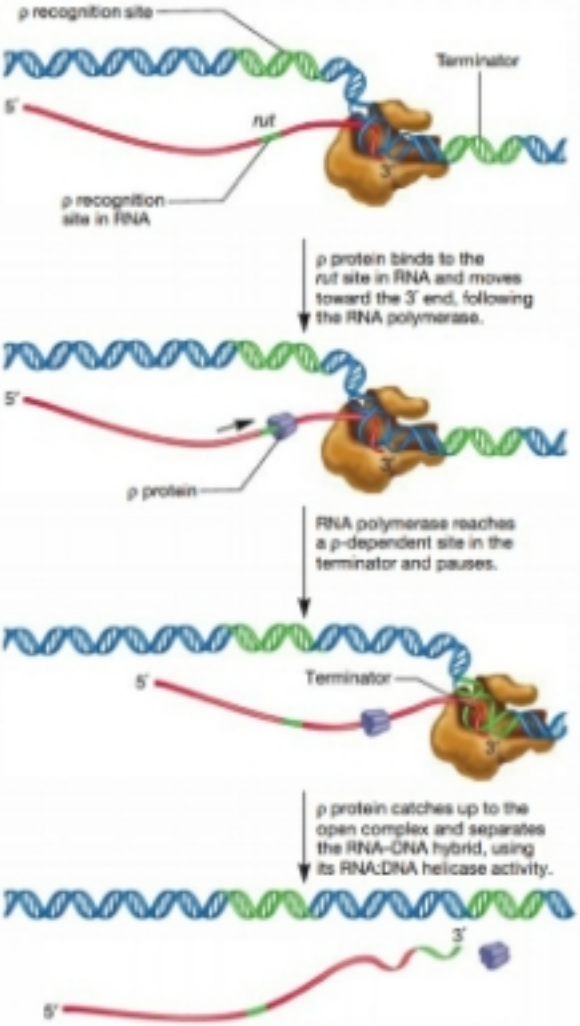



**Figure 13.29 Rho-Factor (ρ)-Dependent Termination of Transcription.** The *rut* site stands for *rho* utilization site.

### *Retrieve, Infer, Apply*

1. Outline the transcription cycle observed in bacteria. In which steps do the following function: RNA polymerase core enzyme, sigma factor, and RNA polymerase holoenzyme?
2. What is a polycistronic mRNA? How are polycistronic mRNAs formed?
3. What is a consensus sequence? What is the importance of the existence of different sigma factors and consensus sequences in bacterial promoters?
4. Tabulate the similarities and differences between factor-dependent and factor-independent transcription termination.

2. ***RNA polymerase*** basically catalyses the formation of phosphodiester bonds between successive nucleotides of a polynucleotide chain during synthesis of both DNA and RNA. With sigma factor, it contributes to the synthesis of RNA, because sigma factor recognises promoter site.
3. Enzyme ***RNA polymerase*** catalyses the transfer of nucleoside monophosphate from nucleoside triphosphate (NTP) to 3′ end of the growing polynucleotide chain.

***RNA polymerase*** lacks proof reading 3′→5′ exonuclease activity. Therefore, one error for every $10^4$ to $10^5$ ribonucleotides incorporated is introduced during RNA transcription. But the mistake during RNA transcription is not serious because of its high turnover and Wobble pairing during translation.

## Transcription Unit

The transcription unit is a stretch of nucleotides in DNA that is transcribed into a single functional RNA molecule. It is called an **operon** in prokaryotes and a **gene** in eukaryotes. A typical transcription unit has following essential parts:

- A promoter region
- A start point or initiation sequence
- A coding segment
- A terminator sequence

***RNA polymerase*** binds to the promoter. Transcription begins at the start point, progresses along the length of coding sequence and terminates at the termination sequence or at termination codon. The transcription of mRNA starts in 5′→3′ direction on 3′→5′ strand of DNA.

## Promoter Site or Promoter Region and Functional or Consensus Sequences in Promoter Region in Prokaryotes

Promoter site is a segment of regulatory DNA located just preceding the initiation codon of a cistron or gene, beyond the 3′ end of coding DNA. Enzyme ***RNA polymerase*** binds to this site for the initiation of transcription. This stretch of DNA in front of structure gene is also called **upstream region** and the sequence of bases in this region is referred to **upstream sequence**. It is a highly variable site having specific sequence of 20–200 base pairs. In *E. coli*, it consists of about 70 base pairs located before the transcription start site. The promoter region in prokaryotes has two or three important **functional sequences**. These are called **consensus sequences**,

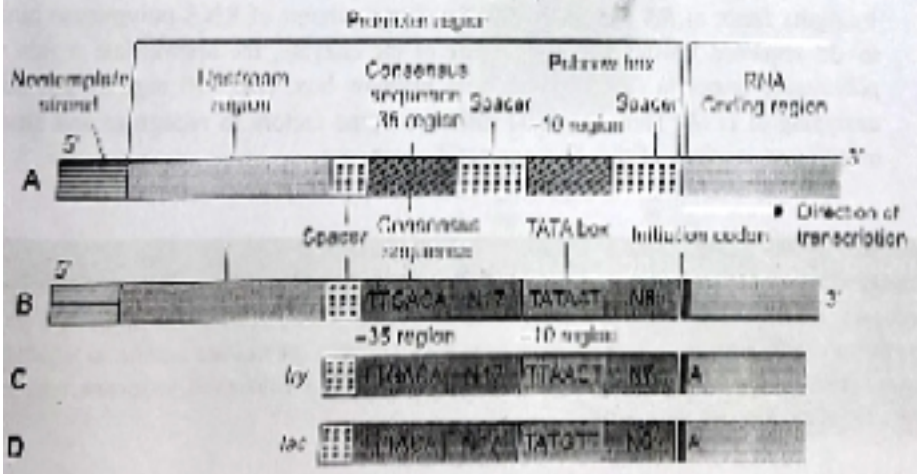



FIGURE 11.3

Base sequence of promoter site in coding or nontemplate DNA of *E. coli*.
A. Different regions of a promoter site;
B. Consensus sequences in promoter region -35 region and pribnow box (or TATA box) -10 region;
C. Promoter region of trytophan gene;
D. Promoter region of lac gene

## 15.3 Mechanisms of Gene Regulation at Transcription Level

### Induction and Repression

A set of genes will be switched on when there is necessity to metabolise a new substrate. When these genes are switched on, enzymes are produced. This phenomenon is called **induction**.

Similarly, when a metabolite needed by bacterium is provided in excess in the medium, bacterium stops its synthesis and the gene associated with its metabolism is turned off. This is called **feedback repression**.

### 1. Induction and Inducible System (Inducible Operon)

In genetic induction the gene or genes are induced or 'switched on' to produce mRNA, needed for the synthesis of required enzyme or enzymes. The substance which induces the gene for protein synthesis or enzyme production is known as **inducer**. The phenomenon of induction can be demonstrated by growing *E. coli* in different nutritional media as follows:

- When *Escherichia coli* are grown in a culture medium containing glycerol, they produce all those enzymes, which are needed for the breakdown of glycerol. The synthesis of all other enzymes is maintained at the minimum level.
- When these *E. coli* are grown in a medium containing lactose, the synthesis of enzyme ***β-galactosidase*** increases manyfolds. This enzyme hydrolyses lactose into glucose and galactose. The synthesis of other enzymes is reduced considerably.
- When *Escherichia coli* are grown on glucose medium, they contain just traces of ***β-galactosidase*** (an enzyme that hydrolyses lactose into glucose and galactose). If these bacteria are transferred into a medium containing lactose, the concentration of this enzyme increases manyfold that enables them to metabolise lactose. Here substrate lactose has acted as inducer activating the specific gene to synthesise the required enzyme.

In the above experiments substrates glycerol and lactose act as inducers for the synthesis of required enzymes. The enzymes whose synthesis can be induced by adding substrate are called **inducible enzymes**. The gene complex responsible for the synthesis of inducible enzymes is called '**inducible system**'.

Analysis of total mRNA present in the cells before and after addition of lactose shows that no *lac* mRNA is present before lactose is added to the medium. Addition of lactose triggers synthesis of *lac* mRNA. Here lactose is an **inducer** and genes associated with the enzymes of lactose metabolism system form **lac inducible system**.

### 2. Repression or Repressible System

In repression the activity of gene or genes is suppressed and the synthesis of specific protein is stopped or reduced. The substance which stops or suppresses protein synthesis is known as **repressor**.

For example, in minimal culture medium, *E. coli* cells can synthesise all the enzymes needed for the synthesis of different amino acids. But when a particular amino acid is added in the culture medium, the enzymes needed for its synthesis are not produced. When histidine is added in the bacterial culture medium, the production of histidine-synthesising enzymes is stopped. The enzymes, whose synthesis is checked by the addition of end product, are known as **repressible enzymes**. The genes associated with these enzymes form **repressible system** and the substance that represses the synthesis of repressible enzymes is called a **repressor**.

**Francois Jacob** in collaboration with **Jacques Monod** presented 'Operon concept' and concept of messenger RNA in 1961. He shared the 1965 **Nobel Prize** along with **Lwoff**.

Francois Jacob and Jacques Monod served heroically during World War II, Jacob as a member of the Free French Forces, injured seriously in Normandy in August 1944 and **Monod** as a leader of the Paris Resistance Force. **Monod** had already obtained his Ph.D but Jacob had to wait until 1947 to gain his MD from the university of Paris. Both spent the major part of their career at the Pasteur Institute, the famous Paris Research Centre set up around Louis Pasteur in the late nineteenth century and still one of the most influential European Laboratories.

Jacques Lucien Monod
(Born on 9th June 1910)

Francois Jacob
(Born on 17th June 1920)

**Francois Jacob** contributed to some of the early works on bacteriophages with **Andre Lwoff** and **Elie Wollman** prior to his collaboration with Monod that led to the Operon theory and the concept of messenger RNA in 1961. The two shared the 1965 Nobel Prize with Lwoff. Both authored a number of good books.

As a matter of fact, some enzyme are normally present in the cell but cease to be synthesised when high concentrations of their end products are present. These end products are called **corepressors**. The regulator gene produces a substance called **aporepressor**. The aporepressor unites with corepressor to form the functional **repressor molecule**. This repressor molecule inhibits transcription of mRNA from all those structural genes which are responsible for the synthesis of enzymes of synthetic pathway. Most repressible enzyme systems occur in anabolic pathway.

**TABLE 15.2: Differences between induction and repression**

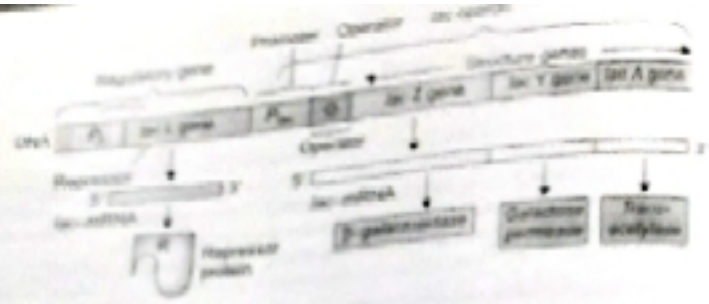
| Induction | Repression |
|---|---|
| 1. Induction involves switching on the specific gene to produce RNA for the synthesis of enzyme required. | 1. Repression is switching off the gene or operon which was actively producing an enzyme that is no more required by the cell. |
| 2. It is responsible for starting transcription and translation. | 2. It stops transcription and consequently the translation. |
| 3. It needs an **inducer**. | 3. It needs a **suppressor** which acts as a **feedback inhibitor**. |
| 4. When a new metabolite enters the system, it needs some new enzymes for its metabolism. Therefore, concerned genes are switched on. | 4. When end products of a metabolic reaction accumulate in the system further reaction is not needed and so also the associated enzymes. Hence, concerned genes are turned off. |
| 5. Inducer does not permit the repressor to bind with operator gene. | 5. Repressor is helped by a corepressor to combine with the operator gene. |
| 6. Induction is associated with catabolic pathway. | 6. Repression acts in an anabolic pathway. |

**1. The Lactose System in *E. coli*:** **Jacob** and **Monod** (1961), while studying catabolism of lactose in *E. coli*, suggested that the action of most genes is regulated at the transcription level by **induction** and **repression** phenomena. Three enzymes, namely, ***β galactosidase***, ***galactose permease*** and ***thiogalactoside transacetylase*** are needed for lactose catabolism. The genes or cistrons (DNA segments) for the above three enzymes are represented by *lac* Z, *lac* Y and *lac* A. These are called **structure 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**.

### Structure of Lac Operon System

**Francois Jacob** and **Jacques Monod** explained regulation of the activity of these lac genes by **operon model**. The operon model for lactose catabolism is called **lac operon**. It consists of two kinds of genes.

FIGURE 15.2
Genetic map of lac operon in E. coli




1. **Structure Genes:** These genes are segments of DNA that are associated with the synthesis of those enzymes that are needed for catabolism of lactose. In **lac operon** there are three structure genes:
   - ***lac Z* gene** for *β-galactosidase enzyme*
   - ***lac Y* gene** for *galactoside permease enzyme* and
   - ***lac A* gene** for *thiogalactoside transacetylase*

   These genes produce a single polycistronic mRNA molecule as shown in Fig. 15.2.
2. **Control Genes:** These genes control the activity of structure genes and lie immediately adjacent to the structure genes. There are three control genes, namely:
   - **Operator gene (O):** It lies just before, i.e. upstream of first structural gene and overlaps the promoter sequence. It controls transcription of mRNA from structure genes. It is under the control of repressor molecule, produced by the regulator or repressor gene.
   - **Promoter gene (P):** 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 structure genes. It controls rate of mRNA synthesis.
   - **Repressor or regulator gene (L):** It lies outside the operon and produces a **repressor substance.** This repressor binds to operator gene and suppresses the transcription of structure genes.

| | |
|---|---|
| **Gene Expression** | Activating a gene to make a protein |
| **Regulatory proteins** | Proteins bind to regulatory elements located near promoters and interact with RNA polymerase that regulate DNA transcription<br>(i) Activators: Regulatory proteins that promote transcription by helping RNA polymerase interact with promoter.<br>(ii) Repressors: Regulatory proteins that prevent transcription by preventing RNA polymerase from transcribing. |
| **Promoter** | Region of a gene where a RNA polymerase binds to initiate transcription of the gene. |
| **RNA polymerase** | An enzyme that helps produce RNA during transcription |
| **Operon** | Region of DNA containing genes that code for proteins with a particular function (mainly for prokaryotic cells). |
| **Operator** | Part of the operon where regulatory proteins bind. |

## 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:

1. In the absence of an inducer-lactose, the regulator gene L produces a repressor protein (R) which binds strongly to the operator site and prevents its transcription. As a result, the structure genes do not synthesise mRNA and proteins are not formed.
2. When an inducer-lactose is introduced in the medium, it enters the cell and gets modified in such a way that it binds to the repressor. The repressor now fails to bind to the operator. The operator is, therefore, free and induces the ***RNA polymerase*** to bind to the initiation site on promoter and to transcribe the cistrons into a polycistronic *lac* mRNA. This codes for the three enzymes necessary for lactose catabolism.

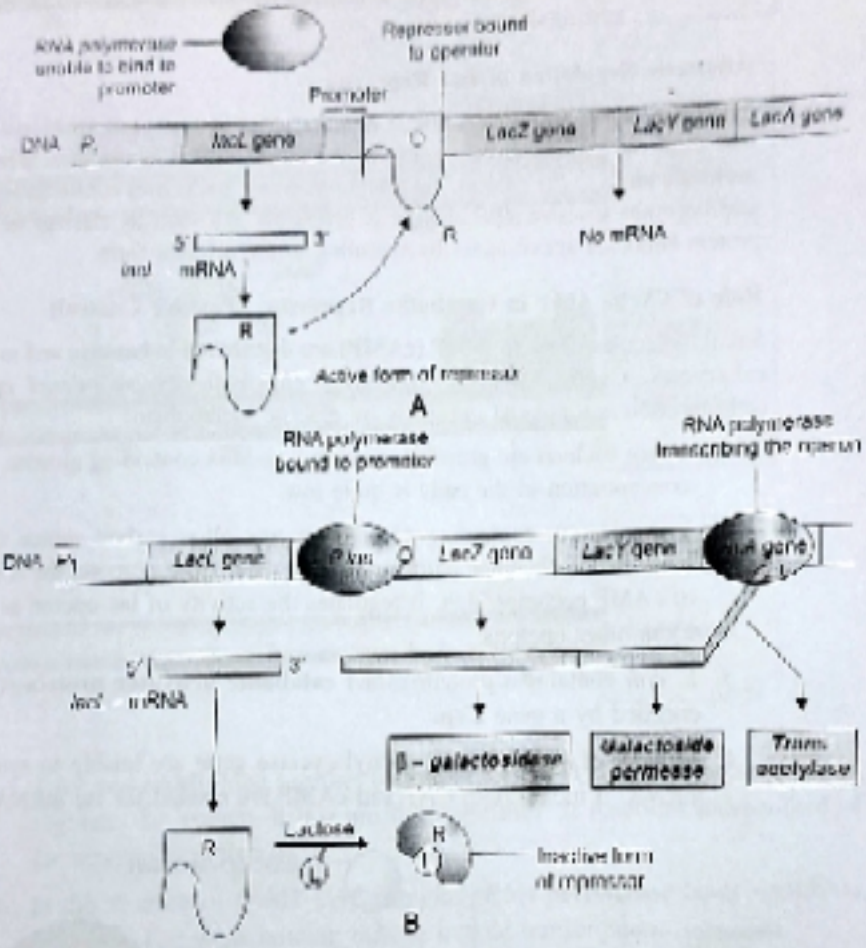



FIGURE 15.3

Function of lac operon in *Escherichia coli*.
A. Lactose absent, repressor active as bound to operator, operon repressed, transcription blocked and no mRNA synthesised.
B. Lactose present, repressor bound to lactose and inactivated, repressor no longer binds the operator, operon derepressed, transcription begins and mRNA synthesised.

## Definition of Operon

An operon is a group of closely linked structure genes (or cistrons) and the associated control genes (operator and promoter genes) which regulate the genetically controlled metabolic activity.

## Structure of Operon

The genes composing operon are classified into two categories:

1. Structure Genes: The structure genes are those segments of DNA which carry codes for the synthesis of proteins. These genes determine the primary structure of polypeptide chain by controlling the sequence of amino acids during protein synthesis.

2. Control Genes: These control the activity of structure genes either by induction or suppression. These genes are:

   - **Regulator gene:** The regulator gene (L) produces some specific enzyme which acts as a **repressor substance**. In case of lac operon, the regulator gene codes for a protein which strongly binds to the operator gene and suppresses its activity.
   - **Promoter gene:** The promoter gene (P) is the DNA segment at which ***RNA-polymerase*** binds and initiates the transcription of the structure genes.
   - **Operator gene:** The operator gene (O) is the segment of DNA, which exercises a control over transcription. The repressor substance produced by the regulator gene binds with the operator gene. It lies close to the structural gene.

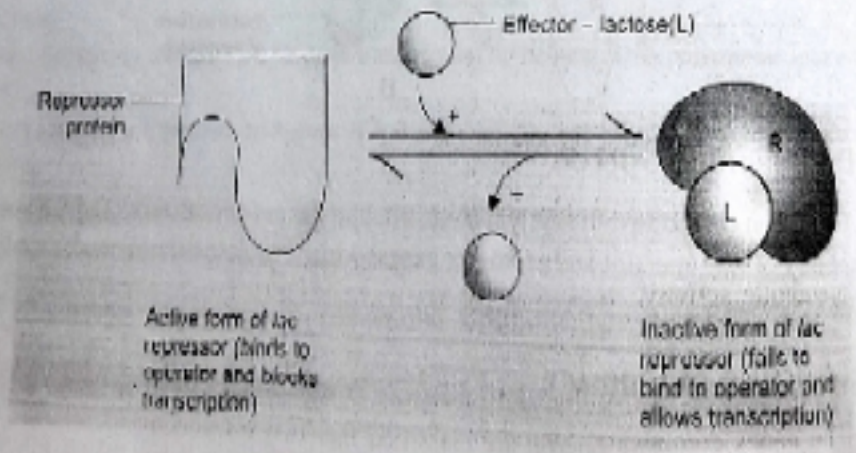
### Allosteric Regulation of *Lac* Repressor

Repressor proteins are allosteric, i.e. these exist in two conformational forms. In one form the protein is active and in the other form it is inactive. When effector molecule binds to the protein, the protein molecule undergoes conformational change and becomes inactive. But change is reversible. As soon as effector moves away, protein becomes active again by returning to the original form.

### Role of Cyclic AMP in Catabolite Repression (Positive Control)

Small molecules of **cyclic AMP (cAMP)** are distributed in bacteria and **multicellular** eukaryotes. Cyclic AMP is synthesised enzymatically by ***adenyl cyclase***. Its concentration is regulated indirectly by glucose metabolism.

1. When bacteria are growing in culture medium containing glucose, the cAMP concentration in the cells is quite low.
2. In a medium containing glycerol or any other carbon source that cannot enter the biochemical pathway used to metabolise glucose, the concentration of cAMP becomes high. It regulates the activity of lac operon as well as of some other operons.
3. *E. coli* contains a protein called **catabolite activator protein (CAP).** It is encoded by a gene **Crp**.
4. Mutants of either Crp or adenyl cyclase gene are unable to synthesise lac mRNA. It means both CAP and cAMP are needed for lac mRNA synthesis.



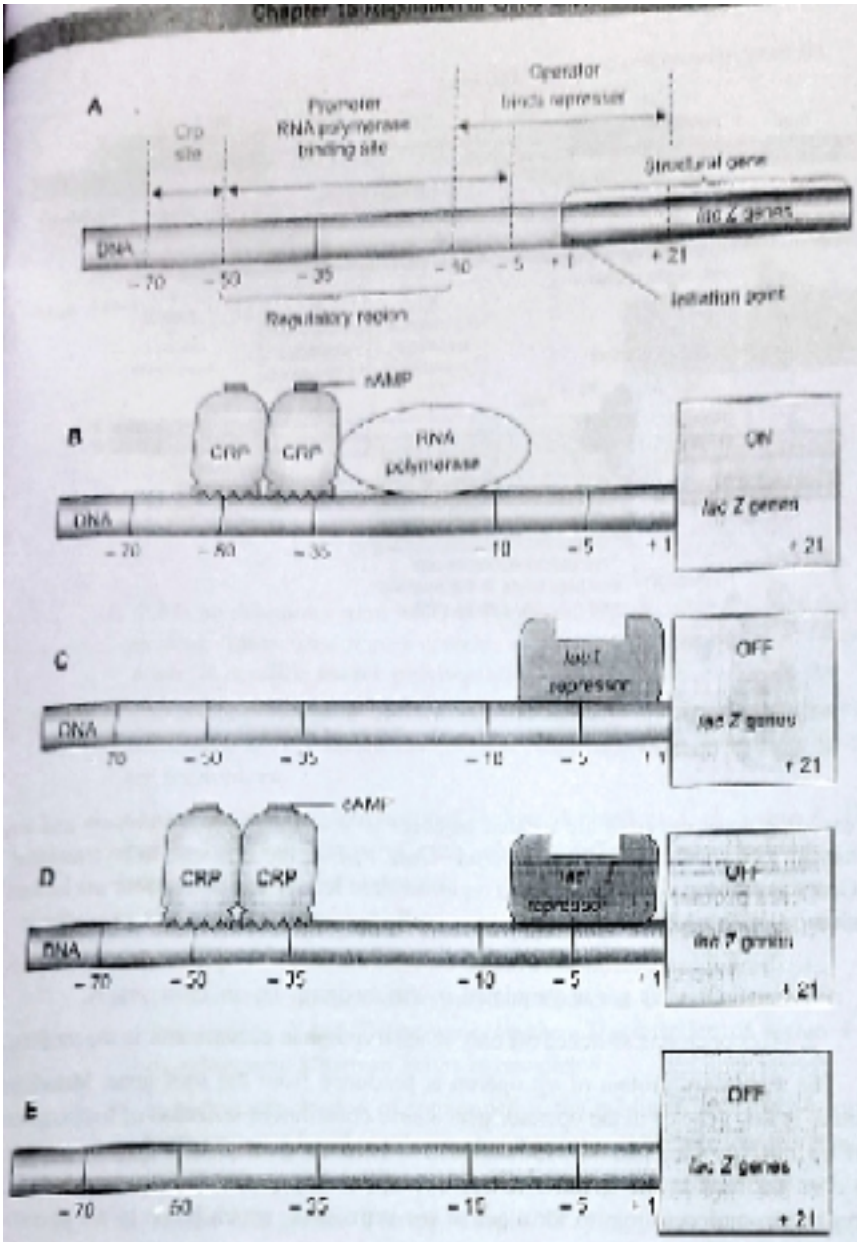



FIGURE 15.5

Role of CRP in positive control of transcription of *lac* operon i.e. (CRP is cAMP – CRP's complex, CRP is cAMP receptor protein):

A. Diagram showing binding sites for various regulatory components on bacterial chromosome near lac Z gene;

B. Glucose present, CRP binds to CRP's site, *lac* I repressor removed, RNA polymerase binds to DNA and *lac* Z gene is switched on;

C. Glucose present, CRP absent, lactose absent, *lac* I repressor binds and *lac* Z gene switched off;

D. Both glucose and lactose are absent, *lac* Z gene remains switched off, though Crp is present;

E. Glucose absent, Crp absent, lactose present hence *lac* repressor removed but *lac* Z gene remains switched off because RNA polymerase does not bind to DNA due to the absence of CRP.

5. *CAP* and *cAMP* bind together forming **cAMP CAP** unit. The complex unit regulates *lac* system. It is a **positive regulator**. It functions independent of lac repression system.

6. In the absence of cAMP-CAP complex, ***RNA polymerase*** binds weakly to promoter. The weak binding fails to initiate transcription.

7. *cAMP Crp* complex binds to a base sequence in DNA in the promoter region to initiate transcription.

## yptophan Operon: A Repressible Operon System

yptophan operon (*trp* operon) in *E. coli* is responsible for controlling synthesis amino acid **tryptophan**. Its synthesis is carried out by five enzymes. The genes

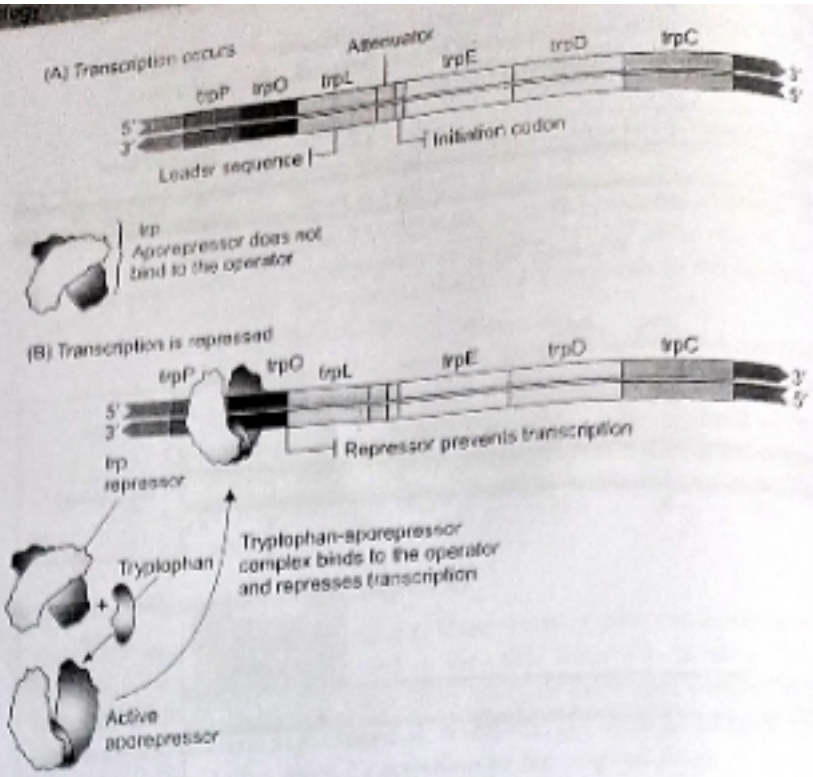



FIGURE 15.6

*Functioning of tryptophan operon of E. coli:*

A. When tryptophan is absent, the repressor does not bind to operator and operon actively transcribes mRNA for required enzymes.

B. When tryptophan is present, it binds to repressor; the activated repressor binds to operator and operon is repressed.

encoding these enzymes are located adjacent to one another in succession and are named *trpE, trpD, trpC, trpB* and *trpA*. *Gene trpE* is the first one to be translated. Genes **promoter, operator** and two regions called **leader** and **attenuator** are located close to gene ***trpE***. The repressor gene ***trpR*** is present far from this gene cluster.

Tryptophan operon consists of:

- a promoter
- an operator
- a leader (trpL)
- an attenuator and 5 structural genes (trpE, trpD, trpC, trpB and trpA)

1. Tryptophan operon is switched on for normal transcription when tryptophan amino acid is not accumulated in the medium, i.e. in short supply.
2. This operon is switched off only when tryptophan accumulates in the medium.

The regulatory protein of *trp* operon is produced from the *trpR* gene. Mutations either in this gene or in the operator gene cause constitutive initiation of transcription of trp mRNA. This protein synthesised from mutant gene is called **aporepressor**. It does not bind to the operator unless tryptophan is present. The aporepressor and tryptophan molecule join to form active *trp* **repressor**, which binds to the operator to switch it off.

## Attenuation: Regulation after the Initiation of Transcription

In most cases regulatory mechanisms control the initiation of transcription. In prokaryotes, some regulatory mechanism exercises control following the initiation. **Charles Yanofsky** (1972) and his colleagues found that *trp* operon has a regulatory site between promoter/operator and the first structural gene *trp E*. This site of DNA is called **leader sequence (L)**. It is non-coding DNA segment which is transcribed into leader mRNA at 5' end of mRNA. This *trp* leader mRNA is 162 nucleotides long. It is sensitive to tryptophan level and exercises an additional control on the transcription of polycistronic trp mRNA. The effect of this control element, i.e. of leader sequence is called **attenuation**, because it reduces or attenuates synthesis of mRNA.

## Structure of Leader mRNA

The 162 base long leader mRNA has following segments:

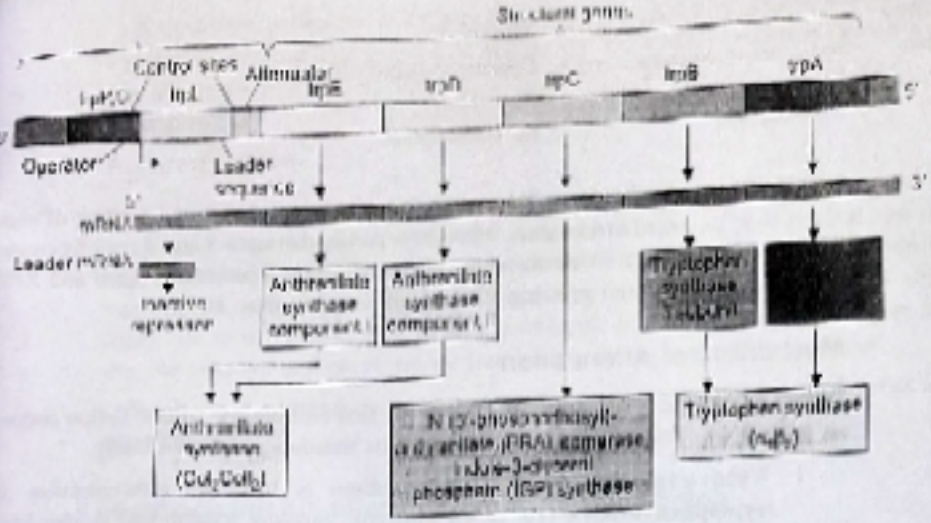



FIGURE 15.7
Structure of tryptophan operon

1. It has an initiation codon **AUG** at 5′ end and a stop codon **UGA** in the same reading frame. This region encodes a polypeptide, consisting of 14 amino acids. It is called **leader polypeptide**.
2. The leader polypeptide mRNA contains two adjacent **tryptophan codons** at position 10 to 11. It means the 10th and 11th amino acids in leader polypeptide are tryptophans.
3. Leader mRNA has four distinct regions denoted as 1, 2, 3 and 4. These regions can base pair with each other in different ways, forming several **hairpin** structures. For example:
   - In a **stable form** the free leader mRNA shows base pairing between 1 and 2 and between 3 and 4 regions.
   - In low concentration of tryptophan, ribosome is stalled in region 1 so that regions 2 and 3 base pair together. Transcription of region 4 and of subsequent structure genes is completed.
   - In high concentration of tryptophan in the medium, the 3 and 4 regions get base paired and synthesis of protein is interrupted on reaching the 2nd region.
4. After region 4, is present a string of eight U's lying just prior to the structure gene *trp E*.

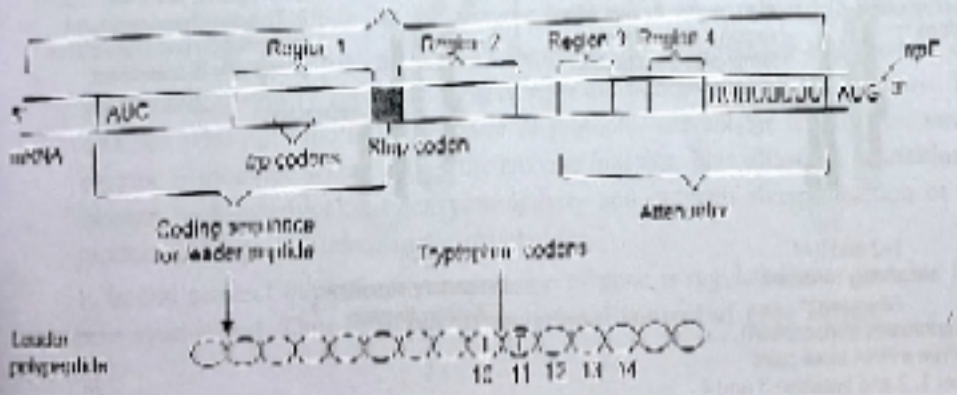



FIGURE 15.8
Structure of leader segment of trp mRNA

with the aid of RNA-protein complexes called snRNPs, assembled into spliceosomes. A fourth class of introns, found in some tRNAs, is the only class known to be spliced by protein enzymes.

- Ribosomal RNAs and transfer RNAs are derived from longer precursor RNAs, trimmed by nucleases. Some bases are modified enzymatically during the maturation process.
- The self-splicing introns and the RNA component of RNase P (which cleaves the 5′ end of tRNA precursors) are two examples of ribozymes. These biological catalysts have the properties of true enzymes. They generally promote hydrolytic cleavage and transesterification, using RNA as substrate. Combinations of these reactions can be promoted by the excised group I intron of *Tetrahymena* rRNA, resulting in a type of RNA polymerization reaction.
- Polynucleotide phosphorylase reversibly forms RNA-like polymers from ribonucleoside 5′-diphosphates, adding or removing ribonucleotides at the 3′-hydroxyl end of the polymer. The enzyme degrades RNA in vivo.

## 26.3 RNA-Dependent Synthesis of RNA and DNA

In our discussion of DNA and RNA synthesis up to this point, the role of the template strand has been reserved for DNA. However, some enzymes use an RNA template for nucleic acid synthesis. With the very important exception of viruses with an RNA genome, these enzymes play only a modest role in information pathways. RNA viruses are the source of most RNA-dependent polymerases characterized so far.

The existence of RNA replication requires an elaboration of the central dogma (Fig. 26–28; contrast this with the diagram on p. 922). The enzymes involved in RNA replication have profound implications for investigations into the nature of self-replicating molecules that may have existed in prebiotic times.

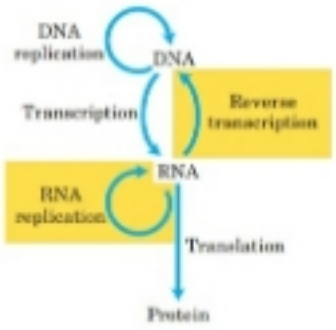



**FIGURE 26–28** Extension of the central dogma to include RNA-dependent synthesis of RNA and DNA.

### Reverse Transcriptase Produces DNA from Viral RNA

Certain RNA viruses that infect animal cells carry within the viral particle an RNA-dependent DNA polymerase called **reverse transcriptase.** On infection, the single-stranded RNA viral genome (~10,000 nucleotides) and the enzyme enter the host cell. The reverse transcriptase first catalyzes the synthesis of a DNA strand complementary to the viral RNA (Fig. 26–29), then degrades the RNA strand of the viral RNA-DNA hybrid and replaces it with DNA. The resulting duplex DNA often becomes incorporated into the genome of the eukaryotic host cell. These integrated (and dormant) viral genes can be activated and transcribed, and the gene products—viral proteins and the viral RNA genome itself—packaged as new viruses. The RNA viruses that contain reverse transcriptases are known as **retroviruses** (*retro* is the Latin prefix for "backward").

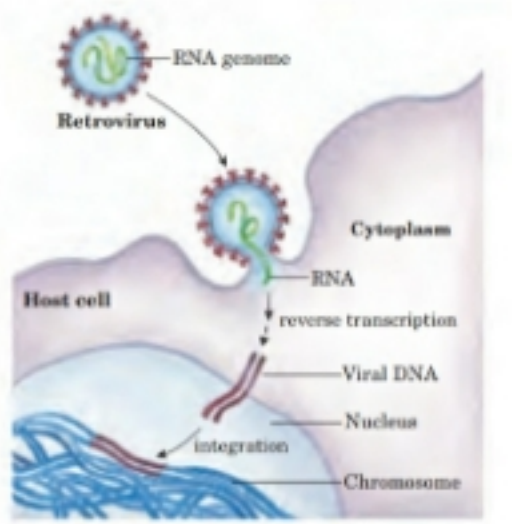



**FIGURE 26–29** **Retroviral infection of a mammalian cell and integration of the retrovirus into the host chromosome.** Viral particles entering the host cell carry viral reverse transcriptase and a cellular tRNA (picked up from a former host cell) already base-paired to the viral RNA. The tRNA facilitates immediate conversion of viral RNA to double-stranded DNA by the action of reverse transcriptase, as described in the text. Once converted to double-stranded DNA, the DNA enters the nucleus and is integrated into the host genome. The integration is catalyzed by a virally encoded integrase. Integration of viral DNA into host DNA is mechanistically similar to the insertion of transposons in bacterial chromosomes (see Fig. 25–43). For example, a few base pairs of host DNA become duplicated at the site of integration, forming short repeats of 4 to 6 bp at each end of the inserted retroviral DNA (not shown).

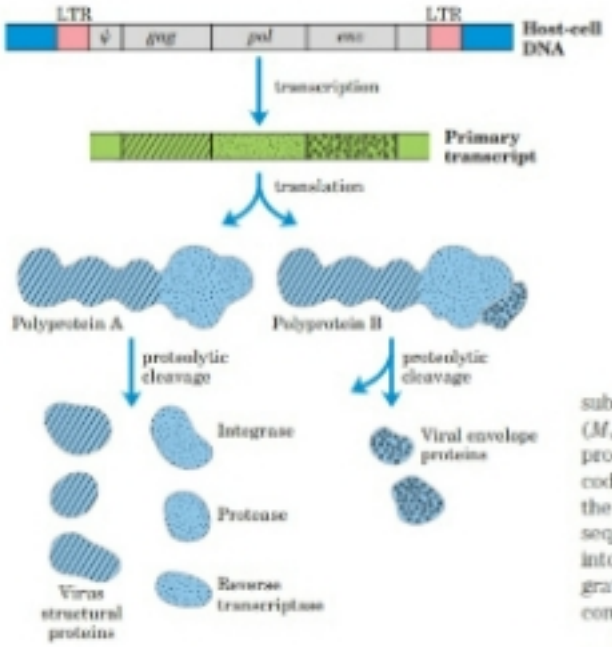



**FIGURE 26–30 Structure and gene products of an integrated retroviral genome.** The long terminal repeats (LTRs) have sequences needed for the regulation and initiation of transcription. The sequence denoted Ψ is required for packaging of retroviral RNAs into mature viral particles. Transcription of the retroviral DNA produces a primary transcript encompassing the *gag*, *pol*, and *env* genes. Translation (Chapter 27) produces a polyprotein, a single long polypeptide derived from the *gag* and *pol* genes, which is cleaved into six distinct proteins. Splicing of the primary transcript yields an mRNA derived largely from the *env* gene, which is also translated into a polyprotein, then cleaved to generate viral envelope proteins.

Howard Temin, 1934–1994

David Baltimore

The existence of reverse transcriptases in RNA viruses was predicted by Howard Temin in 1962, and the enzymes were ultimately detected by Temin and, independently, by David Baltimore in 1970. Their discovery aroused much attention as dogma-shaking proof that genetic information can flow "backward" from RNA to DNA.

Retroviruses typically have three genes: *gag* (derived from the historical designation *g*roup *a*ssociated antigen), *pol*, and *env* (Fig. 26–30). The transcript that contains *gag* and *pol* is translated into a long "polyprotein," a single large polypeptide that is cleaved into six proteins with distinct functions. The proteins derived from the *gag* gene make up the interior core of the viral particle. The *pol* gene encodes the protease that cleaves the long polypeptide, an integrase that inserts the viral DNA into the host chromosomes, and reverse transcriptase. Many reverse transcriptases have two subunits, α and β. The *pol* gene specifies the β subunit ($M_r$ 90,000), and the α subunit ($M_r$ 65,000) is simply a proteolytic fragment of the β subunit. The *env* gene encodes the proteins of the viral envelope. At each end of the linear RNA genome are long terminal repeat (LTR) sequences of a few hundred nucleotides. Transcribed into the duplex DNA, these sequences facilitate integration of the viral chromosome into the host DNA and contain promoters for viral gene expression.

Reverse transcriptases catalyze three different reactions: (1) RNA-dependent DNA synthesis, (2) RNA degradation, and (3) DNA-dependent DNA synthesis. Like many DNA and RNA polymerases, reverse transcriptases contain $Zn^{2+}$. Each transcriptase is most active with the RNA of its own virus, but each can be used experimentally to make DNA complementary to a variety of RNAs. The DNA and RNA synthesis and RNA degradation activities use separate active sites on the protein. For DNA synthesis to begin, the reverse transcriptase requires a primer, a cellular tRNA obtained during an earlier infection and carried within the viral particle. This tRNA is base-paired at its 3′ end with a complementary sequence in the viral RNA. The new DNA strand is synthesized in the 5′→3′ direction, as in all RNA and DNA polymerase reactions. Reverse transcriptases, like RNA polymerases, do not have 3′→5′ proofreading exonucleases. They generally have error rates of about 1 per 20,000 nucleotides added. An error rate this high is extremely unusual in DNA replication and appears to be a feature of most enzymes that replicate the genomes of RNA viruses. A consequence is a higher mutation rate and faster rate of viral evolution, which is a factor in the frequent appearance of new strains of disease-causing retroviruses.

Reverse transcriptases have become important reagents in the study of DNA-RNA relationships and in DNA cloning techniques. They make possible the synthesis of DNA complementary to an mRNA template, and synthetic DNA prepared in this manner, called **complementary DNA (cDNA),** can be used to clone cellular genes (see Fig. 9–14).

of the telomere's single strand into the duplex DNA, perhaps by a mechanism similar to the initiation of homologous genetic recombination (see Fig. 25–31). In mammals, the looped DNA is bound by two proteins, TRF1 and TRF2, with the latter protein involved in formation of the T loop. T loops protect the 3′ ends of chromosomes, making them inaccessible to nucleases and the enzymes that repair double-strand breaks (Fig. 26–35b).

In protozoans (such as *Tetrahymena*), loss of telomerase activity results in a gradual shortening of telomeres with each cell division, ultimately leading to the death of the cell line. A similar link between telomere length and cell senescence (cessation of cell division) has been observed in humans. In germ-line cells, which contain telomerase activity, telomere lengths are maintained; in somatic cells, which lack telomerase, they are not. There is a linear, inverse relationship between the length of telomeres in cultured fibroblasts and the age of the individual from whom the fibroblasts were taken: telomeres in human somatic cells gradually shorten as an individual ages. If the telomerase reverse transcriptase is introduced into human somatic cells in vitro, telomerase activity is restored and the cellular life span increases markedly.

Is the gradual shortening of telomeres a key to the aging process? Is our natural life span determined by the length of the telomeres we are born with? Further research in this area should yield some fascinating insights.

## Some Viral RNAs Are Replicated by RNA-Dependent RNA Polymerase

Some *E. coli* bacteriophages, including f2, MS2, R17, and Qβ, as well as some eukaryotic viruses (including influenza and Sindbis viruses, the latter associated with a form of encephalitis) have RNA genomes. The single-stranded RNA chromosomes of these viruses, which also function as mRNAs for the synthesis of viral proteins, are replicated in the host cell by an **RNA-dependent RNA polymerase (RNA replicase).** All RNA viruses—with the exception of retroviruses—must encode a protein with RNA-dependent RNA polymerase activity because the host cells do not possess this enzyme.

The RNA replicase of most RNA bacteriophages has a molecular weight of ~210,000 and consists of four subunits. One subunit ($M_r$ 65,000) is the product of the replicase gene encoded by the viral RNA and has the active site for replication. The other three subunits are host proteins normally involved in host-cell protein synthesis: the *E. coli* elongation factors Tu ($M_r$ 30,000) and Ts ($M_r$ 45,000) (which ferry amino acyl–tRNAs to the ribosomes) and the protein S1 (an integral part of the 30S ribosomal subunit). These three host proteins may help the RNA replicase locate and bind to the 3′ ends of the viral RNAs.

RNA replicase isolated from Qβ-infected *E. coli* cells catalyzes the formation of an RNA complementary to the viral RNA, in a reaction equivalent to that catalyzed by DNA-dependent RNA polymerases. New RNA strand synthesis proceeds in the 5′→3′ direction by a chemical mechanism identical to that used in all other nucleic acid synthetic reactions that require a template. RNA replicase requires RNA as its template and will not function with DNA. It lacks a separate proofreading endonuclease activity and has an error rate similar to that of RNA polymerase. Unlike the DNA and RNA polymerases, RNA replicases are specific for the RNA of their own virus; the RNAs of the host cell are generally not replicated. This explains how RNA viruses are preferentially replicated in the host cell, which contains many other types of RNA.

## RNA Synthesis Offers Important Clues to Biochemical Evolution

The extraordinary complexity and order that distinguish living from inanimate systems are key manifestations of fundamental life processes. Maintaining the living state requires that *selected* chemical transformations occur very rapidly—especially those that use environmental energy sources and synthesize elaborate or specialized cellular macromolecules. Life depends on powerful and selective catalysts—enzymes—and on informational systems capable of both securely storing the blueprint for these enzymes and accurately reproducing the blueprint for generation after generation. Chromosomes encode the blueprint not for the cell but for the enzymes that construct and maintain the cell. The parallel demands for information and catalysis present a classic conundrum: what came first, the information needed to specify structure or the enzymes needed to maintain and transmit the information?

The unveiling of the structural and functional complexity of RNA led Carl Woese, Francis Crick, and Leslie Orgel to propose in the 1960s that this macromolecule might serve as both information carrier and catalyst. The discovery of catalytic RNAs took this proposal from


Carl Woese


Francis Crick


Leslie Orgel