TRANSCRIPTION IN PROKARYOTES & EUKARYOTES

Dr. R. Prasad
Department of Zoology,
Eastern Karbi Anglong College,
Sarihajan

Transcription in Prokaryotes

Extensive studies on *E. coli* have provided the model for subsequent investigations of transcription in eukaryotes. *E. coli* was also the first organism in which mRNA was discovered, *RNA polymerase* was purified and studied and basic mechanisms by which transcription is regulated was elucidated by pioneering experiments in which regulated gene expression allows the cell to respond to variations in the environment such as changes in the availability of nutrients. *E. coli* has thus, provided the foundation for studies of the far more complex mechanisms that regulate gene expression in eukaryotic cells. Transcription in eukaryotes is similar to bacterial transcription in that the basic principles of transcription known in prokaryotes also apply to eukaryotes.

Transcription involves four stages:

1. Binding of *RNA polymerase* to a promoter sequence
2. Initiation of RNA synthesis
3. Elongation of the RNA chain
4. Termination of RNA synthesis

An overview of all the four stages of transcription has been depicted (Fig.12.10).

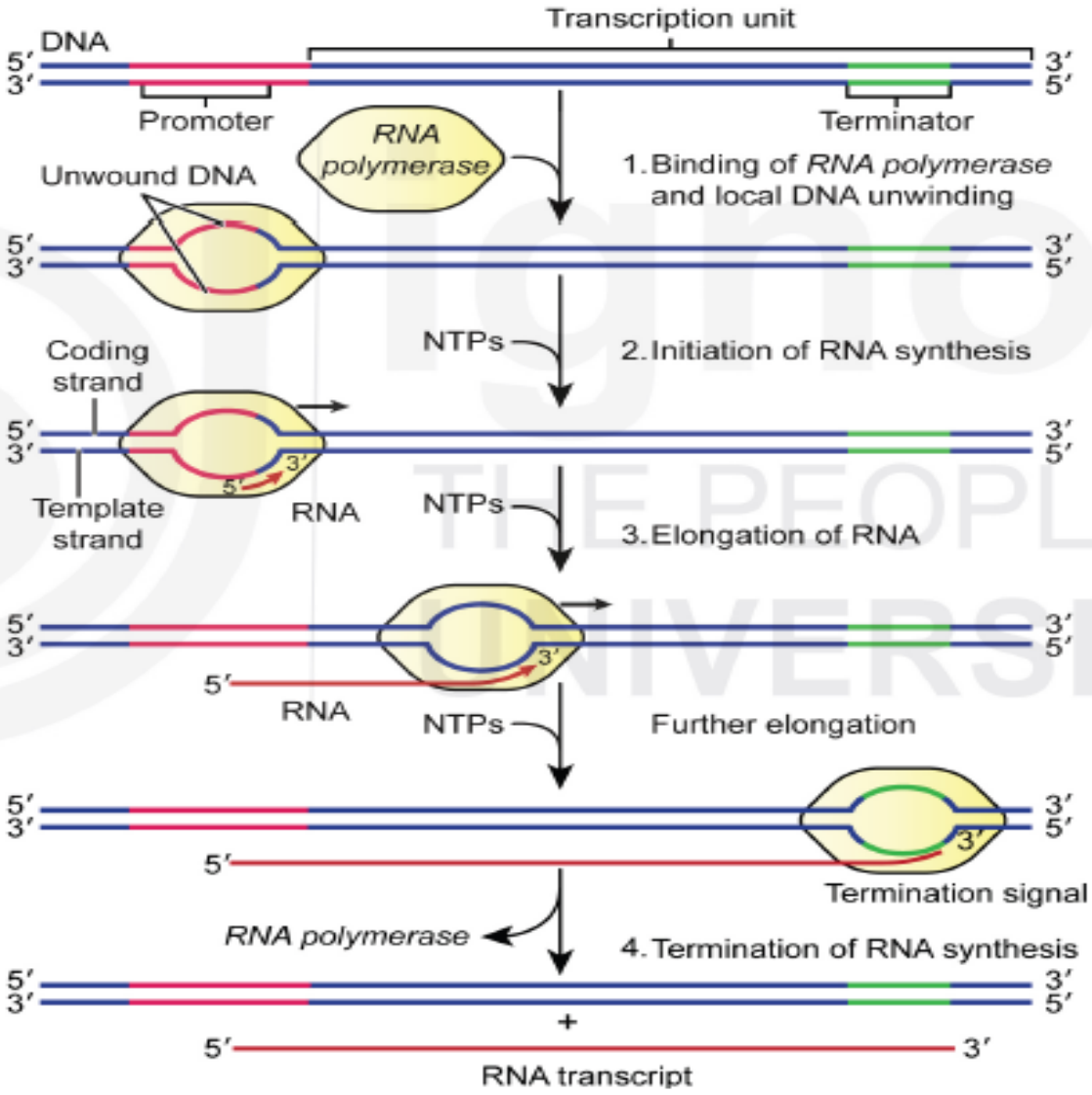

Fig 12.10: An overview of transcription showing the four main stages. Note that *RNA polymerase* moves along the template strand of the DNA in the 3′ →5′ direction and the newly formed RNA molecule grows in 5′ →3′ direction. NTPs- Nucleoside triphosphates

Binding of *RNA Polymerase* to a Promoter

The first step of RNA synthesis is the binding of *RNA polymerase* to a DNA promoter site- a specific sequence made up of several dozen base pairs. This binding determines which DNA strand will serve as the template strand, and the point from where RNA synthesis would begin.

Bacterial Promoters

In bacterial cells promoters are usually positioned adjacent to an RNA-coding sequence. Although the nucleotide sequence of many promoters of *E. coli* and other bacteria vary in sequence, short stretches of nucleotides are common to many. Furthermore, the spacing and location of the sequences relative to the transcription start site are similar in most promoters (Fig. 12.11). The most common consensus sequence found in almost all bacterial promoters is centered about 10 bp upstream of the start site, called the **-10 consensus sequence**, and popularly called as **Pribnow box.** It is simply written as TATAAT.

5′ –T A T A A T – 3′

3′ –A T A T T A – 5′

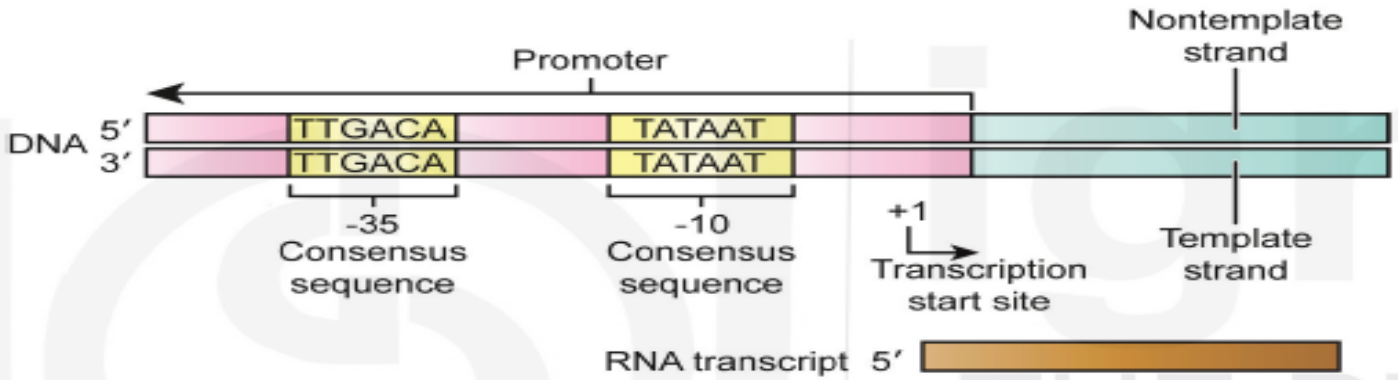

Fig. 12.11: Organization of a bacterial promoter sequence showing the transcription start point (designated + 1 or usually A) and two consensus sequences (the six nucleotides -10 sequence and the six nucleotides -35 sequence).

Another consensus sequence common to most bacterial promoters is TTGACA. It lies approximately 35 nucleotides upstream of the start site and is termed as -35 consensus sequence. The nucleotides on either side of the -10 and -35 consensus sequence or those between them vary considerably from promoter to promoter, suggesting that these nucleotides are not very important in promoter recognition.

During binding of *RNA polymerase* to promoter, the sigma factor associates with the core enzyme to form a holoenzyme which binds to the -35 and -10 consensus sequence in the DNA promoter. Although it binds only to the nucleotides of consensus sequences, the enzyme extends from -50 to +20 when bound to the promoter. The holoenzyme initially binds weakly to the promoter but then undergoes a change in structure that allows it to bind more tightly and unwind the double-stranded DNA. DNA unwinding begins within -10 consensus sequence and extends downstream for about 14 nucleotides, including the start site (from nucleotides -12 to +2).

Initiation of RNA Synthesis

After positioning itself over the start site for transcription (at position + 1), *RNA polymerase* unwinds the DNA to produce single - stranded template. The orientation and spacing of consensus sequence on a DNA strand determines which strand will be the template for transcription thereby also determining the direction of transcription. The RNA molecule is synthesized when *RNA polymerase* pairs the base on a ribonucleoside triphosphate with its complementary base at the start site on the DNA template strand as the nucleotide is added to the 3´ end of the growing RNA molecule. Unlike DNA synthesis, primer is not required to initiate the synthesis of the 5´ end of the RNA molecule. A phosphodiester bond is incorporated at the 3´end of the growing RNA molecule by cleaving two of the three phosphate groups from the ribonucleoside triphosphate. However, there is no cleavage of phosphate groups in the first ribonucleoside. This is because 5´ end of the first ribonucleoside does not take part in the formation of a phosphodiester bond.

After several abortive attempts of generating short RNA transcripts, from 2 to 6 nucleotides in length, *RNA polymerase* eventually synthesizes an RNA molecule from 9 to 12 nucleotides in length. With this initiation, the *RNA polymerase* now enters the elongation phase.

Elongation

Once the initiation of transcription is completed, there is conformational (shape) change in the *RNA polymerase*. This change helps the enzyme to escape from the promoter and begin transcription downstream. The sigma factor subunit is usually released after initiation. *RNA polymerase* progressively unwinds the DNA at the leading (downstream) edge of the transcription bubble. New nucleotides are joined to the RNA molecule according to the sequence on the template. Thereafter, the enzyme rewinds the DNA at the trailing (upstream) edge of the bubble (see Fig. 12.12). In bacterial cell at 37°C, about 40 nucleotides are added per second and this rate of RNA synthesis is much lower than the 1000 to 2000 nucleotides added per second in DNA synthesis.

Termination of RNA Synthesis

There exist specific sequences of DNA and RNA, both in prokaryotes and eukaryotes which act as transitory transcription pauses. Such RNA pauses help in coordinating RNA processing.

Elongation of the growing RNA chain continues until *RNA polymerase* copies a special sequence, called a termination signal. This triggers the end of transcription. Since most terminators are found upstream of the site at which termination actually takes place, it is believed that transcription does not stop abruptly when *polymerase* reaches a terminator. Instead, transcription stops only after the terminator has been transcribed. There are several overlapping events that take place at the terminator site which bring an end to transcription. *RNA polymerase* stops synthesizing RNA; the RNA molecule is released from *RNA polymerase*; the newly synthesized RNA molecule dissociate fully from the DNA; and *RNA polymerase* detaches itself from the DNA template.

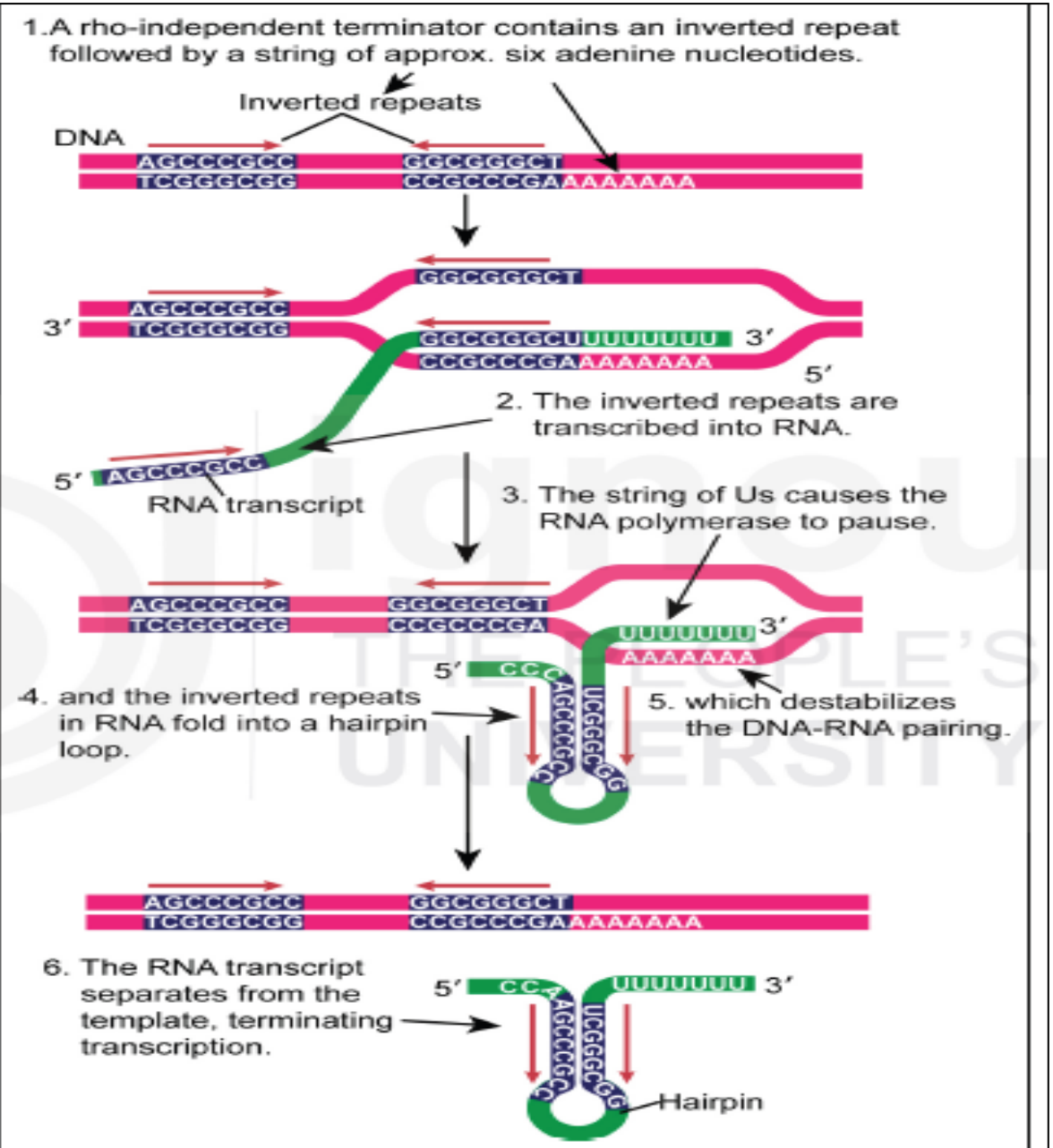

Fig. 12.13: Rho-independent termination in bacteria showing transcription of inverted repeats. The inverted repeats fold into hairpin loop resulting in separation of RNA transcript from the template, terminating transcription.

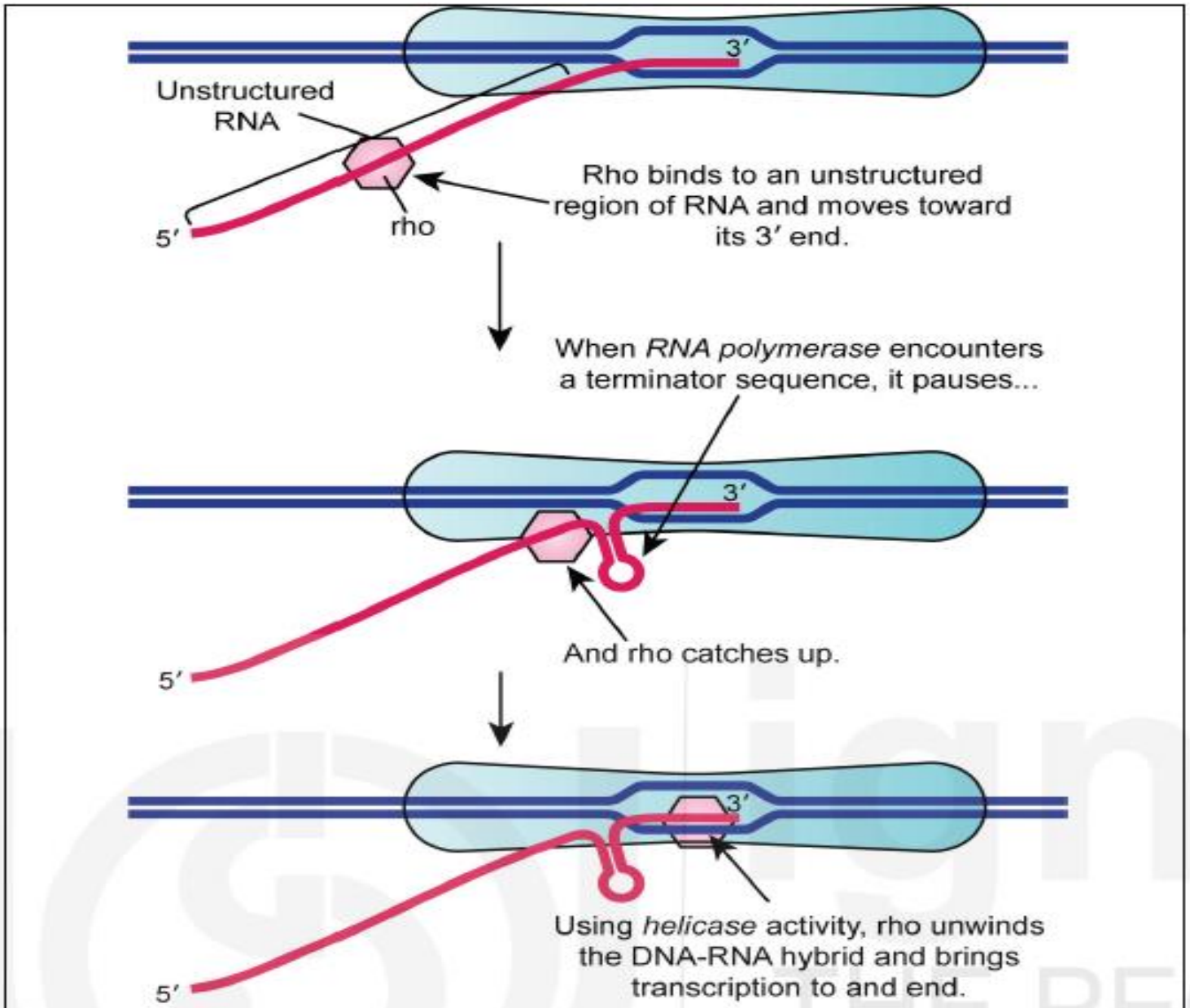

Fig. 12.14: Rho protein dependent termination of transcription in bacterial cells.

The transcription bubble

The transcription bubble is a short stretch of 18 nucleotides unwound DNA within which transcription takes place. RNA is continuously synthesized within this region, with single stranded DNA being used as a template.

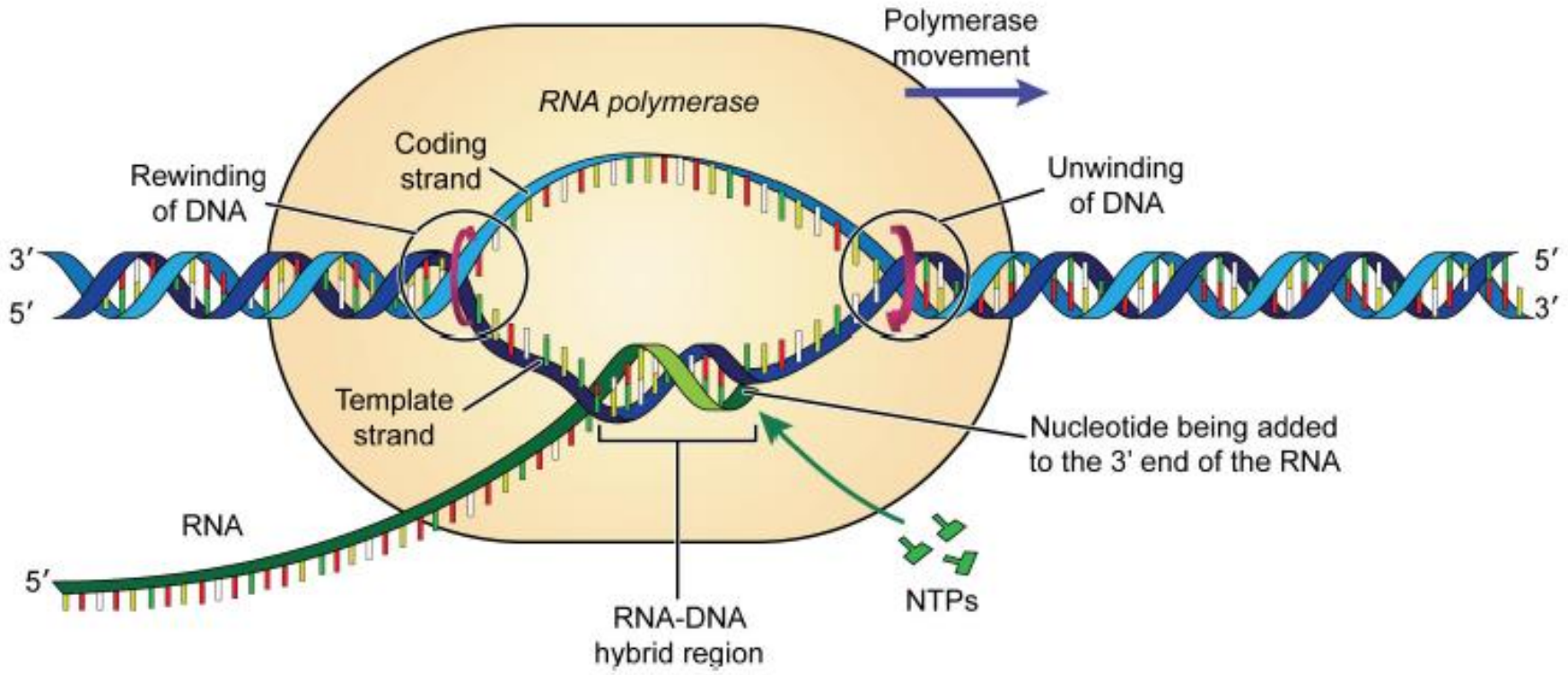

Fig. 12.12: A bacterial elongation complex showing the movement of the enzyme along the template DNA strand from the 3′ towards the 5′ end and RNA strand is elongated in the 5′ →3′ direction. NTPs- Nucleoside triphosphates

Transcription in Eukaryotes

Various steps of transcription in eukaryotes are discussed below.

Binding of *RNA polymerase* to the promoter and initiation of transcription

There are three different kinds of *RNA polymerases* (*RNA polymerase I, RNA polymerase II* and *RNA polymerase III*). These enzymes recognize different types of promoters. Promoter recognition is in fact, carried out by accessory proteins that bind to the promoter and then assign a specific *RNA polymerase* to the promoter. These accessory proteins comprise general transcription factors, which along with *RNA polymerase*, form basal transcription apparatus, which is a group of proteins that assemble near the start site and are sufficient to initiate minimum level of transcription.

Another group of accessory proteins consists of transcriptional activator proteins which bind to specific DNA sequences and bring about higher level of transcription by activating the basal transcription apparatus at the start site.

As *RNA polymerase II* is responsible for the synthesis of mRNA from protein-coding genes, it has been the focus of most studies of transcription in eukaryotes. The promoter of genes transcribed by RNA polymerase II contains several different sequence elements surrounding their transcription sites. These sequences are categorized into two groups:

1. **CORE PROMOTER**
2. **REGULATORY SEQUENCES**

1) Core promoter

There are at least four types of DNA sequences that are involved in core promoter function (see Fig.12.15). These four elements are:

i) a short initiator (Inr) sequence surrounding the transcription start point;

ii) a Goldberg-Hogness, or TATA box which consists of consensus sequence of TATA followed by two or three more A's, usually located 25 to 30 nucleotide upstream (-30) of the start site. The sequence and function of TATA boxes are analogous to those found in the – 10 promoter region of prokaryotic genes. However, unlike prokaryotes, *RNA polymerase II* cannot bind directly to this promoter region;

iii) TFIIB recognition element (BRE), which is located upstream of the transcription start site: and

iv) the downstream promoter element (DPE) located about 30 nucleotides downstream of the start point.

These four elements are organized into two general types of core promoters; TATA- driven promoter which includes an Inr sequence and TATA box with or without associated BRE, and DPE-driven promoters, which contain DPE and Inr sequences, but no TATA box or BRE.

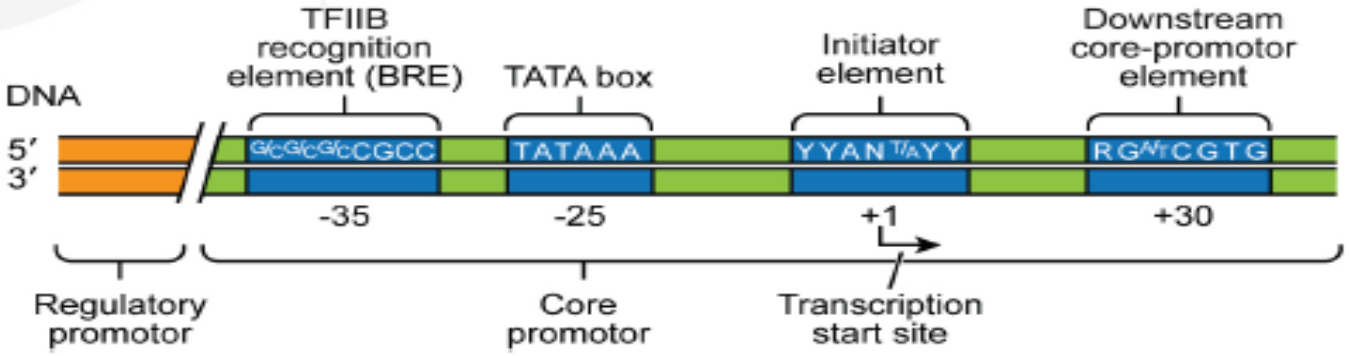

Fig.12.15: The promoters of genes transcribed by *RNA polymerase II* with core promoter and regulatory sequences.

2) Regulatory sequences

Regulatory DNA sequences include **upstream elements**, proximal promoter elements, enhancers and silencers. A core promoter (TATA-driven or DPE-driven) is capable of supporting only a basal (low level) of transcription. However, most protein–coding genes have additional short sequences further upstream called as regulatory sequences that improve promoter's efficiency.

Some of the upstream elements are common to many different genes; examples include CAAT box (consensus sequence GCCCAATCT in animals and yeasts) and the GC box (consensus sequence GGGCGG). The locations of these elements relative to a gene's start point vary from gene to gene. The elements within 100 – 200 nucleotides of the start point are often called **proximal control elements**. Other elements known as **enhancers** and **silencers** have more dramatic effects on eukaryotic gene transcription. As their name suggests, enhancers increase transcription levels and silencers decrease them. They change transcription in response to a cell's requirement for a particular gene product, or at a particular time during the development or place within the organism. Each eukaryotic gene has its own unique arrangement of proximal promoter, enhancer and silencer. The location of these elements can vary from immediately upstream from the promoter to downstream, within, or kilobases away from a gene. Thus, they can modulate transcription even from a distance.

Transcription is initiated through the assembly of transcriptional machinery on the promoter. This machinery consists of *RNA polymerase II* and a series of transcriptional factors that together form a giant complex consisting of 50 or more polypeptides. Assembly of transcription machinery begins when regulatory proteins bind DNA near the promoter and modify the chromatin structure so that transcription can take place. These proteins and other regulatory proteins then recruit basal transcriptional apparatus to the core promoter. The basal transcriptional apparatus consists of *RNA polymerase*, a series of general transcriptional factors, and a complex of proteins known as mediator. The general transcription factor (GTF) includes TFIIA, TFIIB, TFIIE, and TFIIH (TF stands for transcription factor for *RNA polymerase II* and the final letter designates the individual factor). *RNA polymerase II* and transcription factors assemble at the core promoter, forming a **pre-initiation complex** that is analogous to closed complex seen in bacterial initiation. The role of sigma factor that recognizes and binds to the promoter in prokaryotes is replaced by general transcription factor in eukaryotes. Initiation begins with binding of general transcription factor to the promoter in a defined way. TFIID binds directly to the TATA- box sequence. TFIID consists of at least nine polypeptides. One of them is TATA-binding protein (TBP), which recognizes and binds to the TATA consensus sequence. The other transcription factors interact primarily with each other. Hence protein to protein interaction plays a crucial role in the binding stage. After the *RNA polymerase* and transcription factors have assembled on the core promoter, conformational changes take place both in DNA and the *polymerase*. These changes cause 11 to 15 bp of DNA surrounding the transcription start site to separate, thus producing single stranded DNA that will serve as template for transcription. The single-stranded DNA template now positions itself within the active site of RNA polymerase, creating structure called the **open complex**. Once the open complex is formed, the synthesis of RNA begins as phosphate groups are cleaved off nucleoside triphosphate and nucleotides are joined together to form an RNA molecule (see Fig. 12.16). As in bacterial transcription, there are several abortive transcriptions before *polymerase* initiates the synthesis of a full-length RNA molecule.

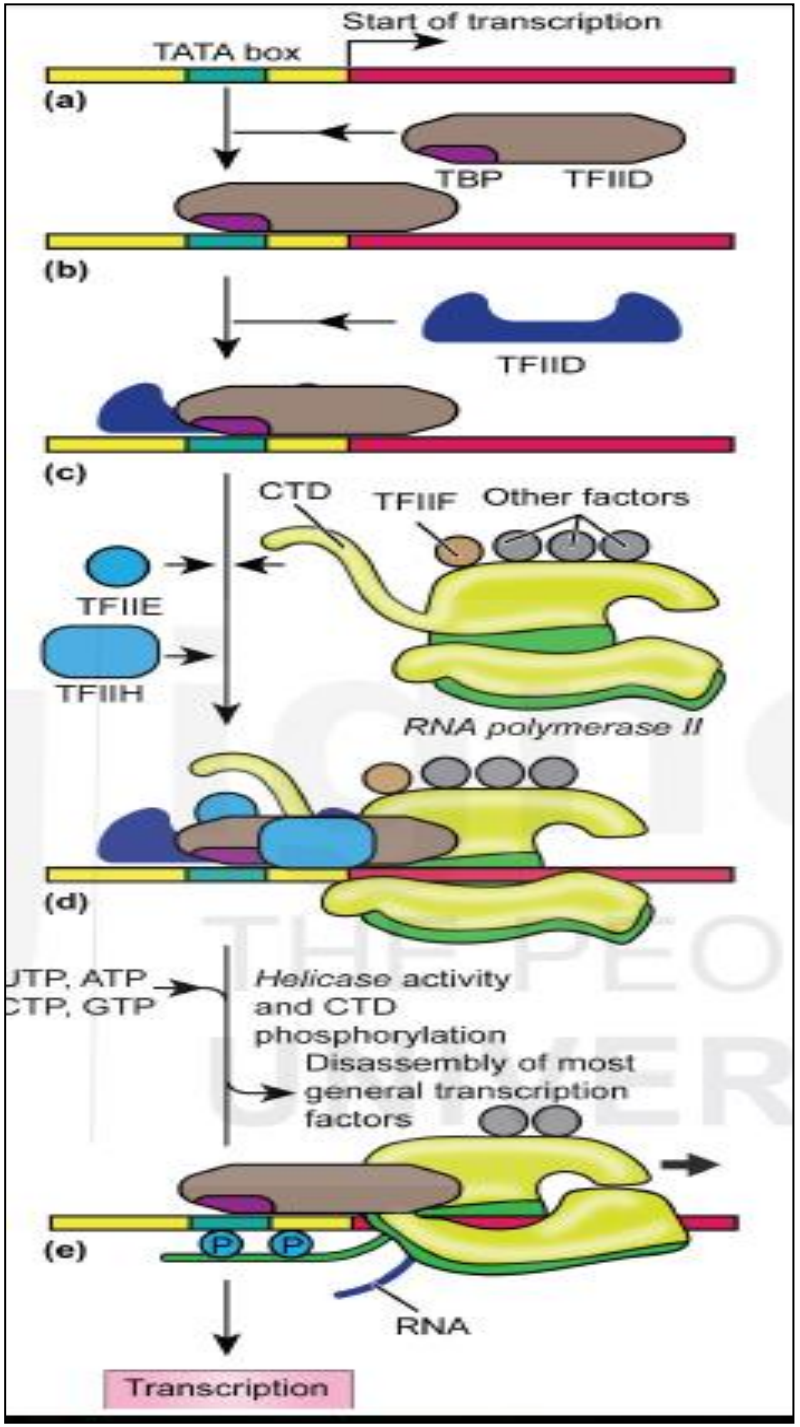

Fig. 12.16: Different steps of Initiation of transcription of a eukaryotic gene by *RNA polymerase II* a) TATA box on the promoter; b) Subunit TBP of TFIID recognizes TATA box and bind, which enables the adjacent binding of TFIID; c) The rest of the general transcription factors as well as RNA polymerase II itself assemble at the promoter; d) TFIIH utilizes ATP to pull apart DNA double helix at the transcription start point, locally exposing the template strand; e) TFIIH also phosphorylates *RNA polymerase II*, changing its conformation so that *polymerase* is released from the general factor and can begin the elongation phase of transcription. At the site of phosphorylation is a long C – terminal domain (CTD) that extends from the *polymerase* molecule.

The general transcription factors and its roles in initiation of eukaryotic transcription have been summarized in Table 12.3.

Table 12.3: General transcription factors needed for transcription initiation by *polymerase II*.

NAME	NUMBER OF SUBUNITS	ROLES IN INITIATION OF TRANSCRIPTION
TFIID		
TBP subunits	1	recognizes TATA box
TAF subunits	~11	recognizes other DNA sequences near the transcription start point; regulates DNA-binding by TBP
TFIIB	1	recognizes BRE element in promoters; accurately positions *RNA polymerase* at the start site of transcription
TFIIF	3	stabilizes *RNA polymerase* interaction with TBP and TFIIB; helps attract TFIIE and TFIIH
TFIIE	2	attracts and regulates TFIIH
TFIIH	9	unwinds DNA at the transcription start point, phosphorylates Ser5 of the *RNA polymerase* CTD; releases *RNA polymerase* from the promoter

(TFIID is composed of TBP and - 11 additional subunits called TAFs (TBP associated factors); CTD (C-terminal domain)

Elongation and Termination of transcription

Once 30 bp of RNA have been synthesized, *RNA polymerase* leaves the promoter and enters into elongation phase. Most of the transcription factors are detached at the promoter and can quickly assist other *RNA polymerases* to continue transcription. Roger Kornberg (Nobel Prize in Chemistry in 2006) and his colleagues worked out the molecular structure and mode of functions of *RNA polymerase* II during elongation. Transcription bubble is maintained by *RNA polymerase* during elongation. The DNA double helix enters a cleft in the p*olymerase* which is gripped by jaw like extension of the enzyme. The two strands of the DNA are unwound and RNA nucleotides that are complementary to the template strand are added to the growing 3´ –end of the RNA molecule. DNA-RNA hybrid hits a wall of amino acids and bends at almost right angle while passing through the funnels of the *polymerase*. This bend positions the end of the DNA-RNA hybrid at the active site of the *polymerase*, and new nucleotides are added to the three end of the growing RNA molecule. The newly synthesized RNA molecule is separated from the DNA and runs through another groove before exiting from the *polymerase*

The basic differences between prokaryotic and eukaryotic transcription has been summarized in Table 12.4.

Table 12.4: Basic differences between prokaryotic and eukaryotic transcription.

Transcription in prokaryotes	Transcription in eukaryotes
1. There is no spatial separation of transcription and translation since prokaryotes lack nucleus. Transcription is carried out by single *RNA polymerase*.	1. Transcription in eukaryotes occurs in nucleus under the direction of three separate form of *RNA polymerase*. For the mRNA to be translated, it must move out from the nucleus into the cytoplasm.
2. For initiation of transcription of prokaryote genes, DNA remains accessible for *RNA polymerase* and other accessory proteins since DNA is not complexed with histone proteins in highly compressed chromatin.	2. Initiation of transcription in eukaryotic genes requires compact nucleosomes coiling to be uncoiled and DNA to be made accessible to *RNA polymerase* and other accessary proteins. This transition refers to as histone modification and chromatin remodeling

3. 3. For promoter recognition in prokaryotic cells the holoenzyme (*RNA polymerase* plus the sigma factor) recognizes and binds directly to the sequences in the promoter	3. In eukaryotic cells the promoter recognition is carried out by accessary proteins called as general transcription factors that bind to the promoter and then recruit a specific *RNA polymerase* (I, II or III) to the promoter. In addition to promoter other control units in eukaryotes are *enhancer* and *silencer.*
4. In prokaryotes the primary RNA transcripts does not require being processed (removal of non-coding sequences- introns) before being translated because prokaryotic genes mostly lack introns.	4. In eukaryotes the primary transcripts are generally much larger than those that are eventually translated into protein known as pre-mRNA or hnRNA (heterogeneous nuclear RNA) that need to be processed before being transported to the cytoplasm for translation.