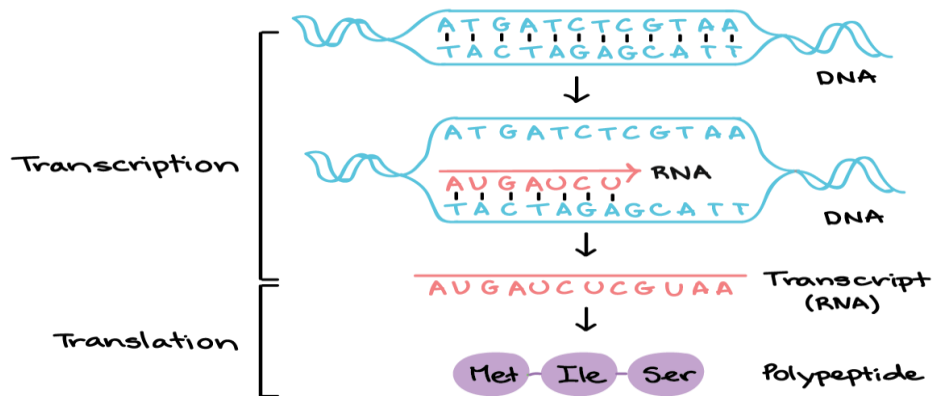


Transcription

Introduction:

Transcription is the process by which the information in a strand of DNA is copied into a new molecule of messenger RNA (mRNA). DNA safely and stably stores genetic material in the nuclei of cells as a reference, or template. Meanwhile, mRNA is comparable to a copy from a reference book because it carries the same information as DNA but is not used for long-term storage and can freely exit the nucleus. Although the mRNA contains the same information, it is not an identical copy of the DNA segment, because its sequence is complementary to the DNA template. Transcription is the first step in gene expression, in which information from a gene is used to construct a functional product such as a protein. The goal of transcription is to make a RNA copy of a gene's DNA sequence. For a protein-coding gene, the RNA copy, or **transcript**, carries the information needed to build a polypeptide (protein or protein subunit). Eukaryotic transcripts need to go through some processing steps before translation into proteins.

Transcription is carried out by an enzyme called RNA polymerase and a number of accessory proteins called transcription factors. Transcription factors can bind to specific DNA sequences called enhancer and promoter sequences in order to recruit RNA polymerase to an appropriate transcription site. Together, the transcription factors and RNA polymerase form a complex called the transcription initiation complex. This complex initiates transcription, and the RNA polymerase begins mRNA synthesis by matching complementary bases to the original DNA strand. The mRNA molecule is elongated and, once the strand is completely synthesized, transcription is terminated. The newly formed mRNA copies of the gene then serve as blueprints for protein synthesis during the process of translation.



The General Process of Transcription:

Transcription, the synthesis of RNA from DNA. Genetic information flows from DNA into protein, the substance that gives an organism its form. This flow of information occurs through the sequential processes of transcription (DNA to RNA) and translation (RNA to protein). Transcription occurs when there is a need for a particular gene product at a specific time or in a specific tissue.

During transcription, only one strand of DNA is usually copied. This is called the template strand, and the RNA molecules produced are single-stranded messenger RNAs (mRNAs). The DNA strand that would correspond to the mRNA is called the coding or sense strand. In eukaryotes (organisms that possess a nucleus) the initial product of transcription is called a pre-mRNA. Pre-mRNA is extensively edited through splicing before the mature mRNA is produced and ready for translation by the ribosome, the cellular organelle that serves as the site of protein synthesis. Transcription of any one gene takes place at the chromosomal location of that gene, which is a relatively short segment of the chromosome. The active transcription of a gene depends on the need for the activity of that particular gene in a specific cell or tissue or at a given time.

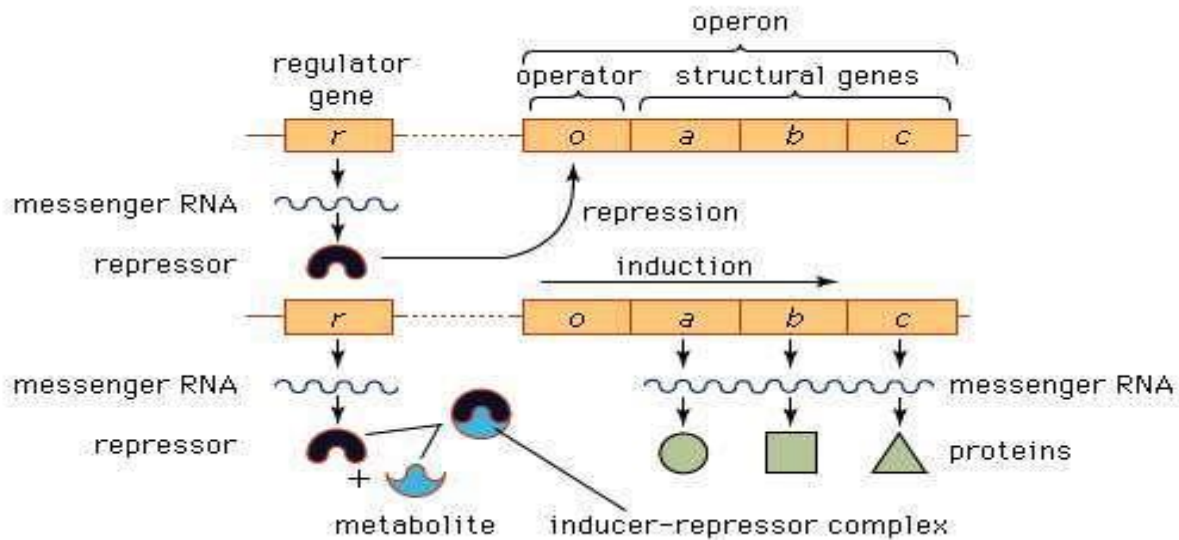
Small segments of DNA are transcribed into RNA by the enzyme RNA polymerase, which achieves this copying in a strictly controlled process.

The first step is to recognize a specific sequence on DNA called a promoter that signifies the start of the gene. The two strands of DNA become separated at this point, and RNA polymerase begins copying from a specific point on one strand of the DNA using a special type of sugar-containing nucleoside called ribonucleoside 5'-triphosphate to begin the growing chain. Additional ribonucleoside triphosphates are used as the substrate, and, by cleavage of their high-energy phosphate bond, ribonucleoside monophosphates are incorporated into the growing RNA chain. Each successive ribonucleotide is directed by the complementary base pairing rules of DNA. For example, a C (cytosine) in DNA directs the incorporation of a G (guanine) into RNA. Likewise, a G in DNA is copied into a C in RNA, a T (thymine) into an A (adenine), and an A into a U (uracil; RNA contains U in place of the T of DNA). Synthesis continues until a termination signal is reached, at which point the RNA polymerase drops off the DNA, and the RNA molecule is released.

Ahead of many genes in prokaryotes (organisms that lack a nucleus), there are signals called “operators” (*see operons*) where specialized proteins called repressors bind to the DNA just upstream of the start point of transcription and prevent access to the DNA by RNA polymerase. These repressor proteins thus prevent transcription of the gene by physically blocking the action of the RNA polymerase. Typically, repressors are released from their blocking action when they receive signals from other molecules in the cell indicating that the gene needs to be expressed. Ahead of some prokaryotic genes are signals to which activator proteins bind to stimulate transcription.

Transcription in eukaryotes is more complicated than in prokaryotes. First, the RNA polymerase of higher organisms is a more complicated enzyme than the relatively simple five-subunit enzyme of prokaryotes. In addition, there are many more accessory factors that help to control the efficiency of the individual promoters. These accessory proteins are called transcription factors and typically respond to signals from within

the cell that indicate whether transcription is required. In many human genes, several transcription factors may be needed before transcription can proceed efficiently. A transcription factor can cause either repression or activation of gene expression in eukaryotes.



RNA polymerase

The main enzyme involved in transcription is **RNA polymerase**, which uses a single-stranded DNA template to synthesize a complementary strand of RNA. Specifically, RNA polymerase builds an RNA strand in the 5' to 3' direction, adding each new nucleotide to the 3' end of the strand.

Both prokaryotes and eukaryotes perform fundamentally the same process of transcription, with the important difference of the membrane-bound nucleus in eukaryotes. With the genes bound in the nucleus, transcription occurs in the nucleus of the cell and the mRNA transcript must be transported to the cytoplasm. In prokaryotes, which lack membrane-bound nuclei and other organelles, transcription occurs in the cytoplasm of the cell.

Prokaryotic Transcription:

Origin of transcription on prokaryotic organisms

Prokaryotes are mostly single-celled organisms that, by definition, lack membrane-bound nuclei and other organelles. The central region of the cell in which prokaryotic DNA resides is called the nucleoid region. Bacterial and Archaeal chromosomes are covalently-closed circles that are not as extensively compacted as eukaryotic chromosomes, but are compacted nonetheless as the diameter of a typical prokaryotic chromosome is larger than the diameter of a typical prokaryotic cell. Additionally, prokaryotes often have abundant plasmids, which are shorter, circular DNA molecules that may only contain one or a few genes and often carry traits such as antibiotic resistance.

Transcription in prokaryotes (as in eukaryotes) requires the DNA double helix to partially unwind in the region of RNA synthesis. The region of unwinding is called a transcription bubble. Transcription always proceeds from the same DNA strand for each gene, which is called the template strand. The RNA product is complementary to the template strand and is almost identical to the other (non-template) DNA strand, called the sense or coding strand. The only difference is that in RNA all of the T nucleotides are replaced with U nucleotides.

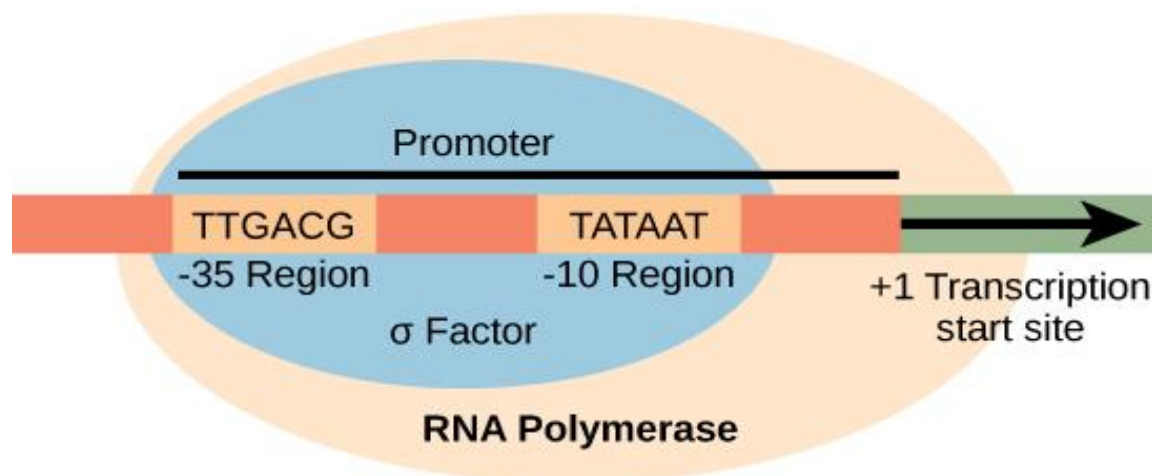
The nucleotide on the DNA template strand that corresponds to the site from which the first 5' RNA nucleotide is transcribed is called the +1 nucleotide, or the initiation site. Nucleotides preceding, or 5' to, the template strand initiation site are given negative numbers and are designated upstream. Conversely, nucleotides following, or 3' to, the template strand initiation site are denoted with "+" numbering and are called downstream nucleotides.

Prokaryotic Promoters and Initiation of Transcription

The nucleotide pair in the DNA double helix that corresponds to the site from which the first 5' mRNA nucleotide is transcribed is called the +1

site, or the initiation site. Nucleotides preceding the initiation site are given negative numbers and are designated upstream. Conversely, nucleotides following the initiation site are denoted with “+” numbering and are called downstream nucleotides.

A promoter is a DNA sequence onto which the transcription machinery binds and initiates transcription. In most cases, promoters exist upstream of the genes they regulate. The specific sequence of a promoter is very important because it determines whether the corresponding gene is transcribed all the time, some of the time, or infrequently. Although promoters vary among prokaryotic genomes, a few elements are conserved. At the -10 and -35 regions upstream of the initiation site, there are two promoter consensus sequences, or regions that are similar across all promoters and across various bacterial species. The -10 consensus sequence, called the -10 region, is TATAAT. The -35 sequence, TTGACA, is recognized and bound by σ . Once this interaction is made, the subunits of the core enzyme bind to the site. The A–T-rich -10 region facilitates unwinding of the DNA template; several phosphodiester bonds are made. The transcription initiation phase ends with the production of abortive transcripts, which are polymers of approximately 10 nucleotides that are made and released.



Prokaryotic RNA Polymerase :

Prokaryotes use the same RNA polymerase to transcribe all of their genes. In *E. coli*, the polymerase is composed of five polypeptide subunits, two of which are identical. Four of these subunits, denoted α , α , β , and β' , comprise the polymerase core enzyme. These subunits assemble each time a gene is transcribed; they disassemble once transcription is complete. Each subunit has a unique role: the two α -subunits are necessary to assemble the polymerase on the DNA; the β -subunit binds to the ribonucleoside triphosphate that will become part of the nascent “recently-born” mRNA molecule; and the β' binds the DNA template strand. The fifth subunit, σ , is involved only in transcription initiation. It confers transcriptional specificity such that the polymerase begins to synthesize mRNA from an appropriate initiation site. Without σ , the core enzyme would transcribe from random sites and would produce mRNA molecules that specified protein gibberish. The polymerase comprised of all five subunits is called the holoenzyme.

Initiation:

Transcription in prokaryotes (and in eukaryotes) requires the DNA double helix to partially unwind in the region of mRNA synthesis. The region of unwinding is called a transcription bubble. The DNA sequence onto which the proteins and enzymes involved in transcription bind to initiate the process is called a **promoter**. In most cases, promoters exist upstream of the genes they regulate. The specific sequence of a promoter is very important because it determines whether the corresponding gene is transcribed all of the time, some of the time, or hardly at all. The structure and function of a prokaryotic promoter is relatively simple. One important sequence in the prokaryotic promoter is located 10 bases before the transcription start site (-10) and is commonly called the TATA box.

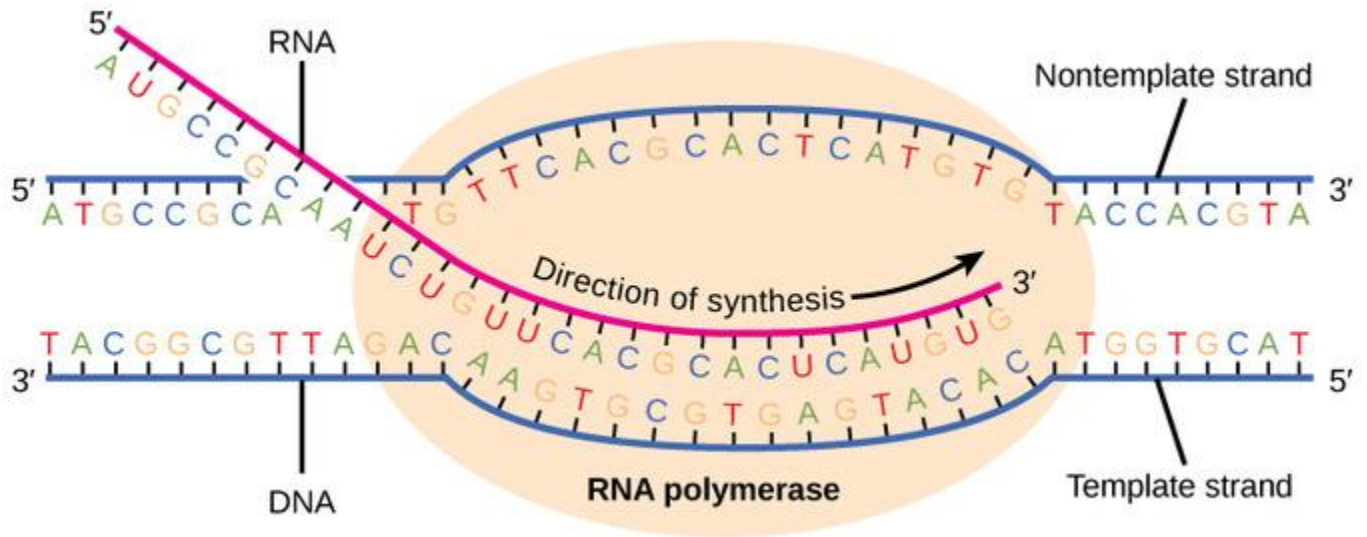
To begin transcription, the RNA polymerase holoenzyme assembles at the promoter. The dissociation of σ allows the core enzyme to proceed along the DNA template, synthesizing mRNA by adding RNA nucleotides according to the base pairing rules, similar to the way a new DNA molecule is produced during DNA replication. Only one of the two DNA strands is transcribed. The transcribed strand of DNA is called the **template strand** because it is the template for mRNA production. The mRNA product is complementary to the template strand and is almost identical to the other DNA strand, called the **non-template strand**, with the exception that RNA contains a uracil (U) in place of the thymine (T) found in DNA. Like DNA polymerase, RNA polymerase adds new nucleotides onto the 3'-OH group of the previous nucleotide. This means that the growing mRNA strand is being synthesized in the 5' to 3' direction. Because DNA is anti-parallel, this means that the RNA polymerase is moving in the 3' to 5' direction down the template strand.

Elongation :

As elongation proceeds, the DNA is continuously unwound ahead of the core enzyme as the hydrogen bonds that connect the complementary base pairs in the DNA double helix are broken. The DNA is rewound behind the core enzyme as the hydrogen bonds are reformed. The base pairing between DNA and RNA is not stable enough to maintain the stability of the mRNA synthesis components. Instead, the RNA polymerase acts as a stable linker between the DNA template and the newly forming RNA strand to ensure that elongation is not interrupted prematurely.

The transcription elongation phase begins with the release of the σ subunit from the polymerase. The dissociation of σ allows the core RNA polymerase enzyme to proceed along the DNA template, synthesizing mRNA in the 5' to 3' direction at a rate of approximately 40 nucleotides per second. As elongation proceeds, the DNA is continuously unwound ahead of the core enzyme and rewound behind it. Since the base pairing between DNA and RNA is not stable enough to maintain the stability of

the mRNA synthesis components, RNA polymerase acts as a stable linker between the DNA template and the nascent RNA strands to ensure that elongation is not interrupted prematurely.



Termination:

Once a gene is transcribed, the RNA polymerase needs to be instructed to dissociate from the DNA template and liberate the newly made mRNA. Depending on the gene being transcribed, there are two kinds of termination signals. One is protein-based and the other is RNA-based. Both termination signals rely on specific sequences of DNA near the end of the gene that cause the polymerase to release the mRNA.

In a prokaryotic cell, by the time transcription ends, the transcript would already have been used to begin making copies of the encoded protein because the processes of transcription and translation can occur at the same time since both occur in the cytoplasm. In contrast, transcription and translation cannot occur simultaneously in eukaryotic cells since

transcription occurs inside the nucleus and translation occurs outside in the cytoplasm.

Depending on the gene being transcribed, there are two kinds of termination signals: one is protein-based and the other is RNA-based.

Rho-dependent termination is controlled by the rho protein, which tracks along behind the polymerase on the growing mRNA chain. Near the end of the gene, the polymerase encounters a run of G nucleotides on the DNA template and it stalls. As a result, the rho protein collides with the polymerase. The interaction with rho releases the mRNA from the transcription bubble.

Rho-independent termination is controlled by specific sequences in the DNA template strand. As the polymerase nears the end of the gene being transcribed, it encounters a region rich in C–G nucleotides. The mRNA folds back on itself, and the complementary C–G nucleotides bind together. The result is a stable hairpin that causes the polymerase to stall as soon as it begins to transcribe a region rich in A–T nucleotides. The complementary U–A region of the mRNA transcript forms only a weak interaction with the template DNA. This, coupled with the stalled polymerase, induces enough instability for the core enzyme to break away and liberate the new mRNA transcript.

Upon termination, the process of transcription is complete. By the time termination occurs, the prokaryotic transcript would already have been used to begin synthesis of numerous copies of the encoded protein because these processes can occur concurrently in the cytoplasm. The unification of transcription, translation, and even mRNA degradation is possible because all of these processes occur in the same 5' to 3' direction and because there is no membranous compartmentalization in the prokaryotic cell. In contrast, the presence of a nucleus in eukaryotic cells prevents simultaneous transcription and translation.

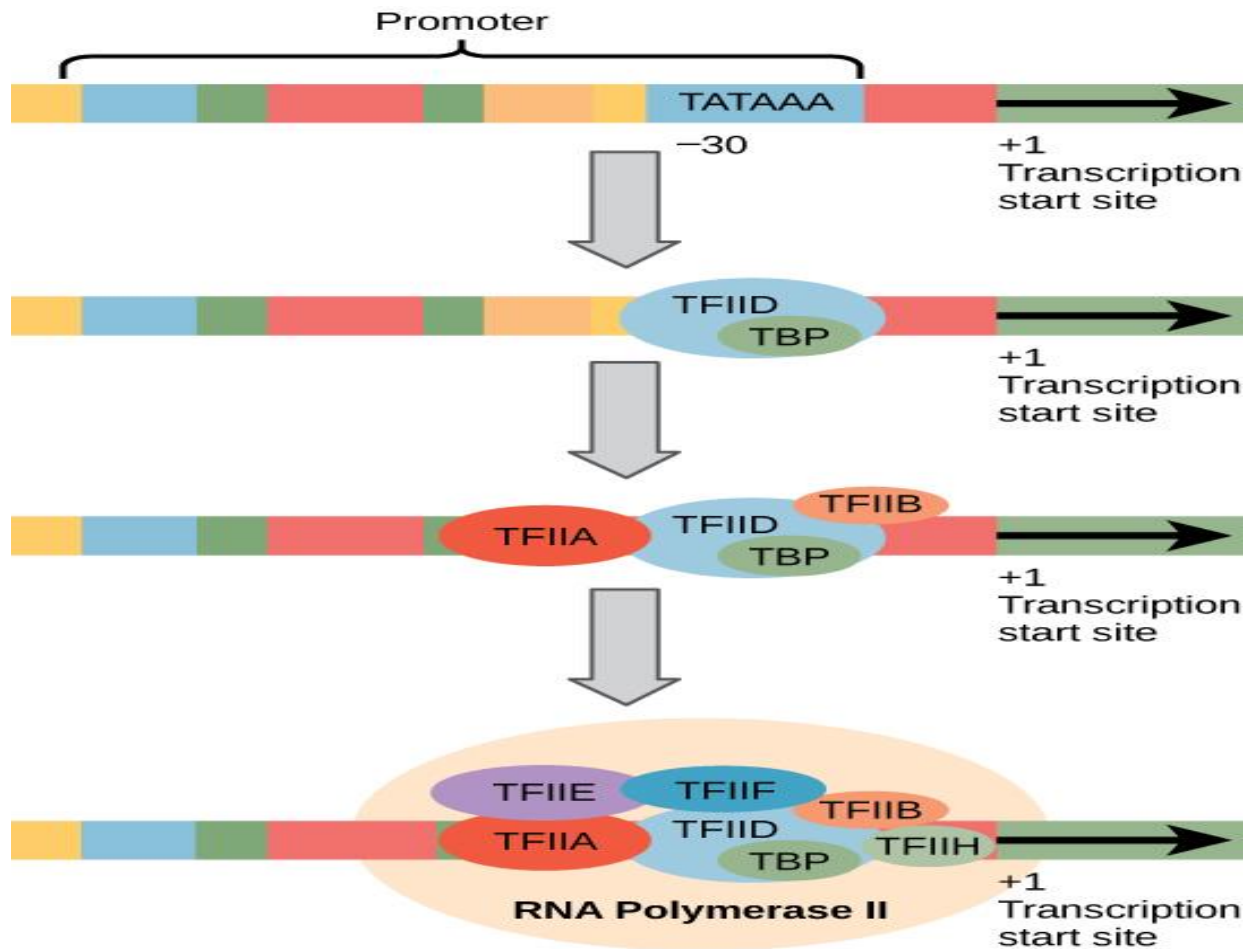
Eukaryotic Transcription:

Initiation of Transcription in Eukaryotes:

Unlike the prokaryotic RNA polymerase that can bind to a DNA template on its own, eukaryotes require several other proteins, called transcription factors, to first bind to the promoter region and then help recruit the appropriate polymerase. The completed assembly of transcription factors and RNA polymerase bind to the promoter, forming a transcription pre-initiation complex (PIC).

The most-extensively studied core promoter element in eukaryotes is a short DNA sequence known as a TATA box, found 25-30 base pairs upstream from the start site of transcription. Only about 10-15% of mammalian genes contain TATA boxes, while the rest contain other core promoter elements, but the mechanisms by which transcription is initiated at promoters with TATA boxes is well characterized.

The TATA box, as a core promoter element, is the binding site for a transcription factor known as TATA-binding protein (TBP), which is itself a subunit of another transcription factor: Transcription Factor II D (TFIID). After TFIID binds to the TATA box via the TBP, five more transcription factors and RNA polymerase combine around the TATA box in a series of stages to form a pre-initiation complex. One transcription factor, Transcription Factor II H (TFIIH), is involved in separating opposing strands of double-stranded DNA to provide the RNA Polymerase access to a single-stranded DNA template. However, only a low, or basal, rate of transcription is driven by the pre-initiation complex alone. Other proteins known as activators and repressors, along with any associated coactivators or corepressors, are responsible for modulating transcription rate. Activator proteins increase the transcription rate, and repressor proteins decrease the transcription rate.



The Three Eukaryotic RNA Polymerases (RNAPs)

The features of eukaryotic mRNA synthesis are markedly more complex than those of prokaryotes. Instead of a single polymerase comprising five subunits, the eukaryotes have three polymerases that are each made up of 10 subunits or more. Each eukaryotic polymerase also requires a distinct set of transcription factors to bring it to the DNA template.

RNA polymerase I is located in the nucleolus, a specialized nuclear substructure in which ribosomal RNA (rRNA) is transcribed, processed, and assembled into ribosomes. The rRNA molecules are considered structural RNAs because they have a cellular role but are not translated into protein. The rRNAs are components of the ribosome and are

essential to the process of translation. RNA polymerase I synthesizes all of the rRNAs except for the 5S rRNA molecule.

RNA polymerase II is located in the nucleus and synthesizes all protein-coding nuclear pre-mRNAs. Eukaryotic pre-mRNAs undergo extensive processing after transcription, but before translation. RNA polymerase II is responsible for transcribing the overwhelming majority of eukaryotic genes, including all of the protein-encoding genes which ultimately are translated into proteins and genes for several types of regulatory RNAs, including microRNAs (miRNAs) and long-coding RNAs (lncRNAs).

RNA polymerase III is also located in the nucleus. This polymerase transcribes a variety of structural RNAs that includes the 5S pre-rRNA, transfer pre-RNAs (pre-tRNAs), and small nuclear pre-RNAs. The tRNAs have a critical role in translation: they serve as the adaptor molecules between the mRNA template and the growing polypeptide chain. Small nuclear RNAs have a variety of functions, including “splicing” pre-mRNAs and regulating transcription factors. Not all miRNAs are transcribed by RNA Polymerase II, RNA Polymerase III transcribes some of them.

Structure of an RNA Polymerase II Promoter

Eukaryotic promoters are much larger and more complex than prokaryotic promoters, but both have a TATA box. For example, in the mouse thymidine kinase gene, the TATA box is located at approximately -30 relative to the initiation (+1) site ([\[link\]](#)). For this gene, the exact TATA box sequence is TATAAAA, as read in the 5' to 3' direction on the nontemplate strand. This sequence is not identical to the *E. coli* TATA box, but it conserves the A–T rich element. The thermostability of A–T bonds is low and this helps the DNA template to locally unwind in preparation for transcription.

A generalized promoter of a gene transcribed by RNA polymerase II is shown. Transcription factors recognize the promoter. RNA polymerase II then binds and forms the transcription initiation complex.

Promoter Structures for RNA Polymerases I and III

In eukaryotes, the conserved promoter elements differ for genes transcribed by RNA polymerases I, II, and III. RNA polymerase I transcribes genes that have two GC-rich promoter sequences in the -45 to +20 region. These sequences alone are sufficient for transcription initiation to occur, but promoters with additional sequences in the region from -180 to -105 upstream of the initiation site will further enhance initiation. Genes that are transcribed by RNA polymerase III have upstream promoters or promoters that occur within the genes themselves.

Elongation:

Following the formation of the preinitiation complex, the polymerase is released from the other transcription factors, and elongation is allowed to proceed as it does in prokaryotes with the RNA polymerase synthesizing pre-mRNA in the 5' to 3' direction. As discussed previously, RNA polymerase II transcribes the major share of eukaryotic genes, so this section will focus on how this polymerase accomplishes elongation and termination.

Although the enzymatic process of elongation is essentially the same in eukaryotes and prokaryotes, the DNA template is more complex. When eukaryotic cells are not dividing, their genes exist as a diffuse mass of DNA and proteins called chromatin. The DNA is tightly packaged around charged histone proteins at repeated intervals. These DNA-histone complexes, collectively called nucleosomes, are regularly spaced and include 146 nucleotides of DNA wound around eight histones like thread around a spool.

For RNA synthesis to occur, the transcription machinery needs to move histones out of the way every time it encounters a nucleosome. This is accomplished by a special protein complex called FACT, which stands for “facilitates chromatin transcription.” This complex pulls histones away from the DNA template as the polymerase moves along it. Once the pre-mRNA is synthesized, the FACT complex replaces the histones to recreate the nucleosomes.

Termination:

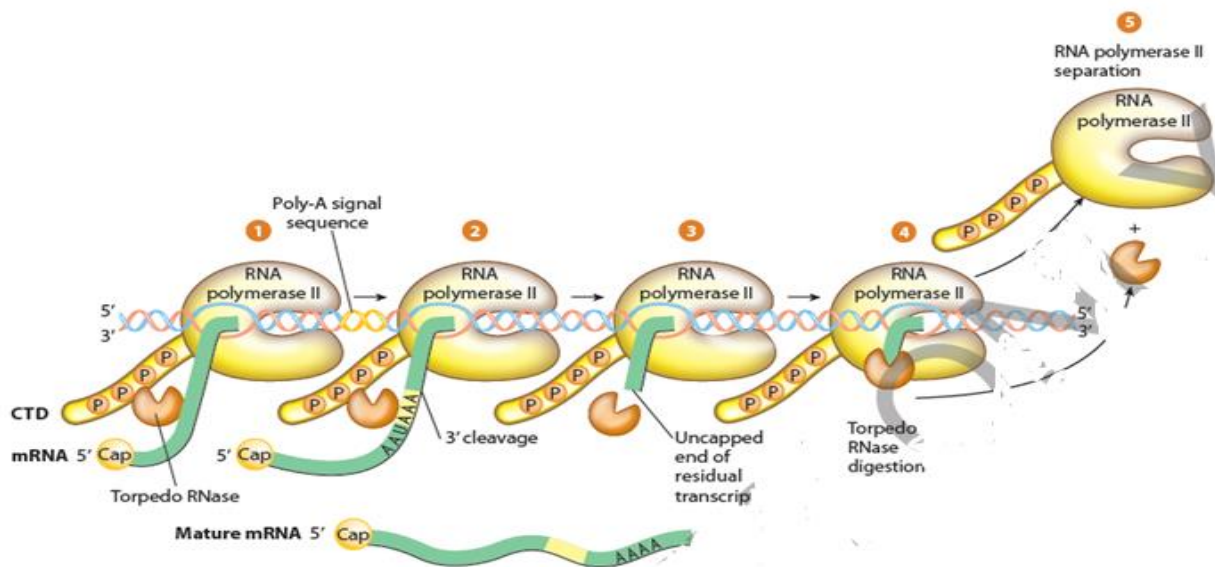
The termination of transcription is different for the three different eukaryotic RNA polymerases.

The ribosomal rRNA genes transcribed by RNA Polymerase I contain a specific sequence of basepairs (11 bp long in humans; 18 bp in mice) that is recognized by a termination protein called TTF-1 (Transcription Termination Factor for RNA Polymerase I.) This protein binds the DNA at its recognition sequence and blocks further transcription, causing the RNA Polymerase I to disengage from the template DNA strand and to release its newly-synthesized RNA.

The protein-encoding, structural RNA, and regulatory RNA genes transcribed by RNA Polymerase II lack any specific signals or sequences that direct RNA Polymerase II to terminate at specific locations. RNA Polymerase II can continue to transcribe RNA anywhere from a few bp to thousands of bp past the actual end of the gene. However, the transcript is cleaved at an internal site before RNA Polymerase II finishes transcribing. This releases the upstream portion of the transcript, which will serve as the initial RNA prior to further processing (the pre-mRNA in the case of protein-encoding genes.) This cleavage site is considered the “end” of the gene. The remainder of the transcript is digested by a 5'-exonuclease (called Xrn2 in humans) while it is still being transcribed by the RNA Polymerase II. When the 5'-exonuclease “catches up” to RNA Polymerase II by digesting away all the

overhanging RNA, it helps disengage the polymerase from its DNA template strand, finally terminating that round of transcription.

In the case of protein-encoding genes, the cleavage site which determines the “end” of the emerging pre-mRNA occurs between an upstream AAUAAA sequence and a downstream GU-rich sequence separated by about 40-60 nucleotides in the emerging RNA. Once both of these sequences have been transcribed, a protein called CPSF in humans binds the AAUAAA sequence and a protein called CstF in humans binds the GU-rich sequence. These two proteins form the base of a complicated protein complex that forms in this region before CPSF cleaves the nascent pre-mRNA at a site 10-30 nucleotides downstream from the AAUAAA site. The Poly(A) Polymerase enzyme which catalyzes the addition of a 3' poly-A tail on the pre-mRNA is part of the complex that forms with CPSF and CstF.



Transcription Factors:

Transcription factors are proteins involved in the process of converting, or transcribing, DNA into RNA. Transcription factors include a wide number of proteins, excluding RNA polymerase that initiate and regulate the transcription of genes. One distinct feature of transcription factors is that they have DNA-binding domains that give them the ability to bind to specific sequences of DNA called enhancer or promoter sequences. Some transcription factors bind to a DNA promoter sequence near the transcription start site and help form the transcription initiation complex. Other transcription factors bind to regulatory sequences, such as enhancer sequences, and can either stimulate or repress transcription of the related gene. These regulatory sequences can be thousands of base pairs upstream or downstream from the gene being transcribed. Regulation of transcription is the most common form of gene control. The action of transcription factors allows for unique expression of each gene in different cell types and during development.

Transcription factor controls the activity of a gene by determining whether the gene's DNA (deoxyribonucleic acid) is transcribed into RNA (ribonucleic acid). The enzyme RNA polymerase catalyzes the chemical reactions that synthesize RNA, using the gene's DNA as a template. Transcription factors control when, where, and how efficiently RNA polymerases function.

Transcription factors are vital for the normal development of an organism, as well as for routine cellular functions and response to disease. Transcription factors are a very diverse family of proteins and generally function in multi-subunit protein complexes. They may bind directly to special "promoter" regions of DNA, which lie upstream of the coding region in a gene, or directly to the RNA polymerase molecule. Transcription factors can activate or repress the transcription of a gene, which is generally a key determinant in whether the gene functions at a given time.

Basal, or general, transcription factors are necessary for RNA polymerase to function at a site of transcription in eukaryotes. They are considered the most basic set of proteins needed to activate gene transcription, and they include a number of proteins, such as TFIIA (transcription factor II A) and TFIIB (transcription factor II B), among others. Substantial progress has been made in defining the roles played by each of the proteins that compose the basal transcription factor complex.

Transcription factors function in the nucleus, where genes are found, and nuclear transport (i.e., import or export) of transcription factors can influence their activity. Another important general mechanism controlling the activity of transcription factors is posttranslational modification such as phosphorylation. Finally, in addition to controlling the genes and transcription of other transcription factors, these protein complexes can also control the genes responsible for their own transcription, leading to complex feedback control mechanisms.

Difference between Prokaryotic and Eukaryotic Transcription:

Prokaryotic Transcription	Eukaryotic Transcription
Transcription and translation occur simultaneously	Transcription and translation don't occur simultaneously.
Prokaryotic transcription occurs in the cytoplasm	Eukaryotic transcription occurs in the nucleus and translation occurs in the cytoplasm.
RNAs are released and processed in the cytoplasm	RNAs are released and processed in the nucleus
RNA polymerases are a complex of	RNA polymerases are a complex of 10

five polypeptides.	-15 polypeptides.
Doesn't require any proteins or other factors for the initiation of transcription	Requires proteins known as transcription factors for the initiation.