



# Transcription

Prof. Mamoun Ahram  
School of Medicine  
Second year, First semester, 2024-2025

# Definition of a gene



- The entire DNA sequence that is necessary for the synthesis of a functional RNA (mRNA, rRNA, tRNA, lncRNA, microRNA, etc.) or a polypeptide, which may become a protein or a functional peptide(s).
  - The DNA sequence encompasses the coding region (that makes the protein), other regulatory sequences like a promoter, an enhancer, etc., or a non-coding region like introns.
- A cistron: an alternative term for a gene.
  - If it encodes one polypeptide from one mRNA, it is monocistronic.
  - If it encodes several or different polypeptides from ONE mRNA molecule, it is polycistronic.



# The general mechanism of transcription

# General description

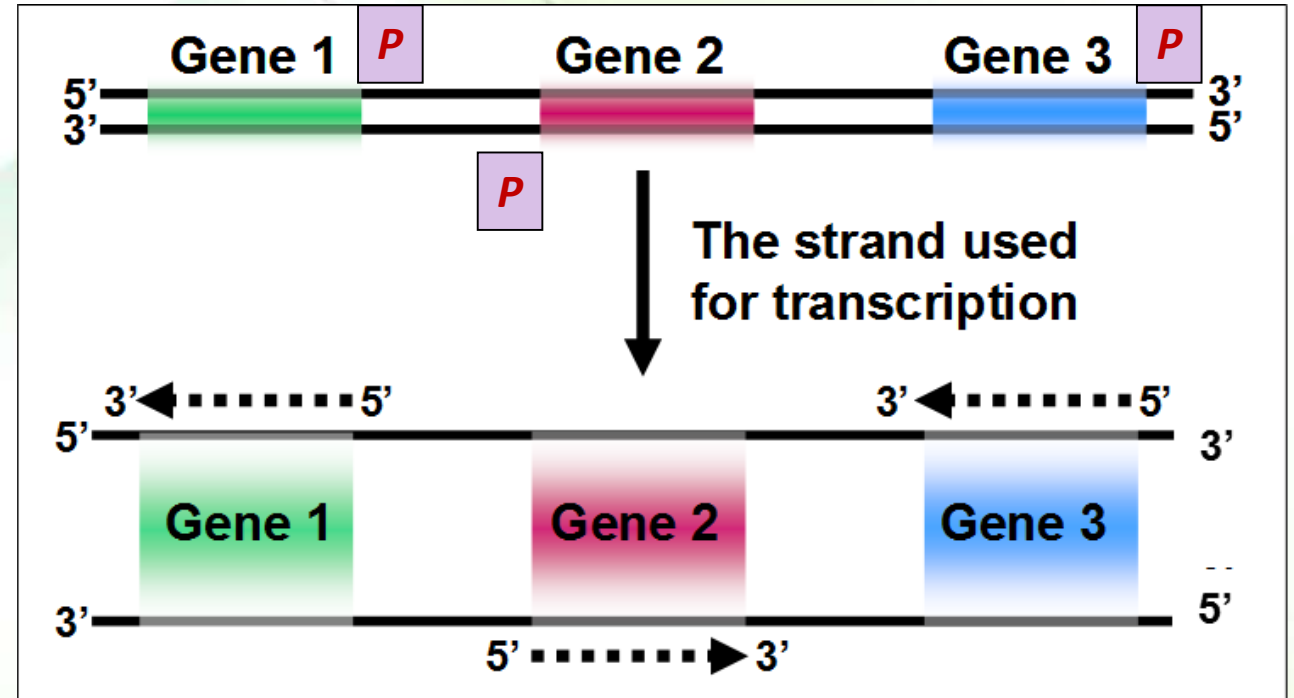


- Transcription is the process of making RNA from DNA.
- **One** of the two strands of the DNA double helix acts as a template for the expression of a particular gene (that is, synthesis of a RNA molecule).
  - Remember? In DNA replication, both strands are the template of the daughter strands.

# Using DNA strands



- Although RNA polymerase can read both DNA strands, it uses one strand for any particular gene in order to make RNA.

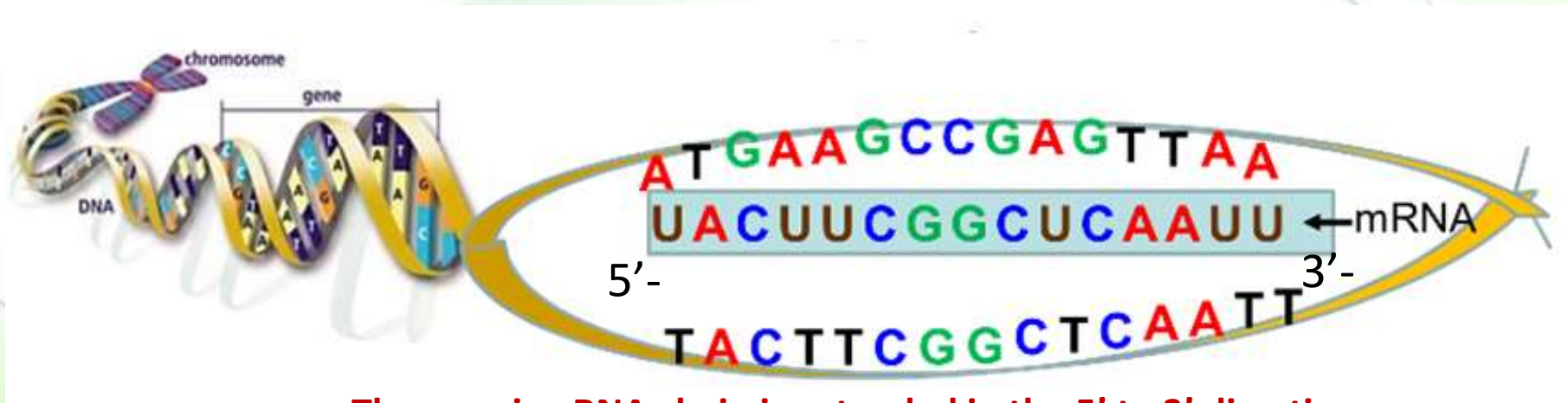


***What does determine which strand is used for transcription?***

# Complementary sequences



- RNA is complementary to its DNA template.
- The RNA chain produced by transcription is also known as the transcript.



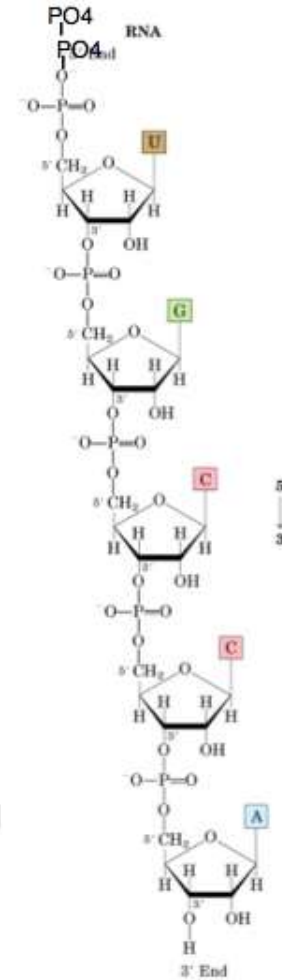
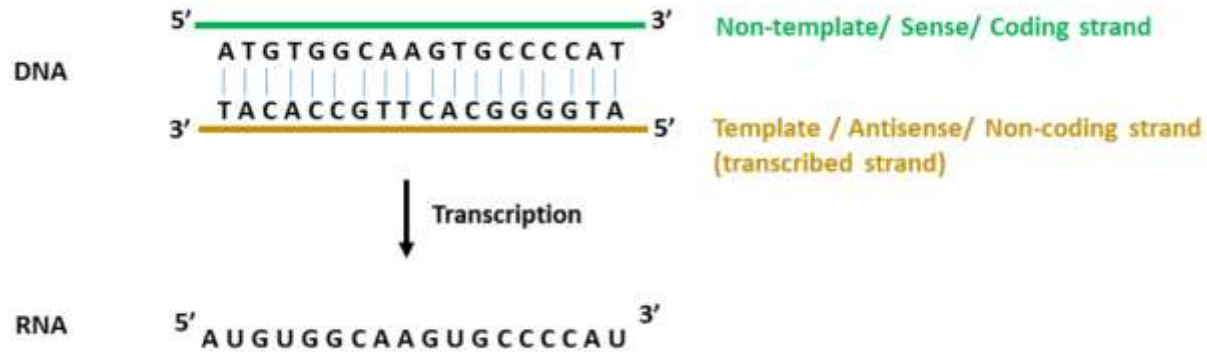
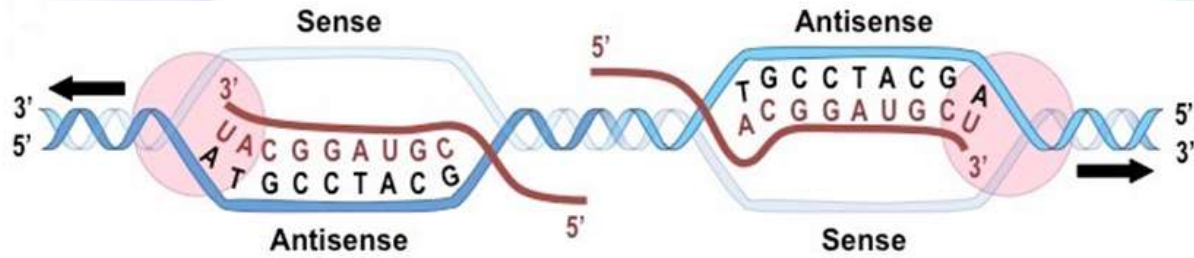
**The growing RNA chain is extended in the 5' to 3' direction.**

# Enzyme and substrate

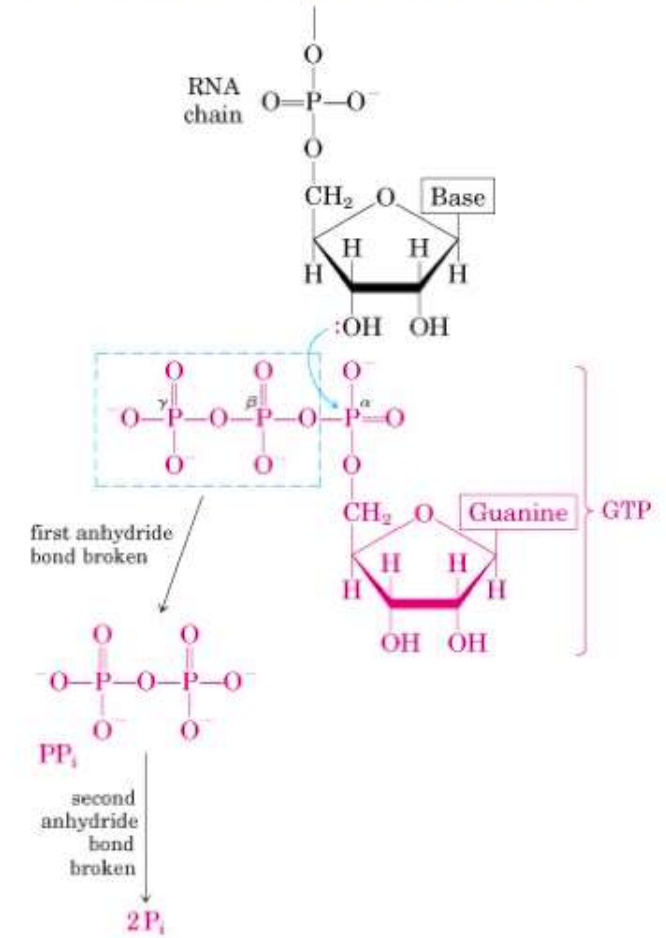


- The enzymes that perform transcription are called RNA polymerases.
- They catalyze the formation of the phosphodiester bonds between two nucleotides.
- RNA polymerase does not require a preformed primer to initiate the synthesis of RNA.
  - Transcription initiates *de novo* at specific sites at the beginning of genes.
- The substrates are nucleoside triphosphates (ATP, CTP, UTP, and GTP).
  - What are substrates for DNA polymerases?
- Hydrolysis of high-energy bonds in NTPs provides the energy needed to drive the reaction forward.

# More clarification and some extra terms



## Phosphodiester bond synthesis





# DNA replication vs. transcription



- The newly synthesized portion of the RNA is bound to the DNA template but is released as RNA extends further.
- RNA polymerase reads the A in DNA and inserts U in the growing chain of RNA rather than T.
- RNA molecules are much shorter than DNA molecules.
- Unlike DNA, RNA does not store genetic information in cells.

# DNA polymerase vs. RNA polymerase

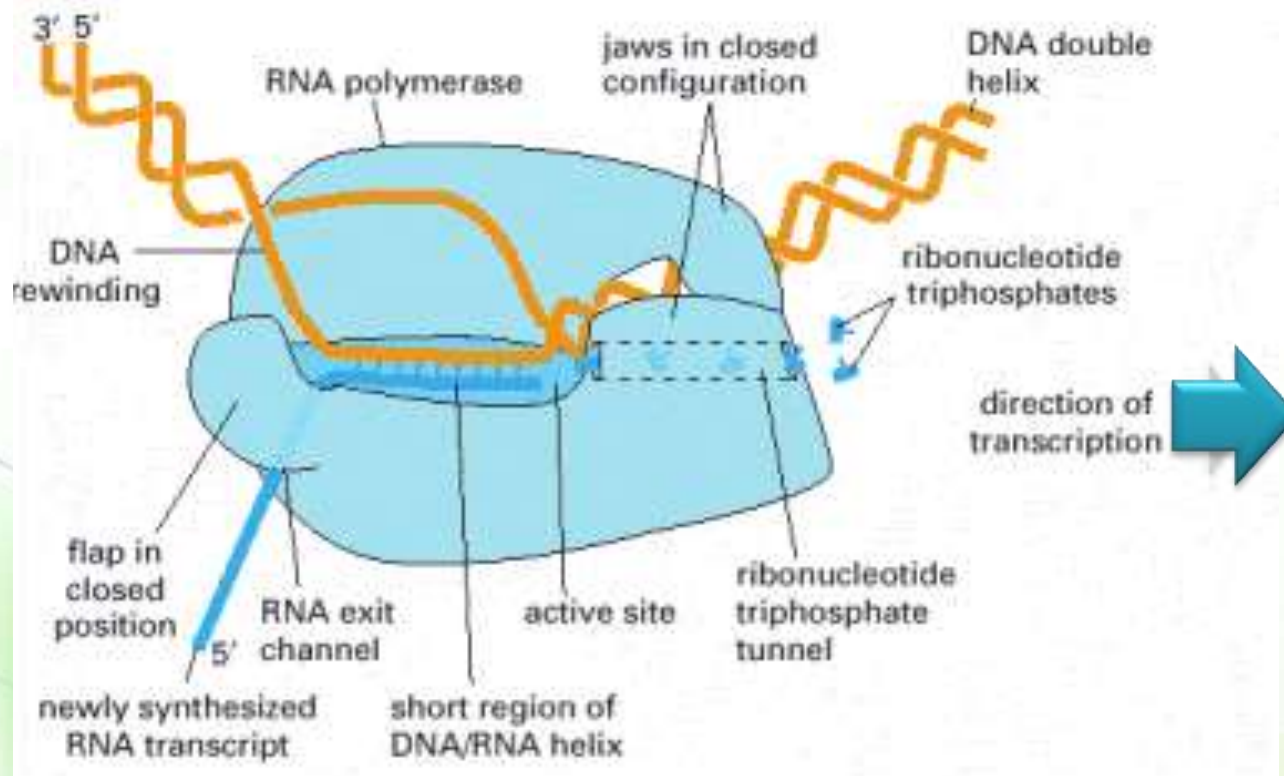


- RNA polymerase catalyzes the linkage of ribonucleotides, not deoxyribonucleotides.
- Unlike DNA polymerases, RNA polymerases can start an RNA chain without a primer.
- RNA polymerases make about one mistake for every  $10^4$  nucleotides.
  - the consequences of an error in RNA transcription are much less significant than that in DNA replication.
- Although RNA polymerases are not as accurate as DNA polymerases, they have a modest proofreading mechanism.

# RNA binding to DNA is temporary



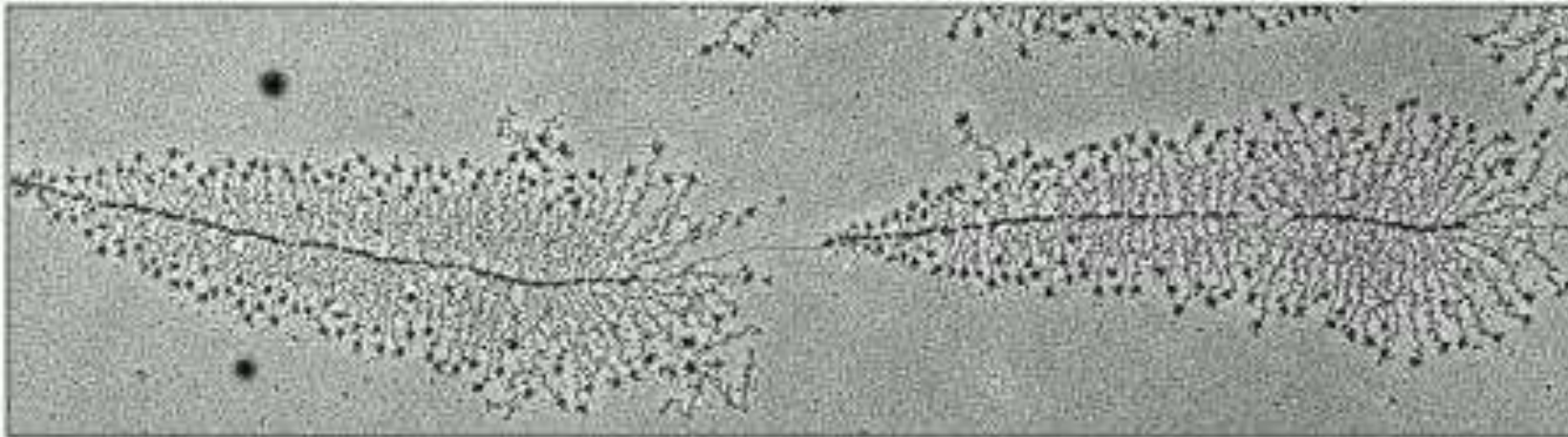
- As RNA is synthesized, it is initially bonded to DNA, but after a short distance, the older polymerized RNA nucleotides are separated, and the newer ones become bonded.



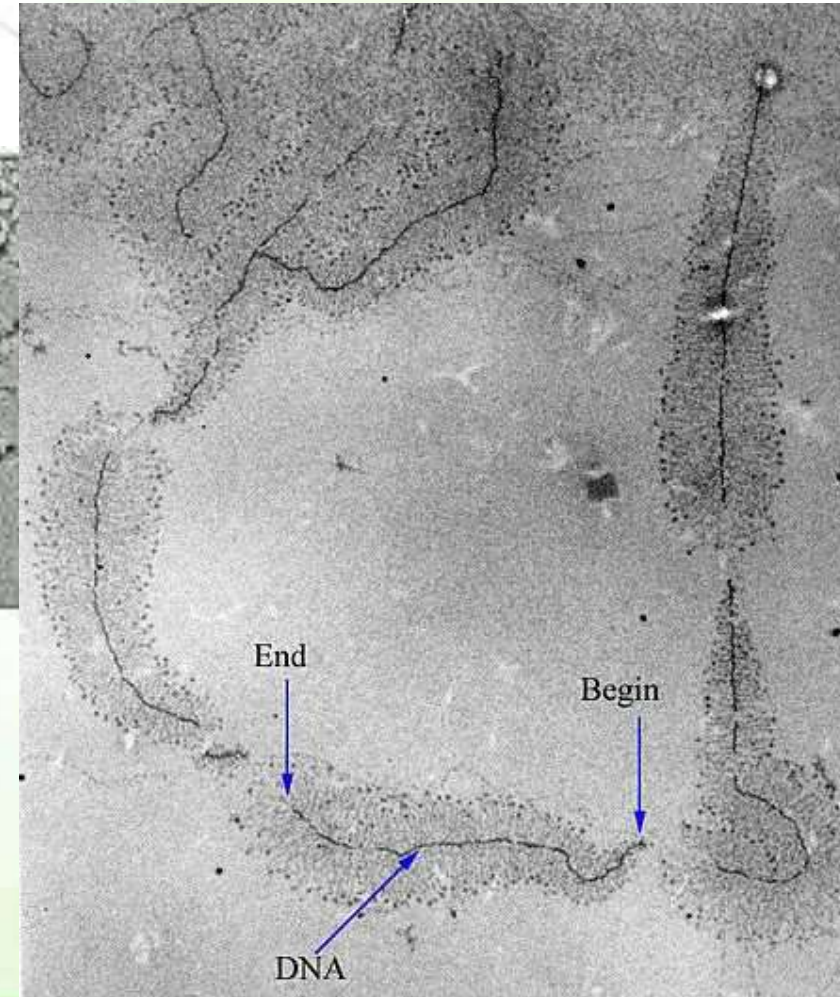
# Polysomes



- This allows the simultaneous synthesis of many RNA chains from the same gene forming structures known as polysomes.



Where is the starting point of transcription? Where is the beginning of the genes?



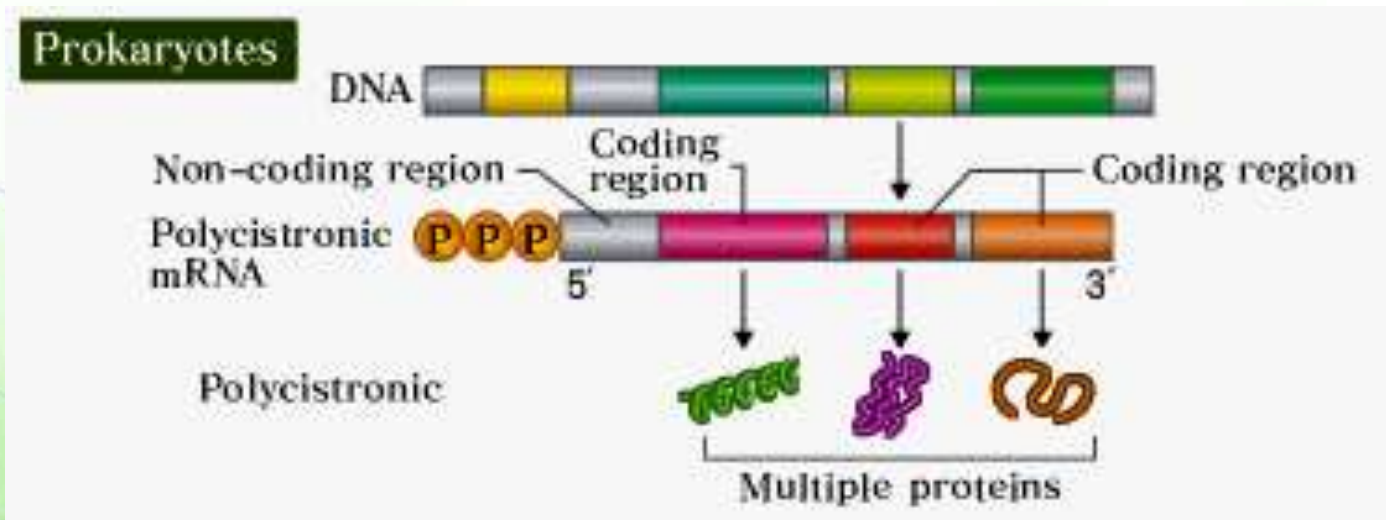


# Transcription in prokaryotes

# Prokaryotic genes (operons)



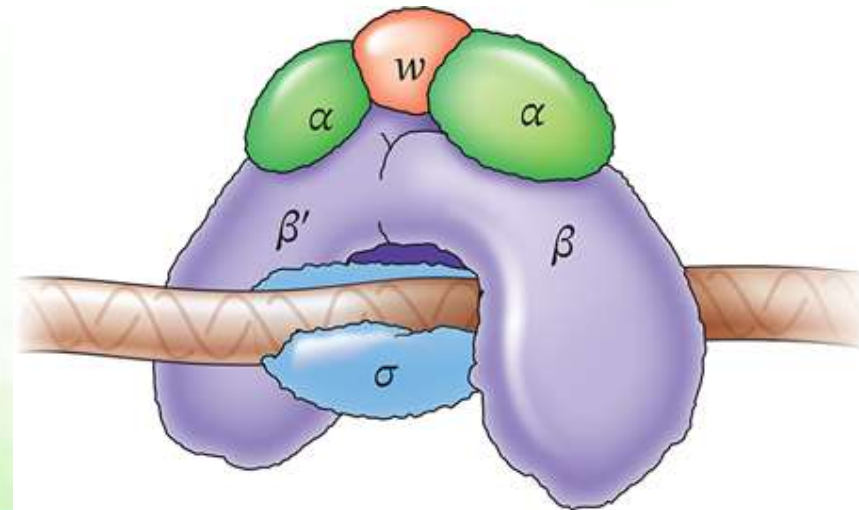
- In bacteria, genes can be **polycistronic** (define!).
- Genes that encode enzymes that are involved in related functions, are often transcribed as one unit from one cistron.
  - Example: the genes encoding the enzymes required to synthesize the amino acid tryptophan are contiguous.
- This cluster of genes comprises a single transcriptional unit referred to as an **operon**.



# The RNA polymerase



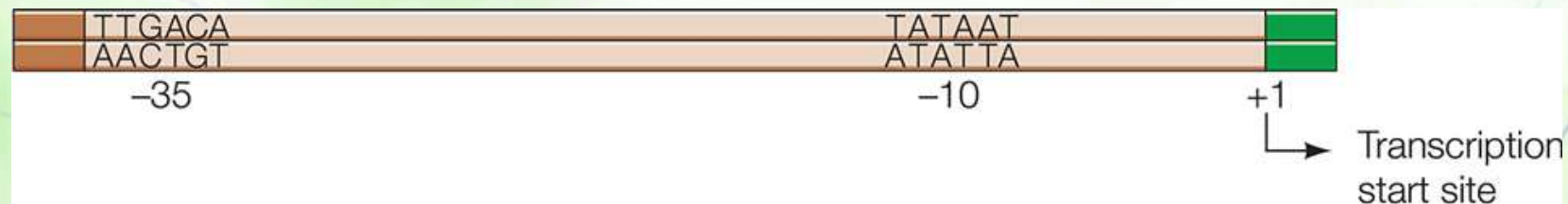
- E. coli RNA polymerase is made up of multiple polypeptide chains or subunits.
- The core polymerase consists of two  $\alpha$ , one  $\beta$ , one  $\beta'$ , and one  $\omega$  subunits.
  - The core polymerase is fully capable of catalyzing the polymerization of NTPs into RNA.
- The enzyme also contains a  $\sigma$  subunit, but it is not required for the basic catalytic activity of the enzyme.



# Consensus sequences (the promoter)



- The DNA sequence to which the RNA polymerase binds to to initiate transcription of a gene is called the **promoter**.
  - A promoter is "upstream" of the transcription initiation site.
- The promoter region upstream of the transcription initiation site contains two sets of sequences that are similar in a variety of genes.
  - **Consensus!**
- They are called the (-10) and (-35) **elements** because they are located approximately 10 and 35 base pairs upstream of the transcription start site.
- The transcription initiation site is defined as the +1 position.

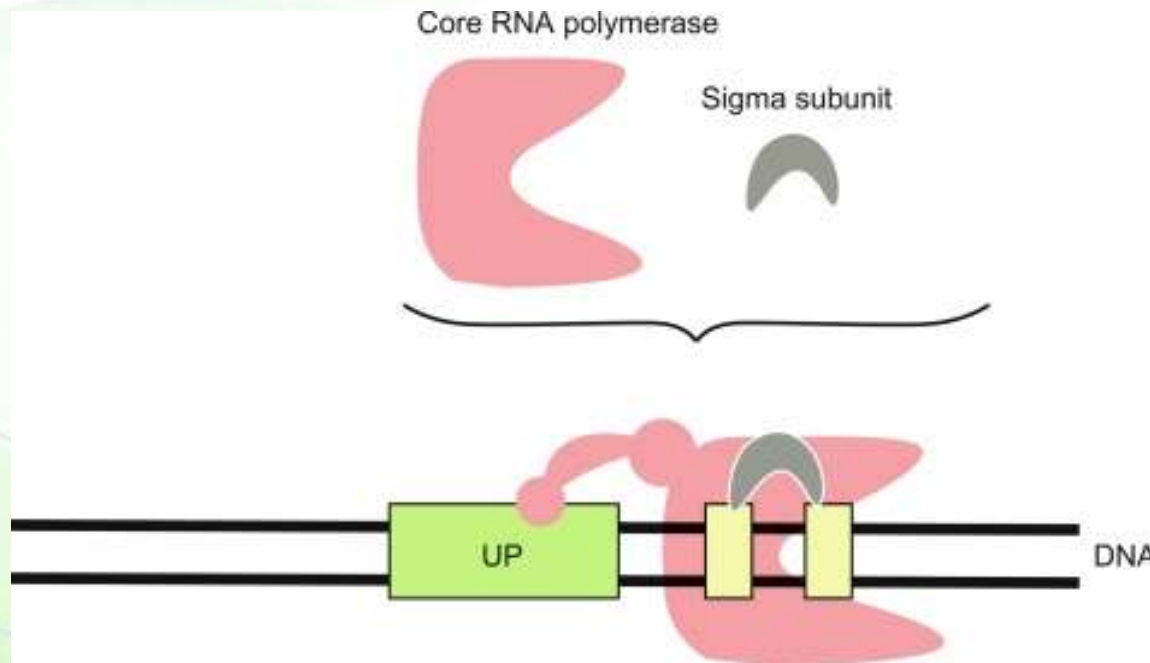




# Role of the $\sigma$ subunit



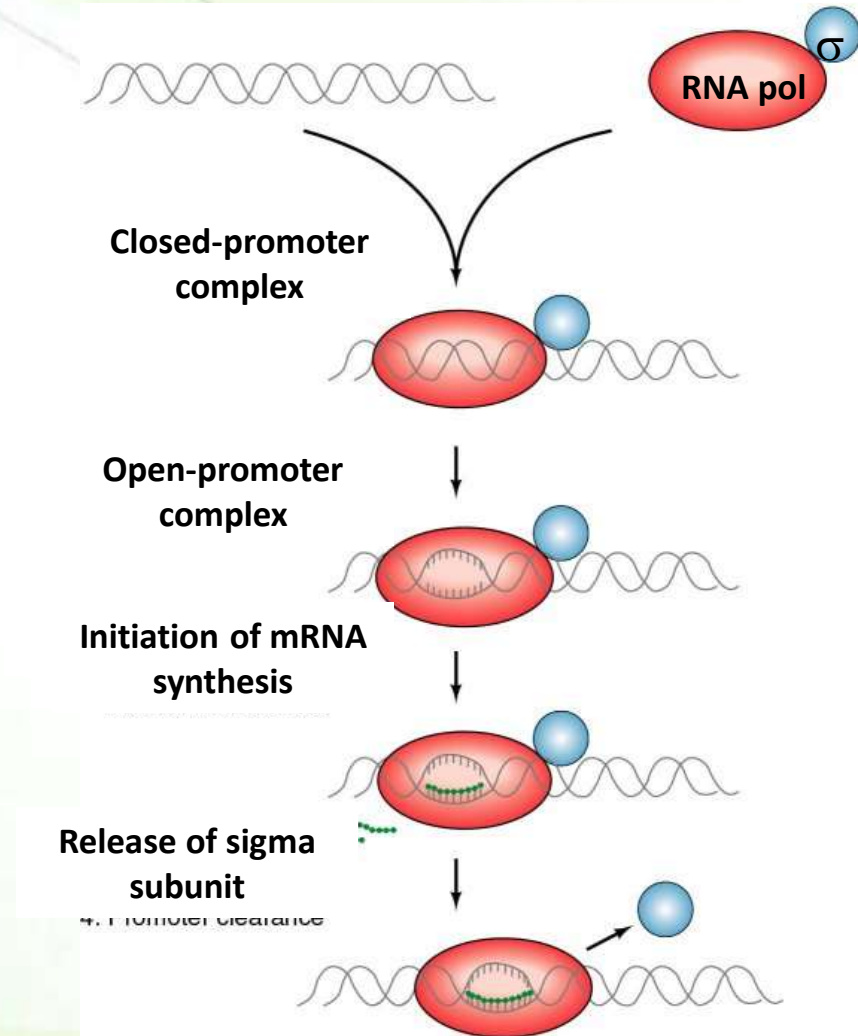
- In the absence of  $\sigma$ , the RNA polymerase binds to DNA with low affinity and nonspecifically.
- The role of  $\sigma$  is to identify and guide the polymerase to the -35 and -10 sequences.



# Mechanism of transcription (initiation)



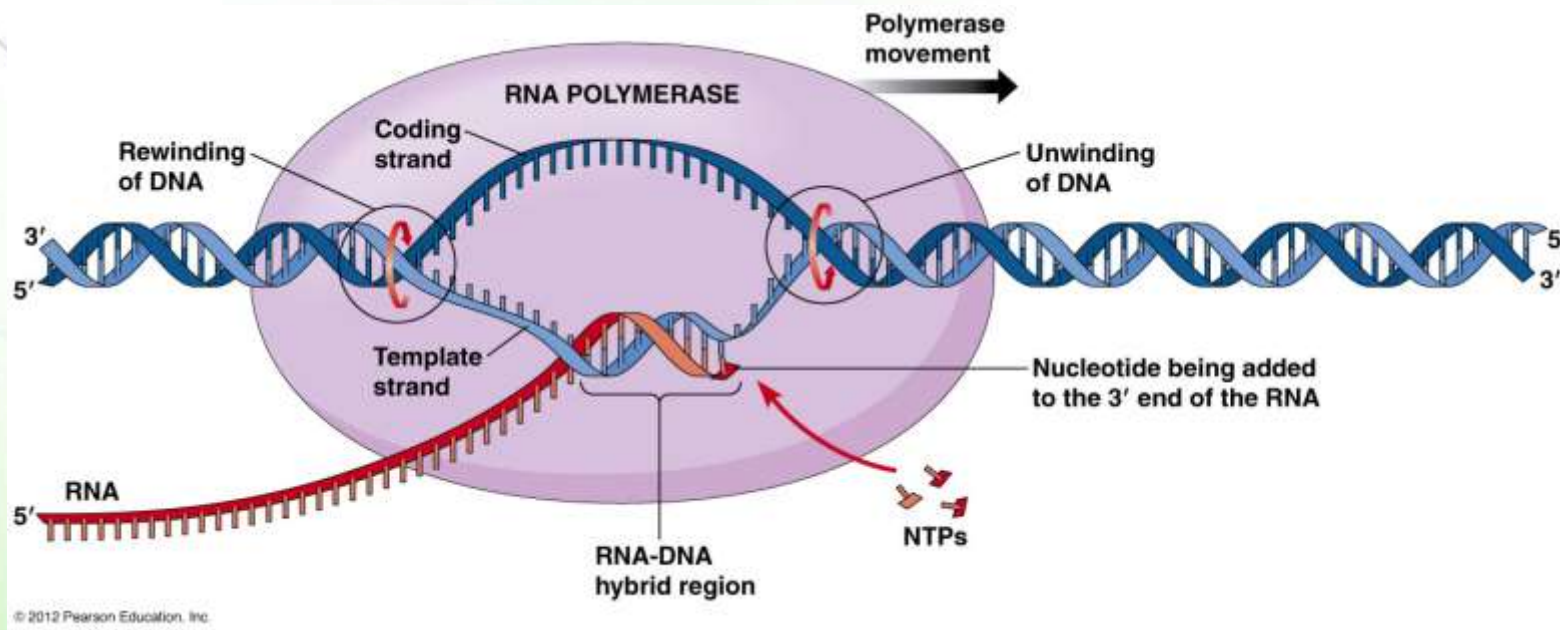
- The RNA polymerase binds to the promoter and opens it (like what?).
- The single-stranded DNA is now available as a template.
- Transcription is initiated by the joining of two NTPs.
- After addition of about 10 nucleotides,  $\sigma$  is released from the polymerase.
- What do you think happens to it?



# Mechanism of transcription (elongation)



- As the polymerase moves forward, it
  - unwinds the template DNA ahead of it (like what?)
  - elongates the RNA
  - rewinds the DNA behind it

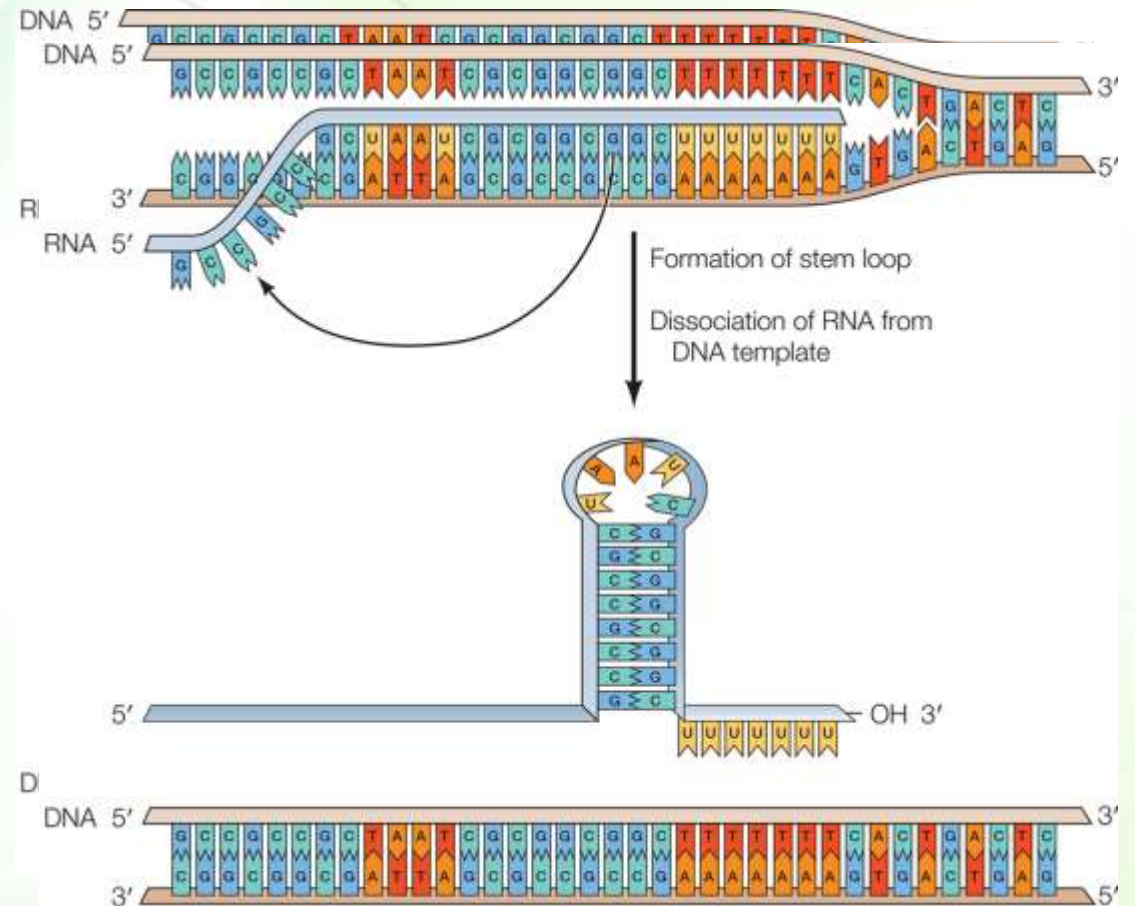
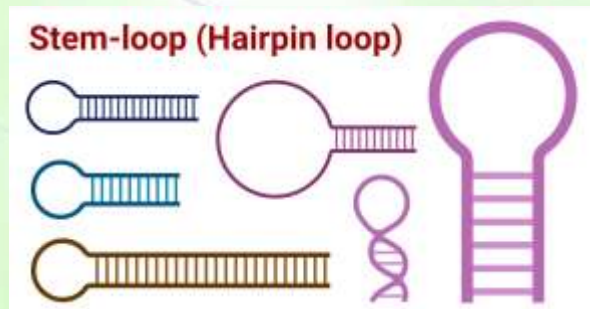


RNA synthesis continues until the polymerase encounters a termination signal where the RNA is released from the polymerase, and the enzyme dissociates from its DNA template.

# Termination sequences



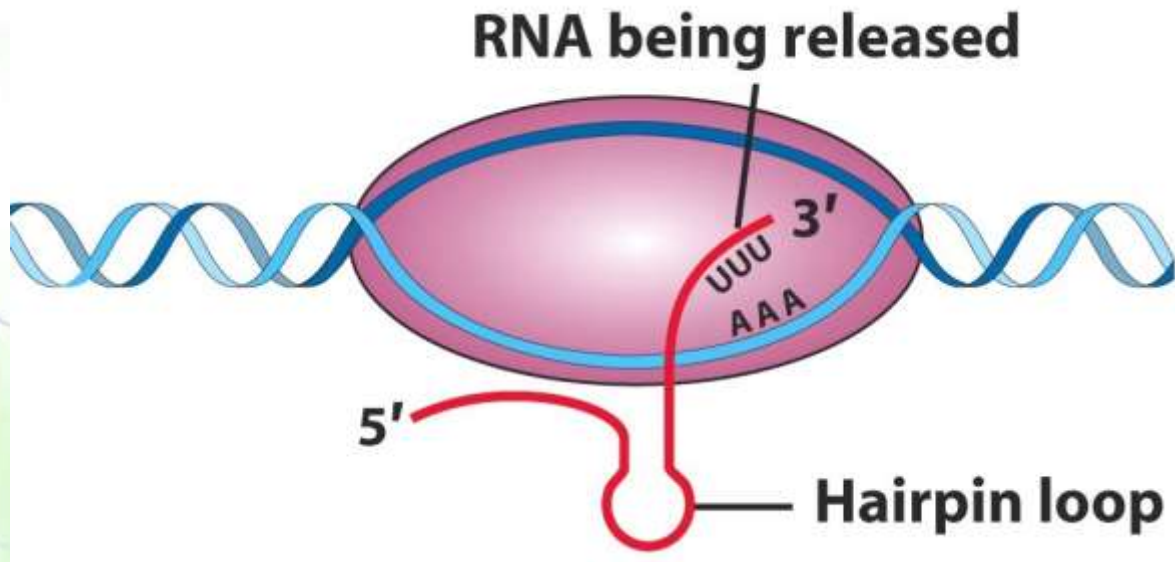
- The most common termination signal among genes in *E. coli* is a stem-loop structure that consists of a symmetrical inverted repeat of a GC-rich sequence followed by A residues (why?).
- Transcription of the GC-rich inverted repeat results in the formation of a stable stem-loop structure.



# The effect of the stem loop structure



- The formation of this structure breaks RNA association with the DNA template, destabilizes the RNA polymerase binding to DNA, and terminates transcription.





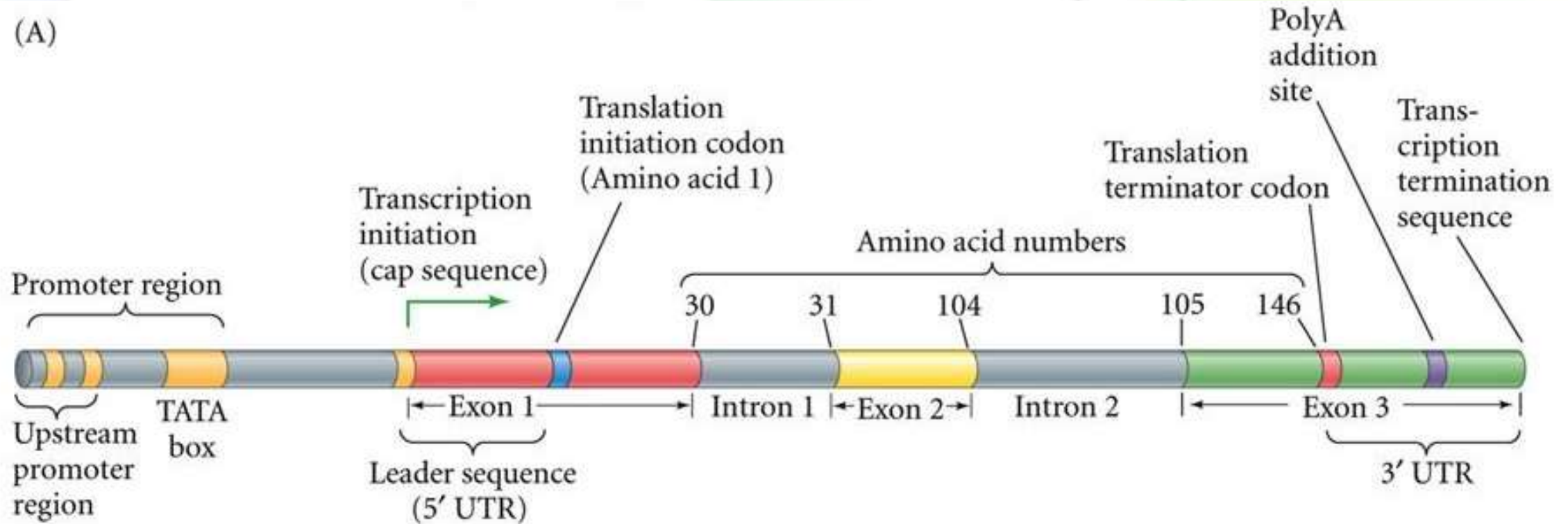


# Transcription in eukaryotes

# Anatomy of a eukaryotic gene



(A)







- In contrast to bacteria, which contain a single type of RNA polymerase, eukaryotic nuclei have three, called RNA polymerase I, RNA polymerase II, and RNA polymerase III
  - RNA polymerase I transcribes rRNA genes.
  - RNA polymerase II transcribes protein-encoding genes (mRNA), long noncoding RNA (lncRNA), and microRNA (miRNA). We will focus on this.
  - RNA polymerase III transcribes tRNA genes and one rRNA gene.

# Eukaryotic RNA polymerases



- Eukaryotic transcription initiation must deal with the packing of DNA into nucleosomes.
- While bacterial RNA polymerase is able to initiate transcription without help from any proteins, eukaryotic RNA polymerases require help from **general transcription factors**.
  - They are "general" because they assemble on all promoters used by RNA polymerase II.
  - They are designated as TFII (for transcription factor for polymerase II), and listed as TFIIA, TFIIB, and so on.

# General transcription factors

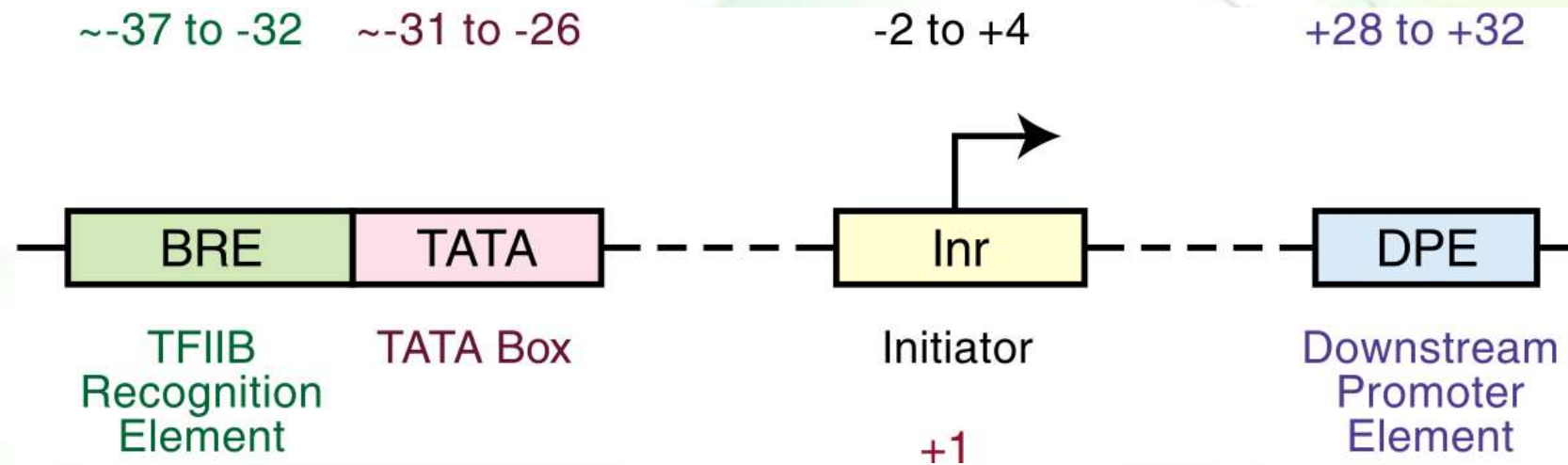


- These general transcription factors
  - help position the RNA polymerase correctly at the promoter (like what?).
  - aid in pulling apart the two strands of DNA to allow transcription to begin (like what?).
  - push the RNA polymerase forward to begin transcription.

# Core components of promoters



- The promoter region in eukaryotic cells is complex.

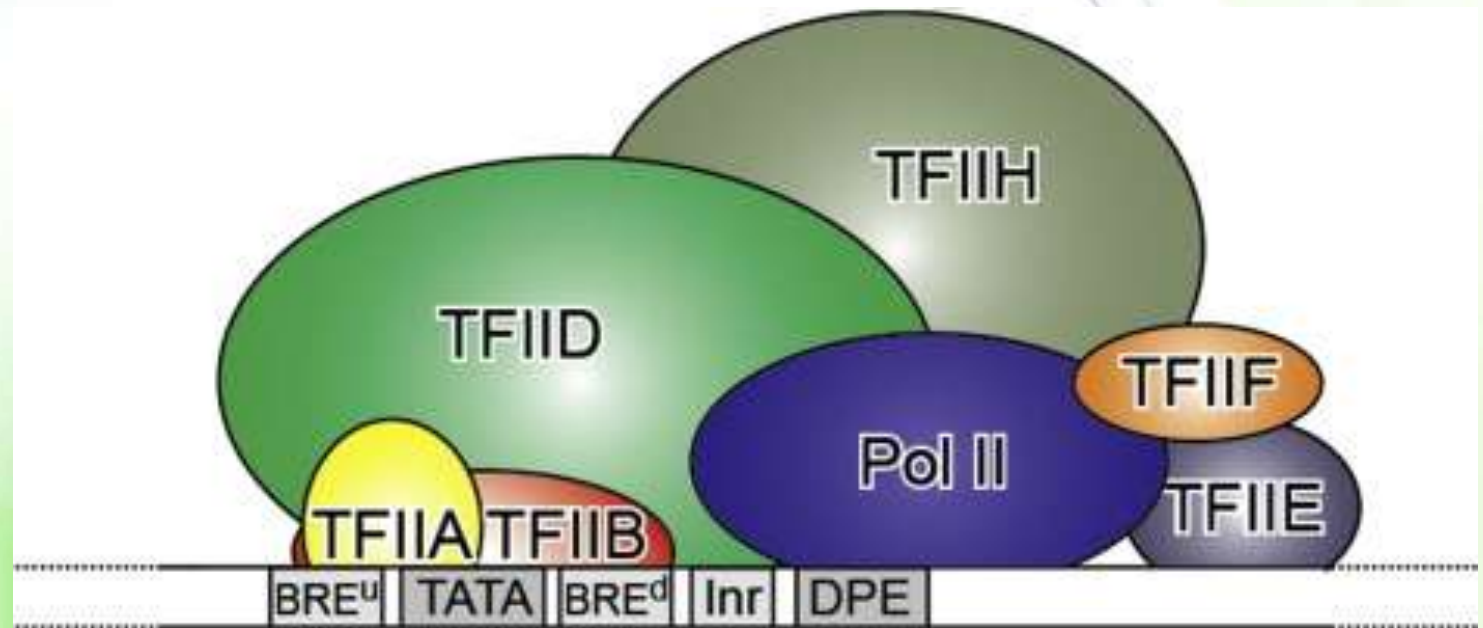


- Not all of these sequences exist at once, but genes can have a combination of these promoter elements.
- This is called the **core promoter region**.

# Formation of preinitiation complex

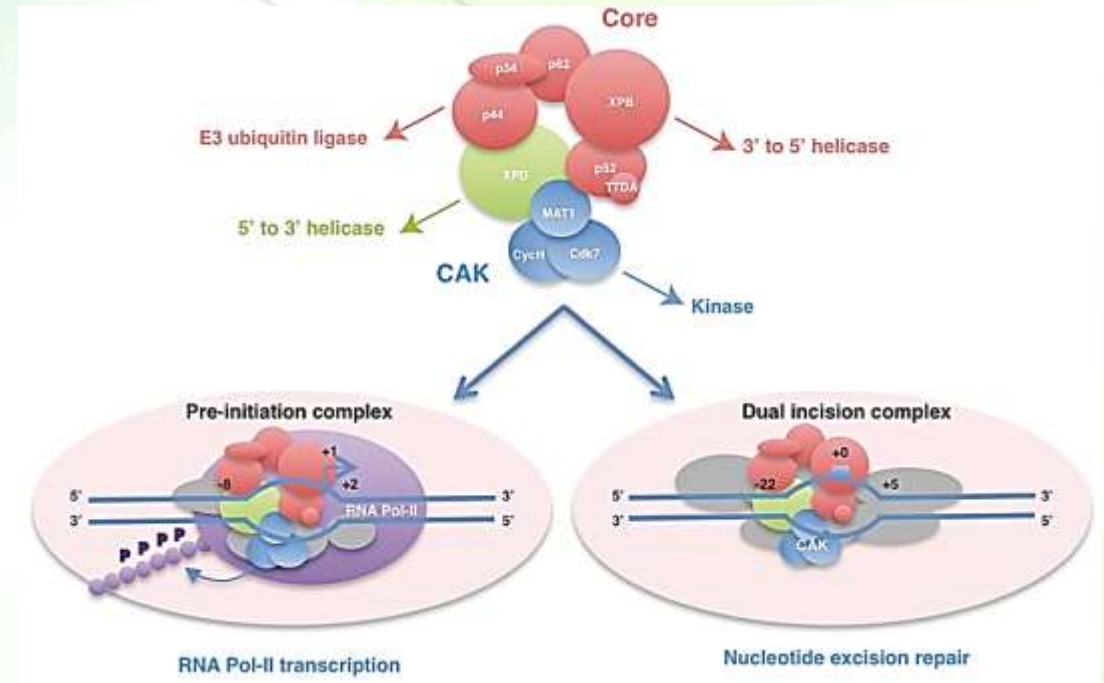


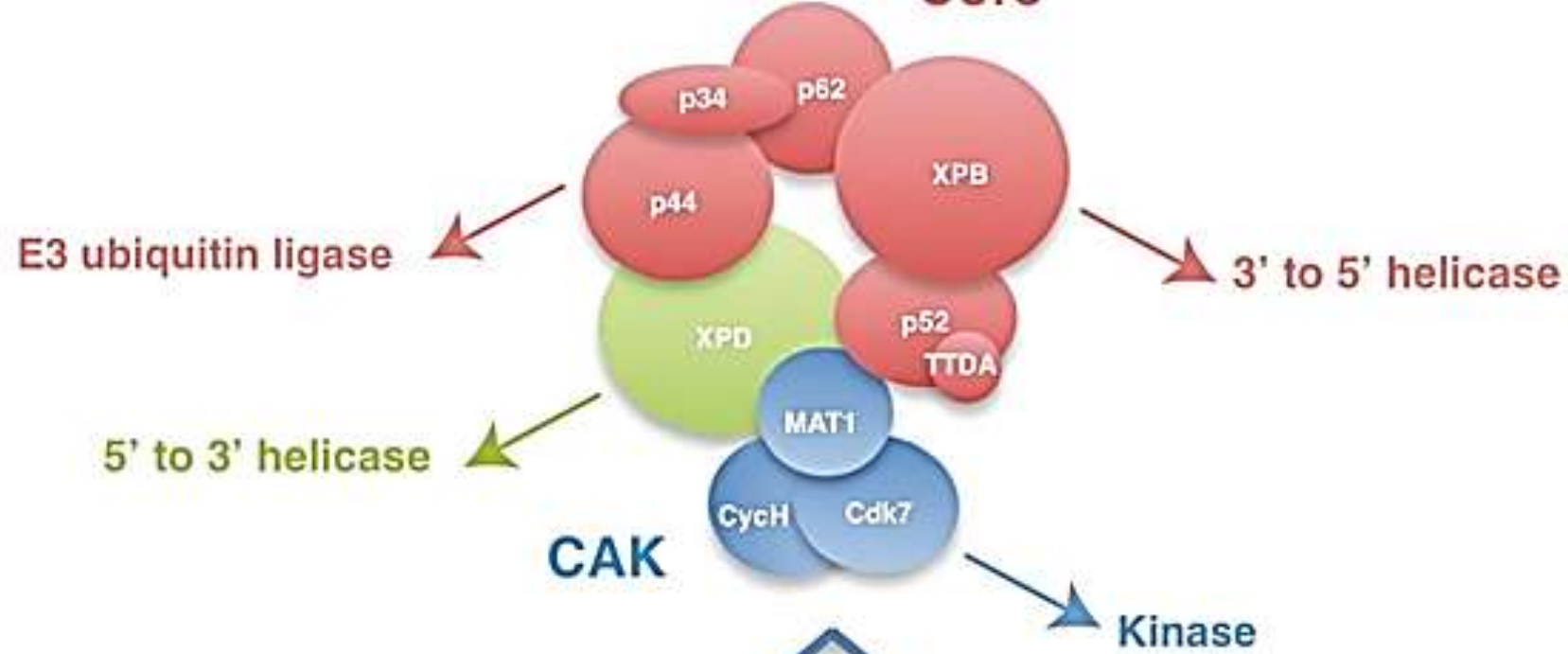
- Formation of a transcription complex is initiated by the binding of the transcription factor TFIID complex.
- Other subunits bind to the promoter, followed by binding of the polymerase.
- TFIIF then binds.



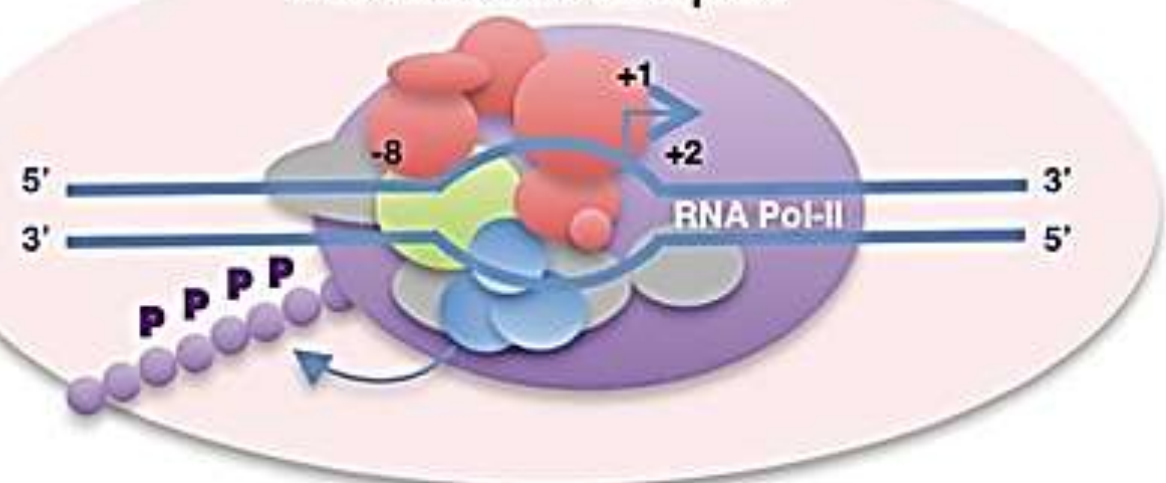


- TFIIH is a multi-subunit factor that has two important enzymatic activities:
  - A helicase activity, which unwind DNA around the transcription start site and is catalyzed by two subunits (XPB and XPD proteins), which are also required for nucleotide excision DNA repair.
  - A kinase activity that phosphorylates the C-terminal domain (CTD) of the largest subunit of RNA polymerase II.
- Phosphorylation releases the polymerase from the preinitiation complex and leads to the initiation of transcription.

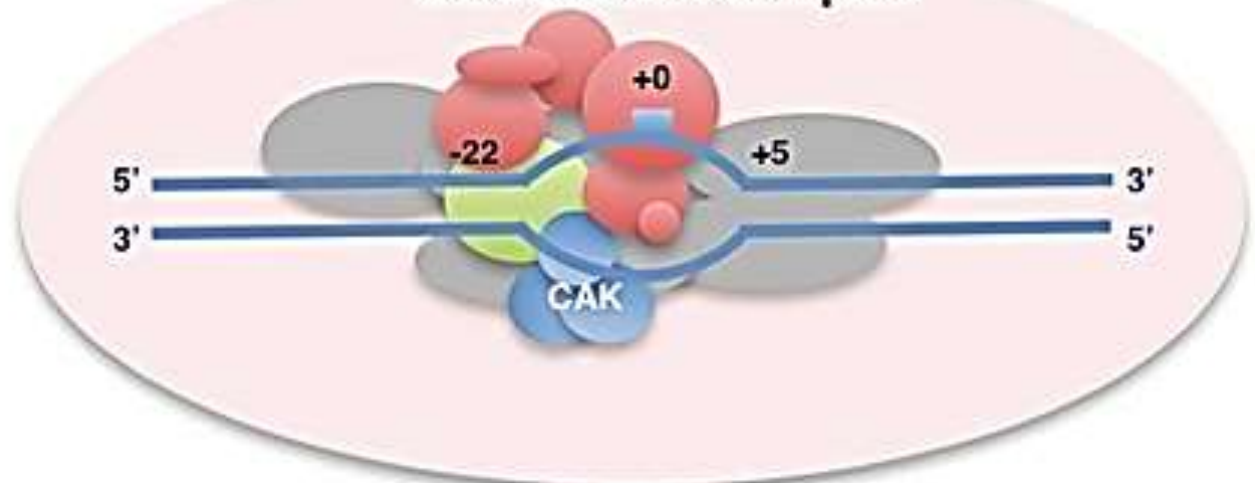




**Pre-initiation complex**



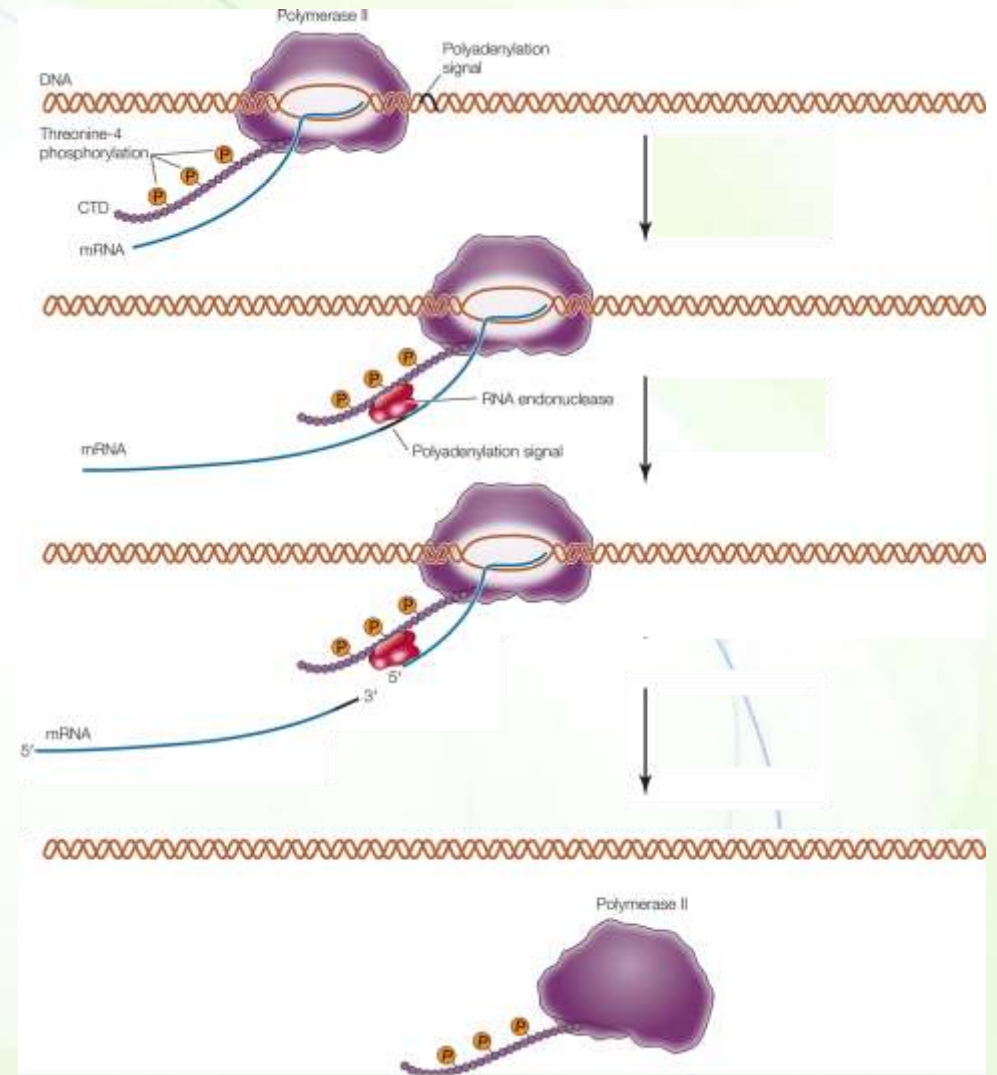
**Dual incision complex**



# Termination of transcription



- Protein-coding genes have a polyadenylation signal at their 3' end.
- The signal is also transcribed and is recognized by an RNA endonuclease that cleaves the polyadenylation signal within the nascent transcript.
- The RNA and the polymerase are released from the DNA template.

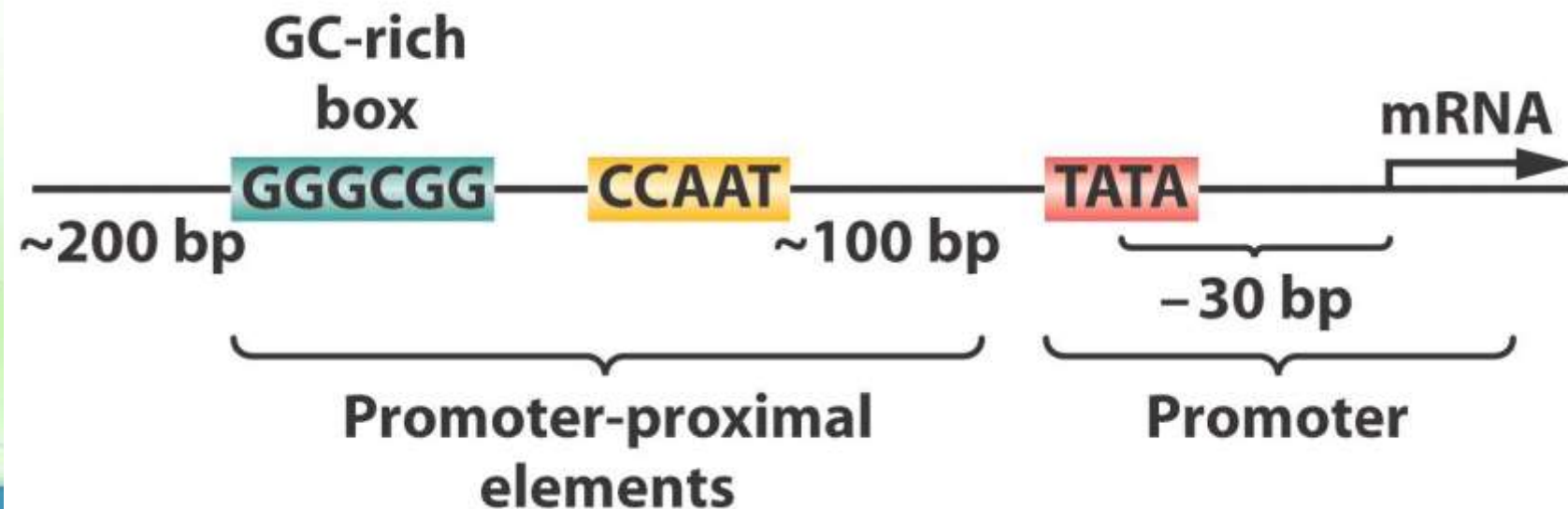




# Promoter-proximal elements



- These are upstream of the core promoter region and sometime called **response elements**.
- They are important for strong expression (versus basal expression).
- They can be shared among different genes (gene-specific) that participate in a similar mechanism or needed for a particular purpose (example: production of enzymes for metabolism of glucose).
  - Alternative to operons!

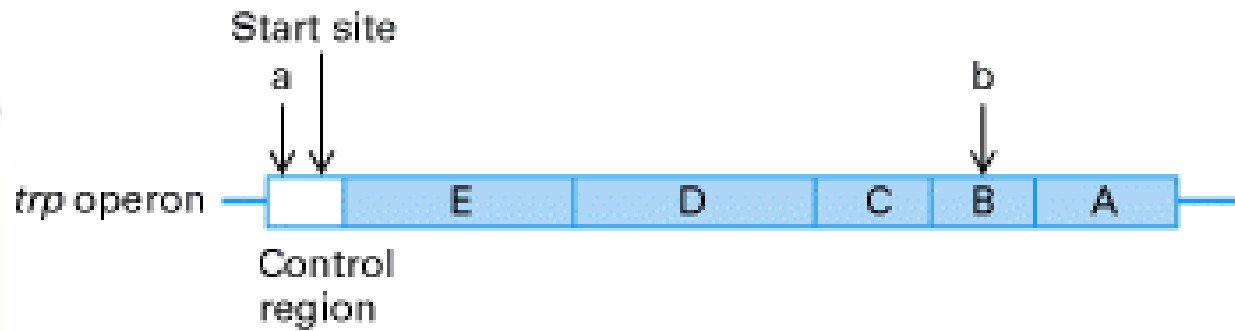


# Operon vs. Proximal-promoter elements



## Prokaryotes

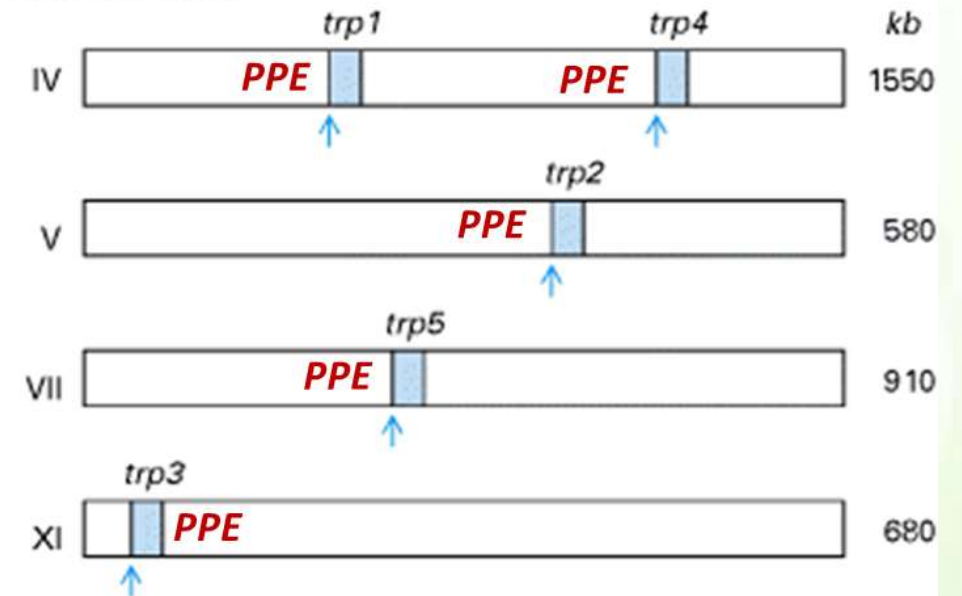
(a) Prokaryotic polycistronic transcription unit



## Eukaryotes

(b) Eukaryotes

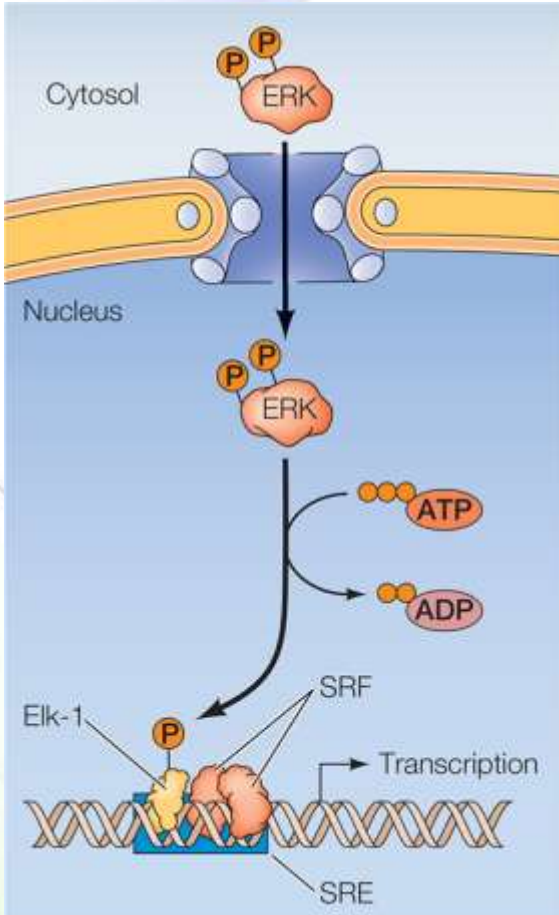
Yeast chromosomes



# Examples of response elements

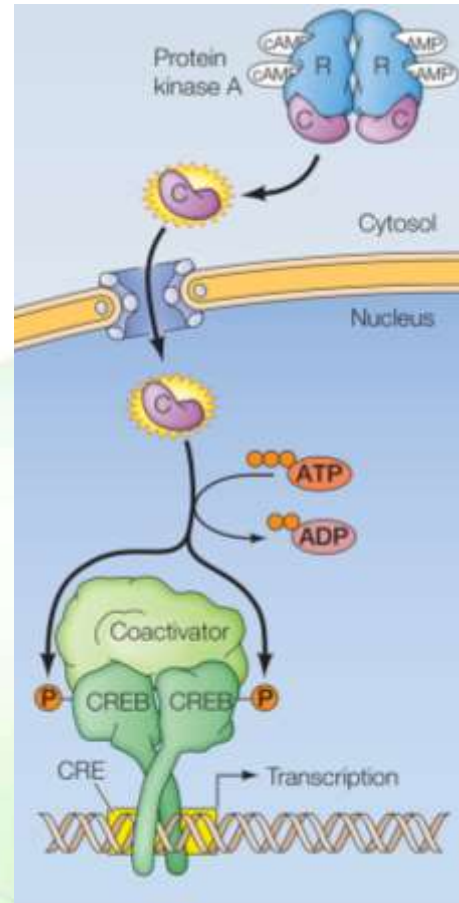


Serum response element (SRE)



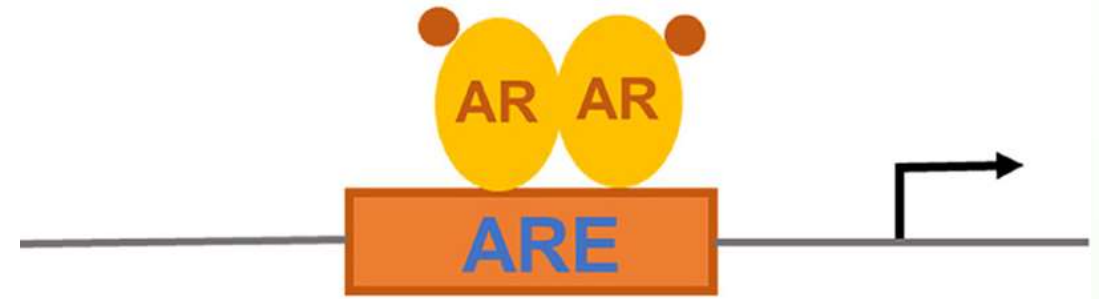
Stimulation of cell proliferation induced by growth factors.

cAMP response element (SRE)



Regulation of proliferation, survival, and differentiation

Androgen response element (ARE)

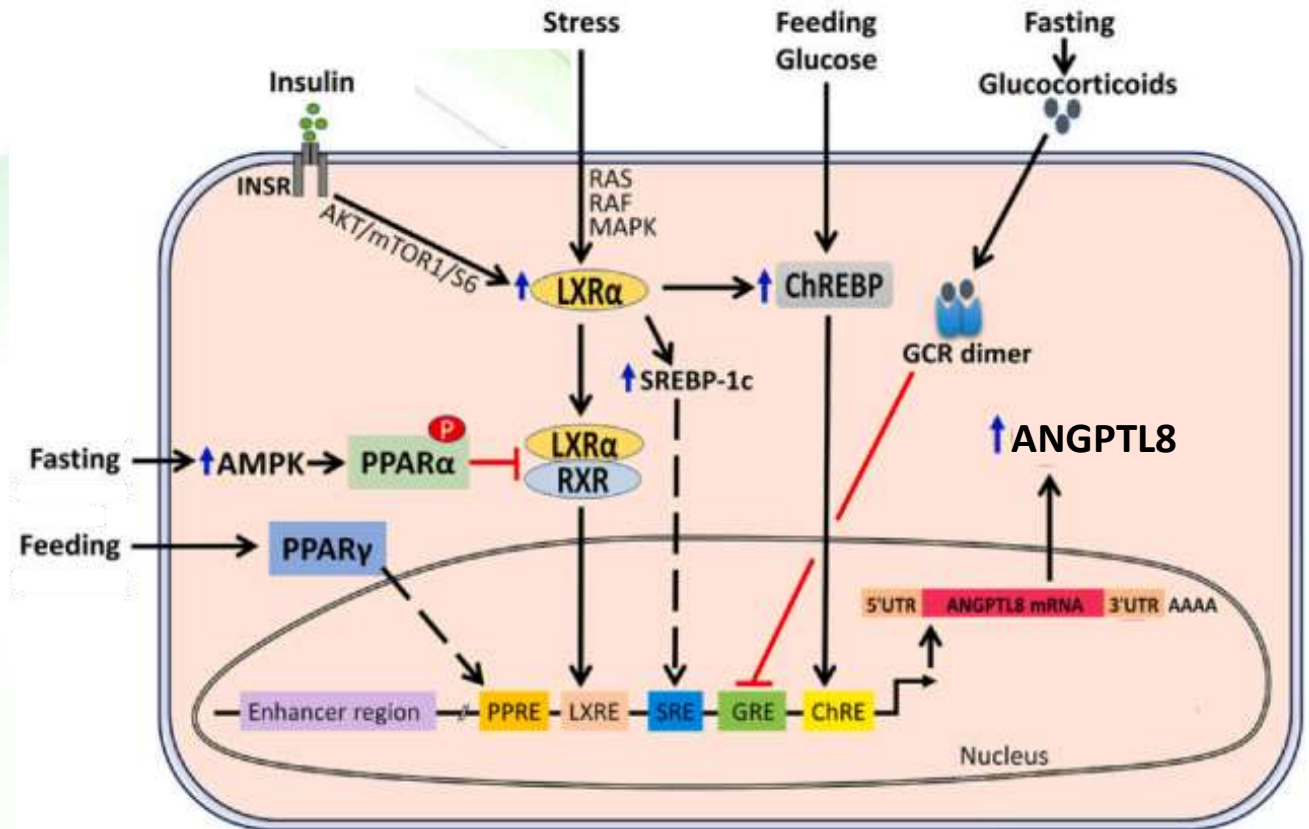


Regulation of metabolism and differentiation

# A single gene, multiple PPEs

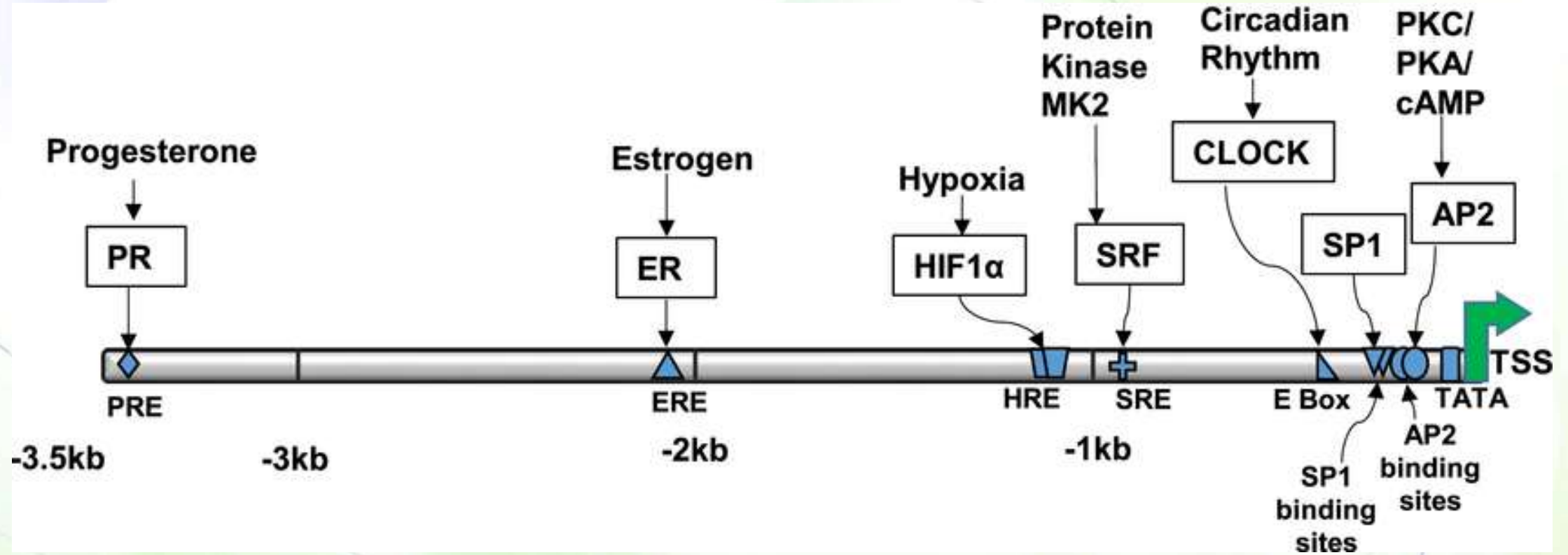


- Angiopoietin-like protein 8 (ANGPTL8) is important for the metabolism of lipoproteins and triglyceride is an inhibitor of lipoprotein lipase (LPL).
- Inducers of anabolism such as feeding, glucose, insulin stimulate transcription factors such as SREBP-1c and ChREBP.
- Inducers of catabolism such as fasting (and glucocorticoids) suppress the gene expression of ANGPTL8.
- *Note: not all PPEs have a positive effects.*

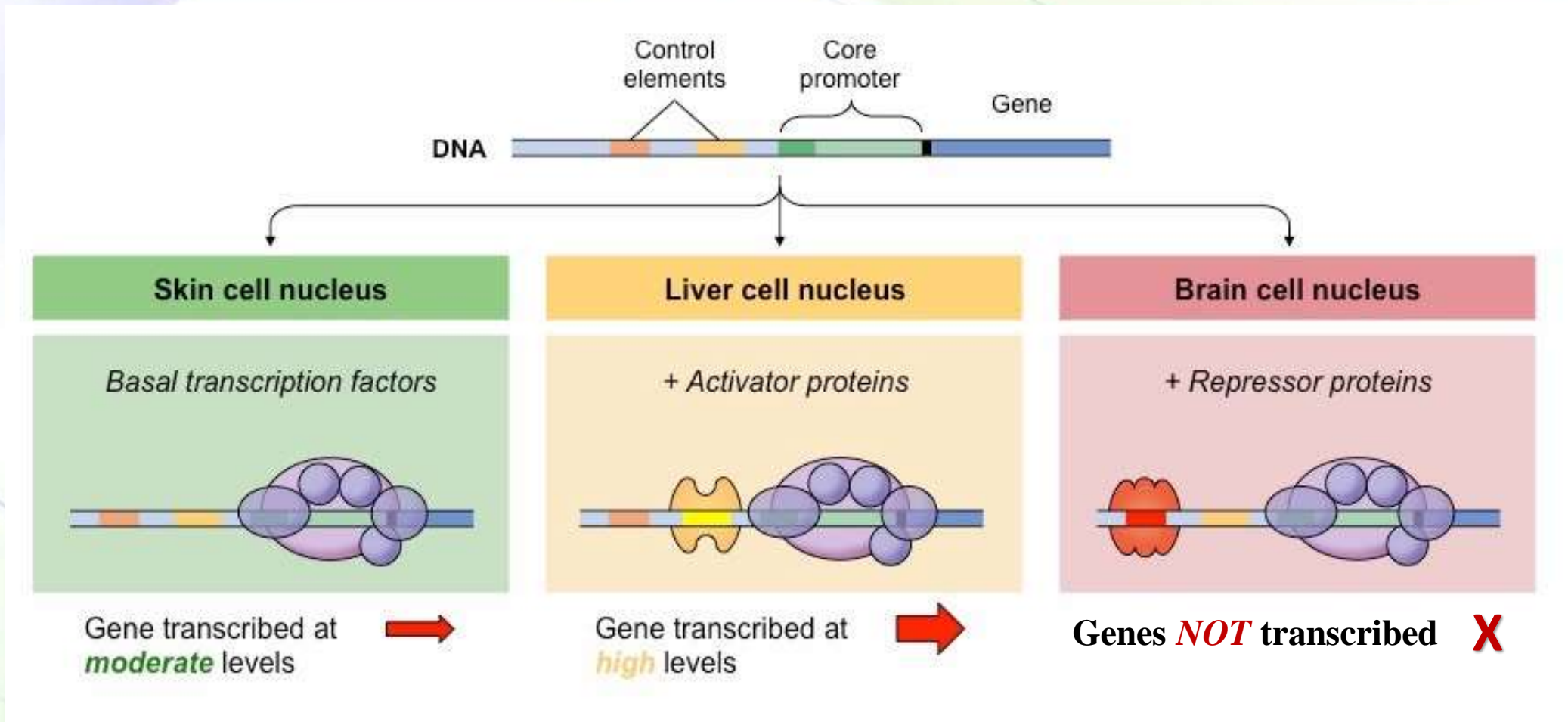


*SREBP-1c: Steroid response element binding protein*  
*ChREBP: Cholesterol response element binding protein*

# 6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3 (PFKFB3)



# Tissue-specific transcription factors

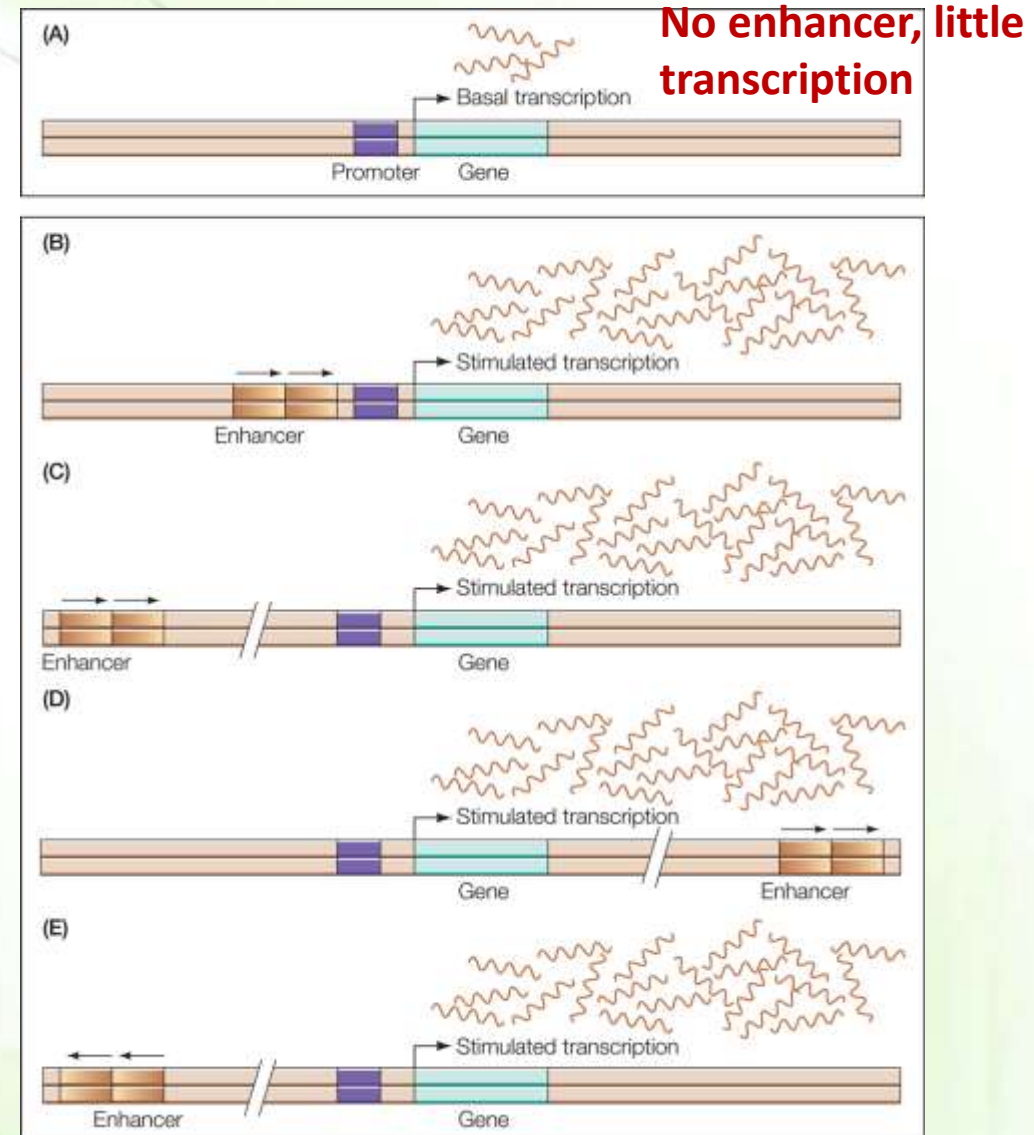


**Differential expression of transcription factors (tissue-specific transcription factors) determines gene expression.**

# Enhancers



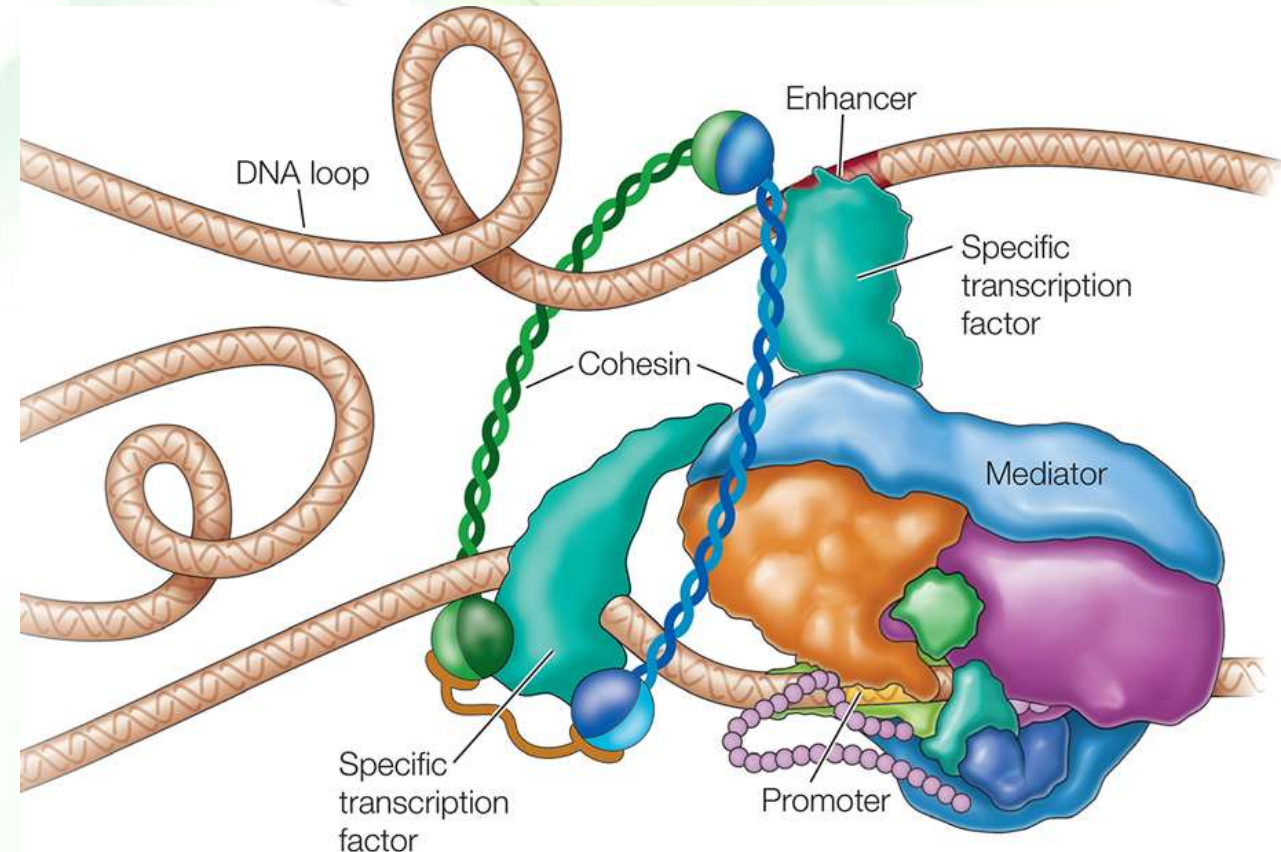
- Many genes are regulated by regulatory sequences called enhancers, which are binding sites for specialized, gene-specific, cell-specific, regulatory transcription factors that regulate RNA polymerase II.
- They can regulate transcription regardless of orientation or location due to DNA looping.
- There are 500,000 to over 1 million enhancers in the human genome, accounting for 10% or more of total genomic DNA.



# DNA looping and cohesin



- Transcription factors bound at an enhancer can interact with a protein called (Mediator) or general transcription factors at the promoter.
- This is due to the ability of DNA to loop.
- DNA looping is stabilized by a protein called cohesin.

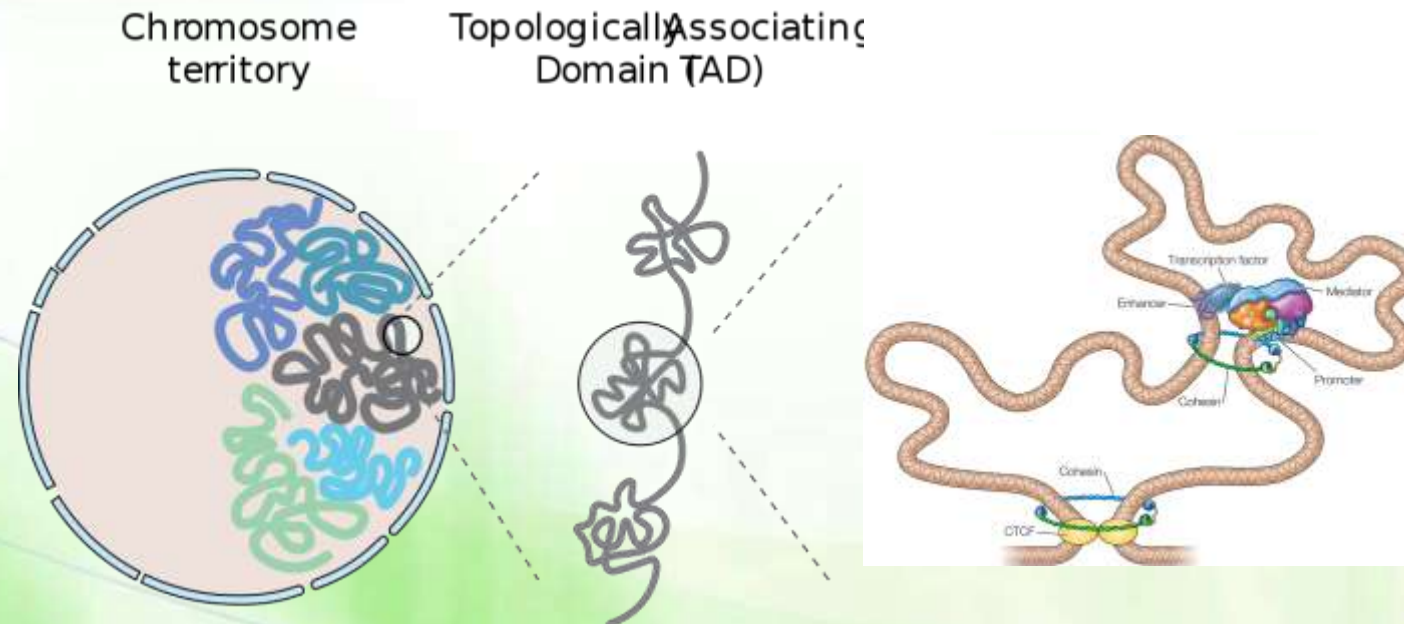




# Enhancers, insulators, TADs, and CTCF



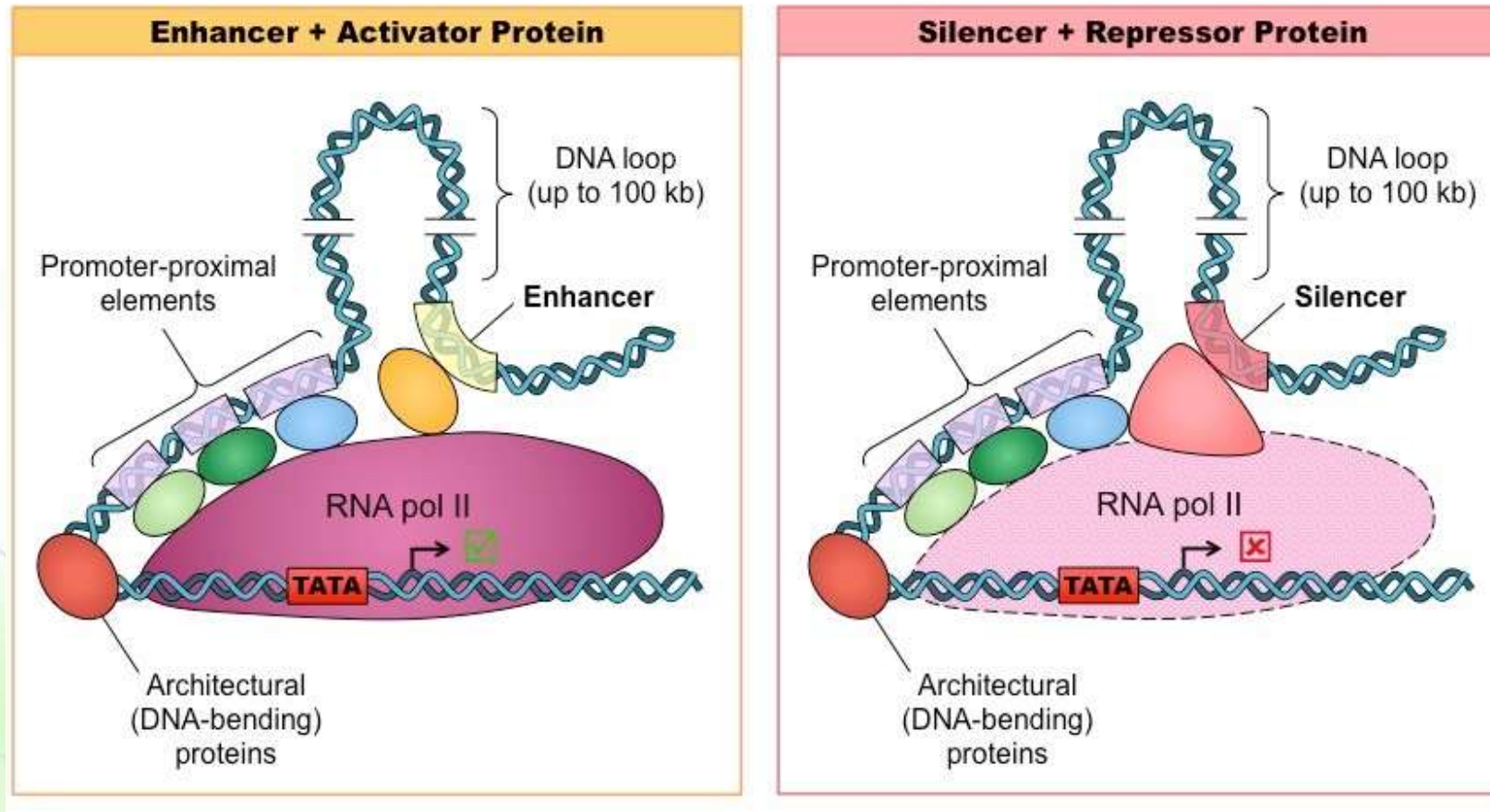
- DNA sequences or elements known as insulators divide the genome into topologically associating domains (TADs) forming loops.
- The boundaries of the loops are stabilized by cohesin and CTCF proteins.
- Enhancers are restricted to interacting with promoters in the same domain.



# Silencers

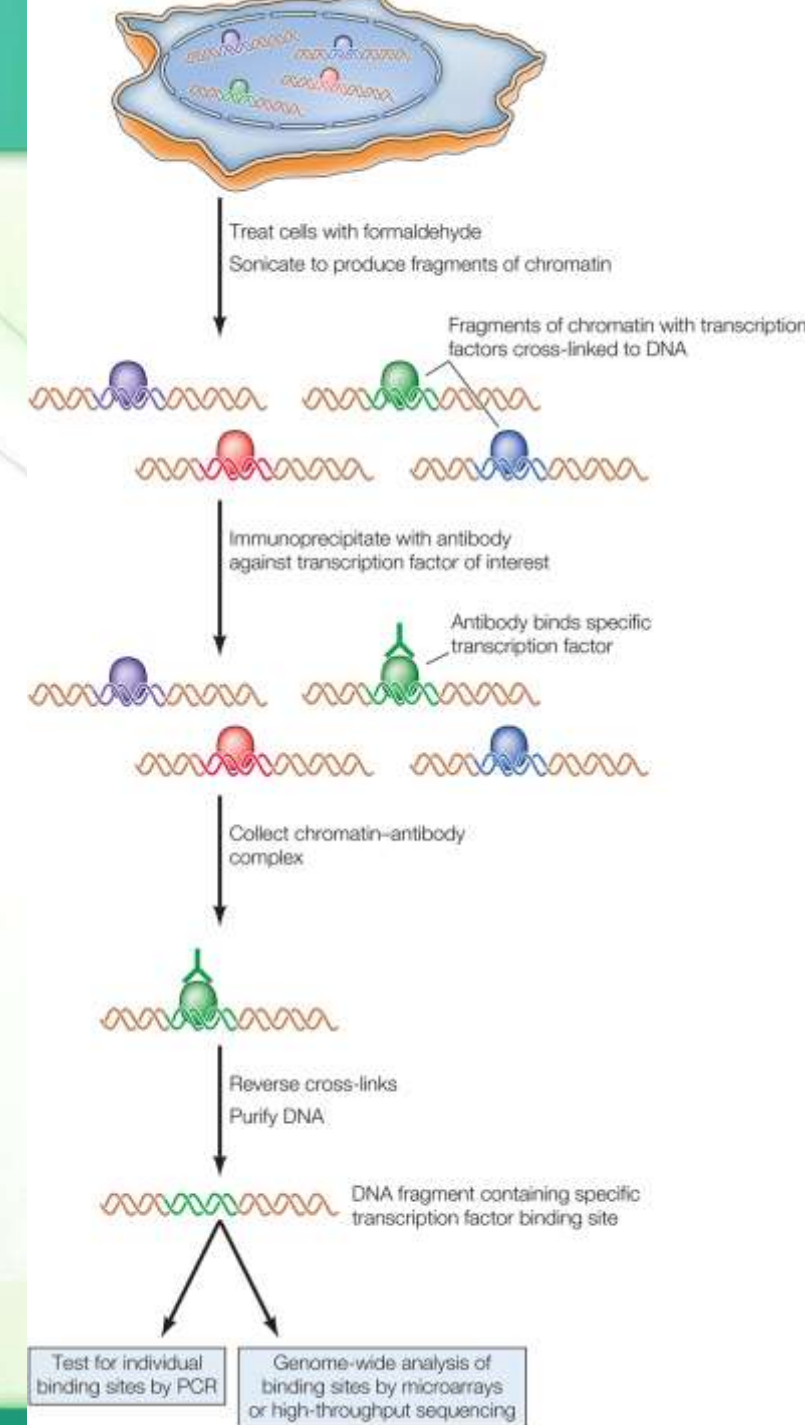


- The opposite of enhancers.



# Chromatin immunoprecipitation

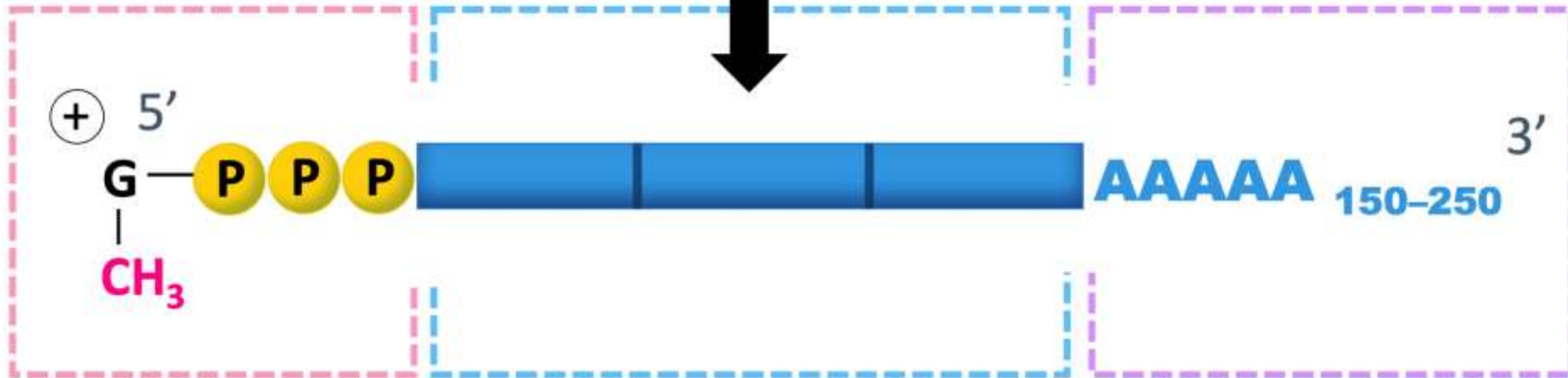
- Transcription factor binding sites can be identified by chromatin immunoprecipitation.
- Proteins bound to DNA are chemically cross-linked to the DNA regions they are bound to.
- DNA is isolated and fragmented.
- The fragments are “immunoprecipitated” with an antibody against a specific transcription factor.
- The cross-links are reversed, and the immunoprecipitated DNA fragments are analyzed by PCR to test for the presence of a specific DNA sequence or by next-generation DNA sequencing microarrays or microarrays to identify all the binding sites for the transcription factor within the genome.





# Eukaryotic RNA processing

# Processing of mRNA in eukaryotes



**Capping**

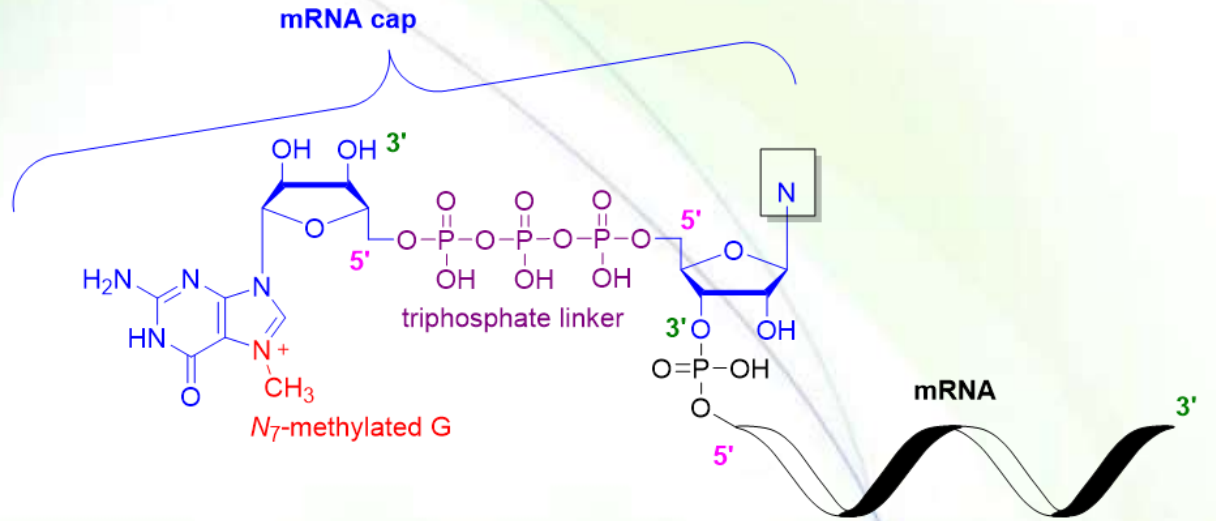
**Splicing**

**Polyadenylation**

# Addition of a cap



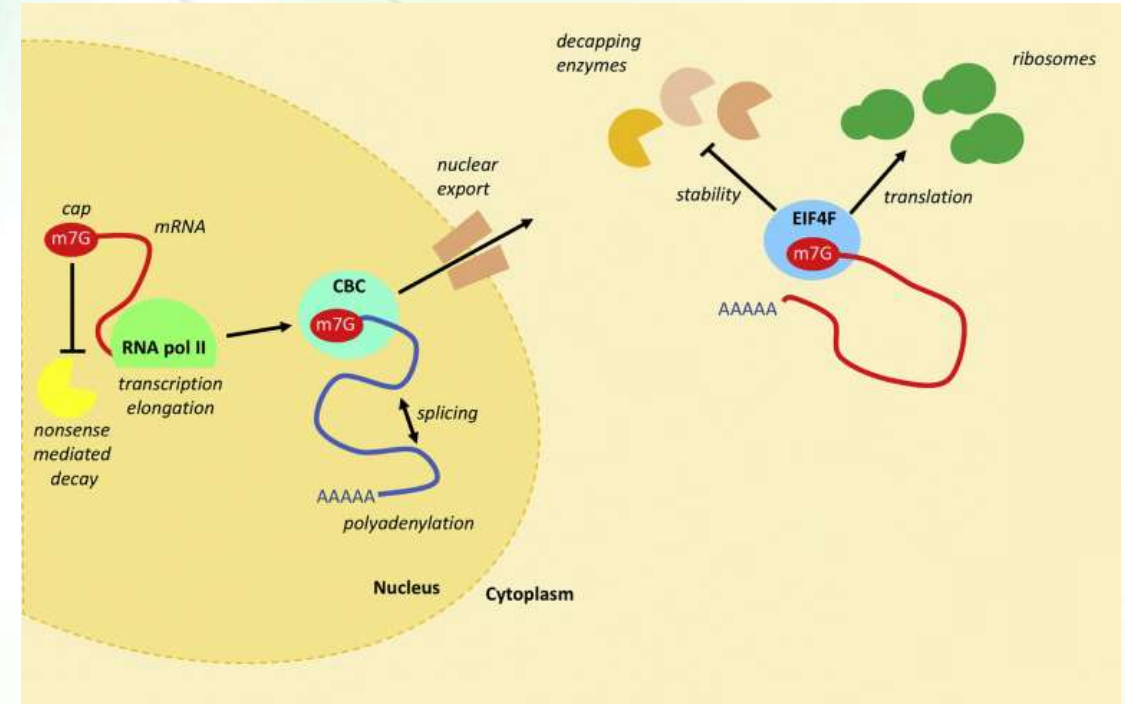
- The first modification comes as soon as RNA polymerase II has produced a few nucleotides of pre-mRNA.
- The 5' end of the new RNA molecule is modified by the addition of a "cap" that consists of a 7-methylguanosine molecule.



# Importance of capping



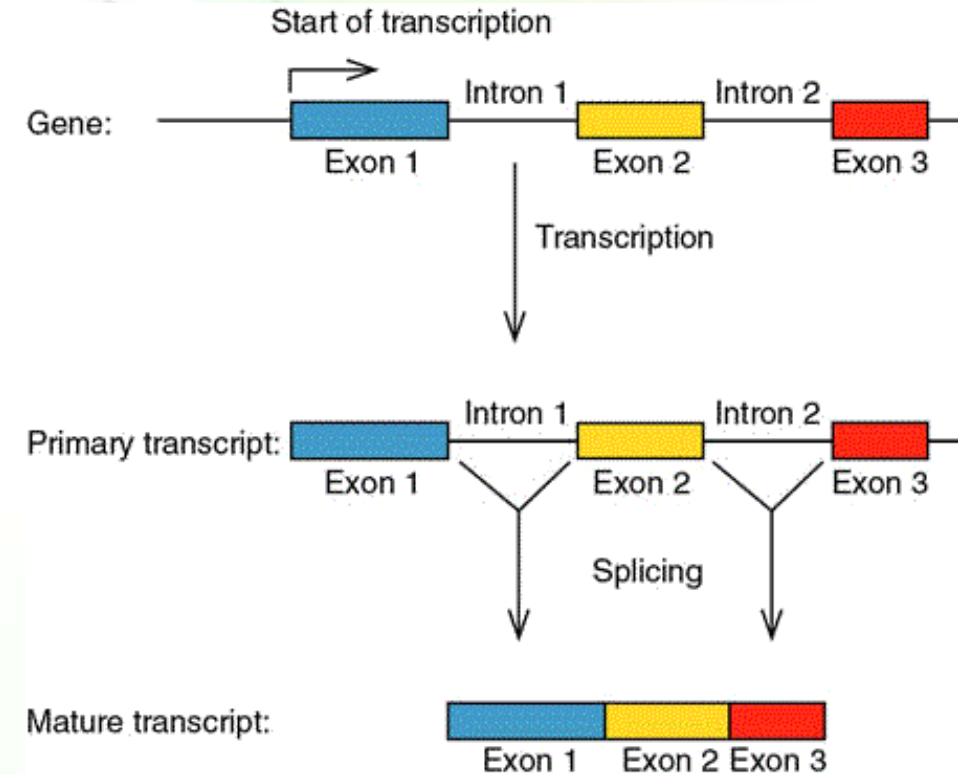
- It differentiates mRNA from other RNA molecules.
- It stabilizes the mRNA.
- It signals the 5' end of eukaryotic mRNAs.
- It recruits proteins necessary for splicing and polyadenylation.
- It helps in exporting RNA to the cytoplasm.
- It helps in the translation of mRNAs to proteins.



# Introns vs. exons and RNA splicing



- The protein-coding genes of eukaryotic cells contain specific DNA sequences known as introns, which are transcribed but not translated.
  - The protein-coding regions are known as exons.
- When RNA is synthesized, the RNA molecule contains both introns and exons and is known as primary transcript or pre-mRNA.
- The intron sequences are removed from the newly synthesized RNA through the process of RNA splicing.
- Now the RNA molecule is known as mRNA (mature transcript).

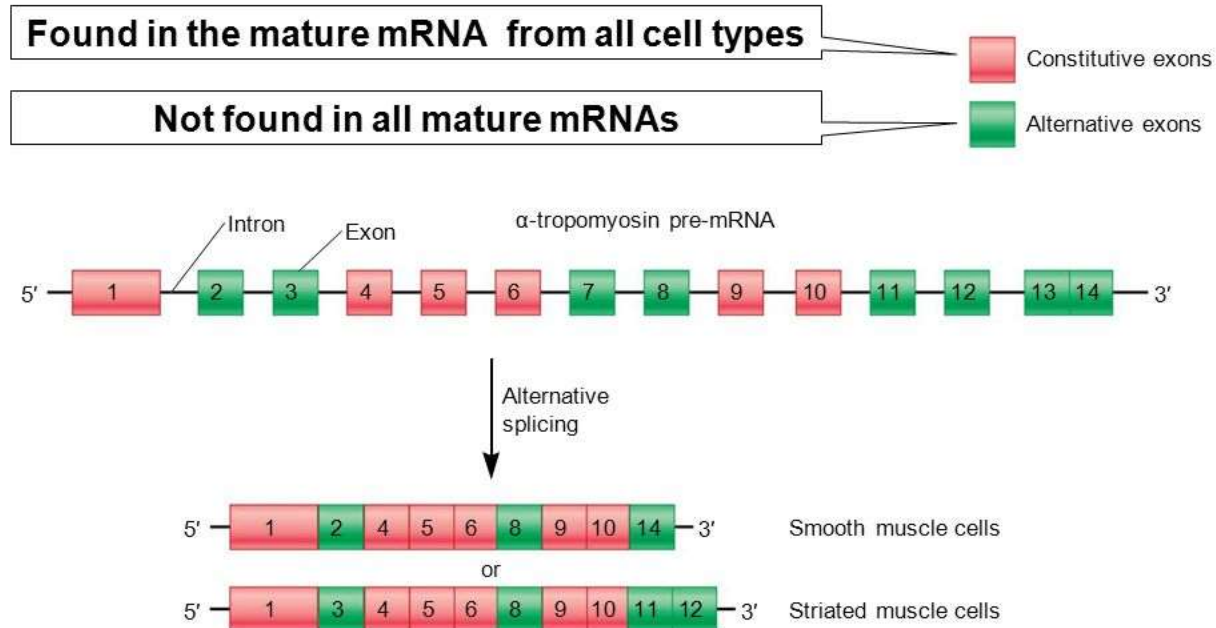




# Alternative splicing



- The transcripts are spliced in different ways to produce different mRNAs and different proteins (known as protein isoforms, which are highly related gene products that perform essentially the same biological function).



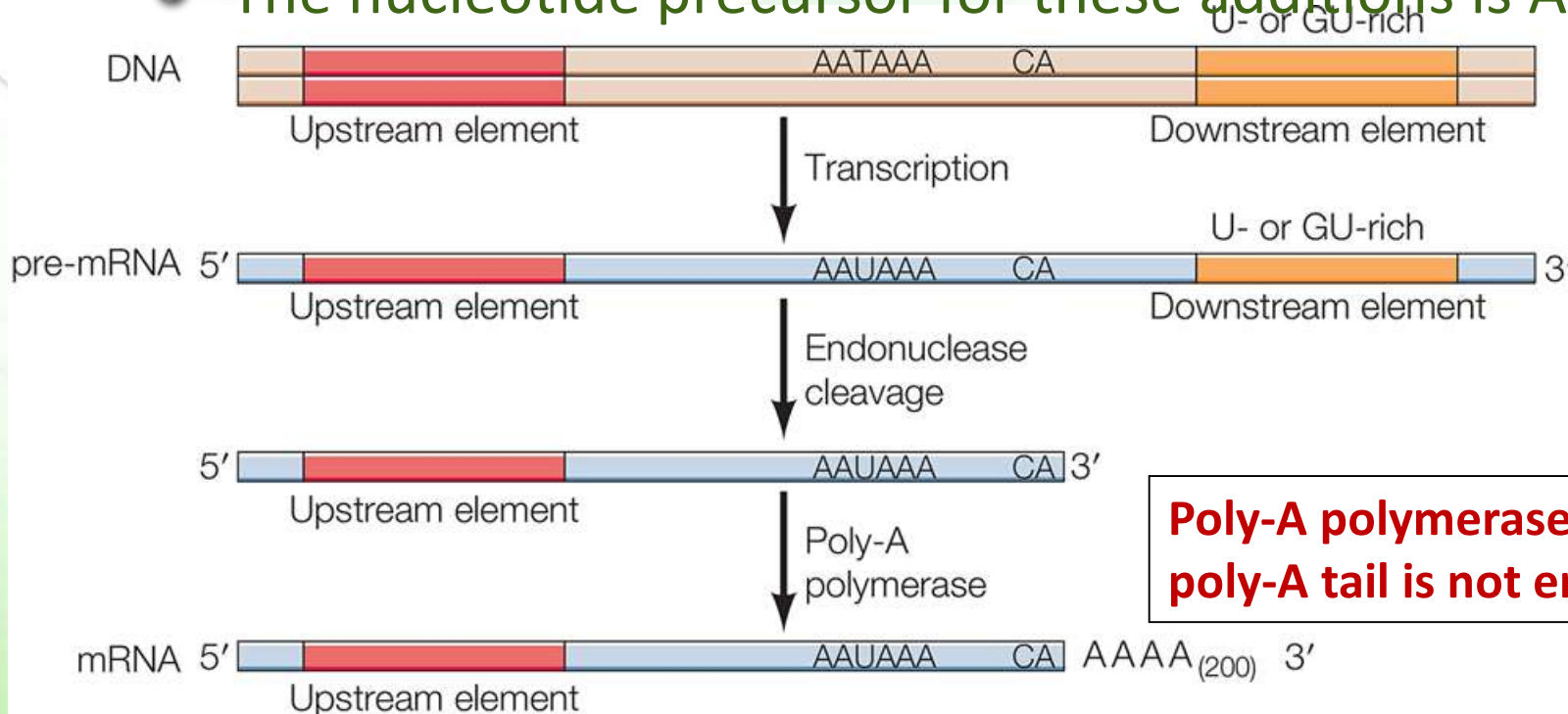
**Note: Exons that are 3' to another exon are never placed 5' to it after splicing.**

**Alternatively spliced versions vary in function to meet the needs of the different cell types**

# Polyadenylation



- A certain sequence in the mRNA (AAUAAA) signals the end of transcription and it is part of the 3' ends of mRNAs.
- The pre-mRNA cleaved after this sequence.
- Poly-A polymerase then adds ~200 A nucleotides to the 3' end.
  - The nucleotide precursor for these additions is ATP.



Why polyA?

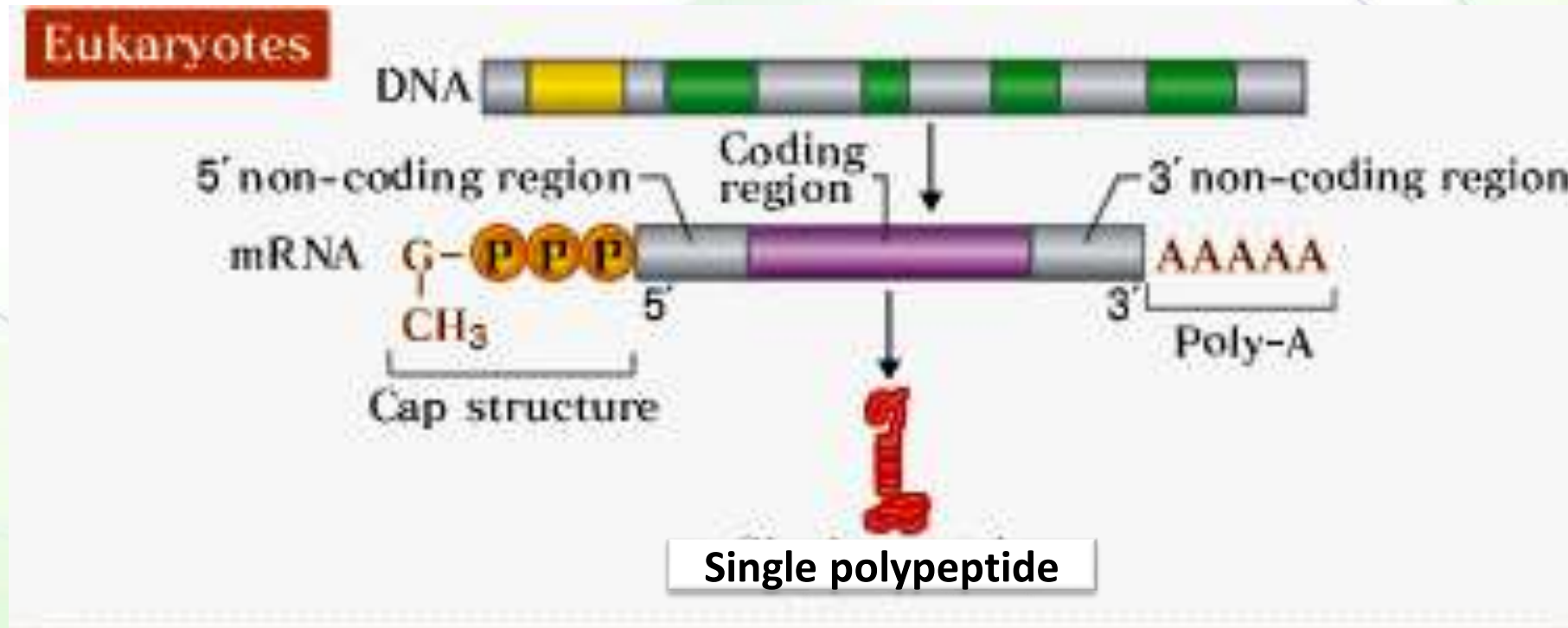
1. mRNA transport from the nucleus to the cytosol.
2. It helps in translation.
3. It stabilizes mRNA.

**Poly-A polymerase does not require a template and the poly-A tail is not encoded in the genome.**

# Eukaryotic genes



- Eukaryotic transcription units produce mRNAs that encode only one protein, thus termed **monocistronic**.





- Transport of mRNA from the nucleus to the cytoplasm, where it is translated into protein, is highly selective- and is associated to correct RNA processing.
- Defective mRNA molecules like interrupted RNA, mRNA with inaccurate splicing, very long mRNA, and so on, are not transported outside the nucleus.

# Degradation of mRNAs



- The half-lives of bacterial mRNA is about 3 minutes.
- The half-lives of eukaryotic mRNAs can be on average 30 minutes but can be longer.
- Degradation of eukaryotic mRNA is initiated by shortening of poly-A tail followed by action of 3'-to-5' exonucleases or decapping (removal of cap) and then 5'-to-3' exonucleases.

