Molecular Biology (7)

Transcription

Mamoun Ahram, PhD
Walhan Al-Shaer, PhD
Second semester, 2019-2020

Resources

- This lecture
- Cooper, Ch. 4, pp. 119-120, Ch. 8, 259-263, 265-269

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 functional peptides.
  - 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 of 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

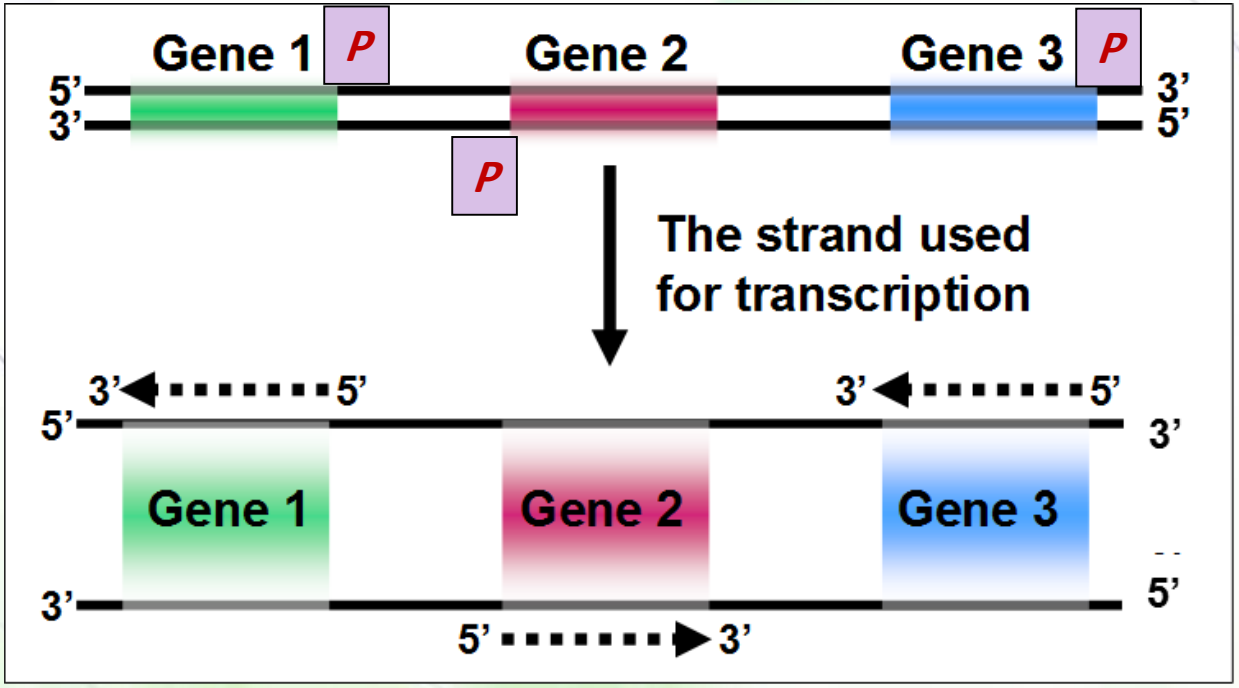
General description

- 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 synthesis of an RNA molecule.
  - *Remember?* In DNA replication, the original strand is the template of the daughter strand.

Using DNA strands

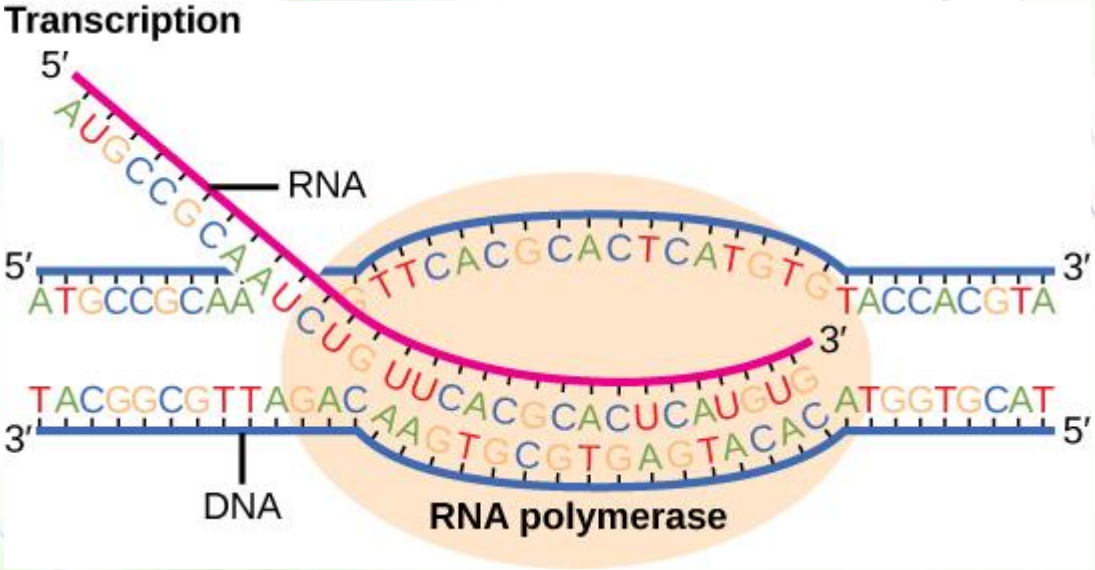
- Although RNA polymerase can read both DNA strands, it uses one strand for any particular gene in order to make RNA.



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.
- 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.

DNA replication vs. transcription

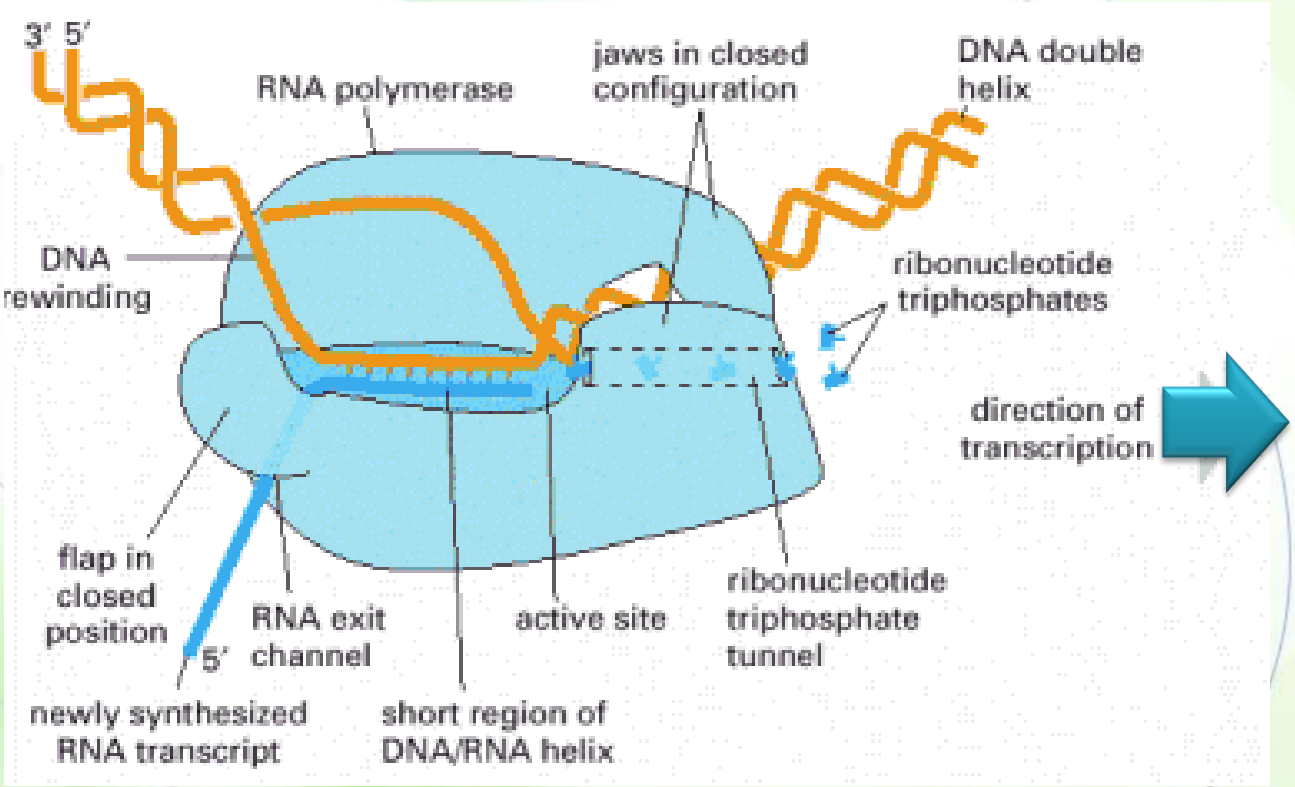
- The RNA strand does not remain hydrogen-bonded to the DNA template strand.
- RNA polymerase read the A in DNA and inserts U in the growing chain of RNA rather than T.
- RNA molecules are much shorter than DNA molecules.
- Unlike DNA, RNA does not store genetic information in cells.

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 the DNA polymerases, they have a modest proofreading mechanism.

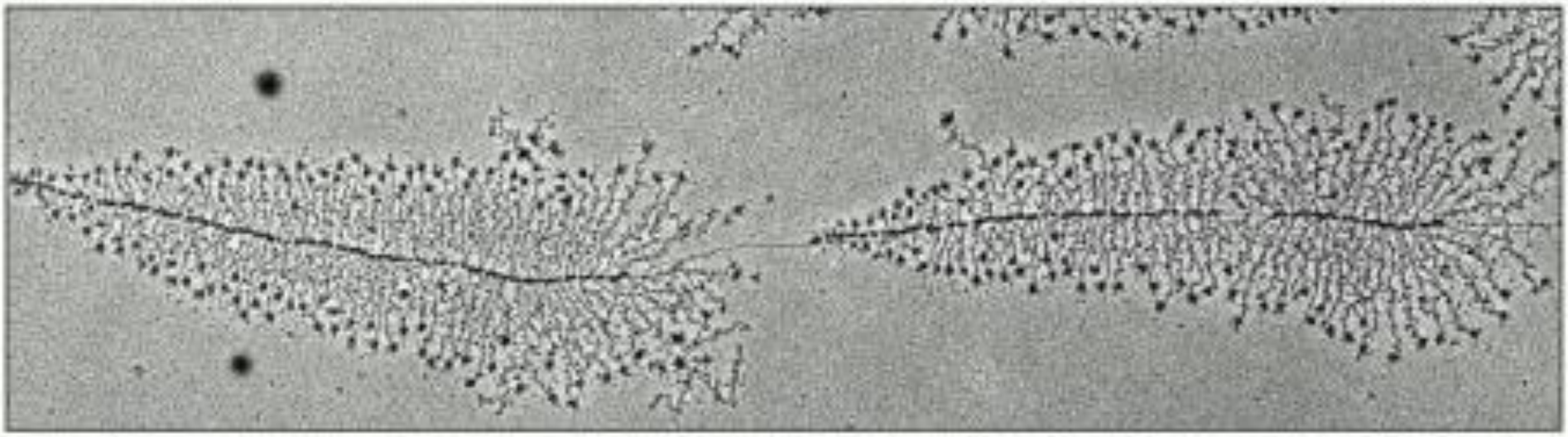
RNA binding to DNA is temporary

- 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.



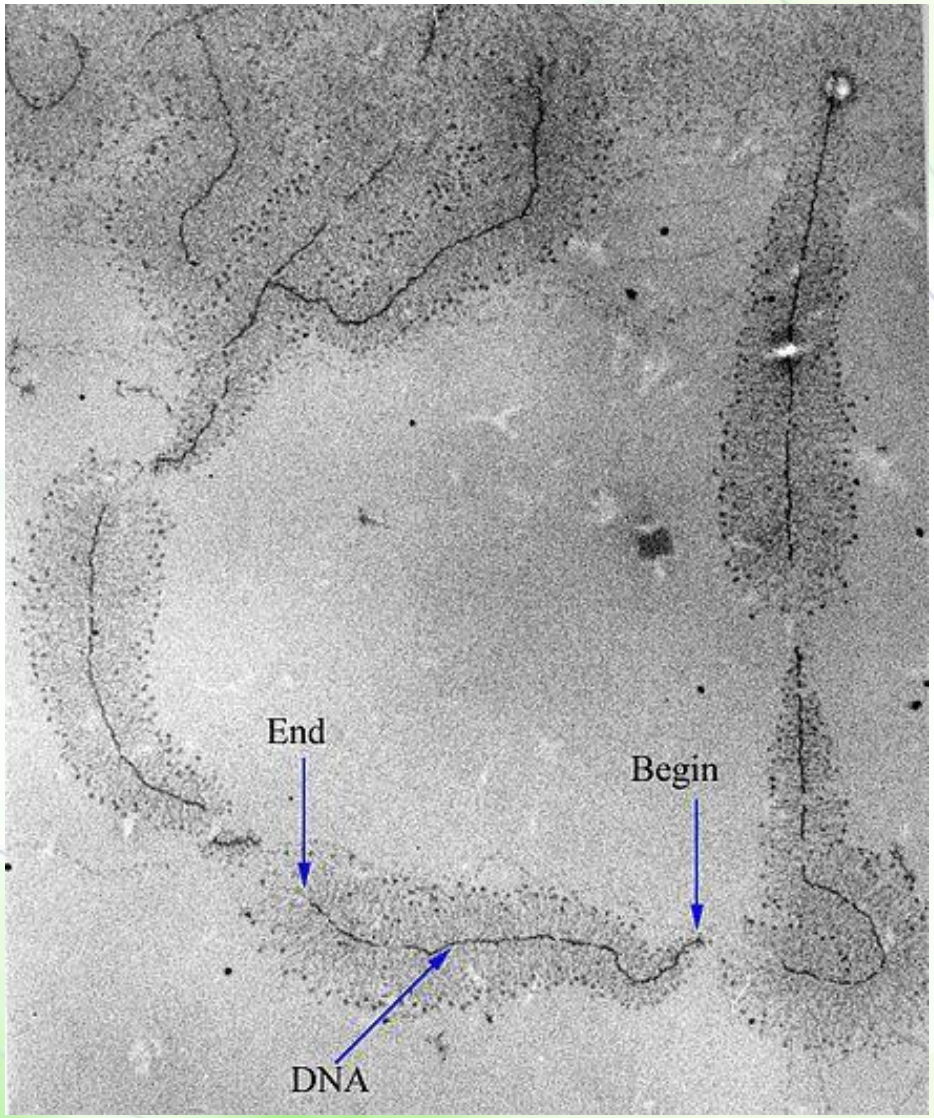
Polyribosomes

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



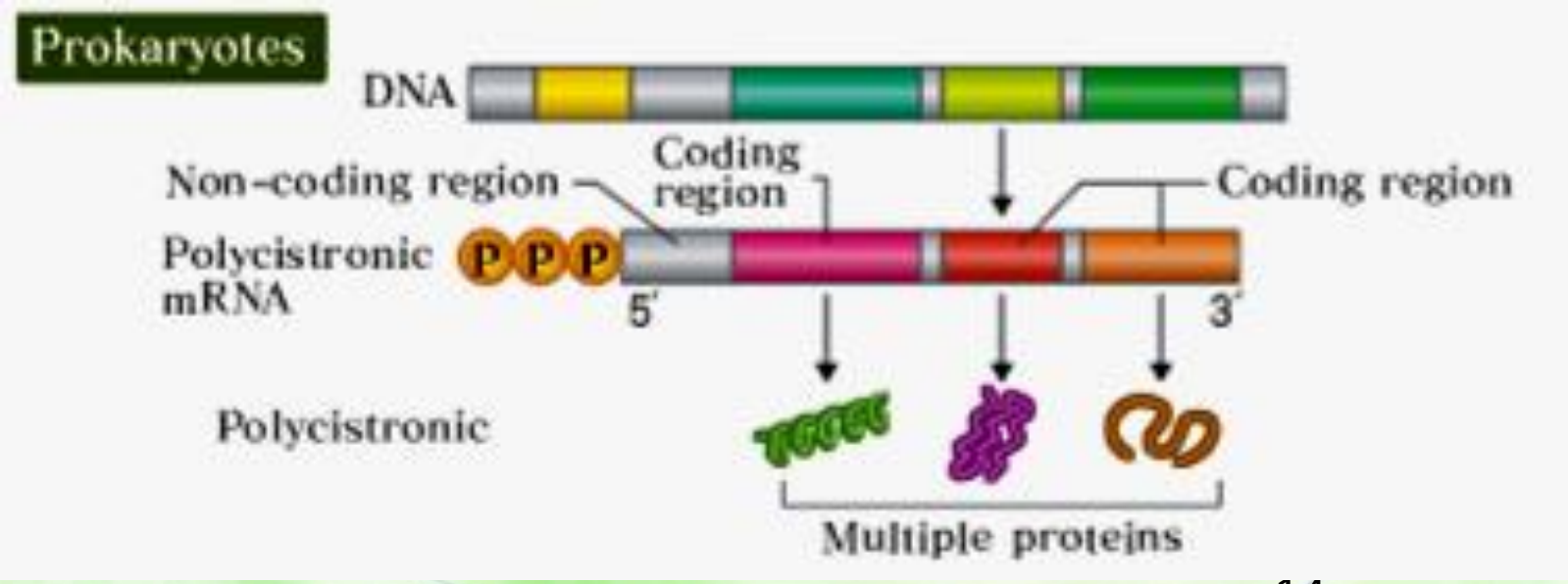
Where is the starting point of transcription?
Where is the beginning of the genes?

How many genes can you see?



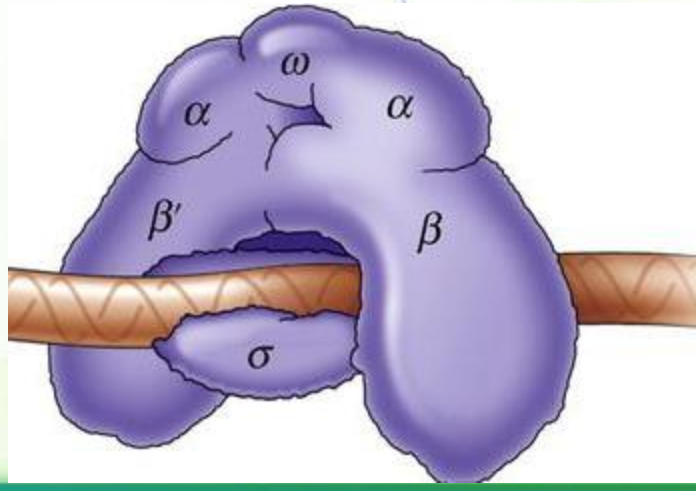
Prokaryotic genes (operons)

- In bacteria, genes can be polycistronic.
- Genes that encode enzymes that are involved in related functions, are often transcribed as one unit from one mRNA.
  - 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.



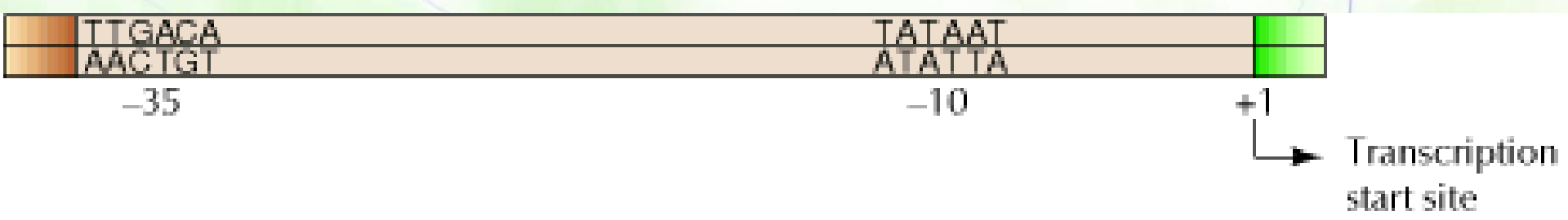
The RNA polymerase

- E. coli RNA polymerase is made up of multiple polypeptide chains or subunits.
- The core polymerase consists of two α, one β, one β', and one ω subunits.
  - The core polymerase is fully capable of catalyzing the polymerization of NTPs into RNA.
- The σ subunit is not required for the basic catalytic activity of the enzyme.



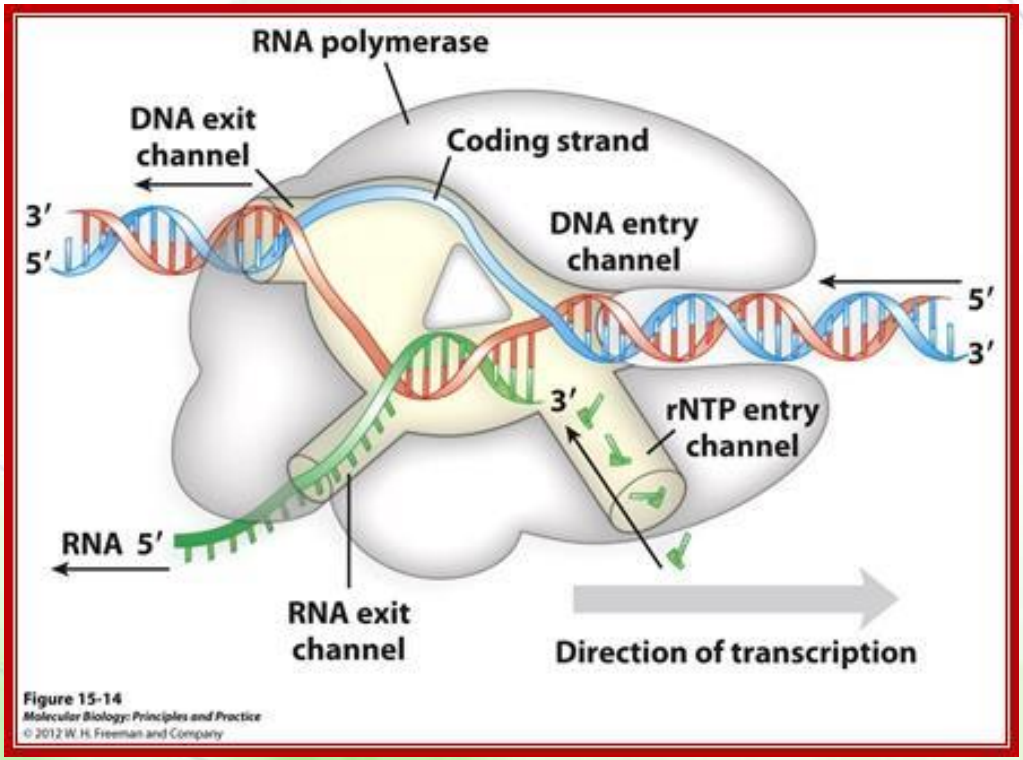
Consensus sequences (the promoter)

- The DNA sequence to which a RNA polymerase binds to initiate transcription of a gene is called the promoter.
  - A promoter is "upstream" of the transcription initiation site.
- The 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 σ subunit

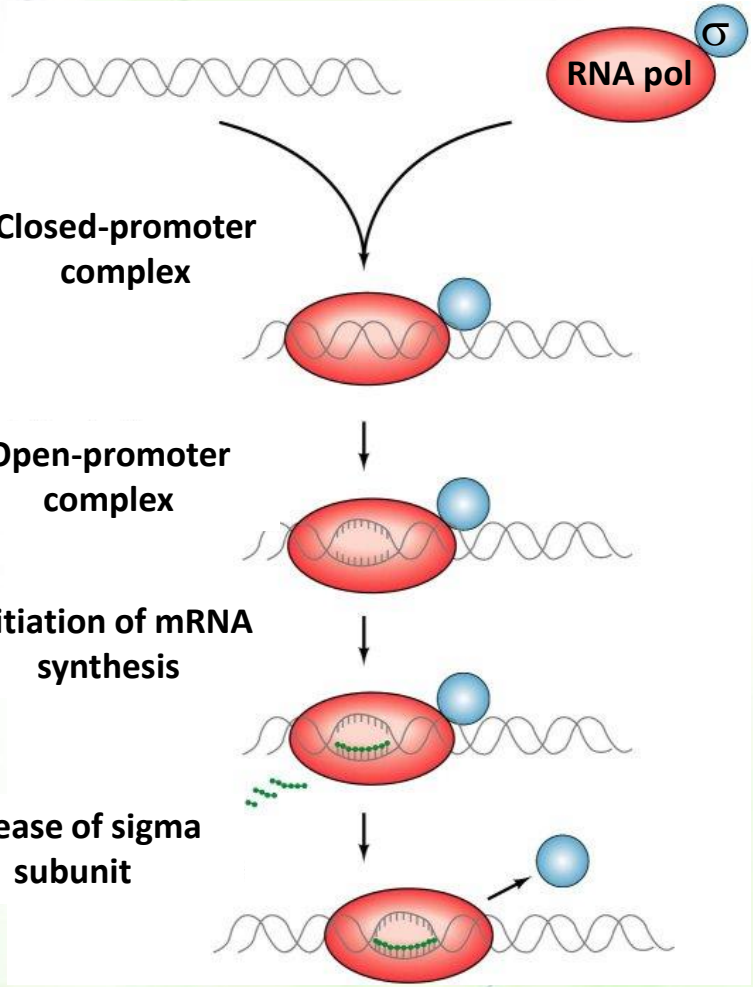
- In the absence of σ , a RNA polymerase binds to DNA with low affinity and nonspecifically.
- The role of σ is to identify and guide the polymerase to the -35 and -10 sequences.



Mechanism of transcription (initiation)

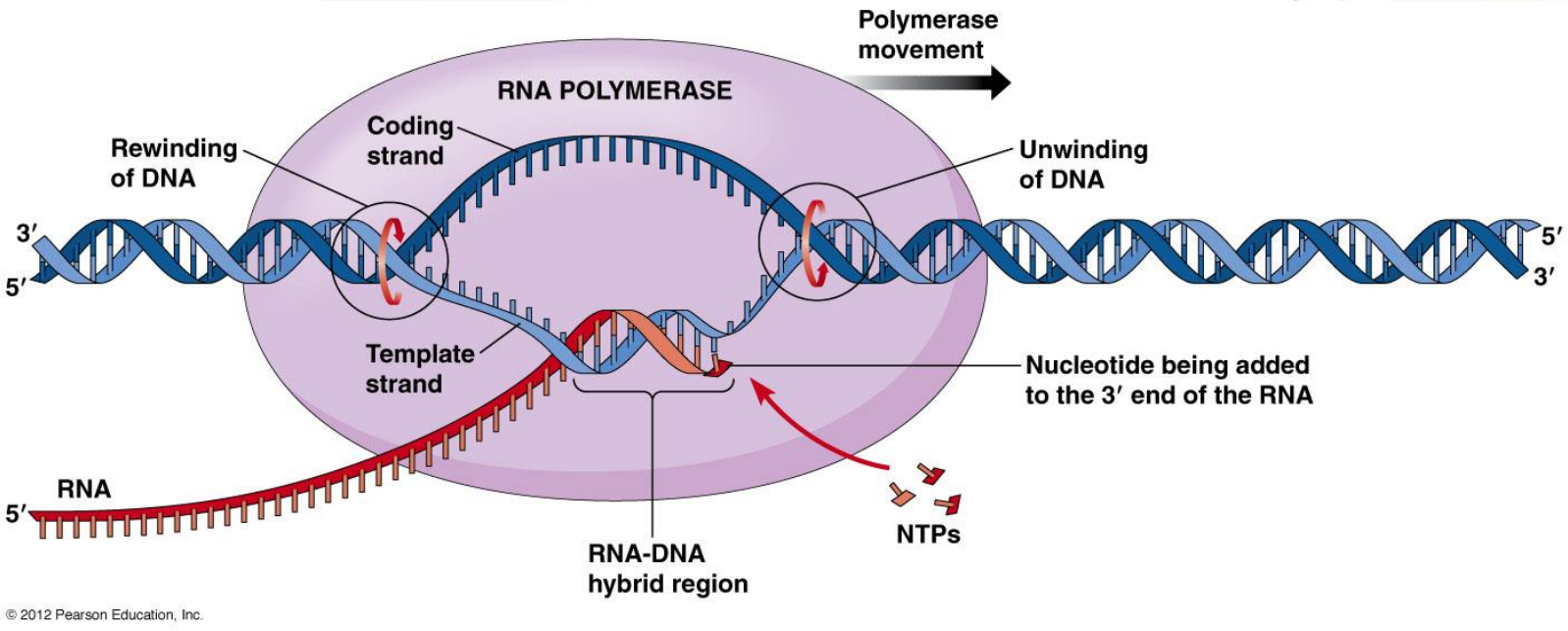
- The RNA polymerase binds to the promoter and opens it *(like what?).*
- The single-stranded DNA is now available as a template.
- Transcription is initiated by the joining of two NTPs.
- After addition of about 10 nucleotides, σ is released from the polymerase.

- *What do you think happens to it?*



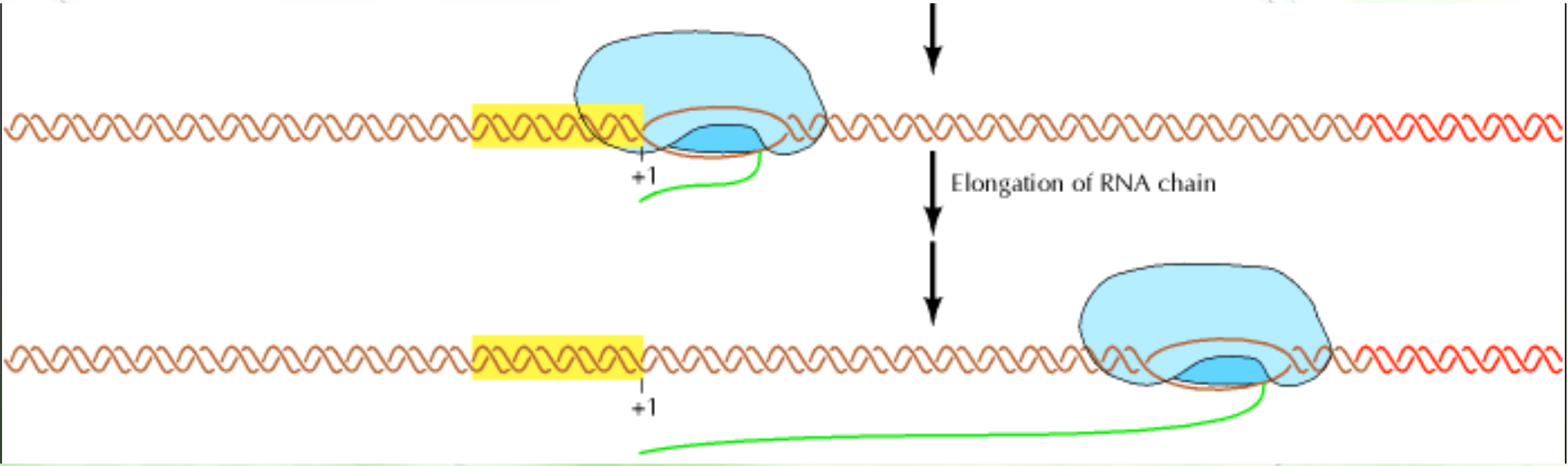
Mechanism of transcription (elongation)

- As the polymerase moves forward, it
  - unwinds the template DNA ahead of it.
  - elongates the RNA
  - rewinds the DNA behind it



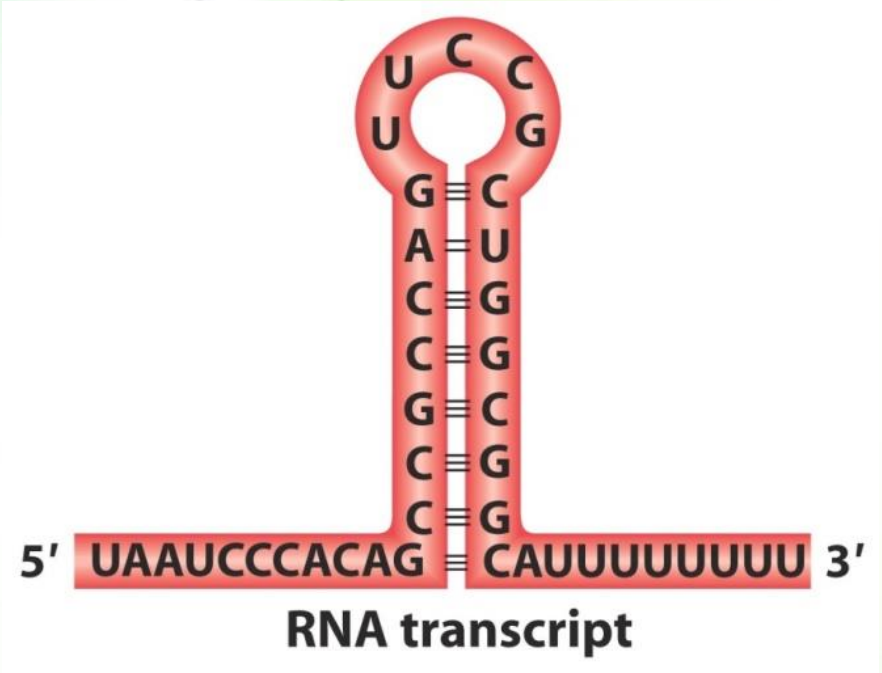
Mechanism of transcription (termination)

- 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.



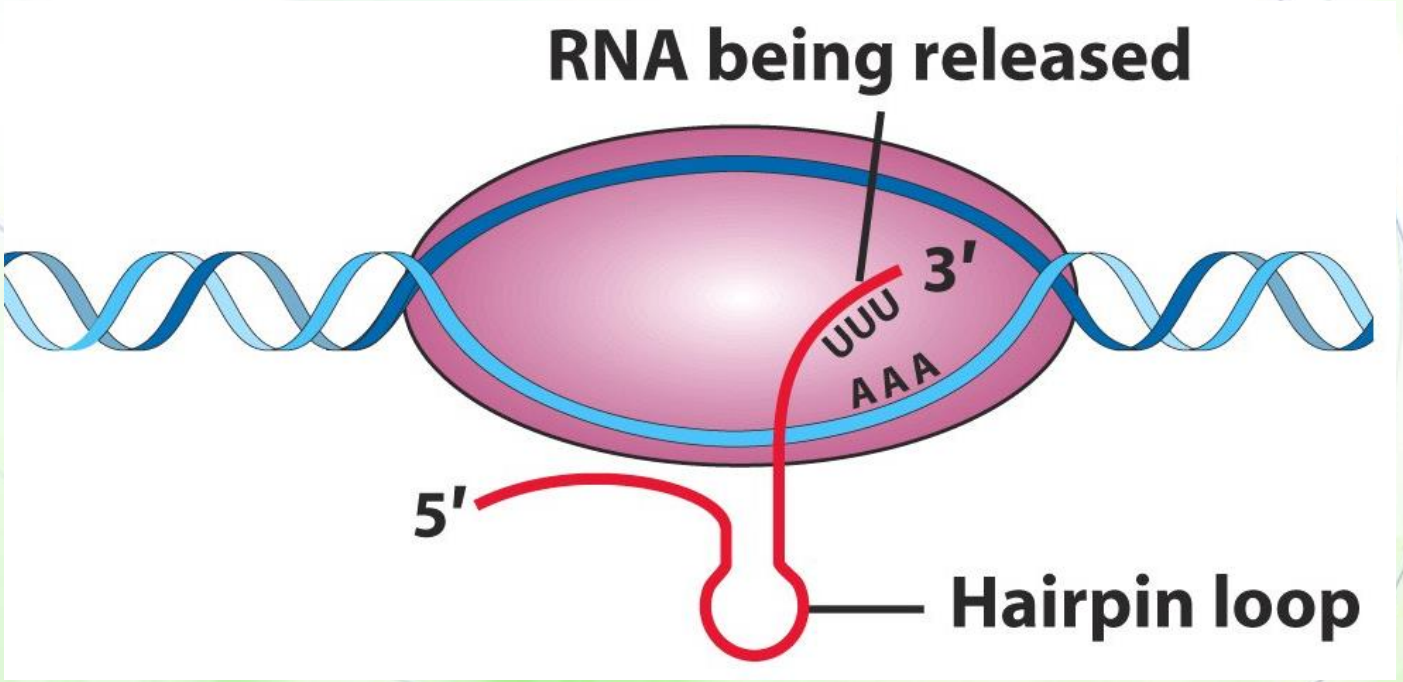
Termination sequences

- The simplest and most common type of termination signal among genes in E. coli consists of a symmetrical inverted repeat of a GC-rich sequence followed by A residues.
- Transcription of the GC-rich inverted repeat results in the formation of a stable stem-loop structure.



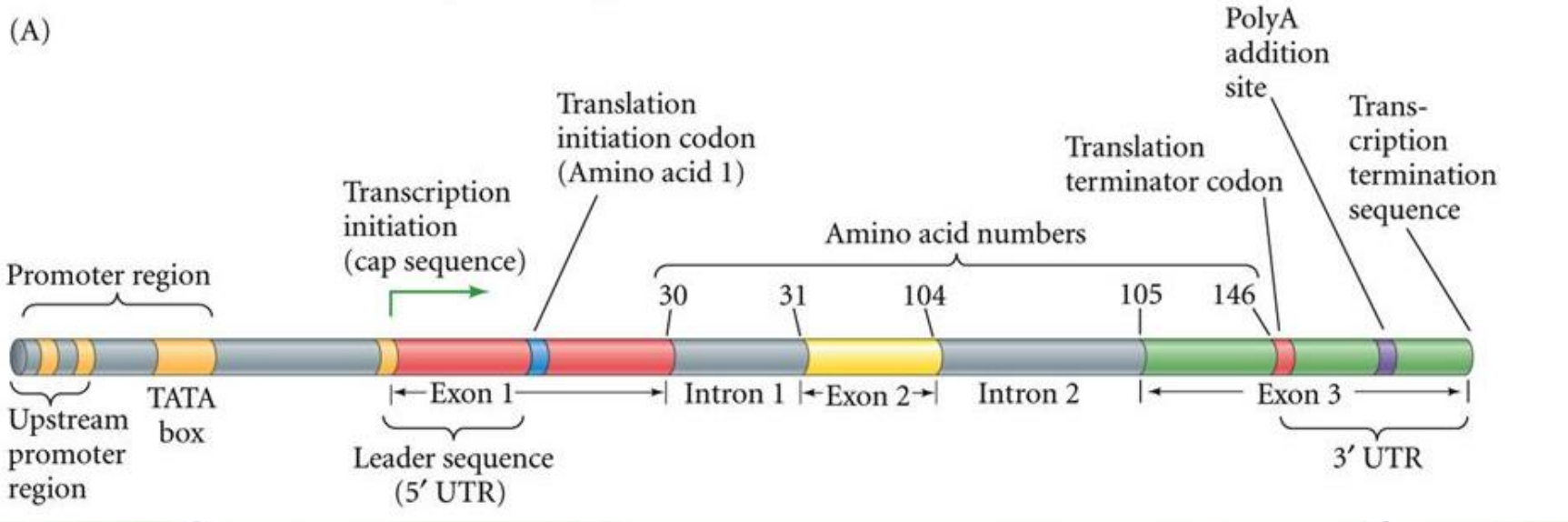
The effect of the stem loop structure

- 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



RNA polymerases

- 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) and microRNA. *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.

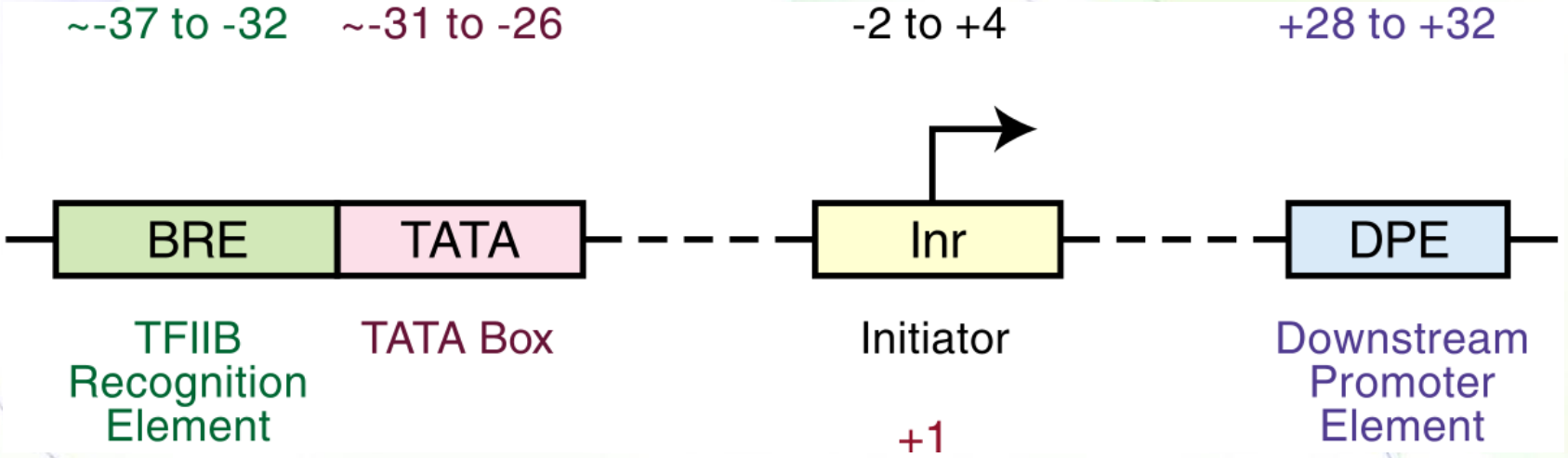
- While bacterial RNA polymerase is able to initiate transcription *without* the help of additional proteins, eukaryotic RNA polymerases cannot.
  - They require help from general transcription factors.
  - They are "general" because they assemble on all promoters used by RNA polymerase II.
  - They 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.
  - aid in pulling apart the two strands of DNA to allow transcription to begin.
  - push the RNA polymerase forward to begin transcription.

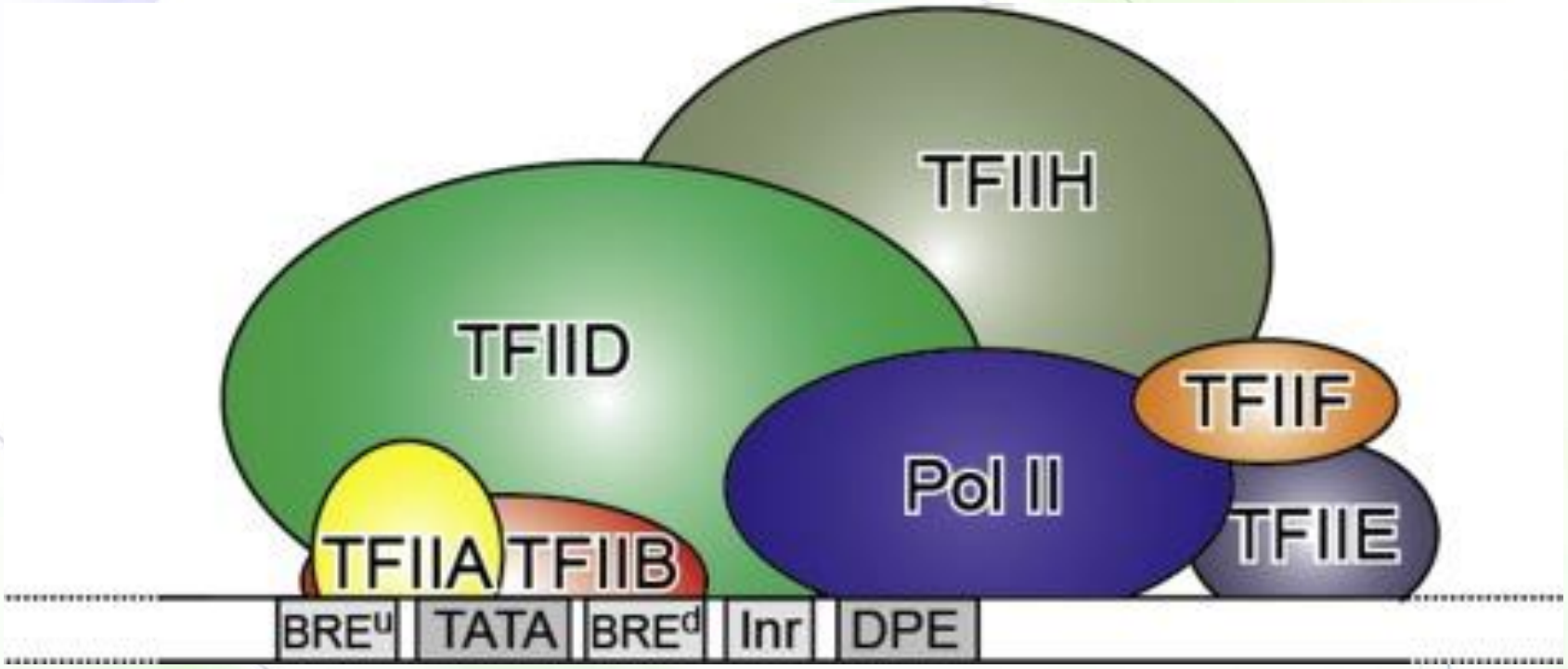
Core components of promoters

- The promoter region in eukaryotic cells is complex.



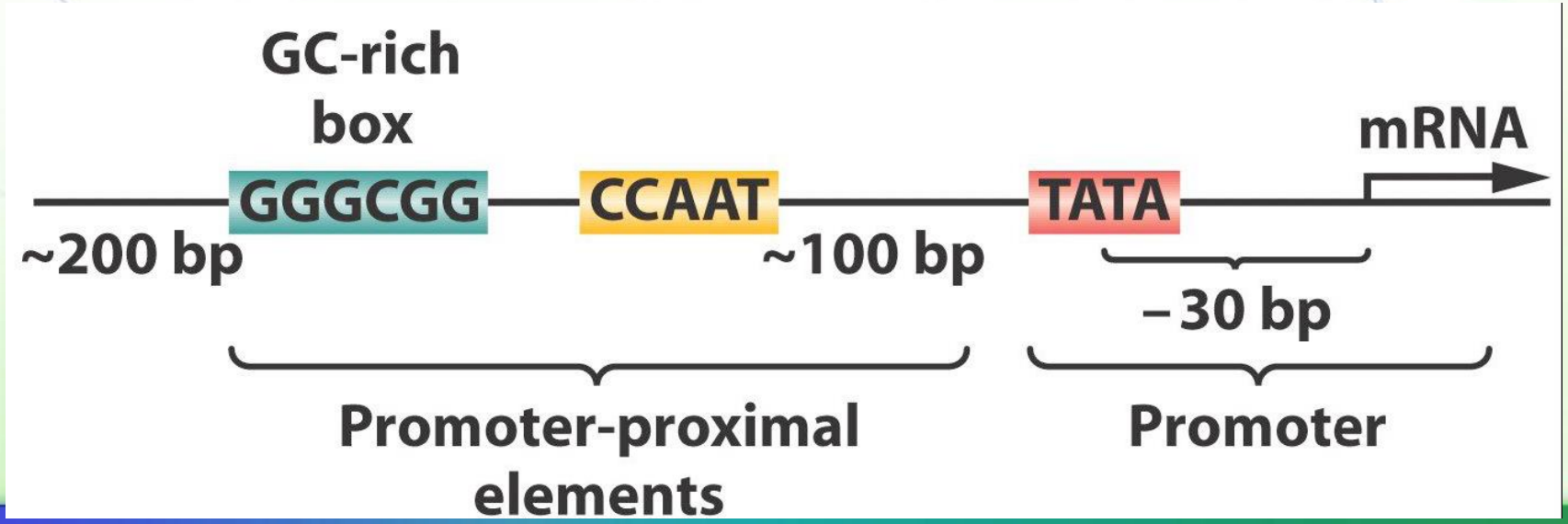
- Not all of these sequences exist at once, but genes can have a combination of these promoter elements.

Formation of preinitiation complex

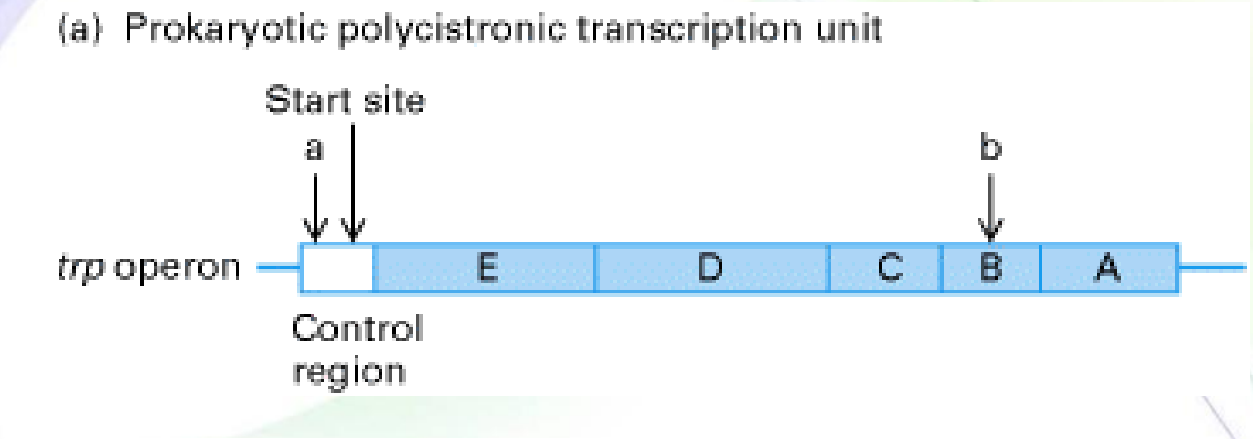


Promoter-proximal elements

- These are upstream of the core promoter region.
- They are important for strong expression (versus basal).
- They are 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!

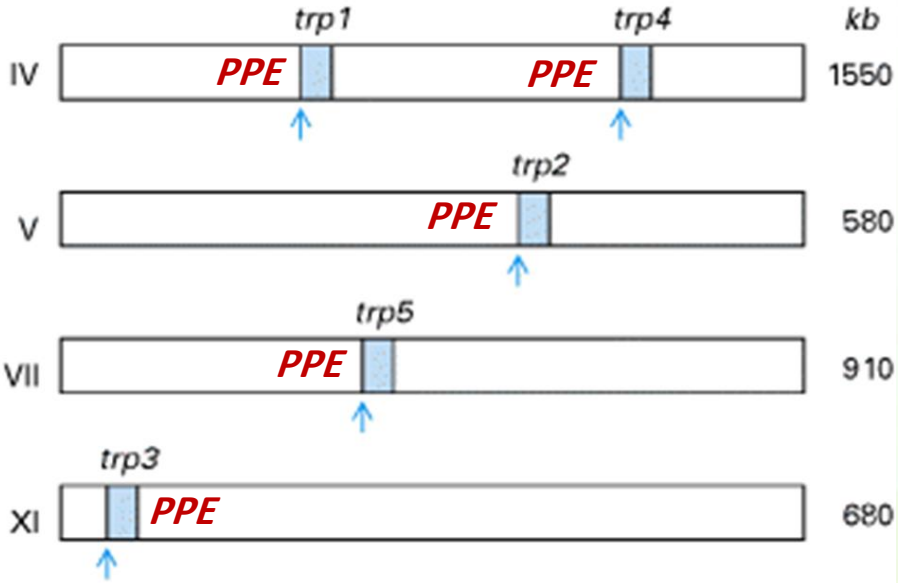


Operon vs. Proximal-promoter elements

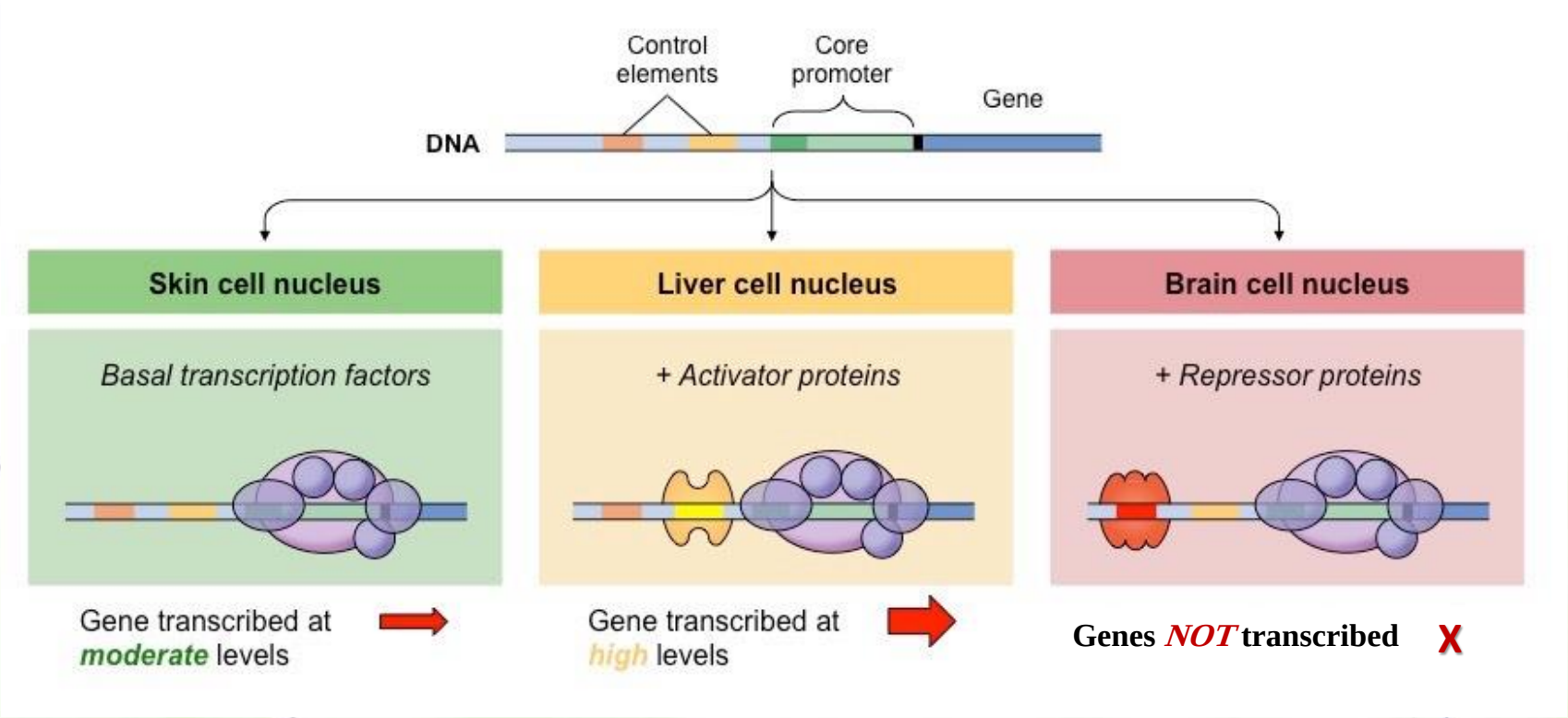


(b) Eukaryotes

Yeast chromosomes

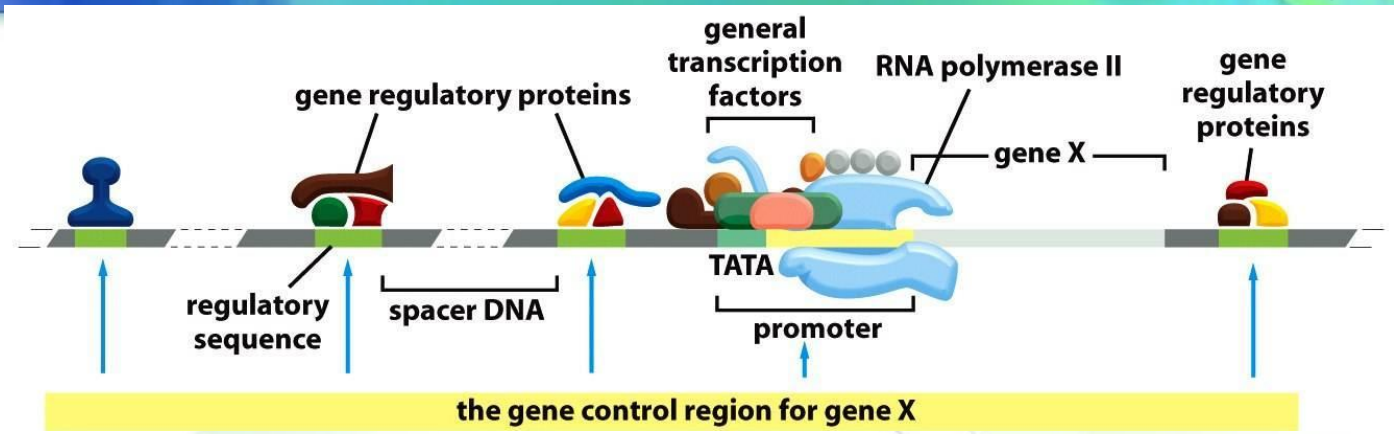


Tissue-specific transcription factors

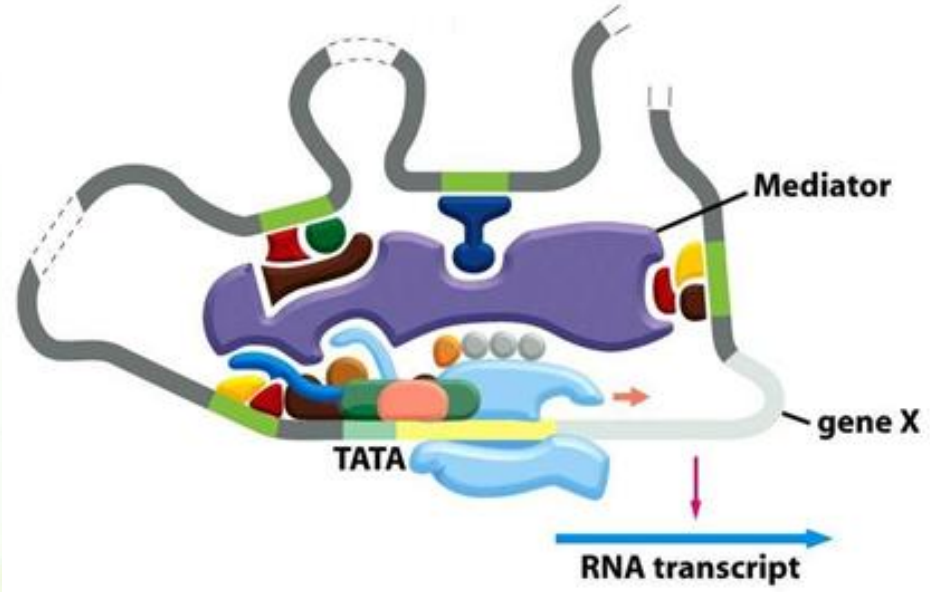


Differential expression of transcription factors (tissue-specific transcription factors) determine gene expression.

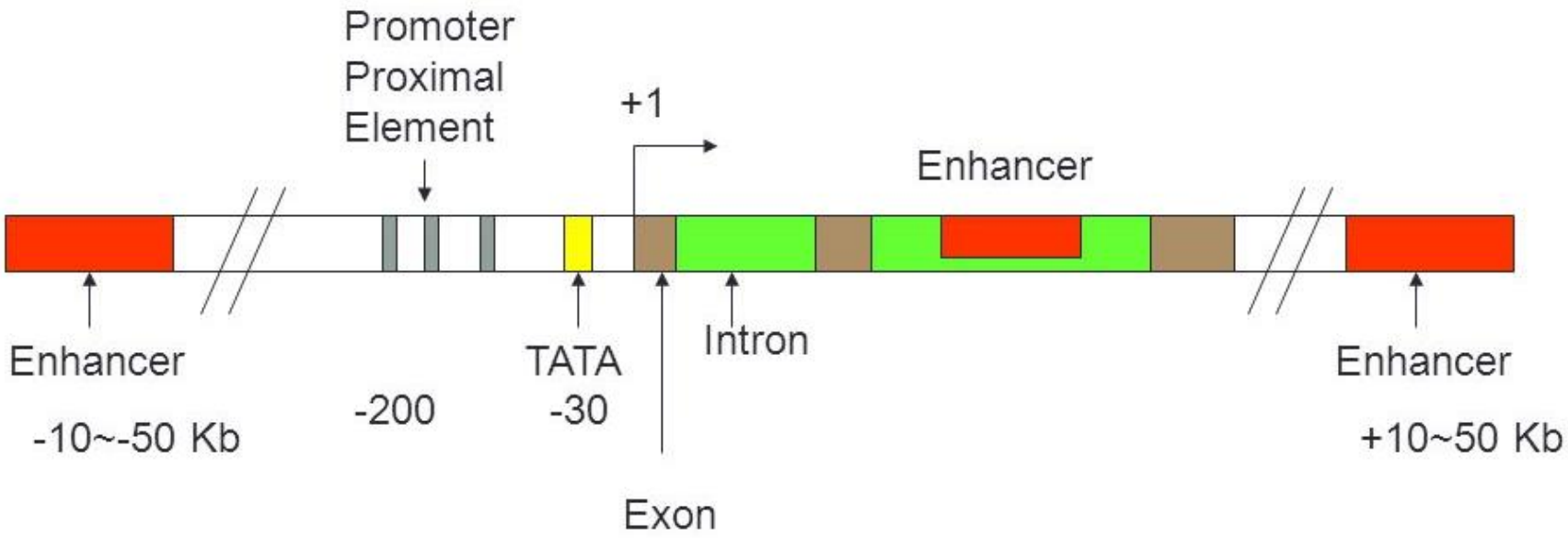
Enhancers



- 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 such as a protein called the *Mediator*.
- They can regulate transcription <u>regardless</u> of orientation or location due to DNA looping.

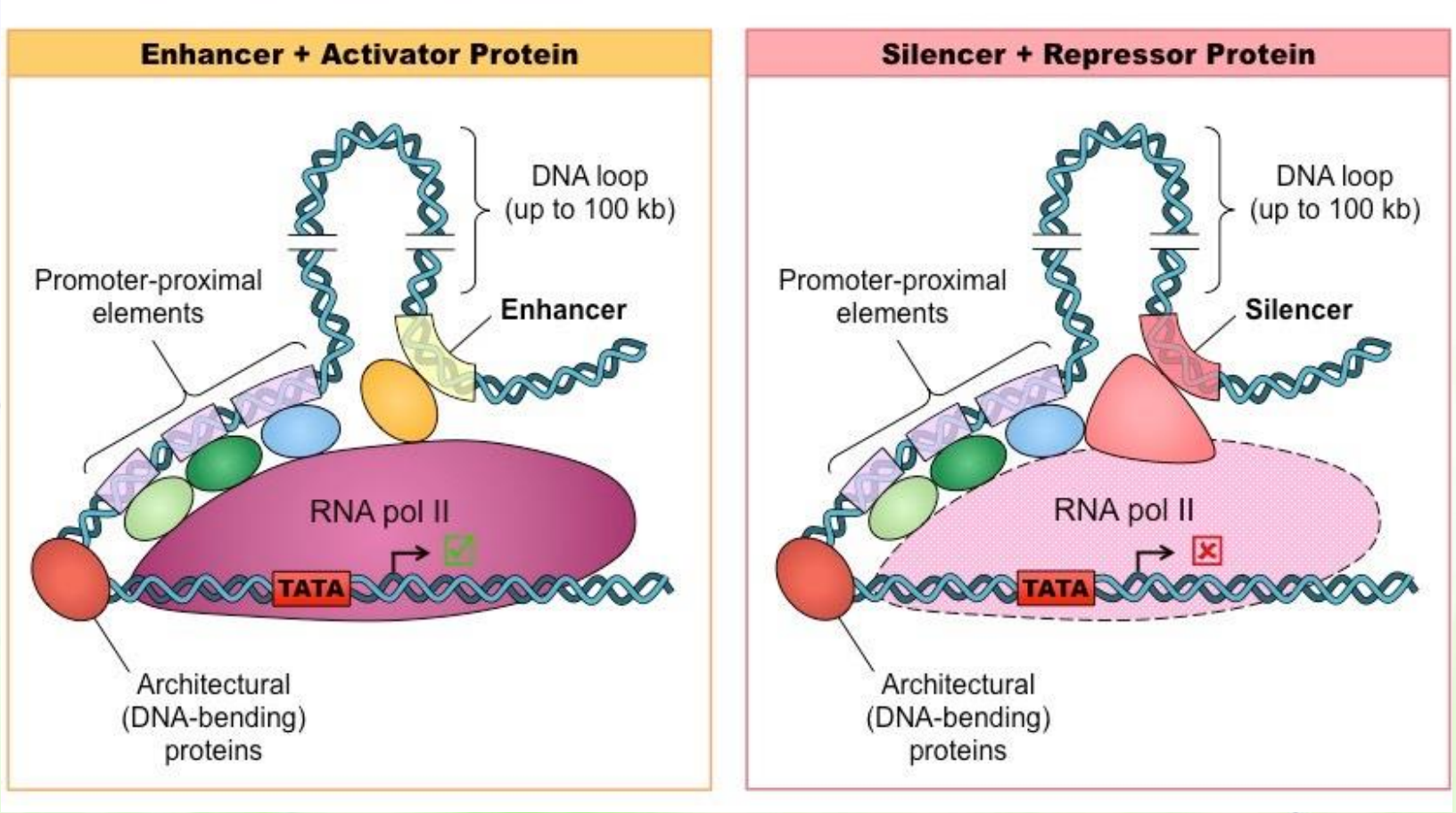


A hypothetical regulatory sequences of a mammalian gene



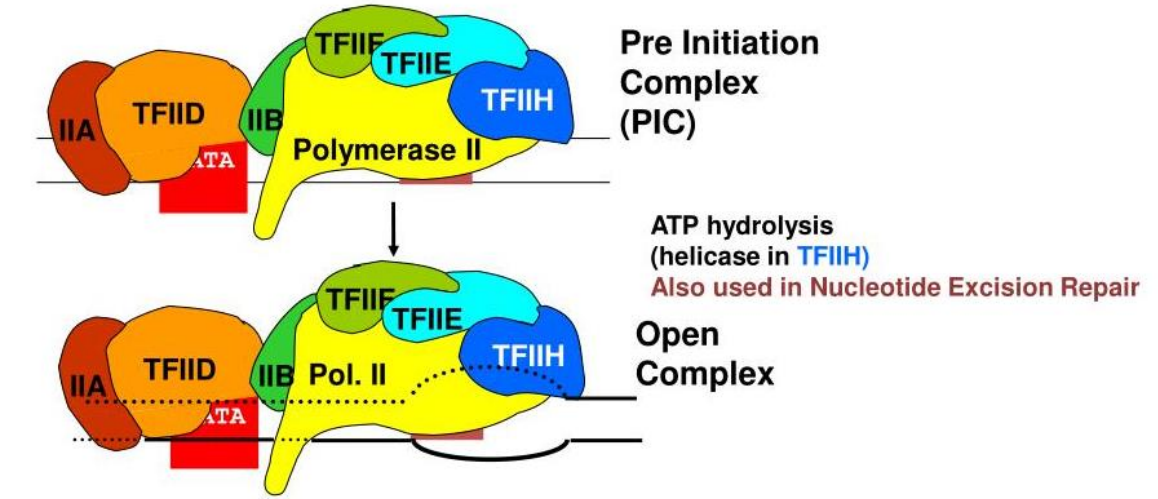
Silencers

- The opposite of enhancers.



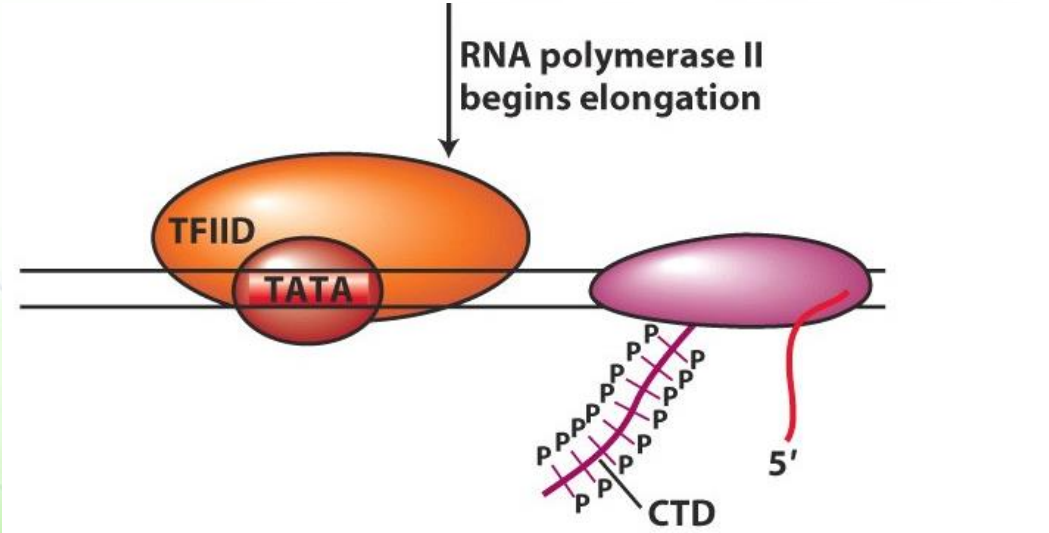
Mechanism of transcription (initiation)

- TFIID binds to the promoter recruiting other proteins and forming the transcription pre-initiation complex.
- A member of this complex is TFIIH, which contains a DNA helicase.
  - TFIIH creates an open promoter exposing the DNA template to the RNA polymerase.



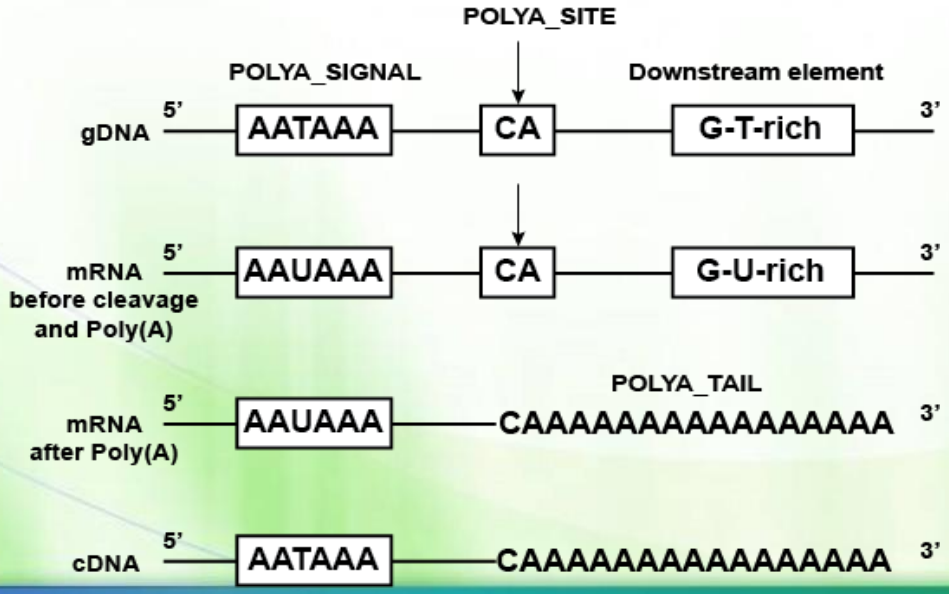
Mechanism of transcription (elongation)

- Movement of the polymerase is activated by the addition of phosphate groups to the "tail" of the RNA polymerase.
- This phosphorylation is also catalyzed by TFIIH, which, also possesses a protein kinase subunits.



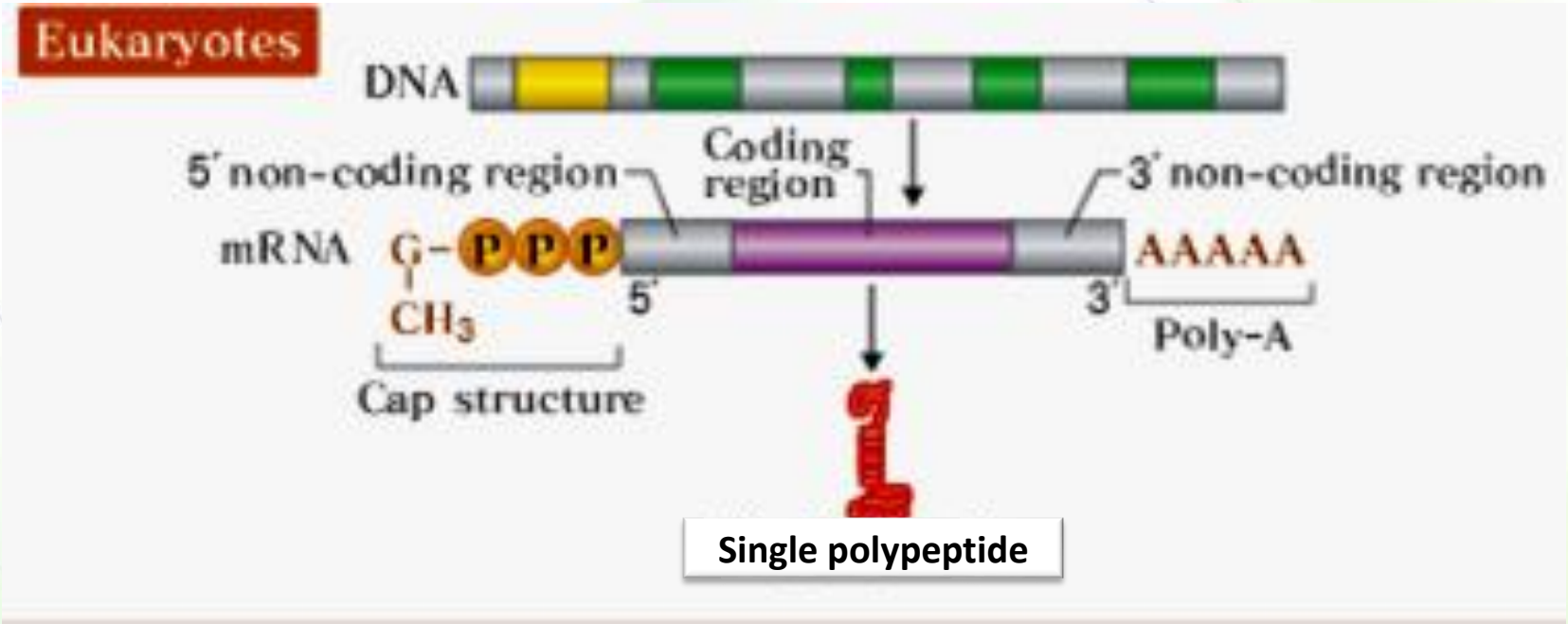
Mechanism of transcription (termination)

- Termination is determined by a consensus sequence for termination in mRNA, which is AAUAAA followed 10-30 nucleotides downstream by a GU-rich sequence.
  - What is the sequence in DNA? Try to write it with the correct directions.
- Termination is coupled to the process that cleaves and polyadenylates the 3'-end of the transcript.



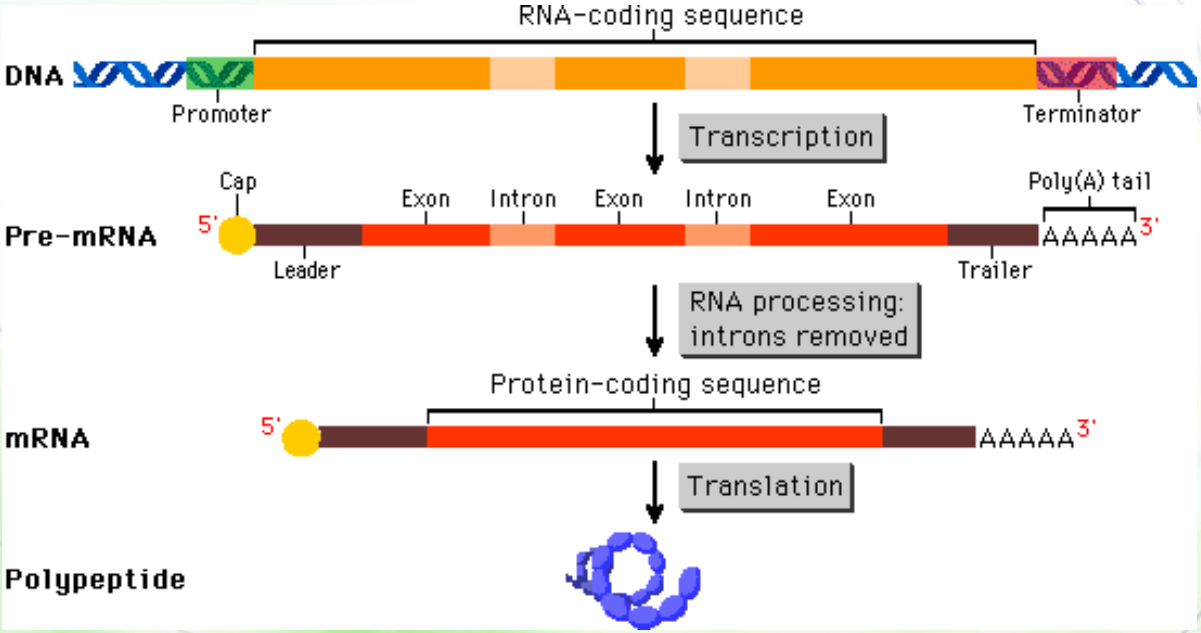
Eukaryotic genes

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



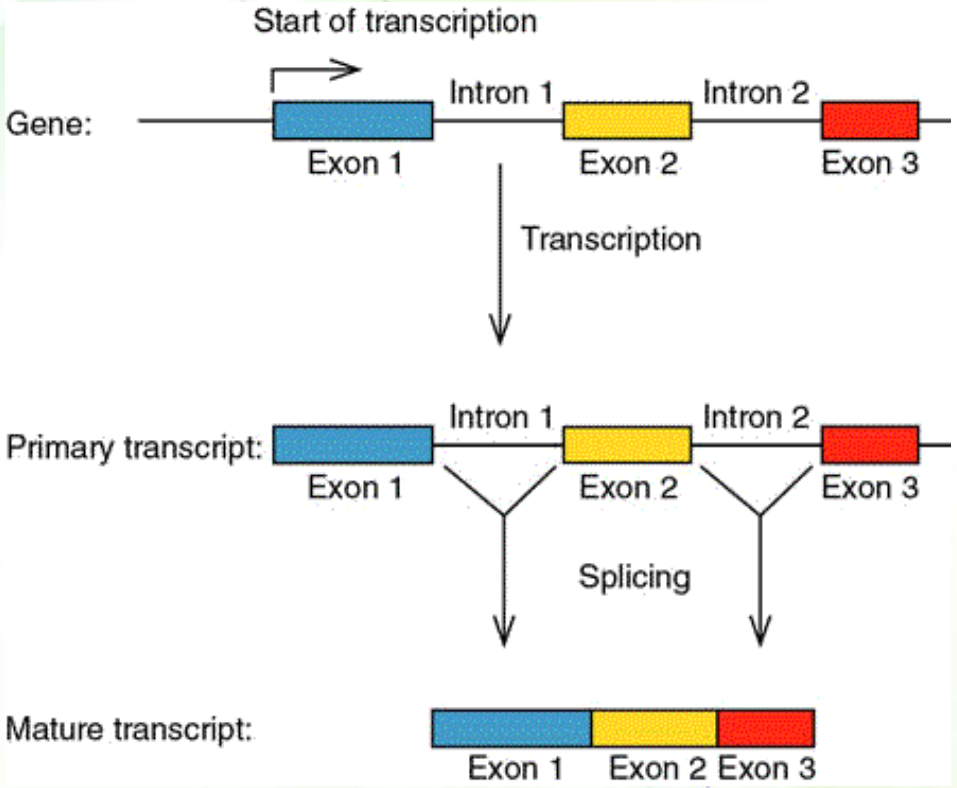
Introns vs. exons

- The genomes of eukaryotic cells contain specific DNA sequences that do not code for proteins known as **introns**.
  - The protein-coding regions are known as exons.
- When RNA is synthesized, the RNA molecule contains both introns and exons and is known as pre-mRNA.



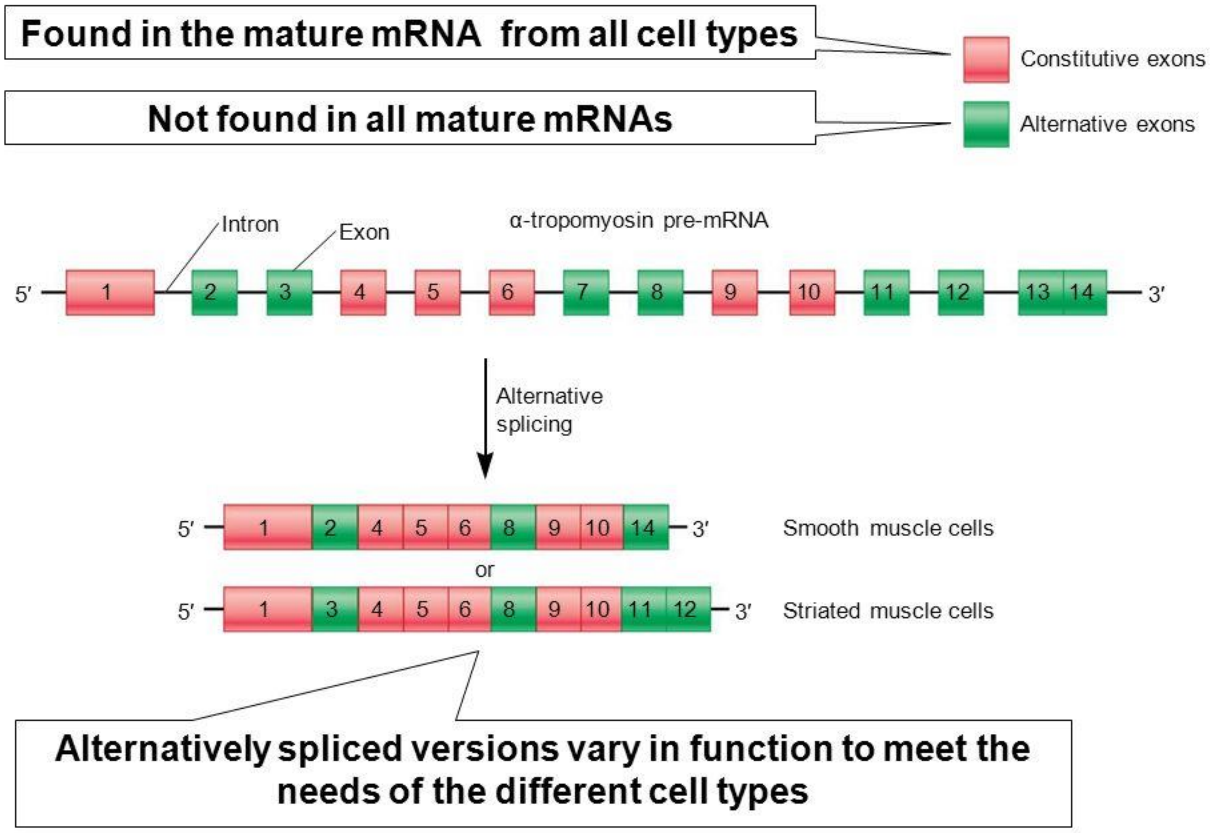
RNA splicing

- 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

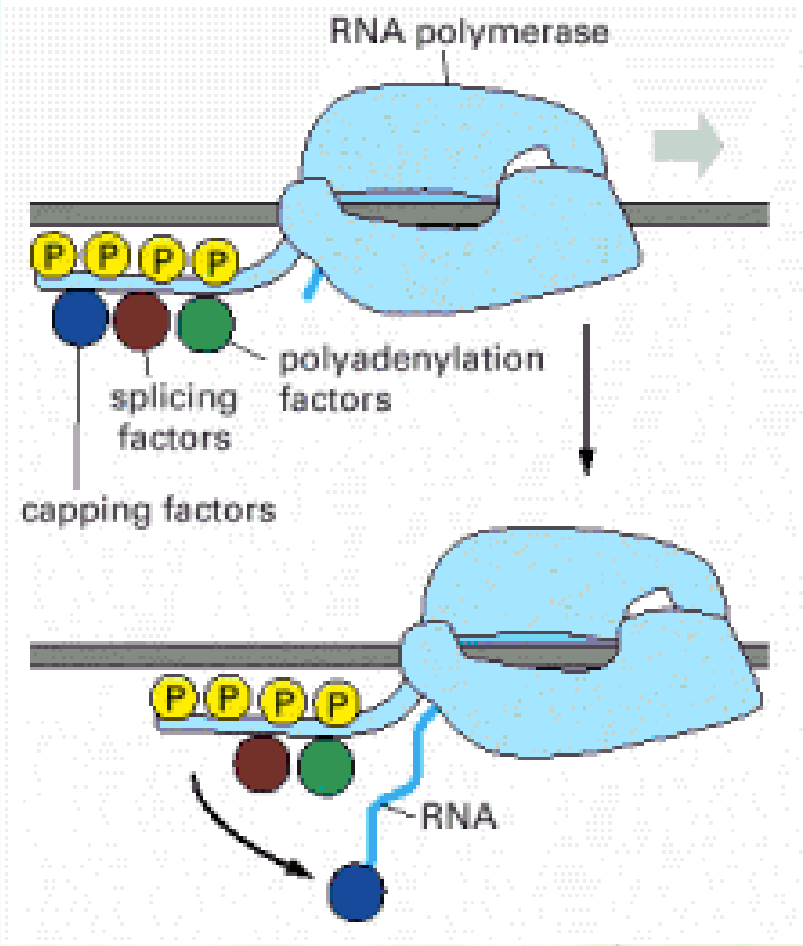
- The transcripts are spliced in different ways to produce different mRNAs and different proteins (known as protein isoforms, which are highly related gene products that perform essentially the same biological function).



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

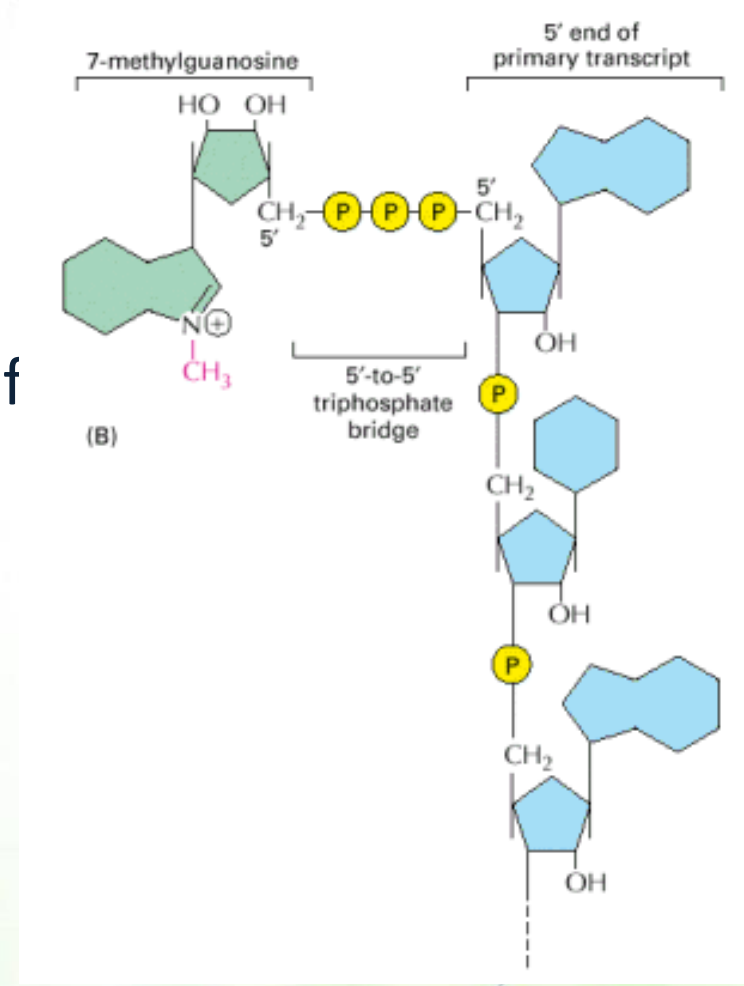
Processing of mRNA in eukaryotes

- mRNA is processed and modified extensively
  - Capping
  - Splicing
  - Polyadenylation
- Some of these processing proteins are associated with the tail of RNA polymerase II.
- These proteins jump from the polymerase tail onto the RNA molecule as it appears.



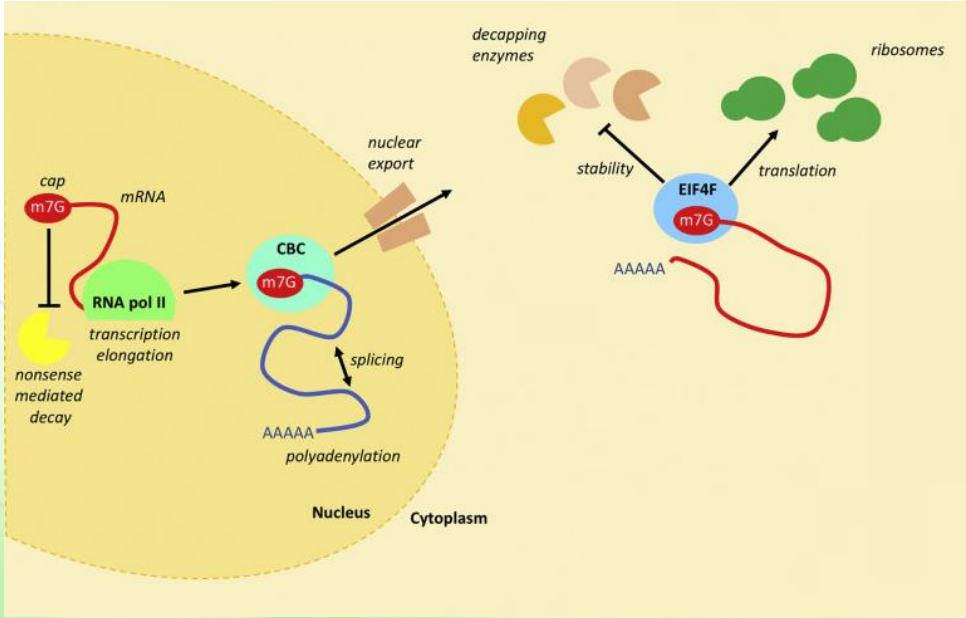
Addition of a cap

- As soon as RNA polymerase II has produced about ~25 nucleotides of pre-mRNA, the 5' end of the new RNA molecule is modified by addition of a "cap" that consists of GTP in reverse orientation.
  - 5’ to 5’ instead of 5’ to 3’.



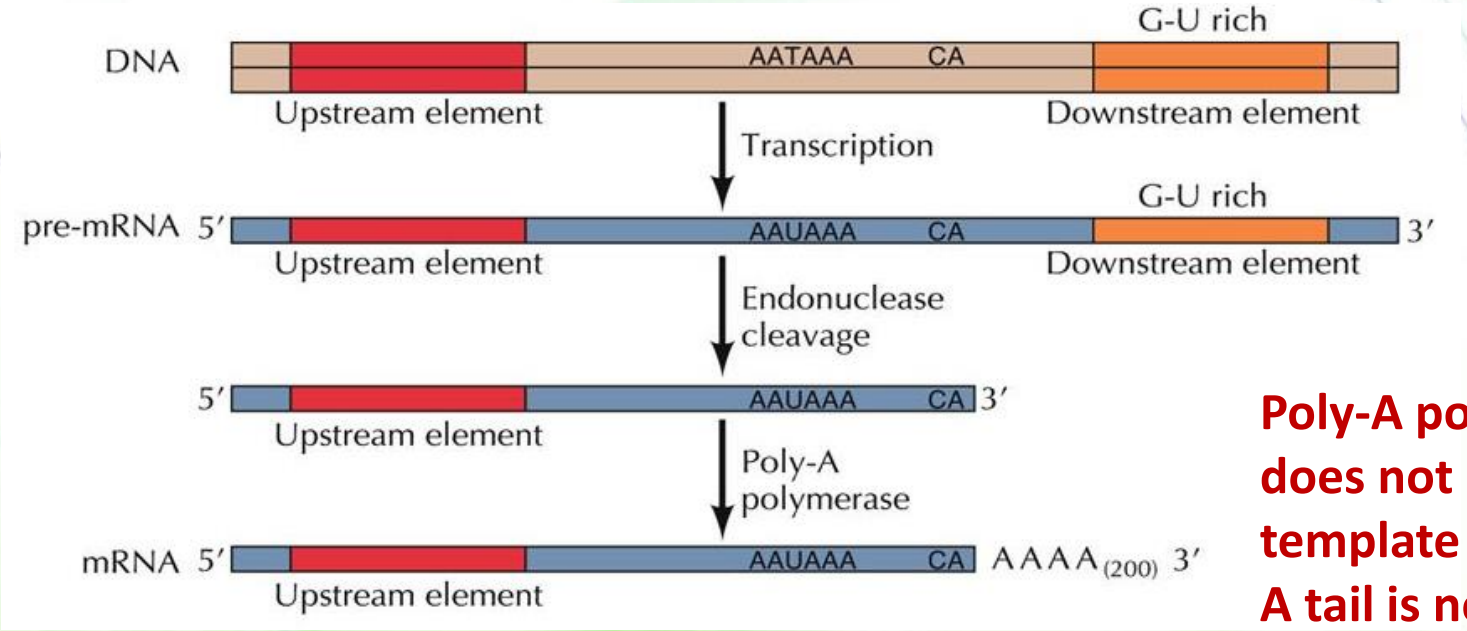
Importance of capping

- It stabilizes the mRNA.
- It signals the 5’ end of eukaryotic mRNAs.
  - This helps the cell to distinguish mRNAs from the other types of RNA molecules, which are uncapped.
- It recruits proteins necessary for splicing and polydenylation.
- It helps in exporting RNA to the cytoplasm.
- It helps in the translation of mRNAs to proteins.



Polyadenylation

- A certain sequence in the mRNA (AAUAAA) in the 3’ ends of mRNAs is recognized by enzymes that cleave it.
- Poly-A polymerase then adds ~200 A nucleotides to the 3’ end.
  - The nucleotide precursor for these additions is ATP.



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

Significance of polyadenylation

- It help in transporting mRNA from the nucleus to the cytosol.
- It helps in translation.
- It stabilize mRNA.

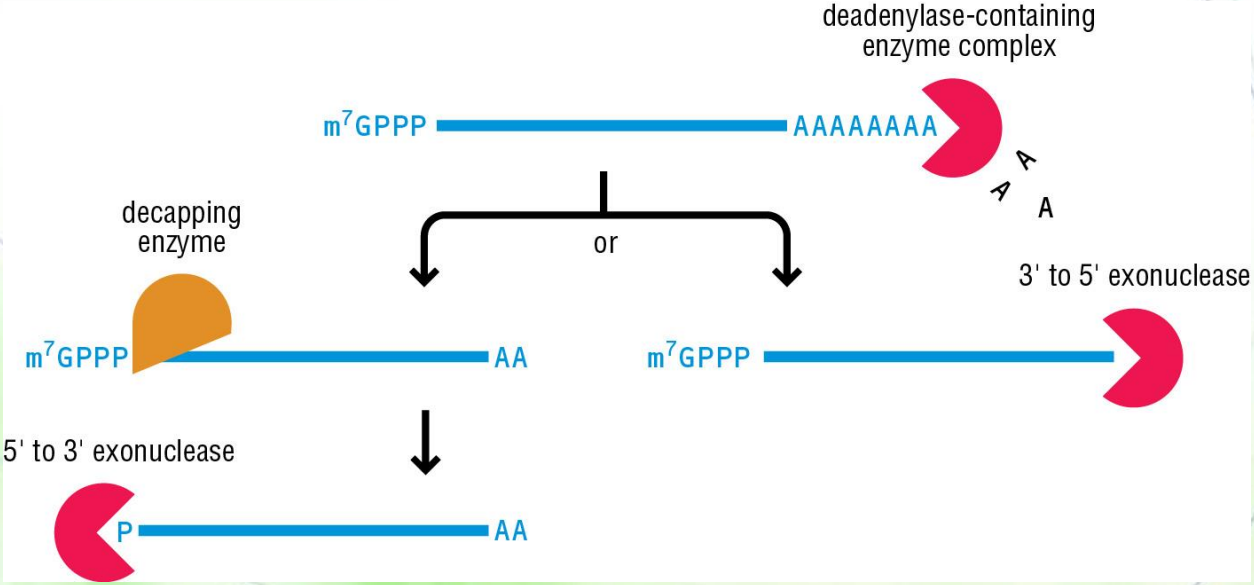
mRNA transport

- 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, and so on, are not transported outside the nucleus.

Degradation of mRNAs

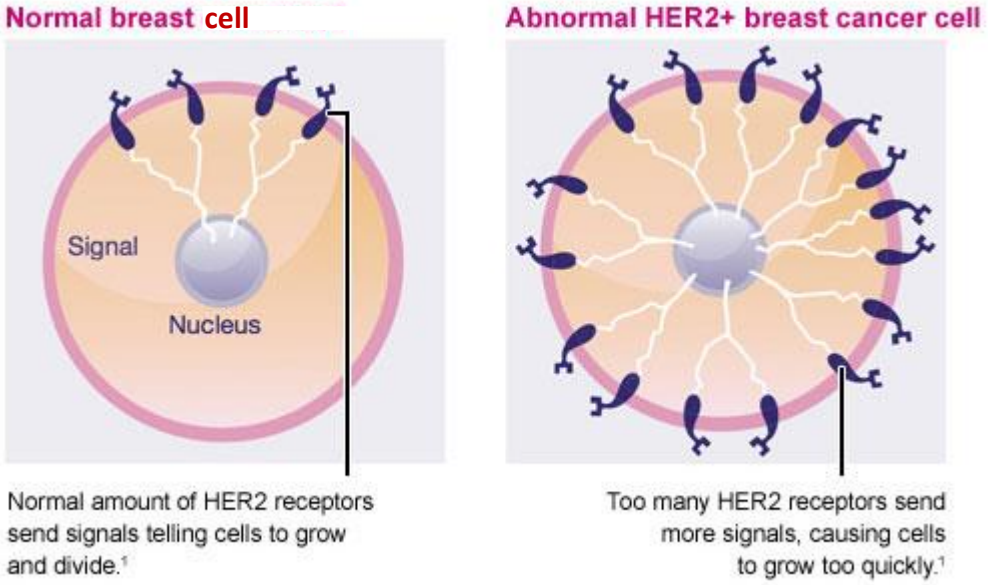
- The vast majority of mRNAs in a bacterial cell are very unstable, having a half-life of about 3 minutes.
- The mRNAs in eukaryotic cells are more stable (up to 10 hours; average of 30 minutes).
- Degradation of eukaryotic mRNA is ainitiated by shortening of poly-A tail followed by action of 3’-to-5’ exonucleases or decapping (removal of cap) and then 5’-to-3’ exonucleases .

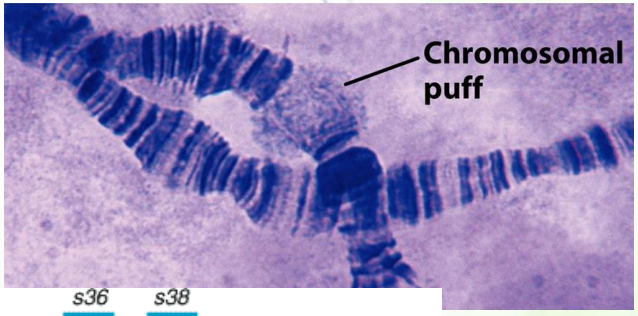


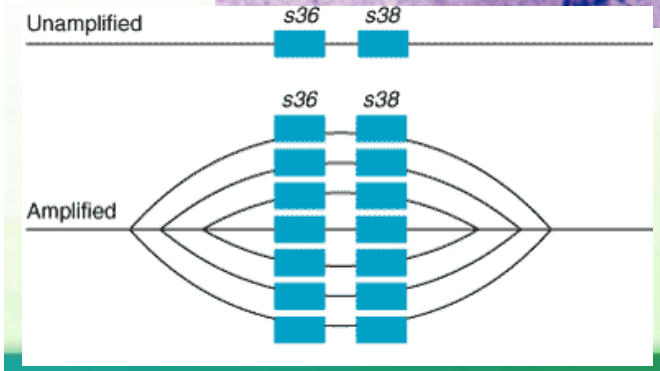
Some phenomena in eukaryotes

Gene amplification

- It is an increase in copy number of a restricted region of a chromosome increasing the quantity of DNA in these regions.
- It is a mechanism that cancer cells use to induce resistance against methotrexate whereby the target gene, dihydrofolate reductase, is amplified.
- It is also a mechanism by which breast tumor cells progress and become more aggressive whereby they amplify the human epidermal growth factor receptor 2 (HER2), which stimulates cell growth.



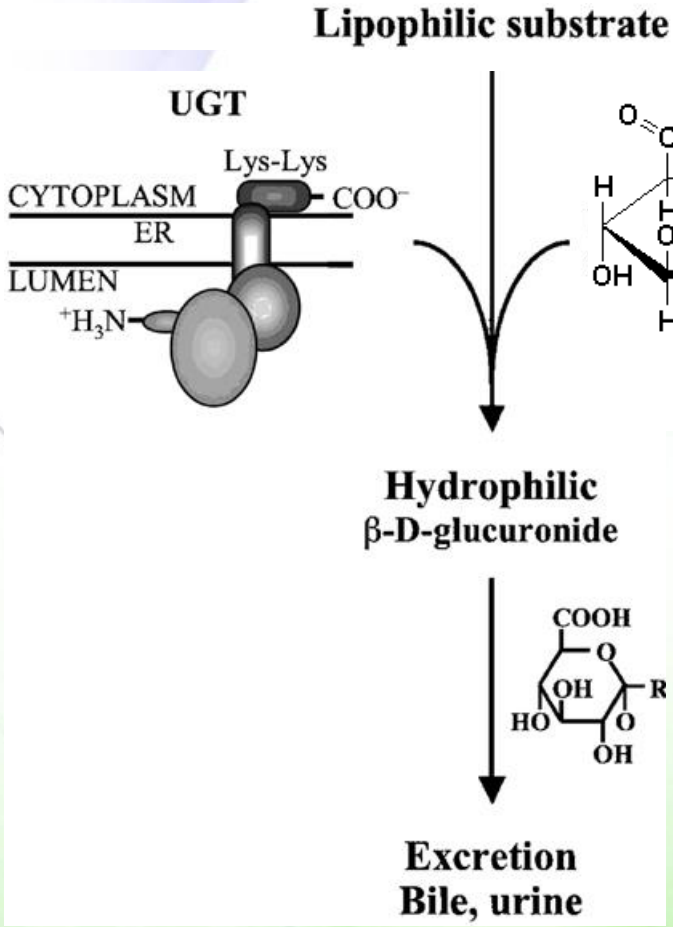


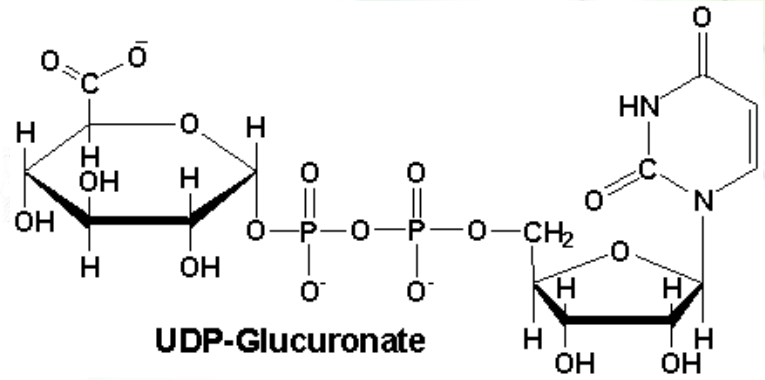


Another phenomenon:
Multiple promoters

An example of alternative splicing:

UDP-glucuronosyltransferase (UGT)





The uridine diphosphate glucuronosyltransferase (UGT) enzymes transfer of glucuronic acid onto xenobiotics and other endogenous compounds making them water soluble and allowing for their biliary or renal elimination.

It has many substrates with different structures

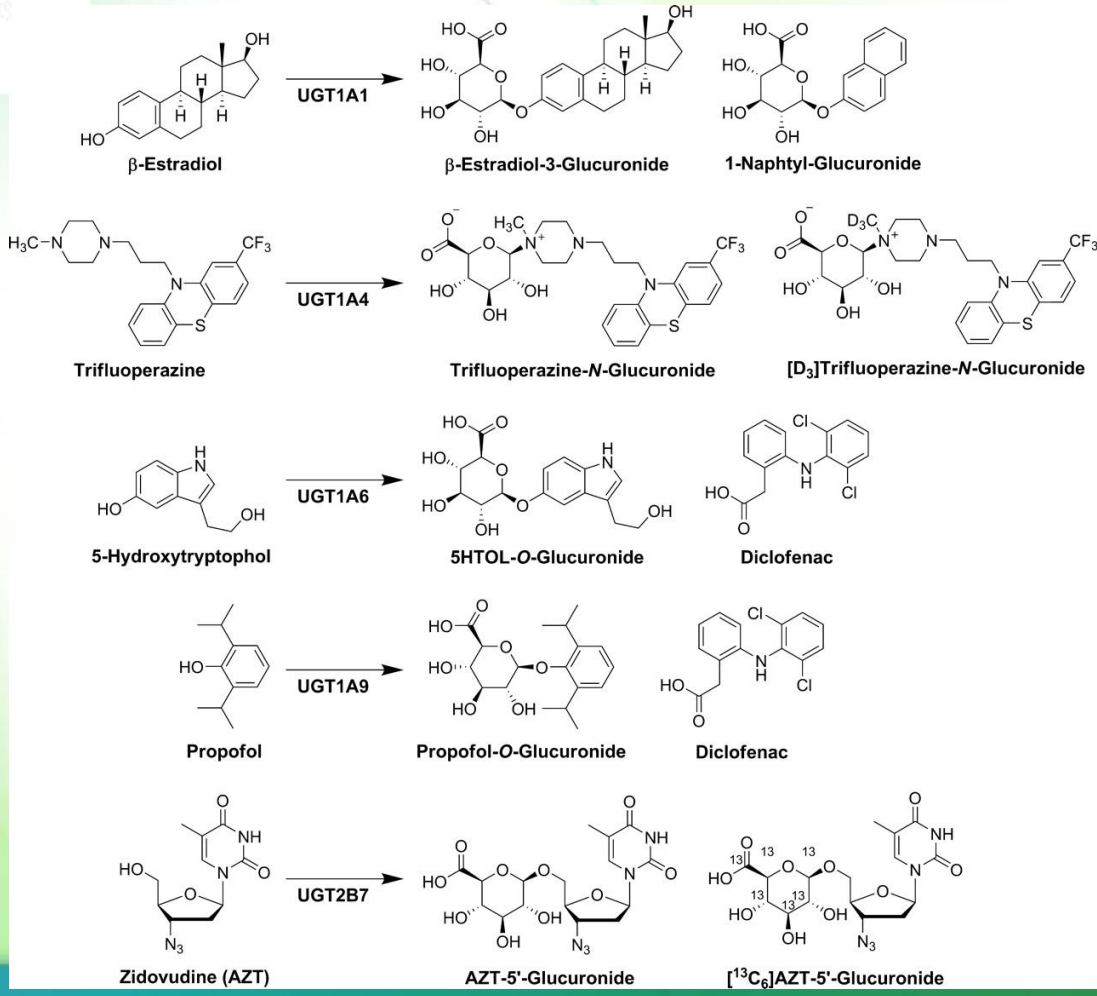
Lipophilic substrate

Therapeutic drugs
Carcinogens
Environmental toxicants
Dietary constituents
Bilirubin

Biliary acids
Steroïds
Retinoic acids
Fatty acids

Substrates
Etoposide
Genistein
Tamoxifen
PCBs
heterocyclic amines
Benzo[a]phrene
Nicotine
Raloxifene

It is a family of enzymes that is responsible for the glucuronidation of hundreds of compounds, including hormones, flavonoids and environmental mutagens.



and reactions are catalyzed in different tissues

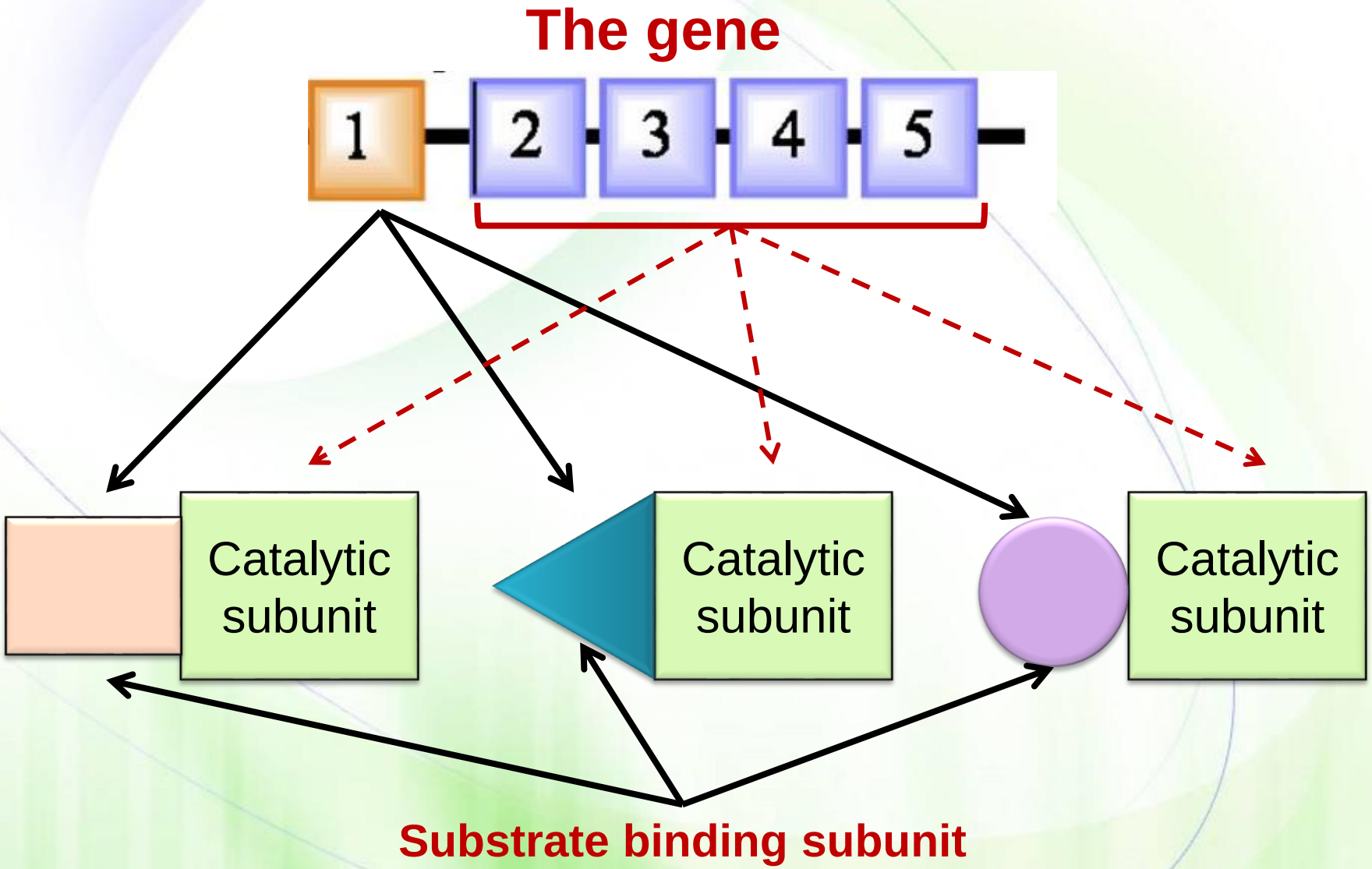
Substrates	Place of reactio
Etoposide	Biliary tissue, colon, intestine, liver, stomach
Genistein	Biliary tissue, colon, liver, stomach
Tamoxifen	Biliary tissue, colon, intestine, liver
PCBs	Biliary tissue, brain, colon, kidney, larynx, liver, lung, stomach
Heterocyclic amines	Esophagus, intestine, kidney, larynx
Benzo[a]phrene	Colon, esophagus, intestine, kidney, larynx
Nicotine	Breast, colon, esophagus, liver, kidney, ovary, prostate, skin, testis
Raloxifene	Biliary tissue, colon, esophagus, intestine, orolaryngeal tissue, stomach

Get this concept, first...

One drill, many flutes

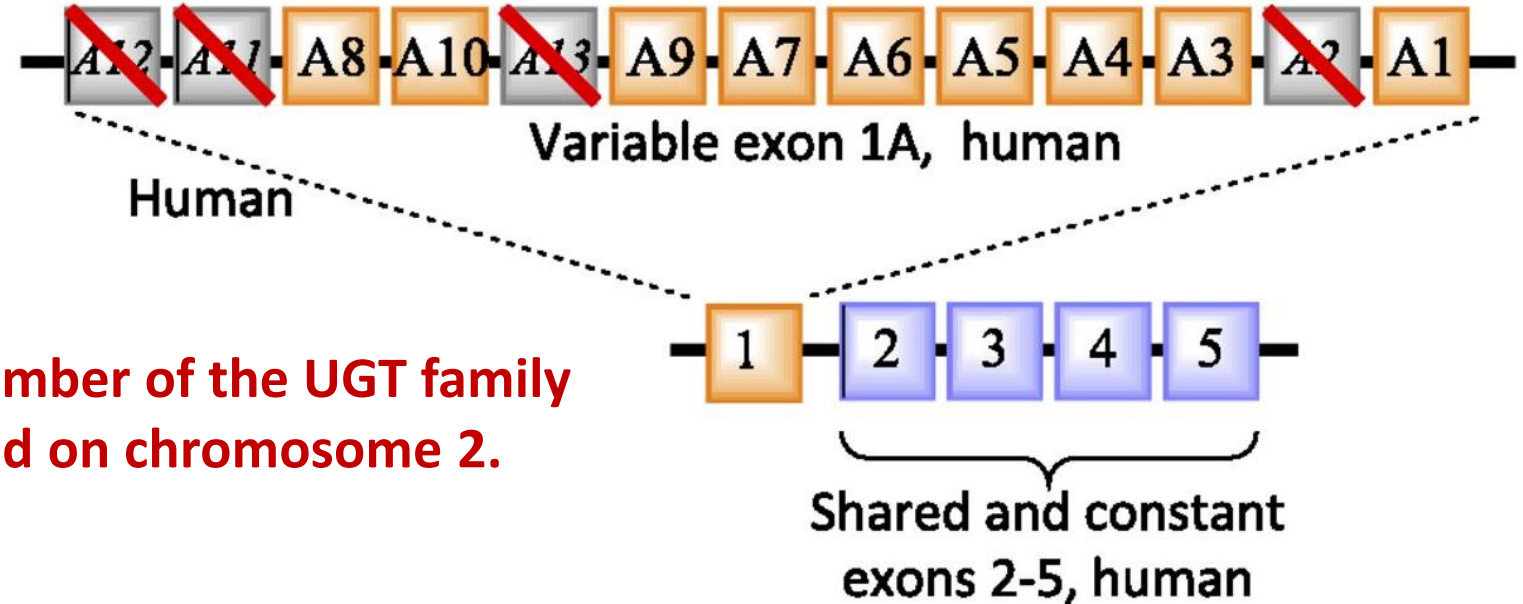
One head, many hats

Then this…


The gene
1
2
3
4
5
Catalytic subunit
Catalytic subunit
Catalytic subunit
Substrate binding subunit

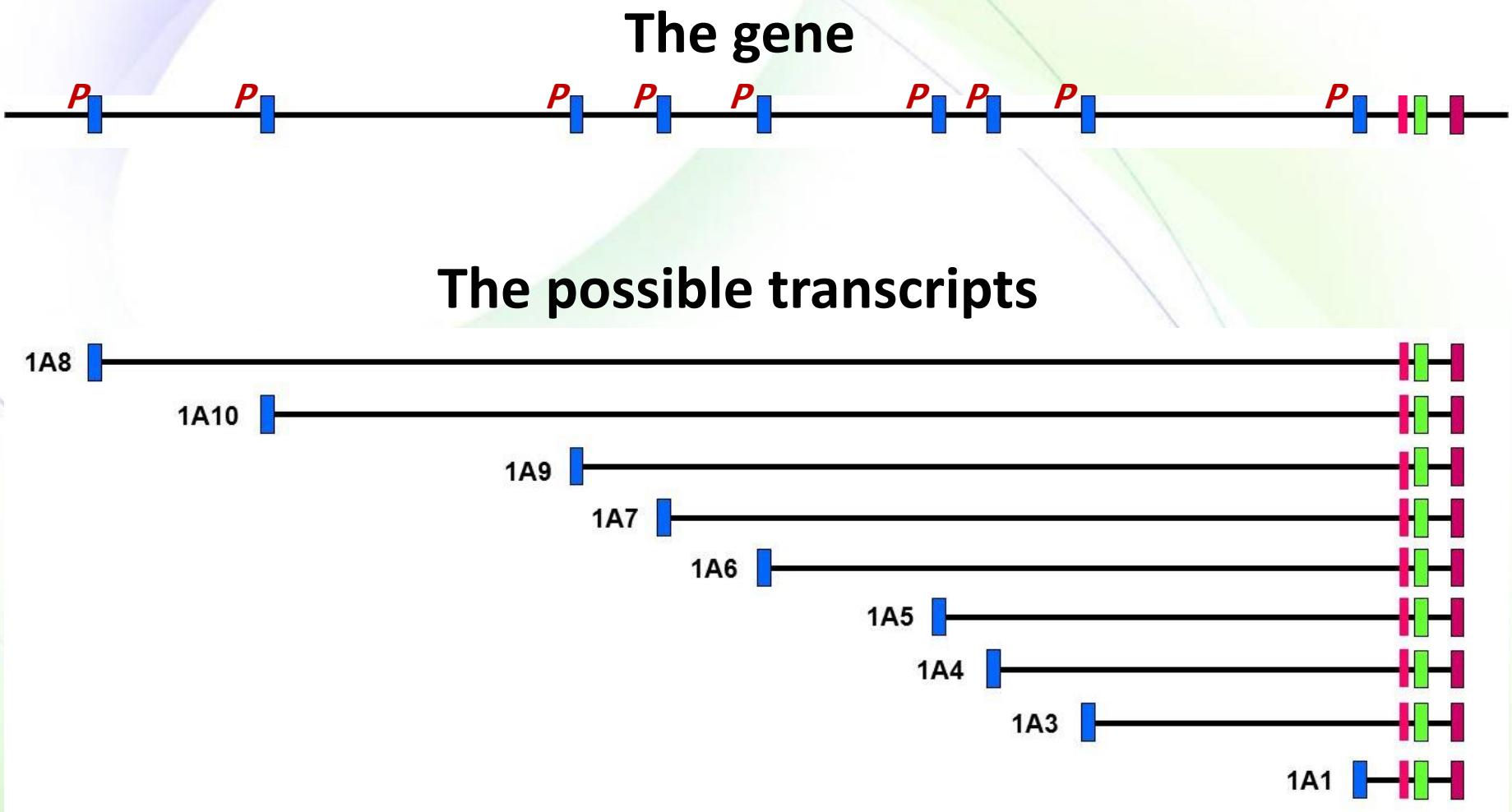
How does it do this?

- Exons 2, 3, 4, and 5 encode the catalytic domain that interacts with UDP-glucuronic acid, but…
- The 5' region of the UGT1A complex contains 9 viable tandemly arrayed first exons an, each with its own promoter.
- The 9 exons determine substrate specificity and one of them is spliced to exon 2 generating 9 possible UGT1A transcripts.



UGT1A is member of the UGT family that is located on chromosome 2.

Splice variants for UGT1A



Gene	Where expressed	Substrates
UGT1A1	Biliary tissue, colon, intestine, liver, stomach	Etoposide
UTG1A3	Biliary tissue, colon, liver, stomach	Genistein
UGT1A4	Biliary tissue, colon, intestine, liver	Tamoxifen
UGT1A6	Biliary tissue, brain, colon, kidney, larynx, liver, lung, stomach	PCBs
UGT1A7	Esophagus, intestine, kidney, larynx	heterocyclic amines
UGT1A8	Colon, esophagus, intestine, kidney, larynx	Benzo[a]phrene
UGT1A9	Breast, colon, esophagus, liver, kidney, ovary, prostate, skin, testis	Nicotine
UGT1A10	Biliary tissue, colon, esophagus, intestine, orolaryngeal tissue, stomach	Raloxifene

Regulation of mRNA stability

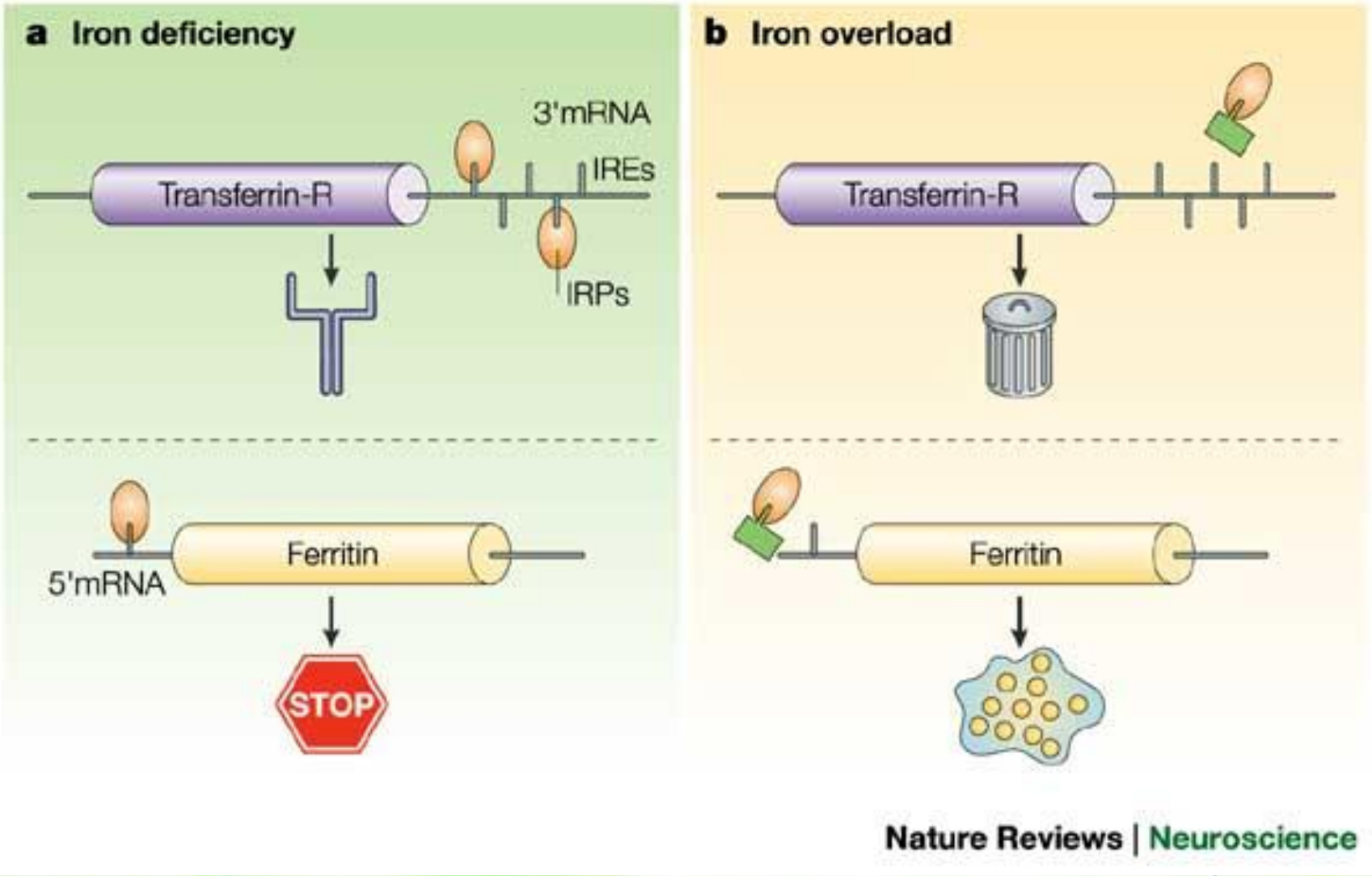
Iron-responsive elements

- In human cells, there are regions of mRNA called iron responsive elements (IREs).
- These regions are contained within the mRNA sequences that code for certain proteins that regulate the levels of iron.
  - Ferritin, transferrin receptor, ferroportin, and DMT1
- Iron-responsive element binding protein (IRE-BP) binds to these mRNA sequences influencing protein expression.

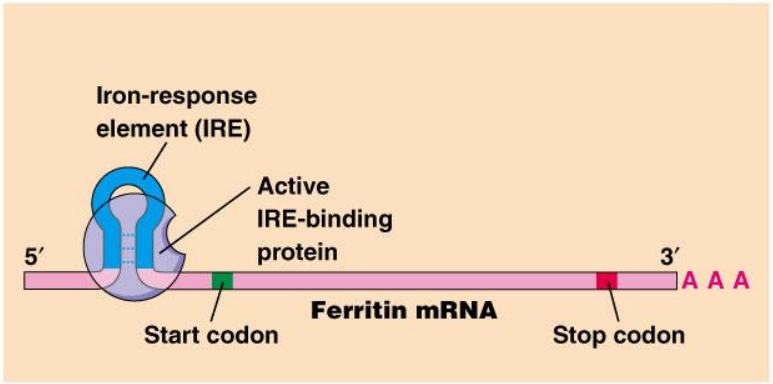
Note:
Liver ferritin stores iron when abundant (in liver)
Transferrin receptor activates iron entry in peripheral cells when needed

Effect on expression

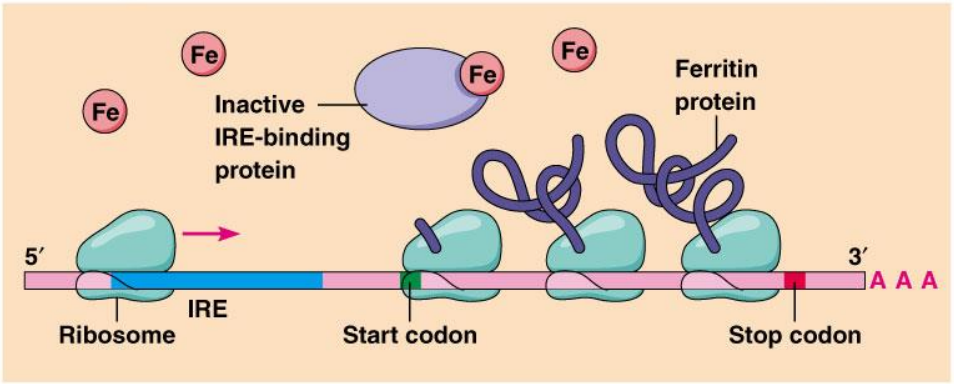
- When iron is abundant, it binds to IRE-BP, disabling the binding of IR-BP to ferritin mRNA
  - This prevents the degradation of the mRNA molecules allowing the production of more ferritin protein
  - Therefore, the iron itself causes the cell to produce more iron storage molecules
- On the other hand, at low iron levels, the IRE-BP will bind to the ferritin mRNA and, thus, the mRNA will be destabilized, making less ferritin protein
- An opposite effect is seen on the stability of transferrin receptor mRNA, which has IRE at the 3'-end.


a Iron deficiency
3'mRNA
IREs
Transferrin-R
IRPs
5'mRNA
Ferritin
STOP
b Iron overload
Transferrin-R
Ferritin
Nature Reviews | Neuroscience

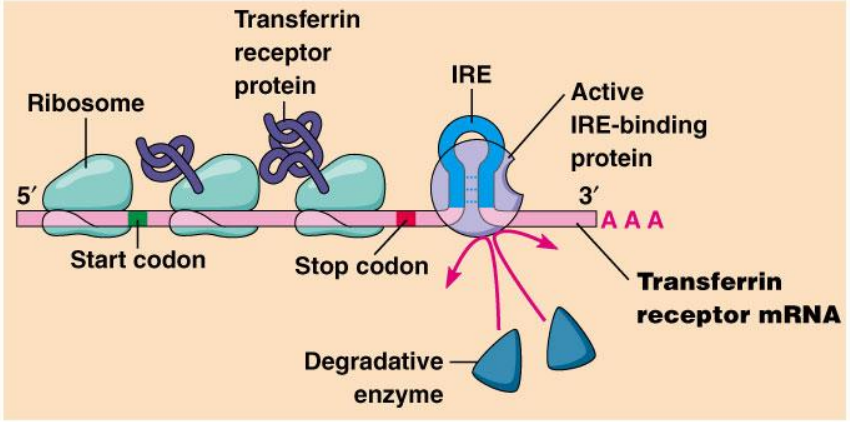
(a) Low iron concentration. IRE-binding protein binds to IRE, so translation of ferritin mRNA is inhibited.



(b) High iron concentration. IRE-binding protein cannot bind to IRE, so translation of ferritin mRNA proceeds.



(a) Low iron concentration. IRE-binding protein binds to the IRE of transferrin receptor mRNA, thereby protecting the mRNA from degradation. Synthesis of transferrin receptor therefore proceeds.



(b) High iron concentration. IRE-binding protein cannot bind to IRE, so mRNA is degraded and synthesis of transferrin receptor is thereby inhibited.