Molecular Biology (6)

Transcription

Mamoun Ahram, PhD
Second semester, 2018-2019

Resources

- This lecture
- Cooper, Chapter 8, p. 158 and 290, pp. 315-317

Definition of a gene

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

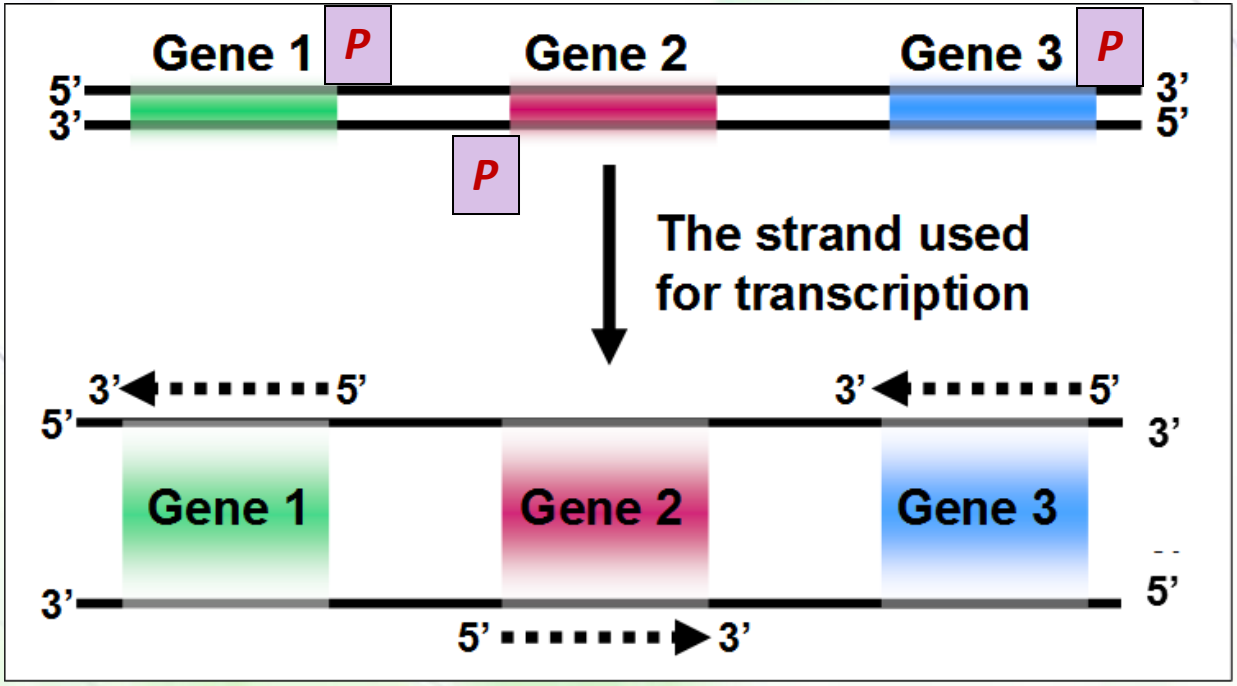
The general mechanism of transcription

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.

Using DNA strands

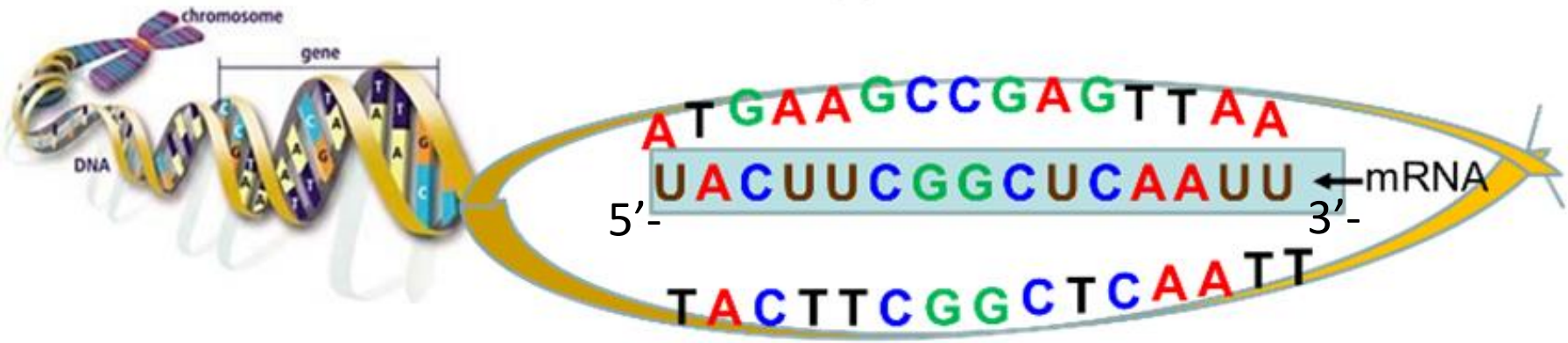
- Although RNA polymerase can read both DNA strands, it uses one strand for any particular gene in order to make RNA.



What does determine which strand is used for transcription?

Complementary sequences

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



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

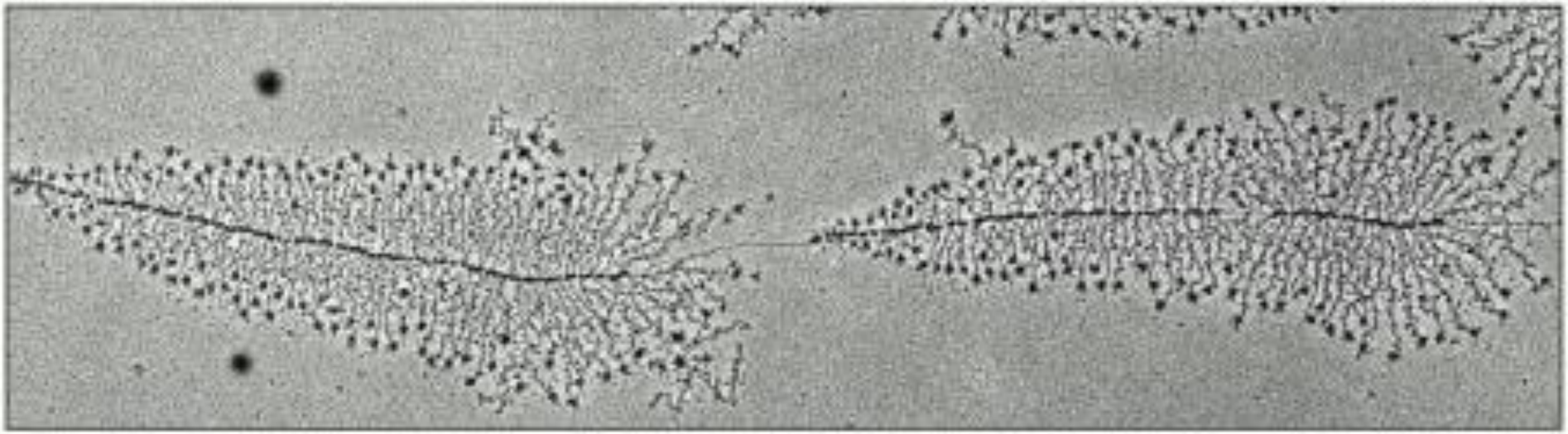
Enzyme and substrate

- The enzymes that perform transcription are called RNA polymerases.
- RNA polymerases catalyze the formation of the phosphodiester bonds between two nucleotides.
- The substrates are nucleoside triphosphates (ATP, CTP, UTP, and GTP).
- A hydrolysis of high-energy bonds in NTP provides the energy needed to drive the reaction forward.

Polysomes

- 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

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

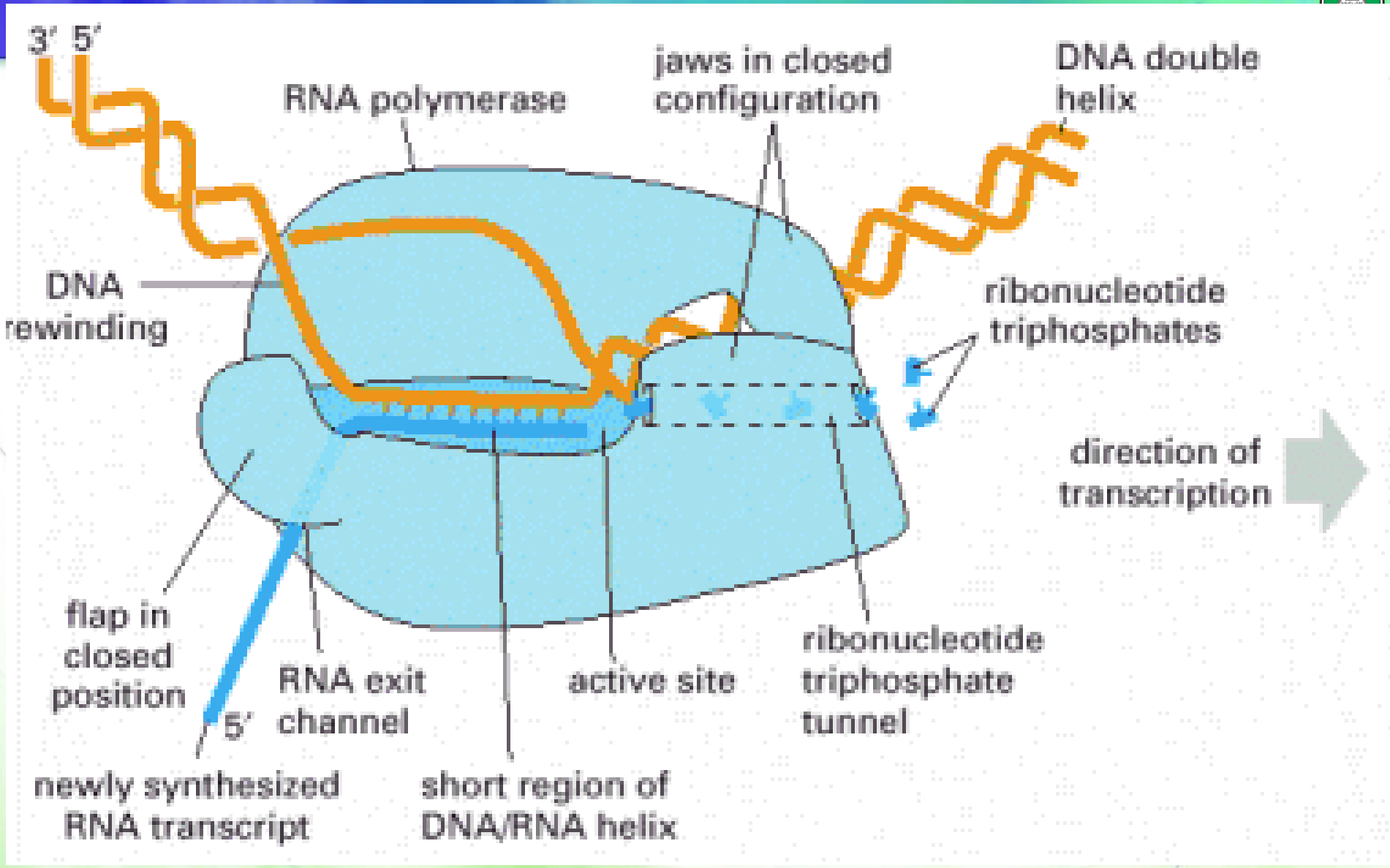


DNA replication vs. transcription

- 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

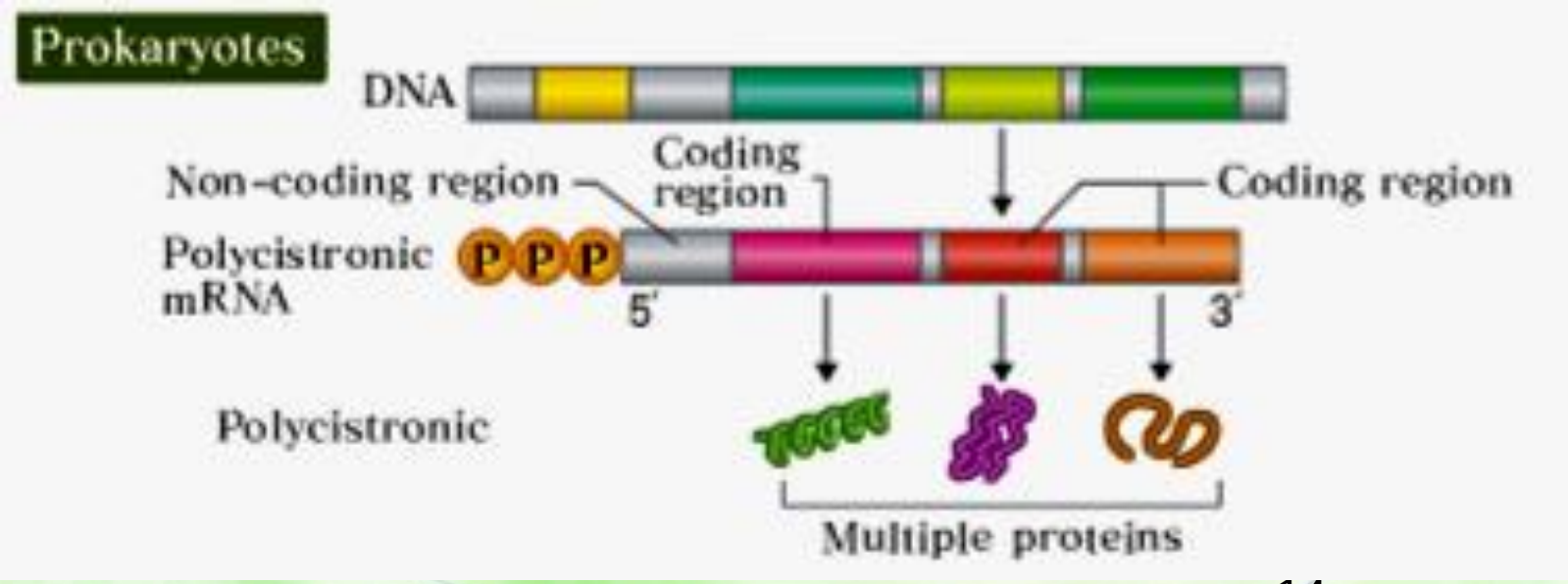
- RNA polymerase catalyzes the linkage of ribonucleotides, not deoxyribonucleotides.
- Unlike DNA polymerases, RNA polymerases can start an RNA chain without a primer.
- RNA polymerases make about one mistake for every 10^4 nucleotides.
  - the consequences of an error in RNA transcription are much less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


3′ 5′
RNA polymerase
jaws in closed configuration
DNA double helix
DNA rewinding
ribonucleotide triphosphates
direction of transcription
flap in closed position
RNA exit channel
active site
ribonucleotide triphosphate tunnel
5′
newly synthesized RNA transcript
short region of DNA/RNA helix

Transcription in prokaryotes

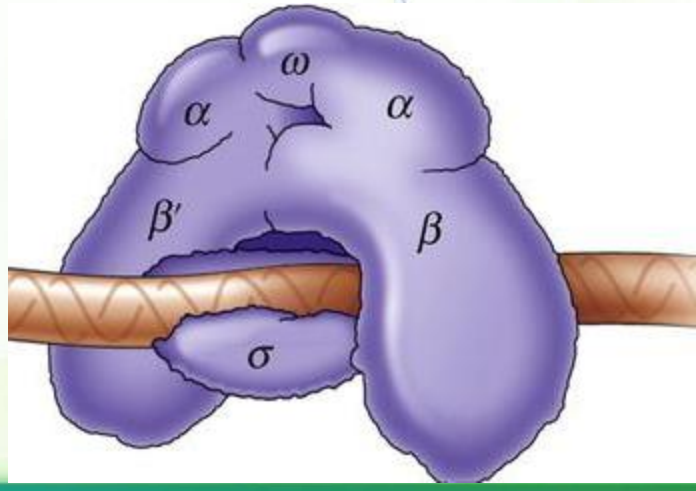
Prokaryotic genes (operon)

- In bacteria, genes can be polycistronic.
- In bacteria, genes that encode enzymes, which are involved in related functions, often are located next to each other
  - Example: the genes encoding the enzymes required to synthesize tryptophan are located in one contiguous stretch.
- This cluster of genes comprises a single transcriptional unit referred to as an operon.



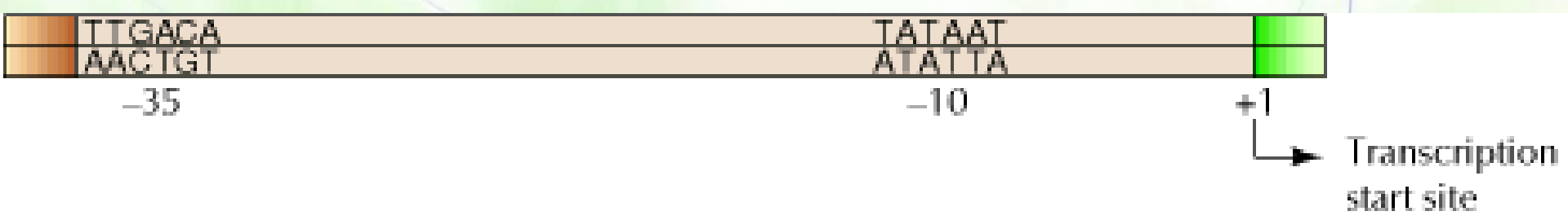
The RNA polymerase

- E. coli RNA polymerase is made up of multiple polypeptide chains.
- 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.



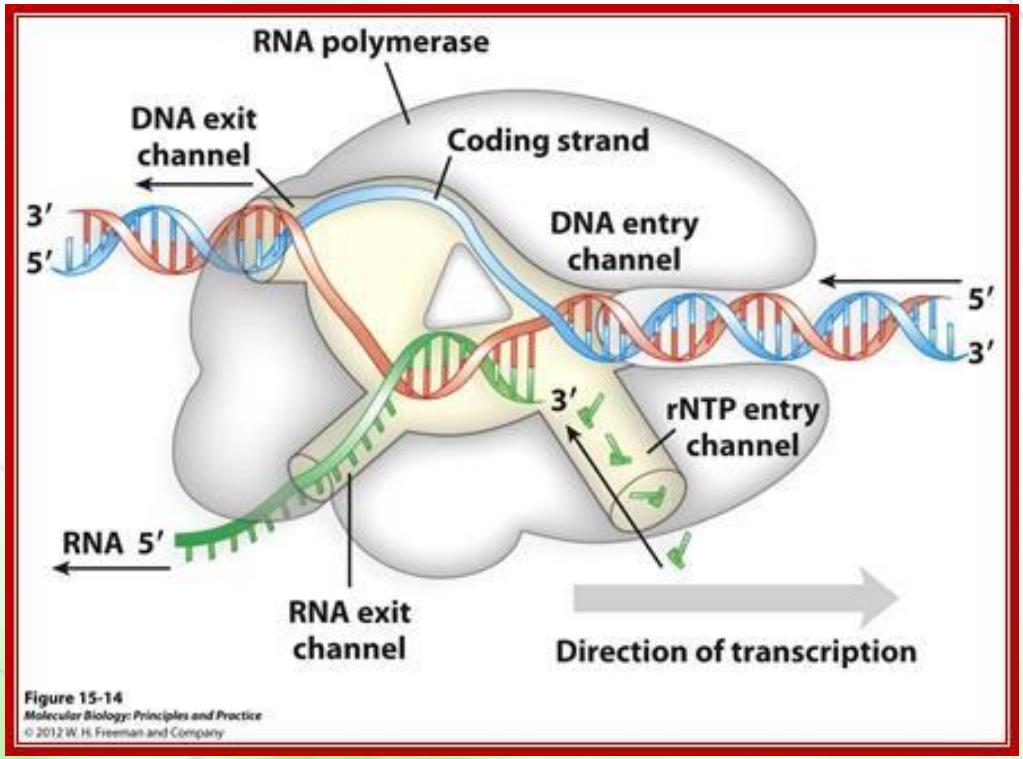
Consensus sequences (the promoter)

- The DNA sequence to which 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.
- 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
  - Open reading frame: DNA sequence that can be transcribed into mRNA from first base to last one.



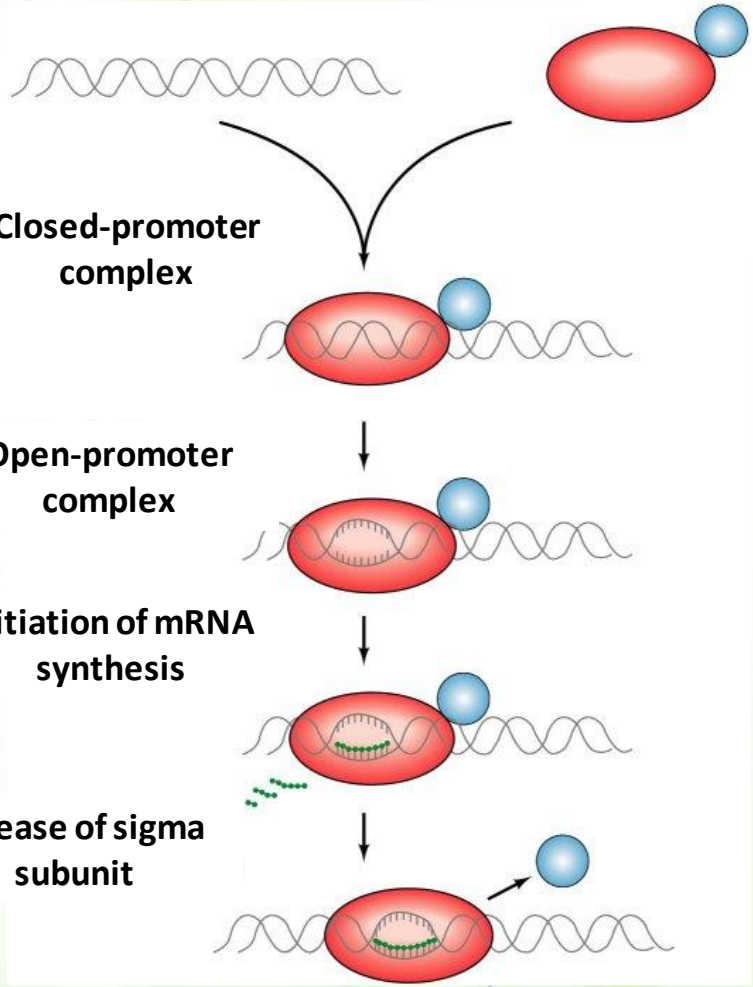
Role of the σ subunit

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



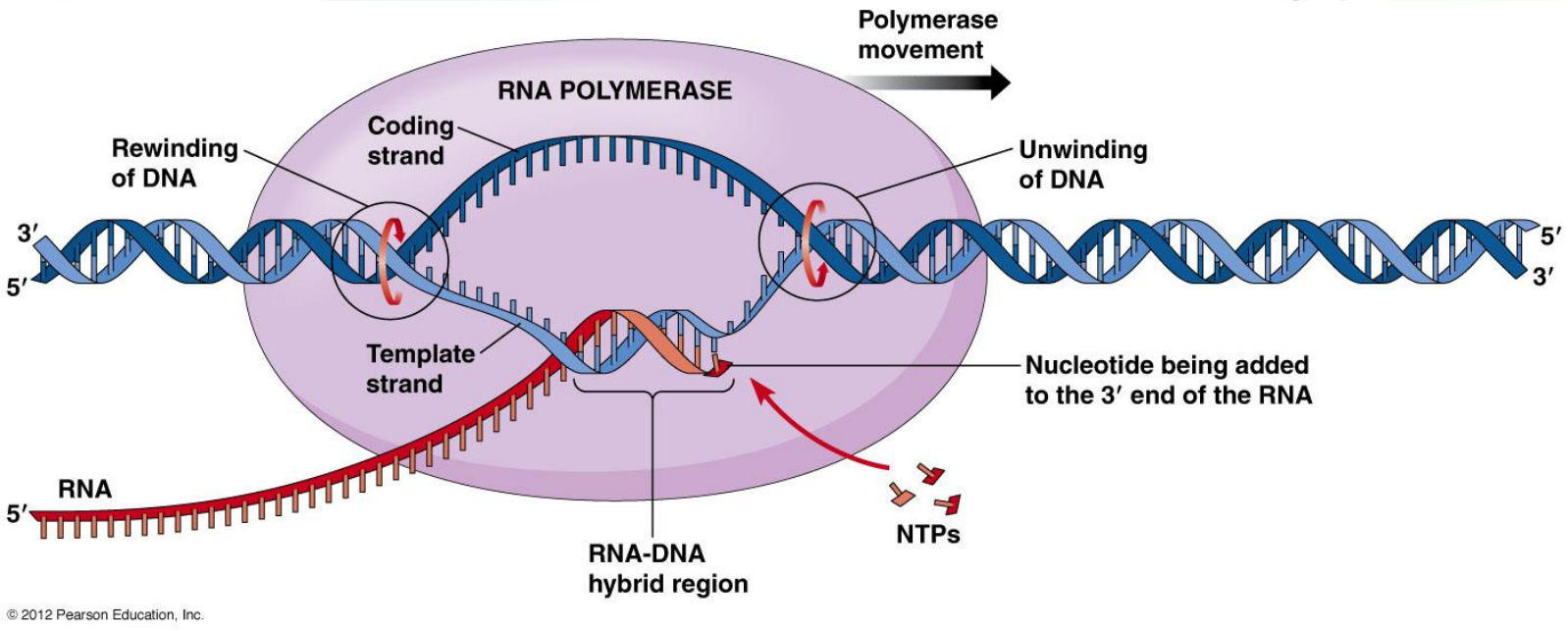
Mechanism of transcription (initiation)

- The complex of the polymerase and a promoter is referred to as a closed-promoter complex.
- The polymerase unwinds a few bases of DNA to form an open-promoter complex.
- Single-stranded DNA is available as a template.
- Transcription is initiated by the joining of two NTPs.
- After addition of about the first 10 nucleotides, σ is released from the polymerase.



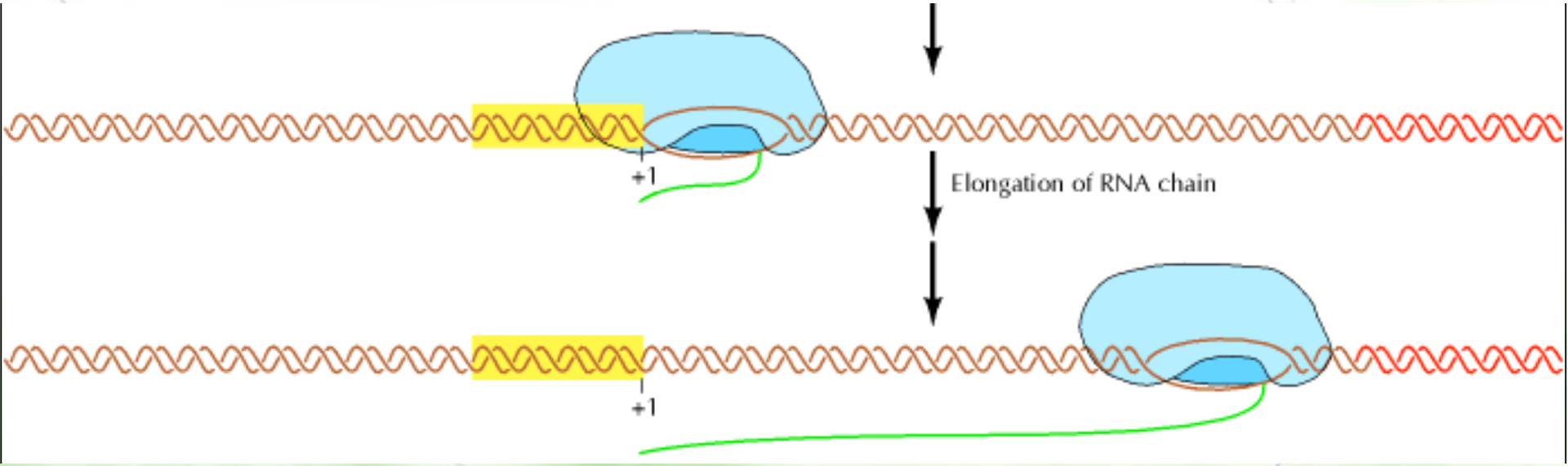
Mechanism of transcription (elongation)

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



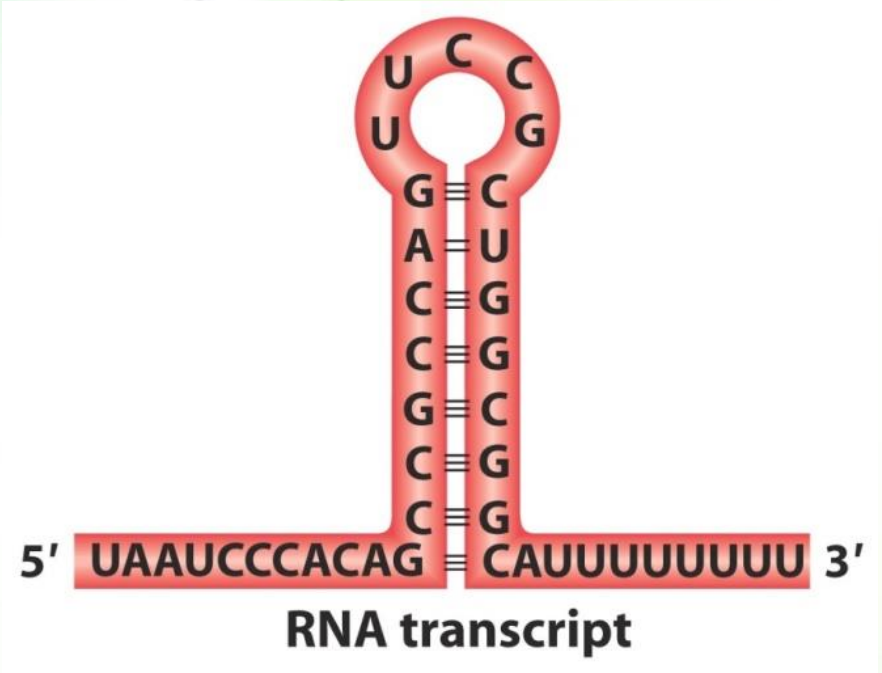
Mechanism of transcription (termination)

- 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.



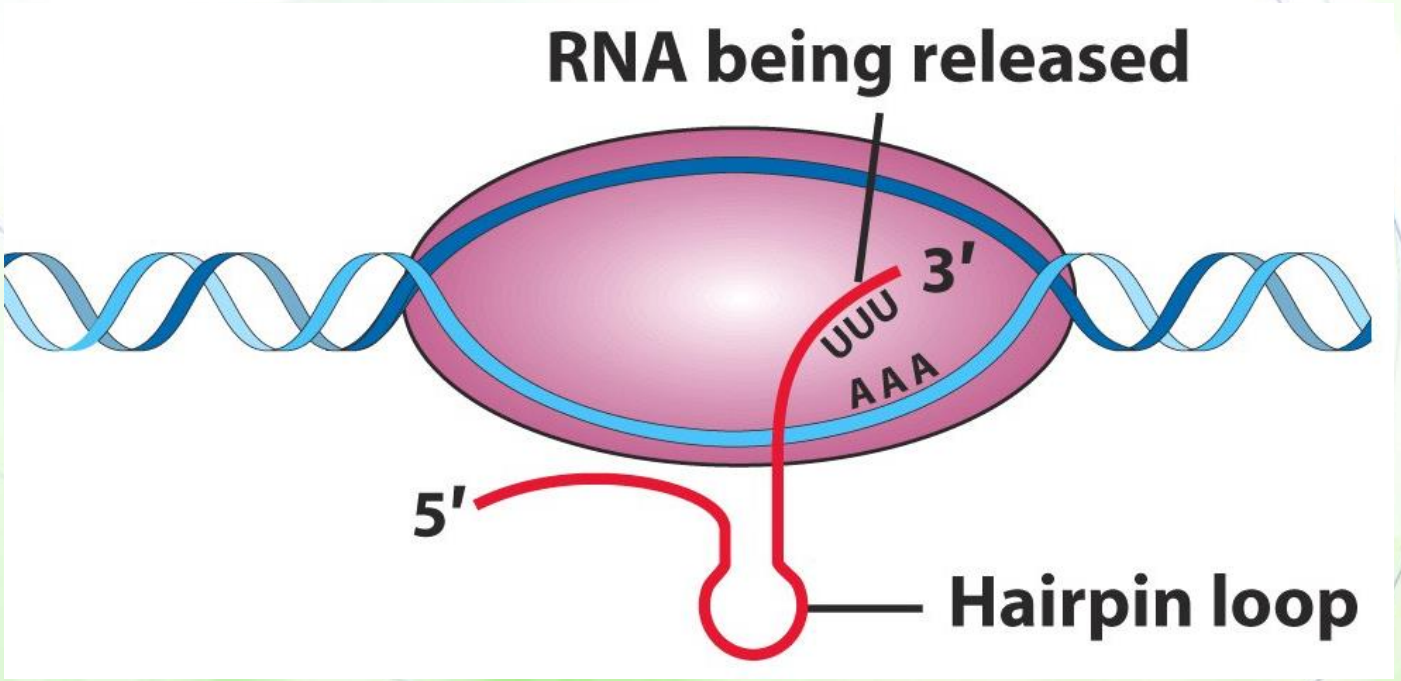
Termination sequences

- The simplest and most common type of termination signal 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.



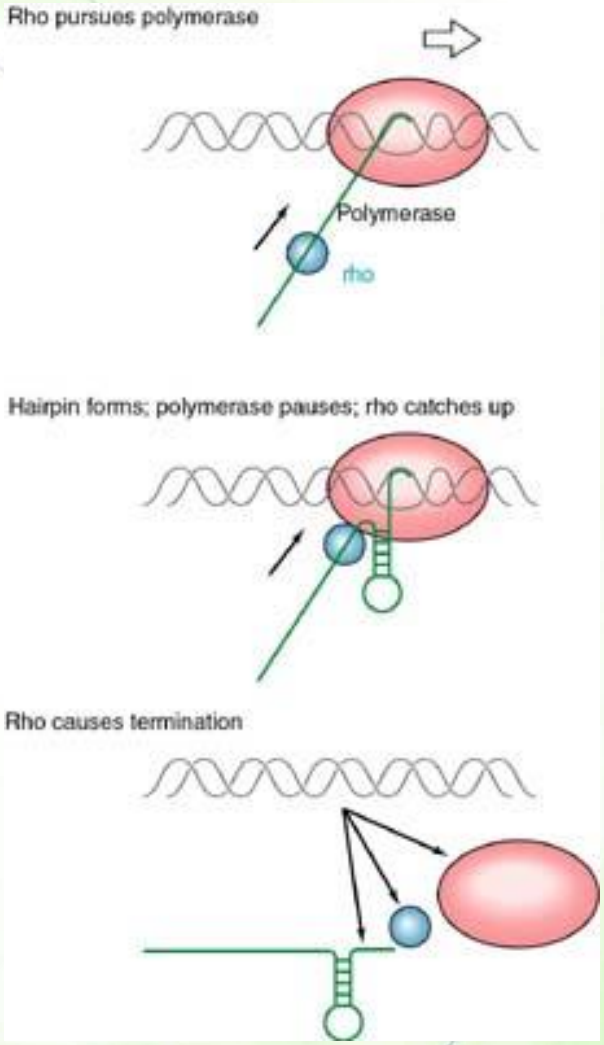
The effect of the stem loop structure

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



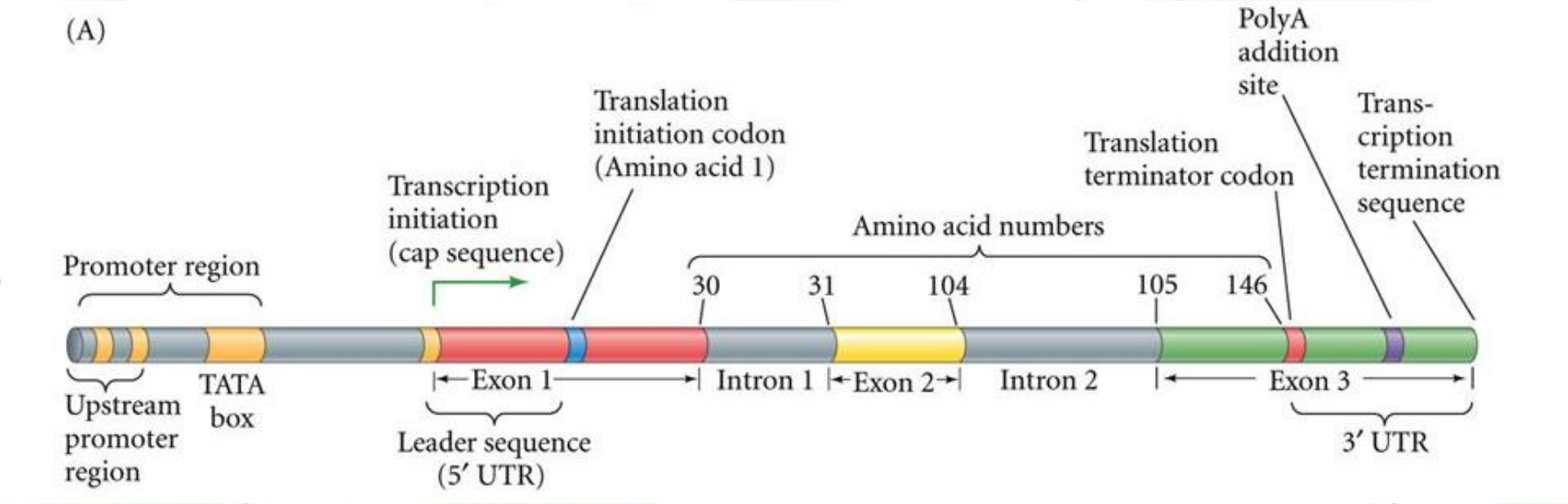
Rho-dependent terminator

- Rho follows the RNA polymerase along the transcript. When the polymerase stalls at a hairpin, Rho catches up and breaks the RNA-DNA base pairs, releasing the transcript.
- Rho-dependent termination signals do not have the string of U residues at the end of the RNA.



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.
  - 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.

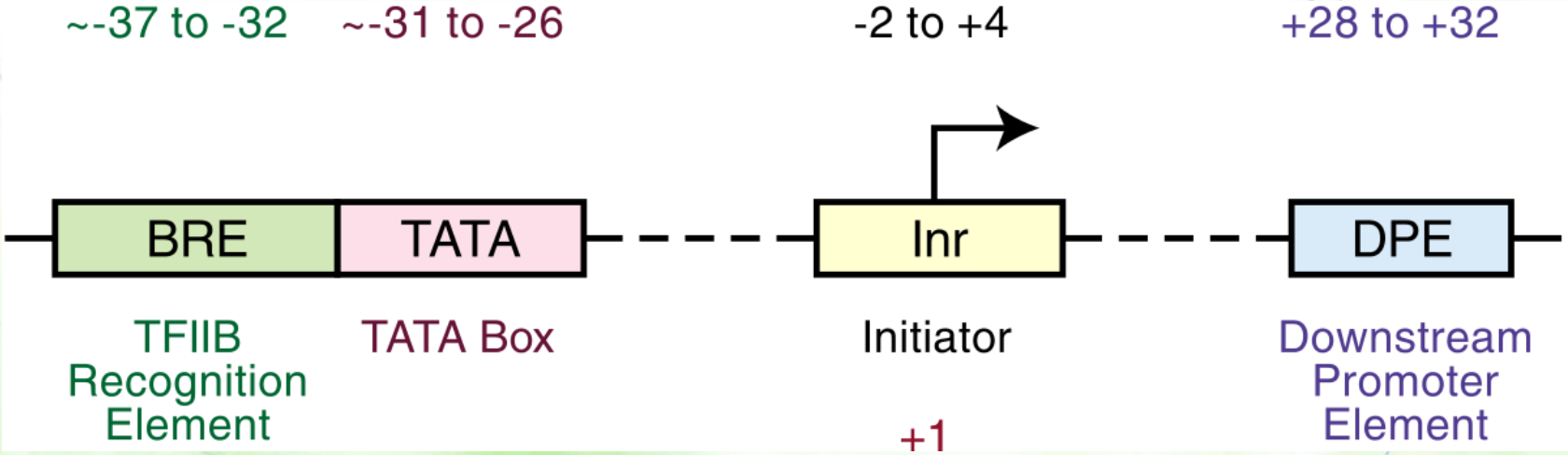
- While bacterial RNA polymerase is able to initiate transcription without the help of additional proteins, eukaryotic RNA polymerases cannot.
  - They require the help of a large set of proteins called general transcription factors.
  - The proteins are "general" because they assemble on all promoters 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.

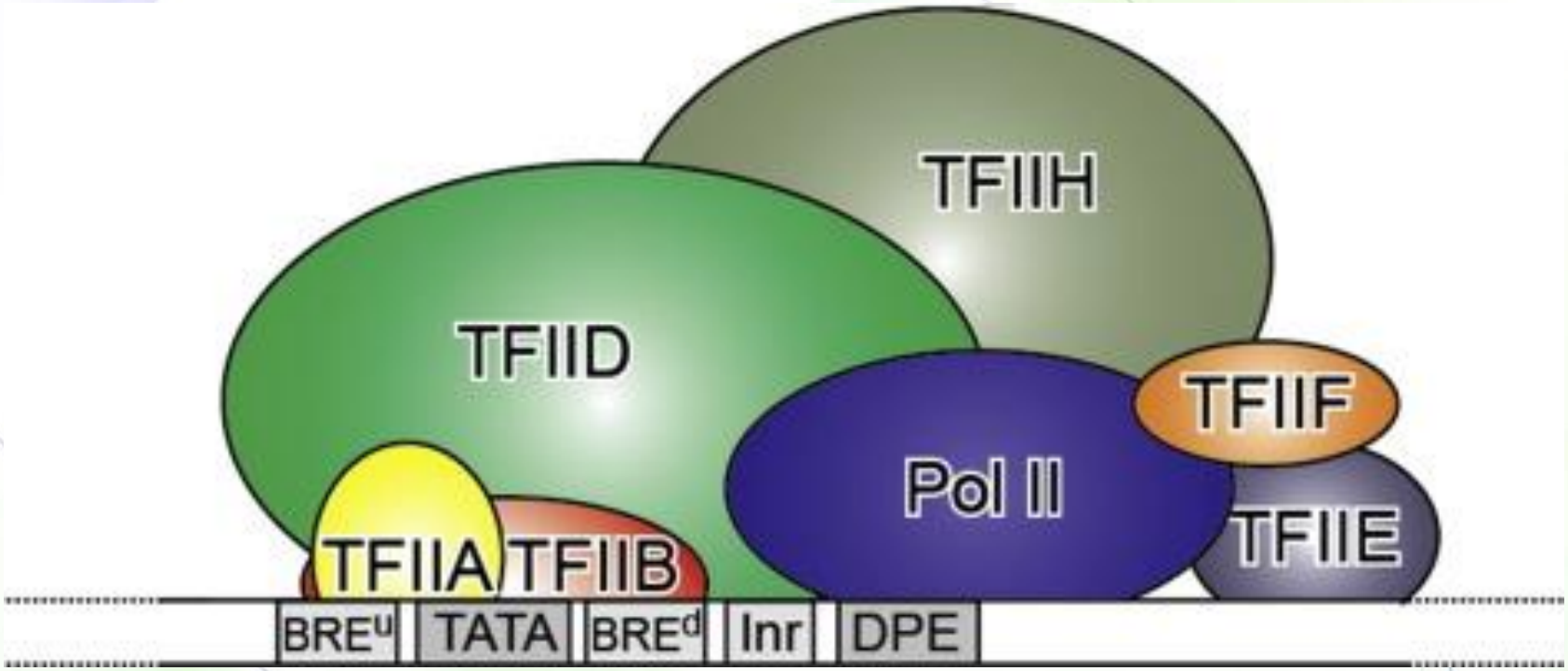
Core components of promoters

- These are important for basal expression.
  - The TATA box, which is the binding site of TFIID.
    - It is commonly found I promoters of genes transcribed by RNA polymerase II.
  - An upstream element (BRE) that is binding site of TFIIB.
  - The initiator element (*Inr)*, which surrounds the +1 site.
  - Multiple downstream elements.



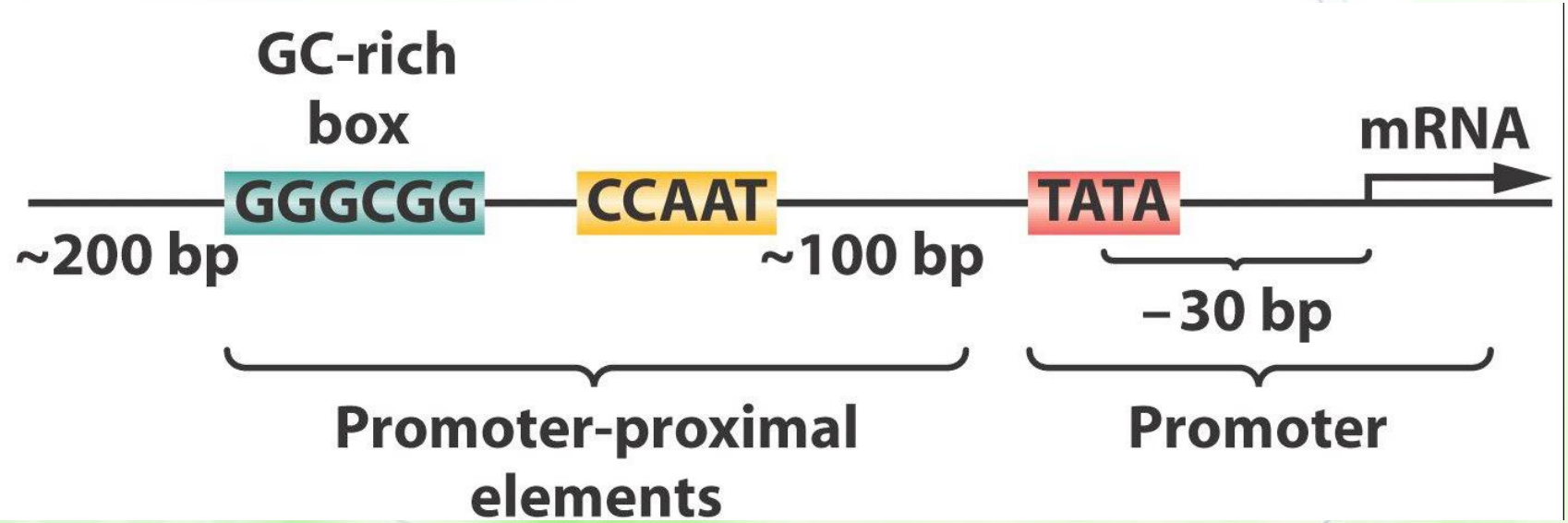
- Not all of them exist at once, but genes can have a combination of these promoter elements.

Formation of preinitiation complex

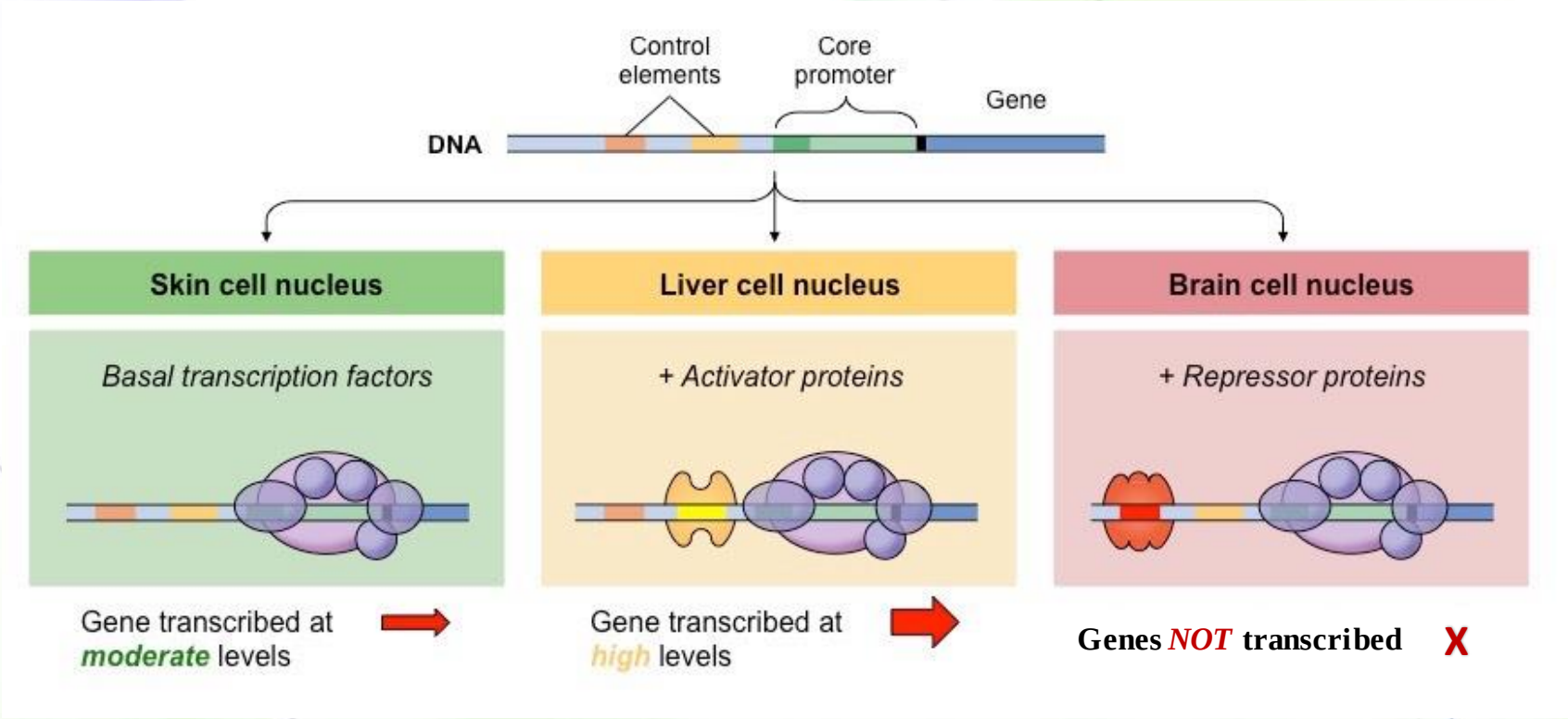


Promoter-proximal elements

- These are upstream of the core promoter region.
- They are important for strong expression.
- They can be gene-specific and bound to regulatory proteins according to cell needs (e.g. energy) or conditions (e.g. temperature).

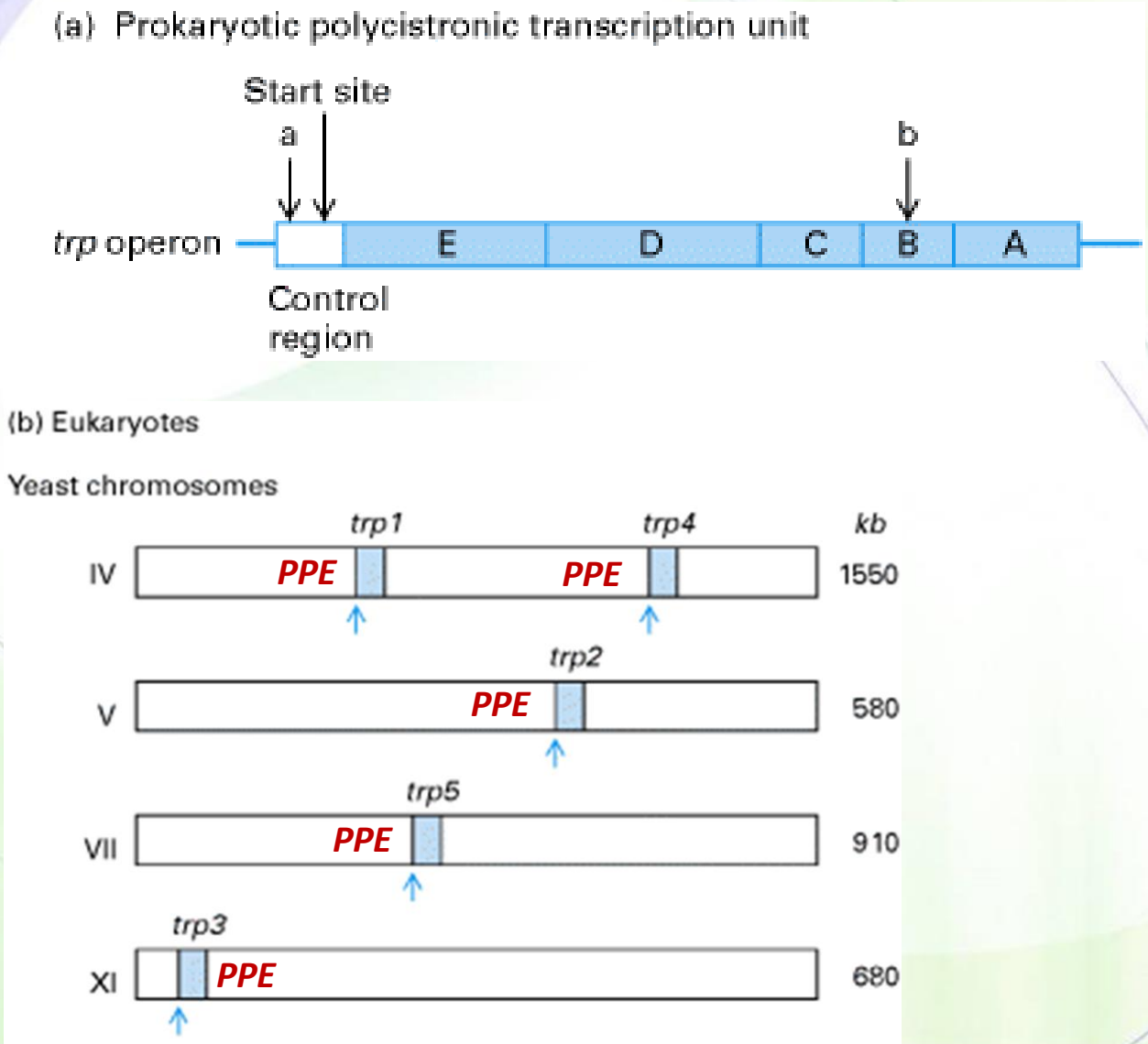


Tissue-specific transcription factors

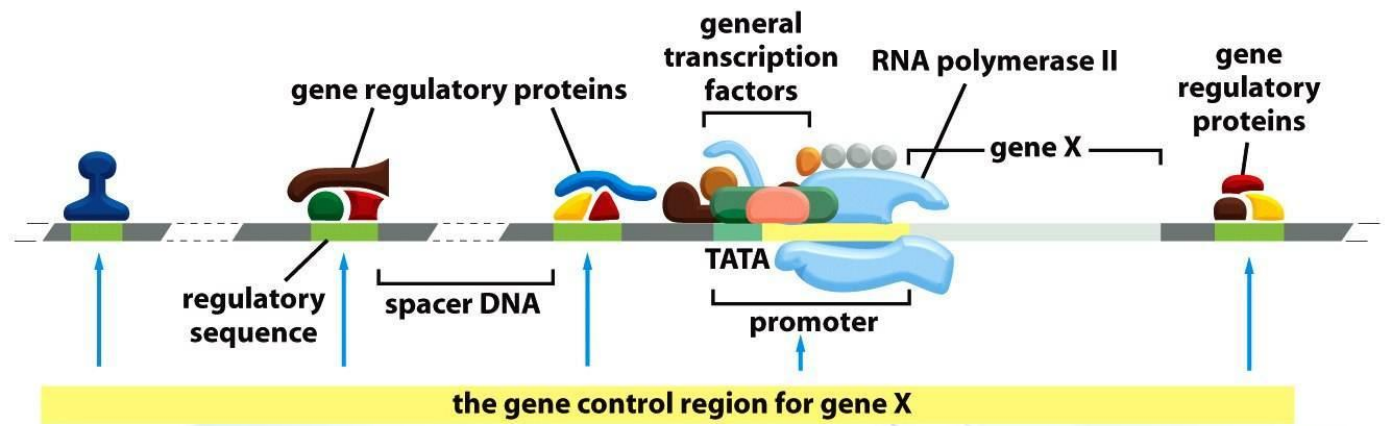


Differential expression of transcription factors (tissue-specific transcription factors) determine gene expression.

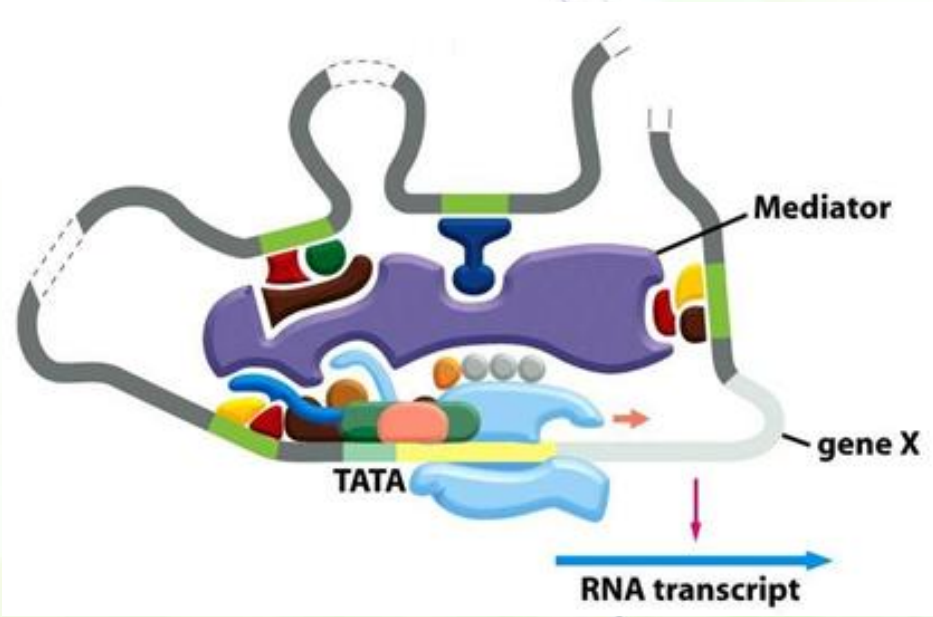
Operon vs. Proximal-promoter elements



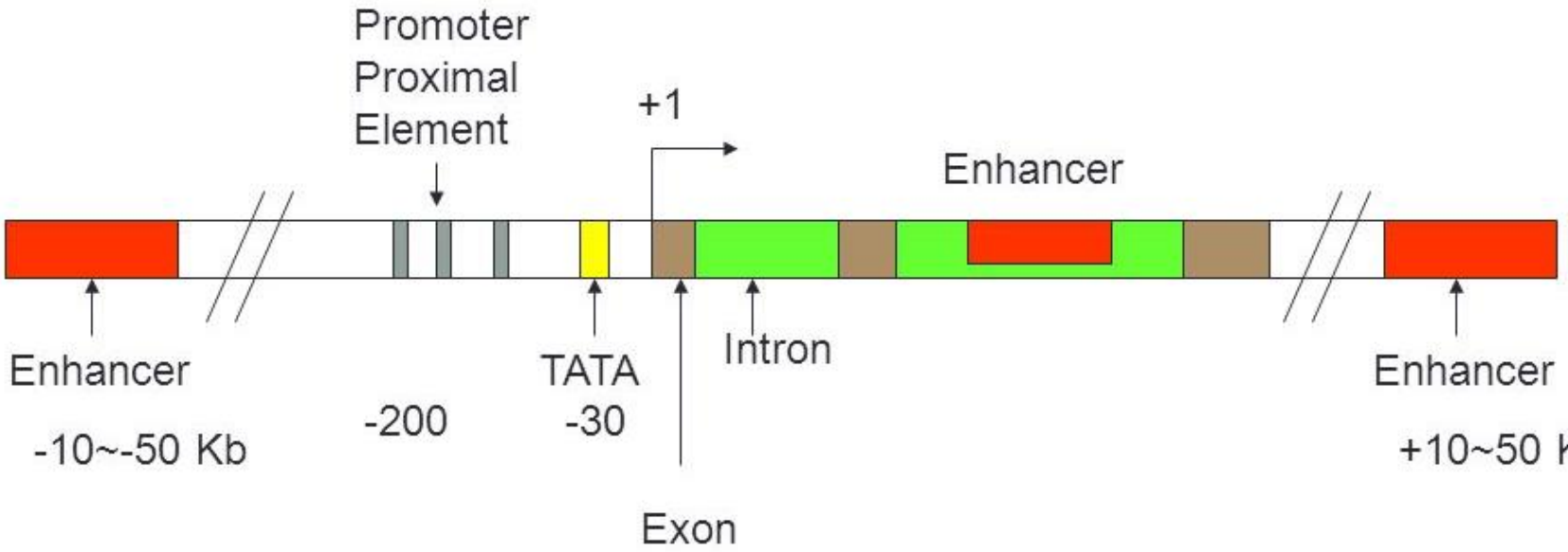
Enhancers



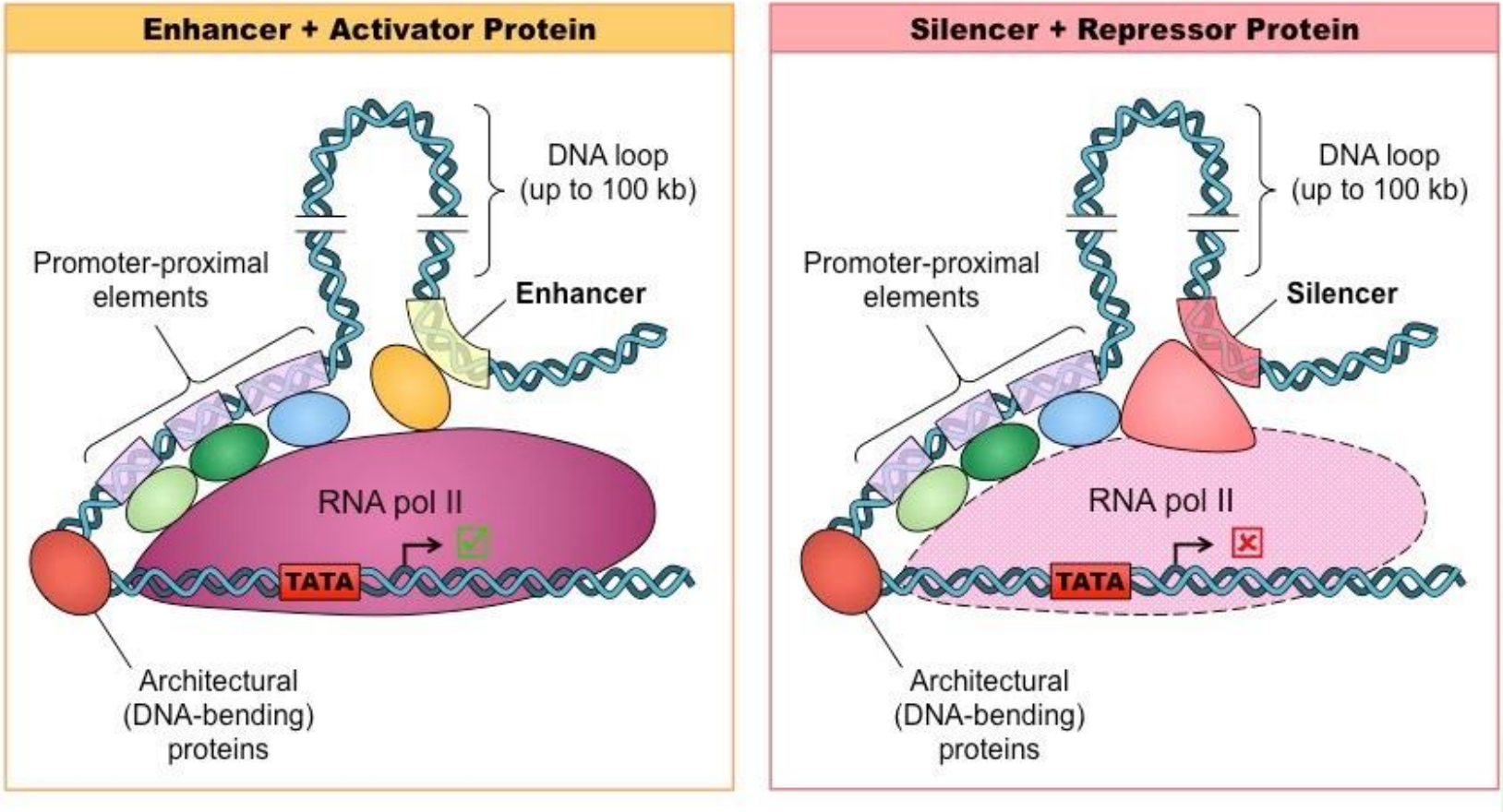
- Many genes are regulated by cis-acting regulatory sequences called enhancers, which are binding sites for **gene-specific** transcription factors that regulate RNA polymerase II such as a protein called the *Mediator*.
- They can regulate transcription <u>regardless</u> of orientation or location due to DNA looping.



A hypothetical mammalian promoter region



All working together

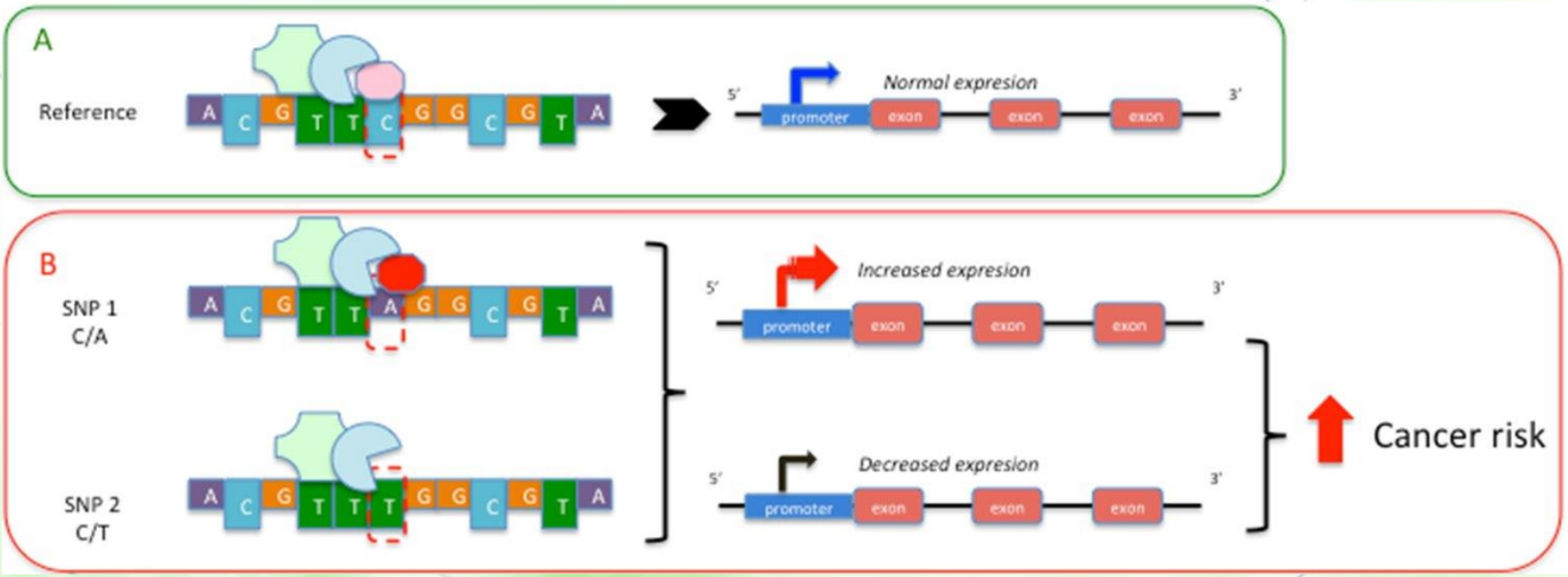


Mechanism of transcription (initiation)

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

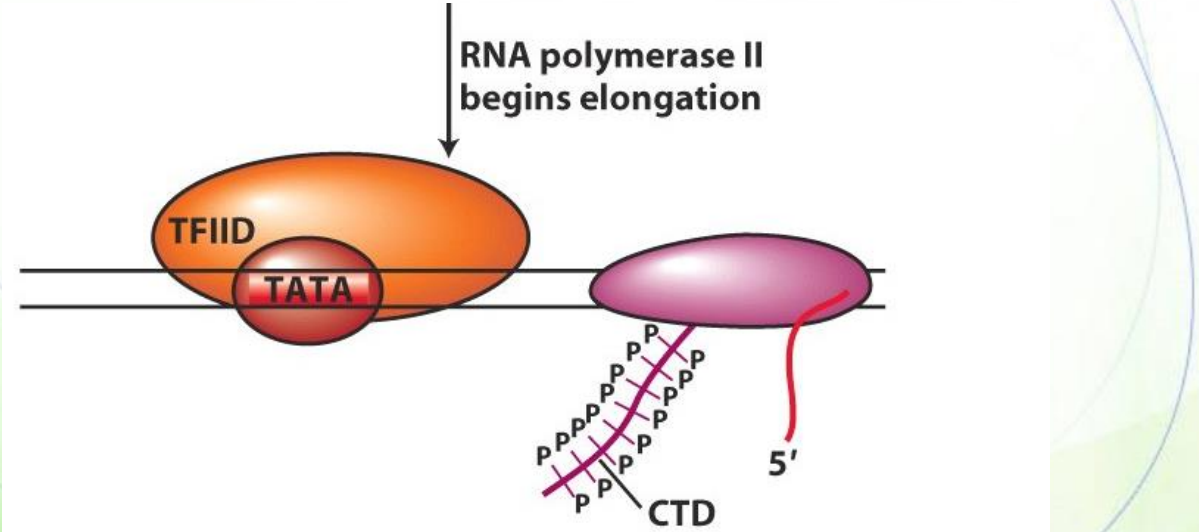
SNPs in promoter

- Single nucleotide polymorphisms (SNP) in promoter region can alter the binding of transcription factors required for the expression of a gene.
- These variations may increase or decrease the expression of the affected gene, which eventually can influence the risk of developing a disease.



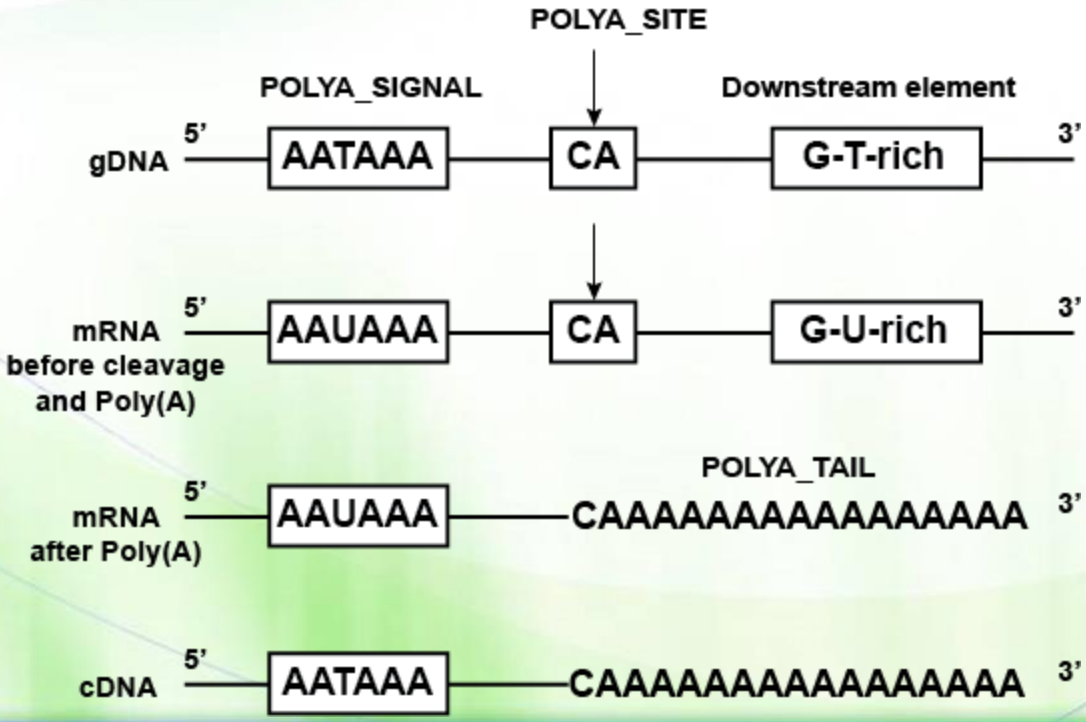
Mechanism of transcription (elongation)

- 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 contains a protein kinase subunits.



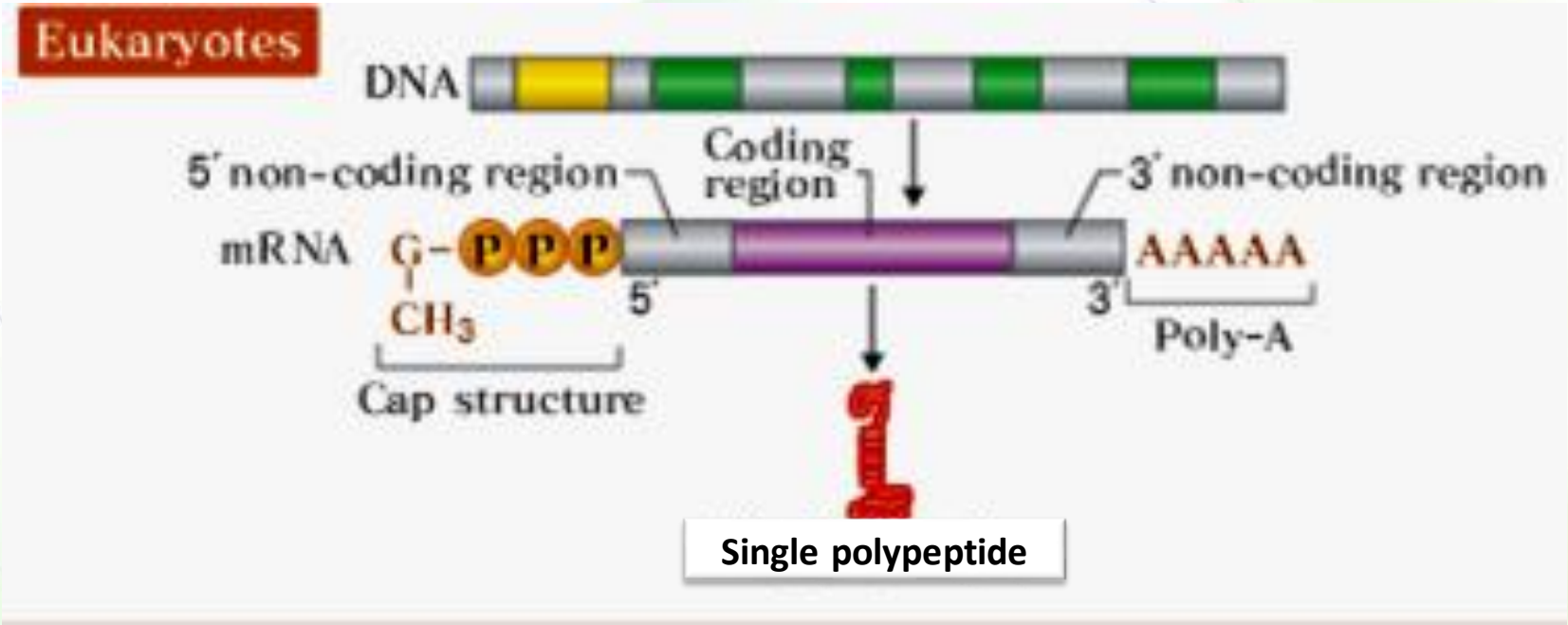
Mechanism of transcription (termination)

- Termination is determined by a consensus sequence for termination, which is AAUAAA followed 10-30 nucleotides downstream by a GU-rich sequence.
- Termination is coupled to the process that cleaves and polyadenylates the 3'-end of the transcript.



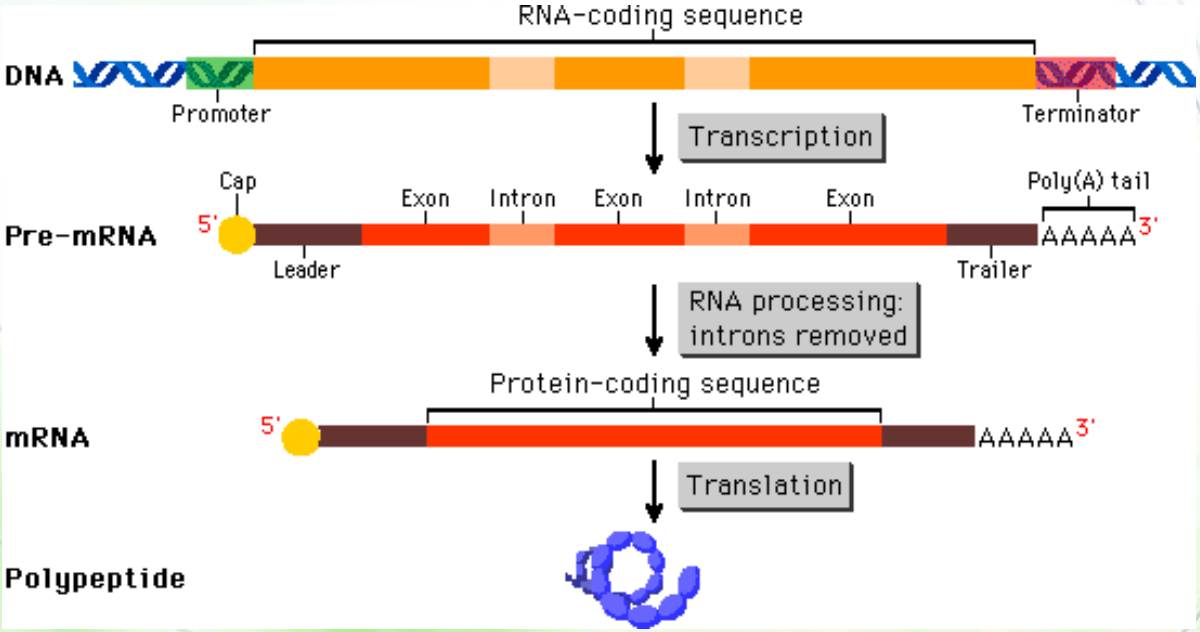
Eukaryotic genes

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



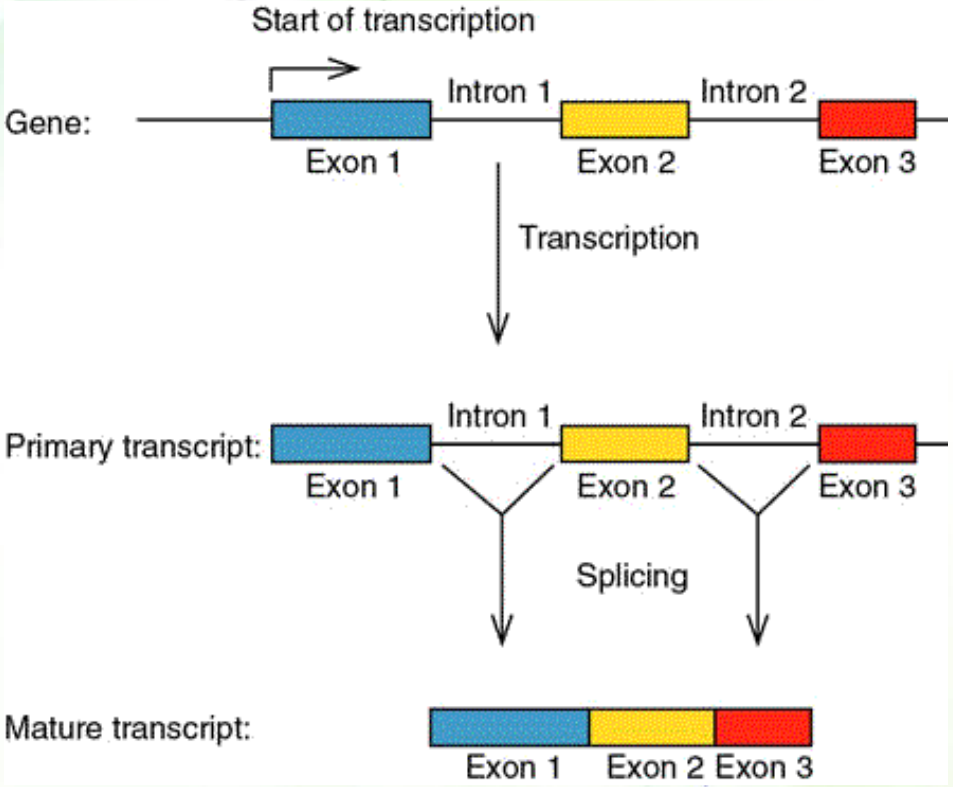
Introns vs. exons

- 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).

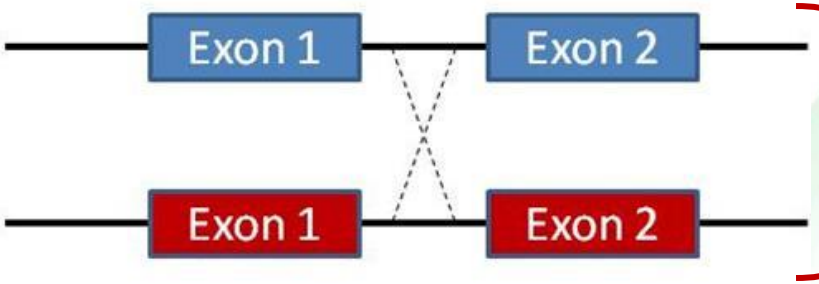


Significance of introns

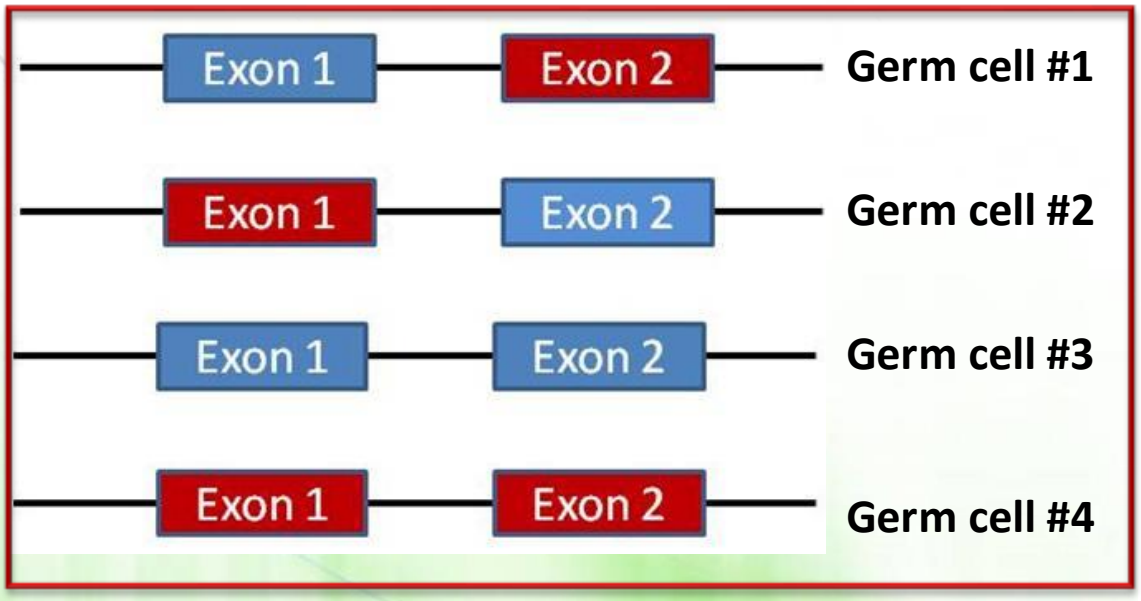
- They can encode functional RNAs such nucleolar RNA that function in ribosomal processing as well as microRNAs.
- They contain regulatory sequences of gene expression.
- Variation among individuals can be generated via genetic recombination.
- The exon-intron arrangement may facilitate the emergence of new proteins via Alternative splicing.

Introns and genetic recombination

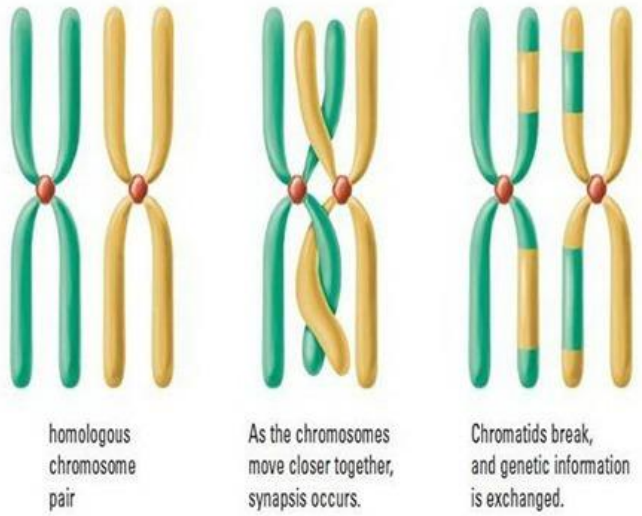
During production of hapoloid gametes



Homologous chromosomes in a diploid precursor germline cell before meiosis

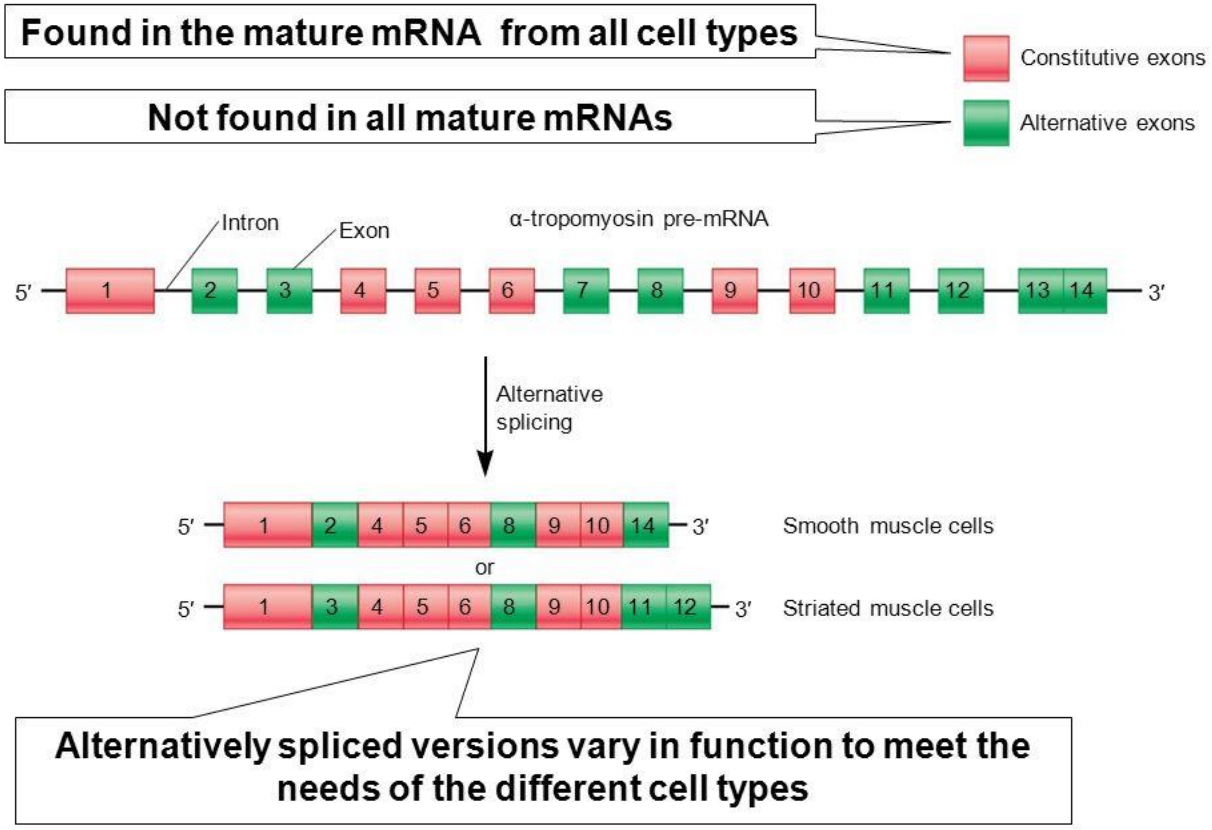


The possible cells



Alternative splicing

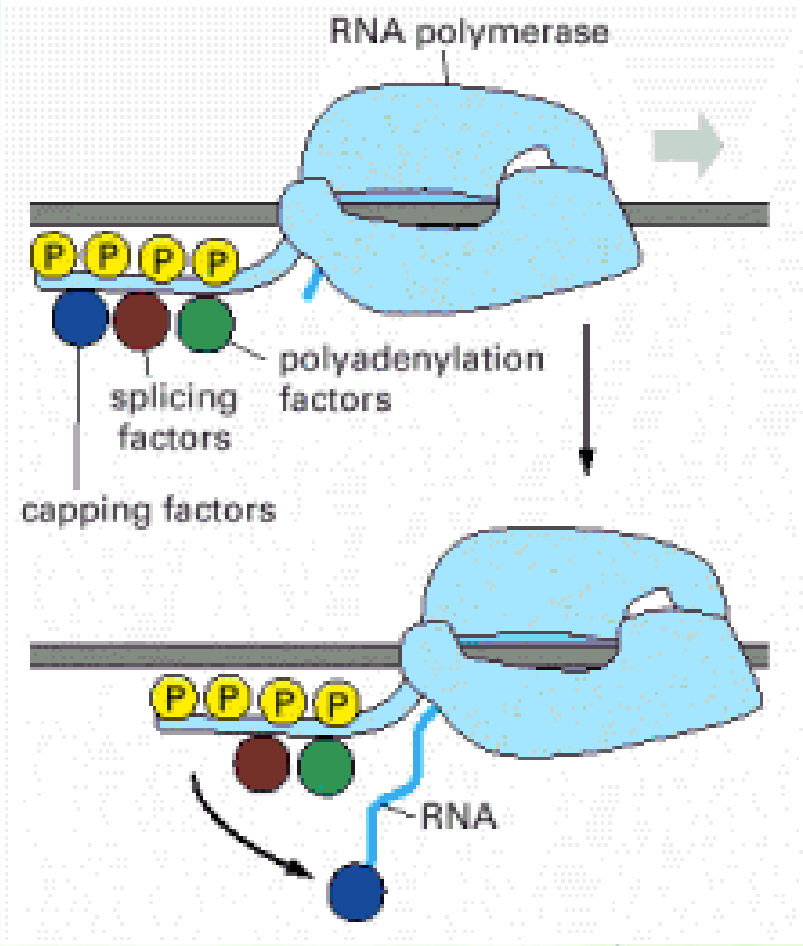
- The transcripts are spliced in different ways to produce different mRNAs and different proteins (known as protein isoforms, which are highly related gene products that perform essentially the same biological function).



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

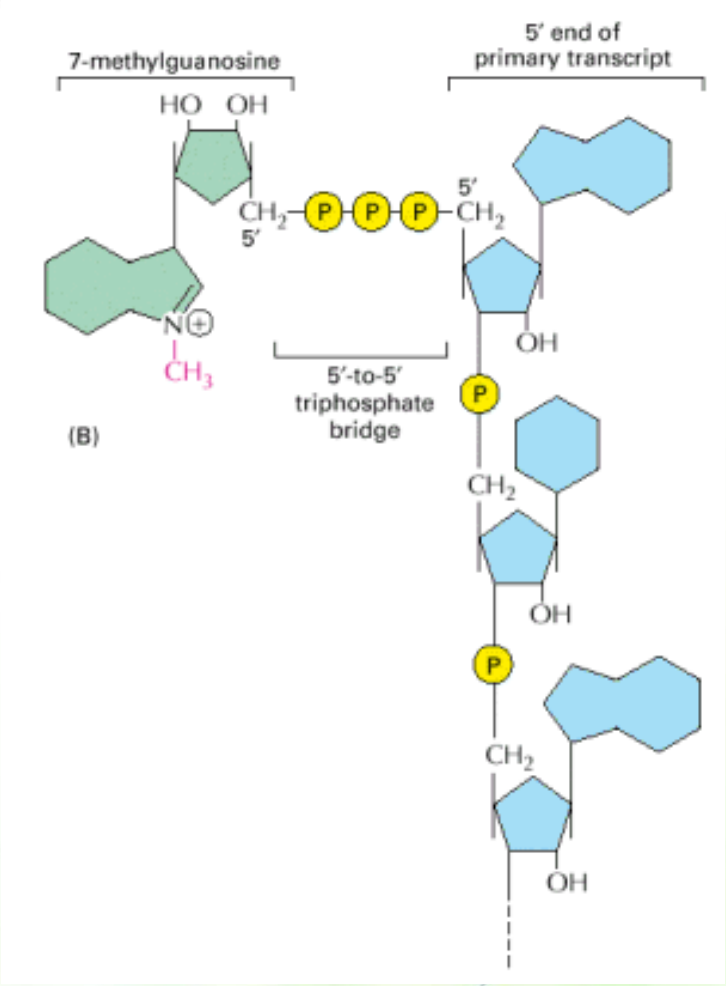
Processing of mRNA in eukaryotes

- 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 appear.



Addition of a cap

- As soon as RNA polymerase II has produced about 25 nucleotides of RNA, 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’.

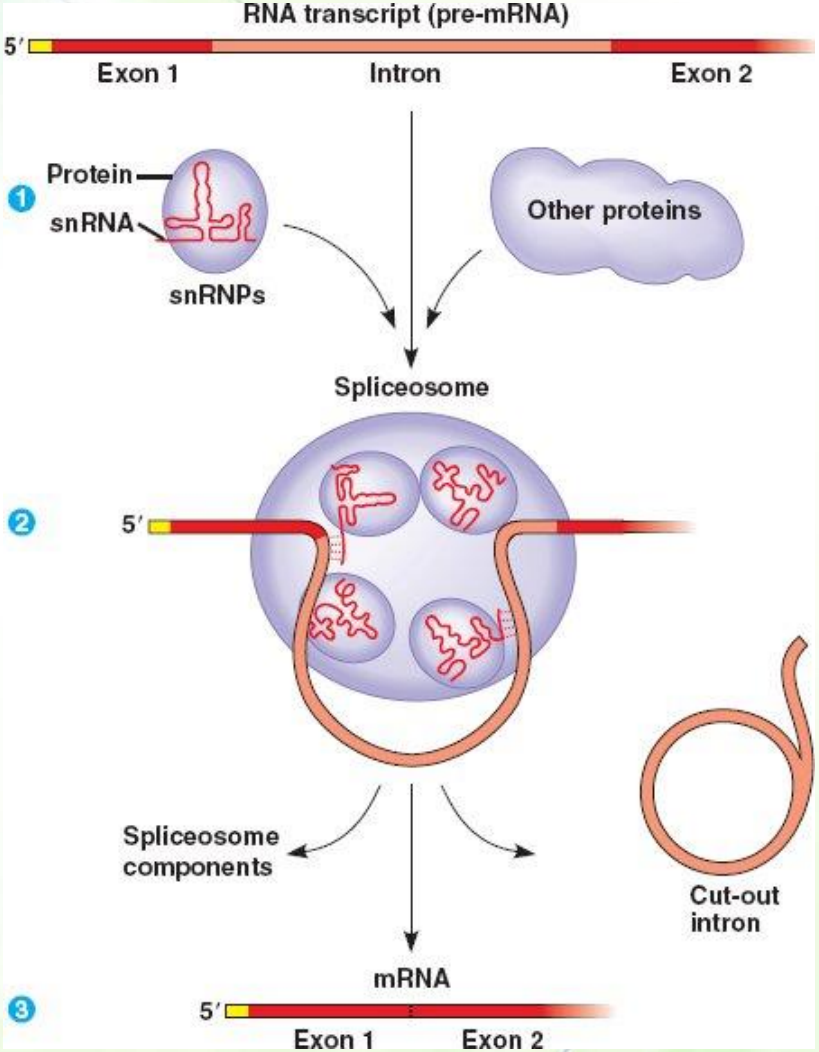


Importance of capping

- The cap 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.
- In the nucleus, the cap binds a protein complex called CBC (cap-binding complex), which helps the RNA to be exported into the cytoplasm.
- The 5’-methyl cap also has an important role in the translation of mRNAs to proteins.

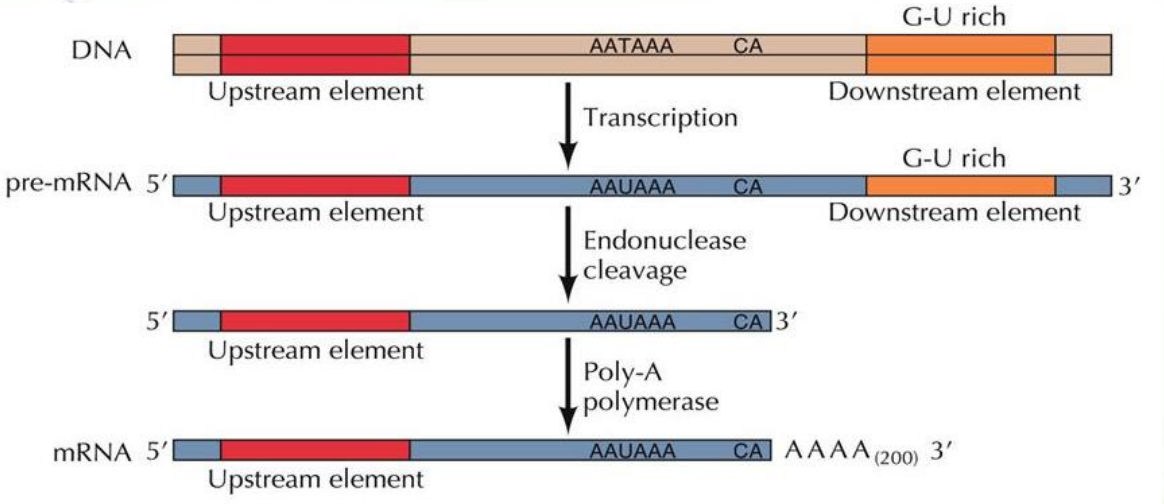
RNA splicing

- RNA splicing is catalyzed by spliceosomes.
  - They are composed of small nuclear ribonucleoprotein (snRNP), which are a complex of proteins and small nuclear RNAs (snRNAs) .
  - The catalytic site itself is largely formed by RNA molecules instead of proteins.
- Splicing is accurate.



Polyadenylation

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



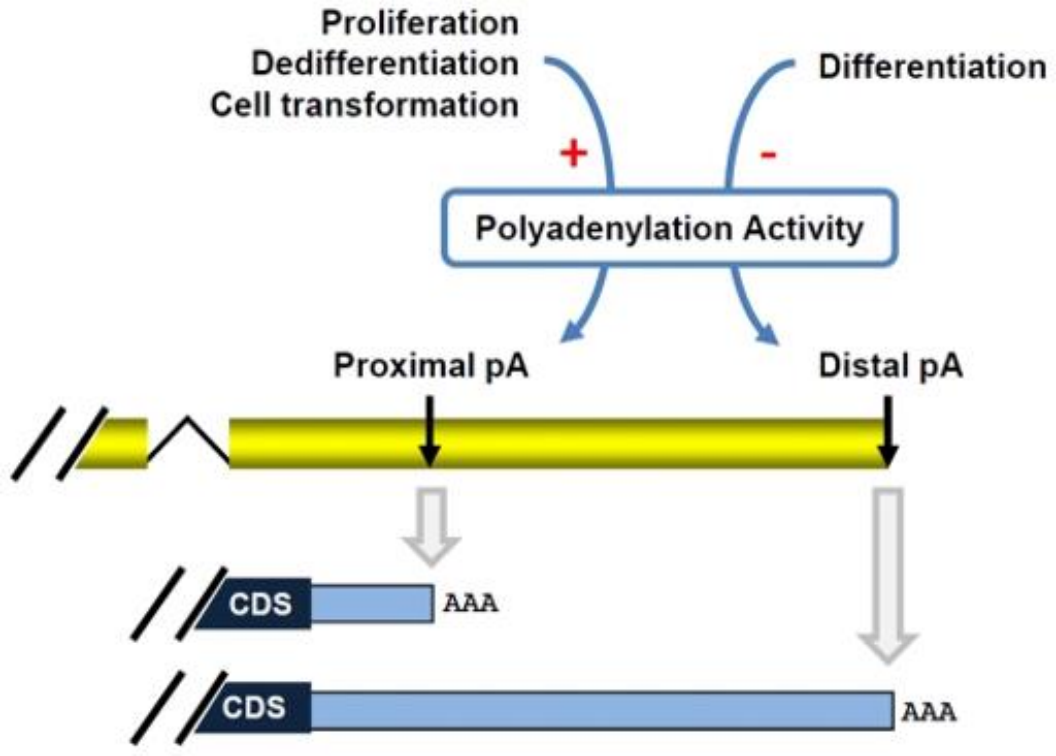
Poly-A polymerase does not require a template and the poly-A tail is not encoded in the genome.

Significance of polyadenylation

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

Alternative polyadenylation

- Some protein-coding genes have more than one polyadenylation site, producing mRNAs with different lengths of a noncoding sequence at the 3'-end called the 3'-untranslated region (3'-UTR).



Alternative polyadenylation

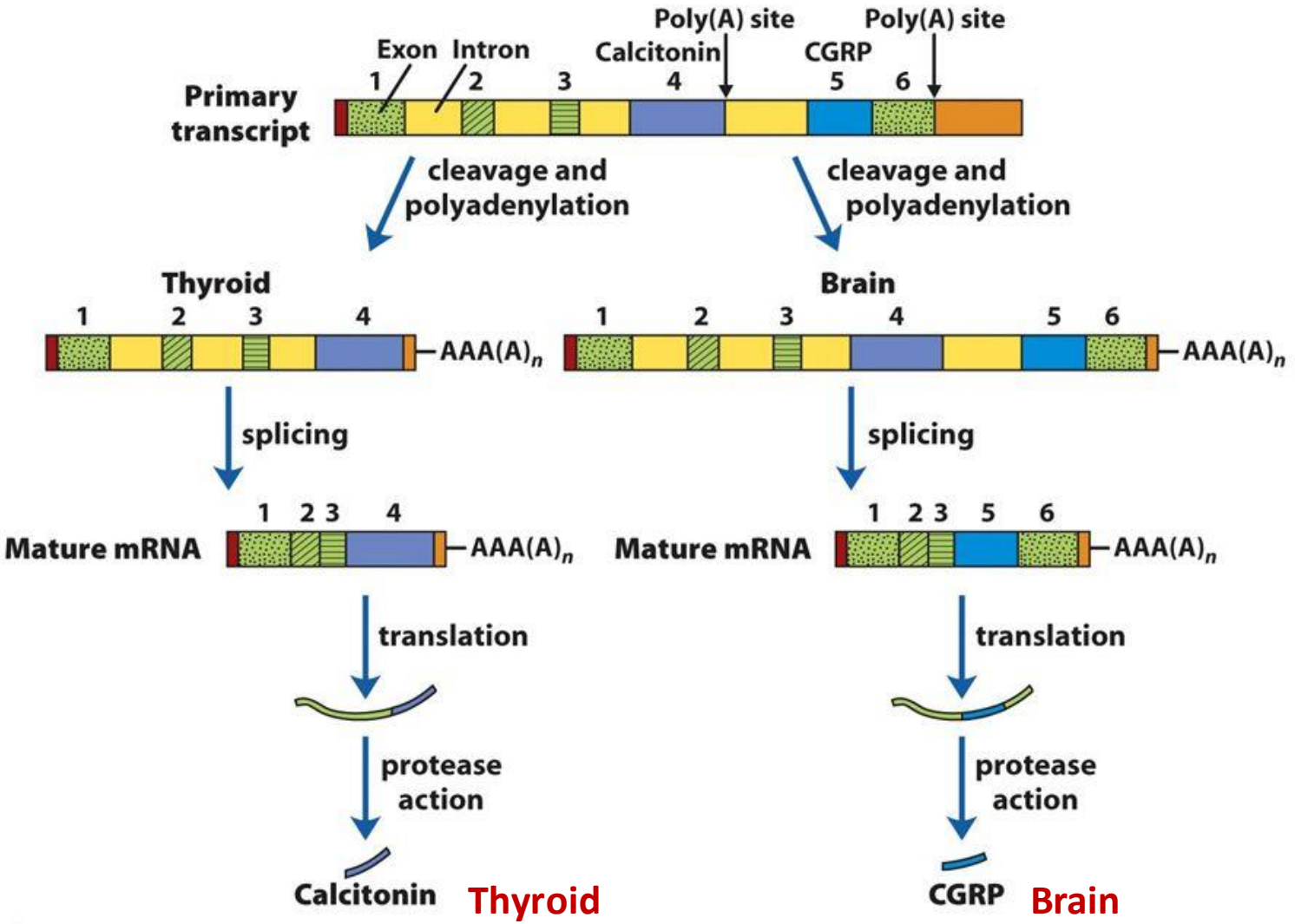
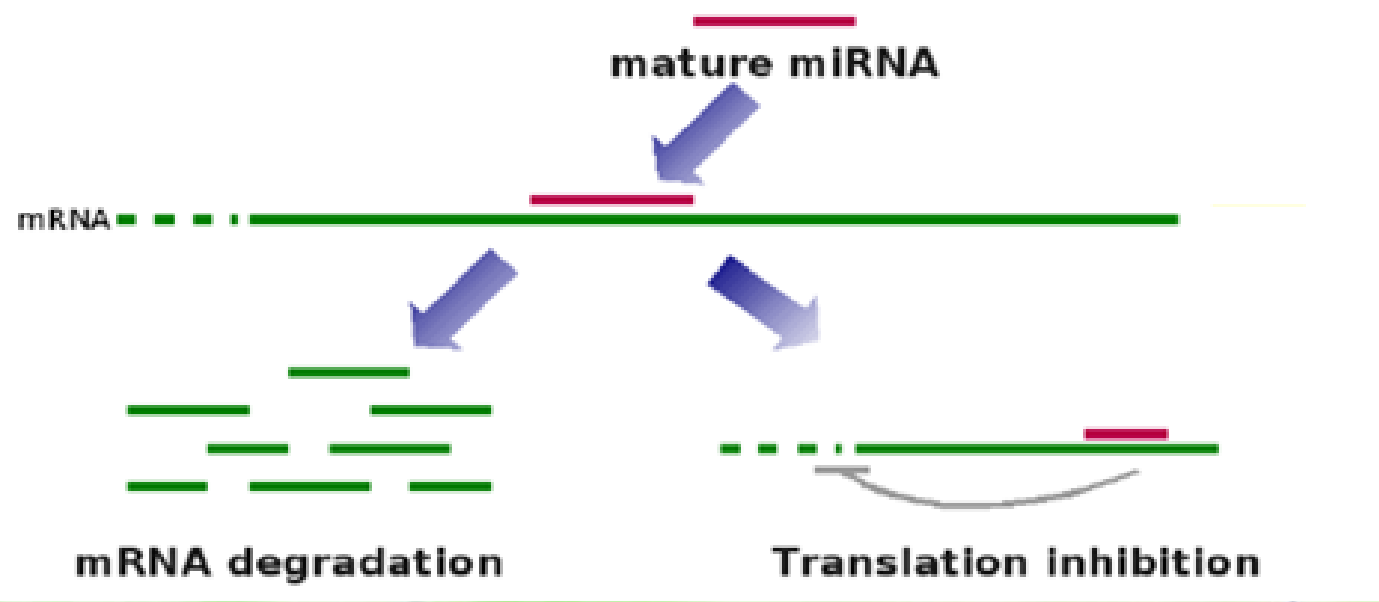


Figure 26-21
Lehninger Principles of Biochemistry, Fifth Edition

What are microRNA molecules?

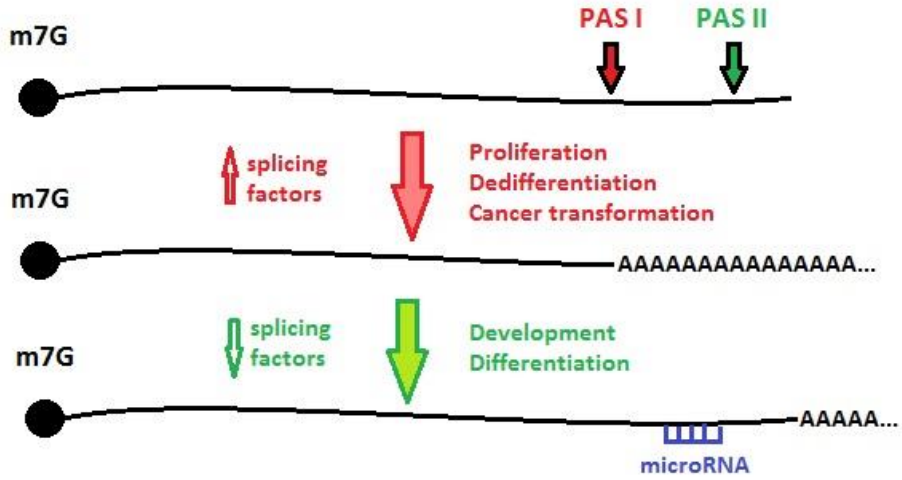
- They are RNA molecules that are transcribed, but not translated.
- They bind to complementary sequences of mRNA causing their degradation or blockage of their use in translation (protein synthesis).



Alternative polyadenylation

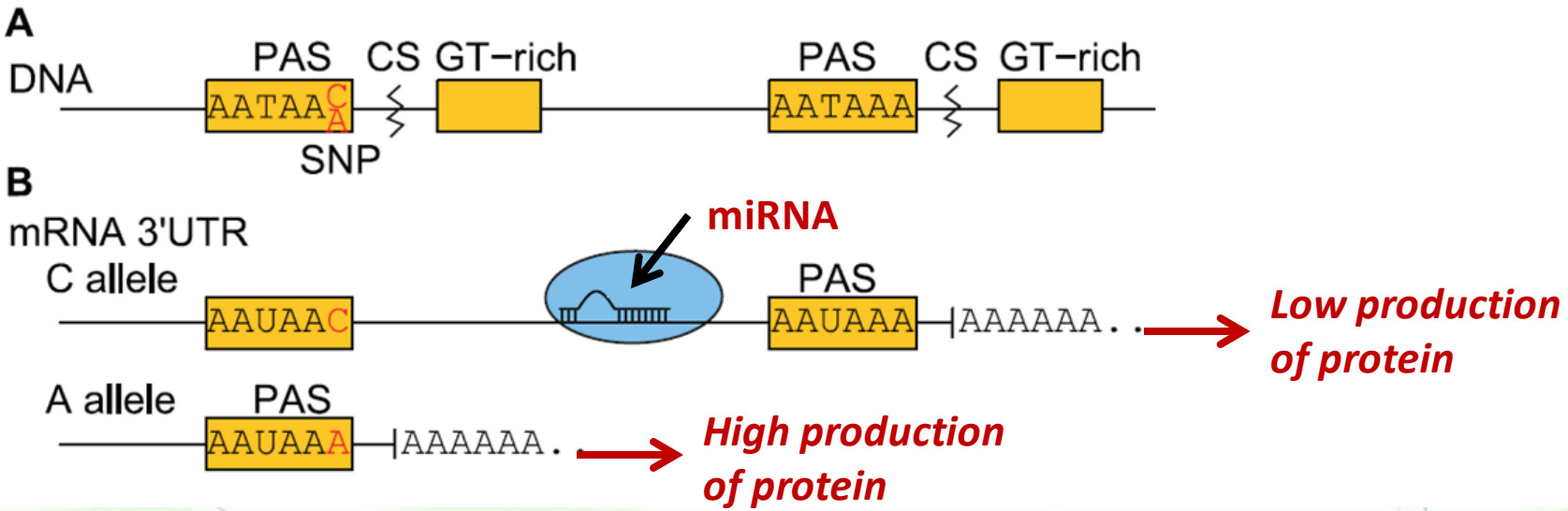
- The choice of poly(A) site can be influenced by extracellular stimuli that regulate the expression of the proteins that take part in polyadenylation.
- Having a shorter transcripts would remove regulatory elements in the 3‘-UTR and influence the half-lives of mRNA and, hence the amount of generated proteins.

- Example: Longer 3’-UTR would contain binding sites for microRNAs at the 3‘-UTR, which tend to repress translation and promote degradation of the mRNAs.



SNPs and alternative polyadenylation

The presence of SNPs within the polyadenylation signal can also alter the length of the mRNA and , hence, protein amount in cells.



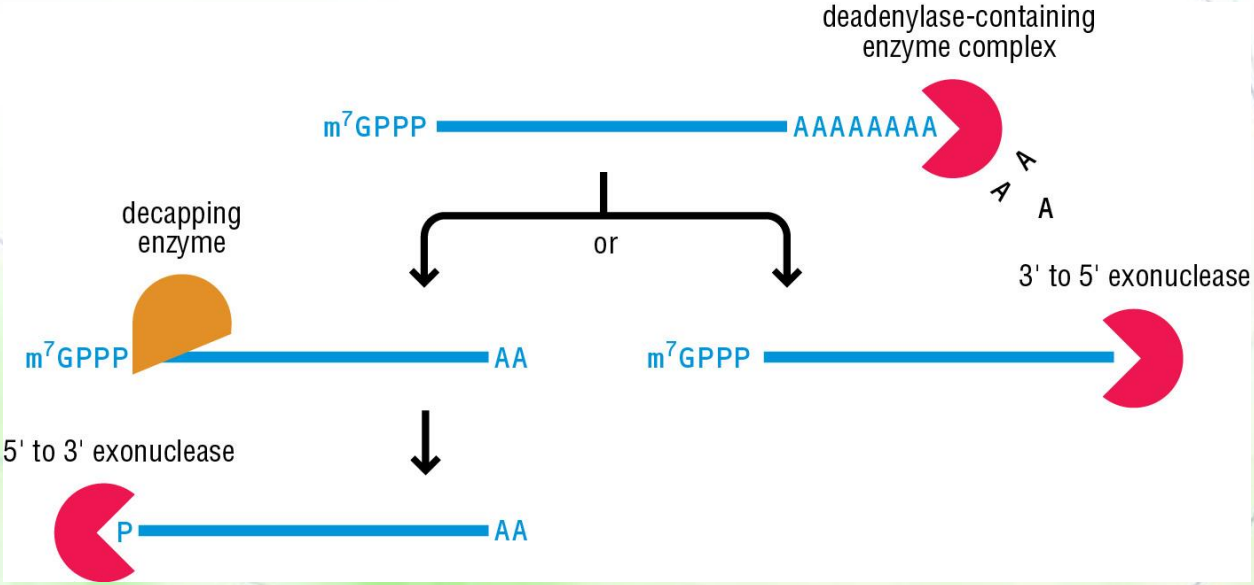
mRNA transport

- 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

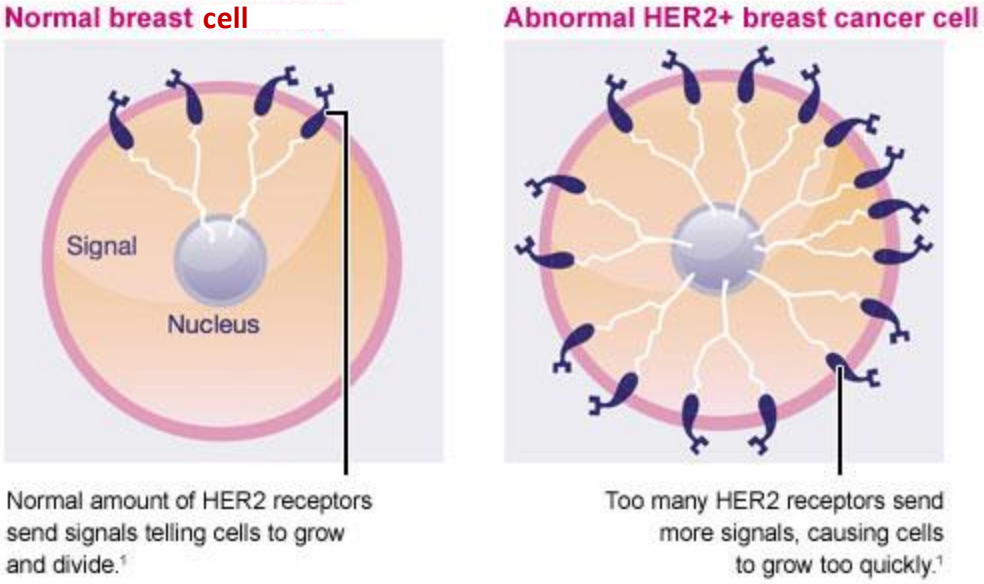
- The vast majority of mRNAs in a bacterial cell are very unstable, having a half-life of about 3 minutes.
- The mRNAs in eukaryotic cells are more stable (up to 10 hours; average of 30 minutes).
- Degradation is ainitiated by shortening of poly-A tail. This is followed by action of exonucleases or followed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

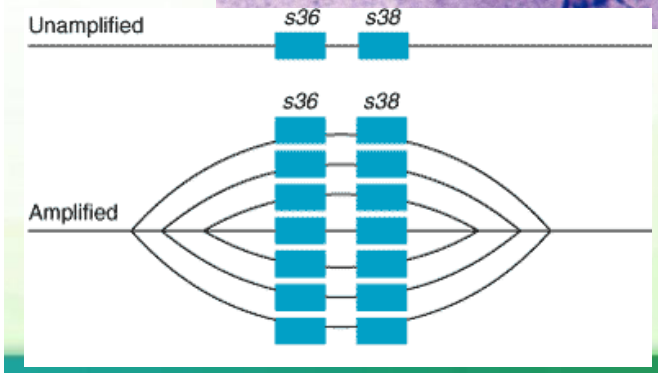


Some phenomena in eukaryotes

Gene amplification

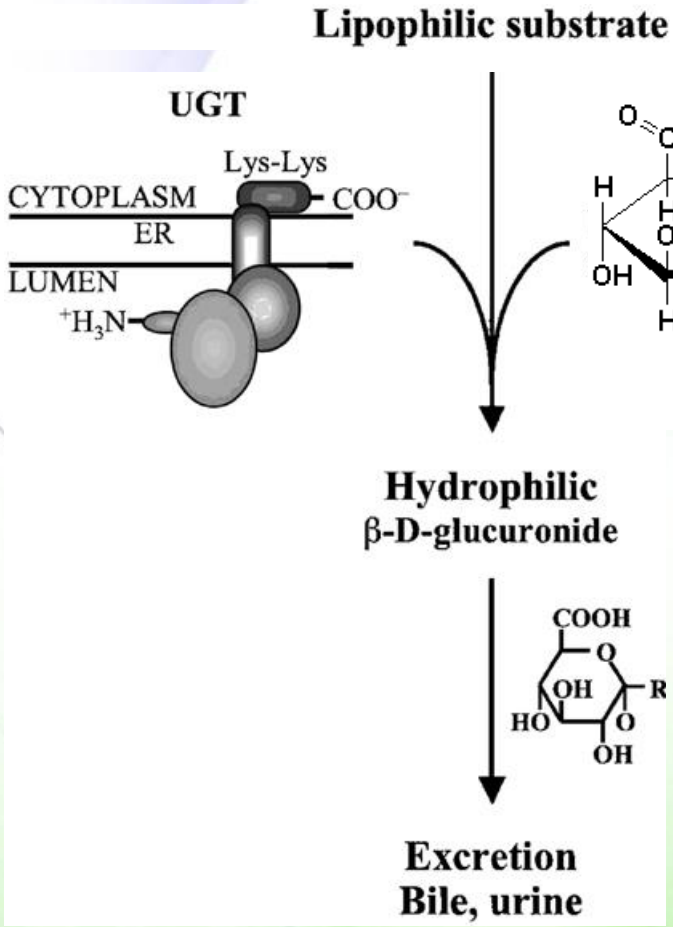
- It is an increase in copy number of a restricted region of a chromosome increasing the is the quantity of DNA in these regions.
- It is a mechanism that cancer cells use to escape resistance from methotrexate whereby the target gene, dihydrofolate reductase, is amplified.
- It is also a mechanism by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

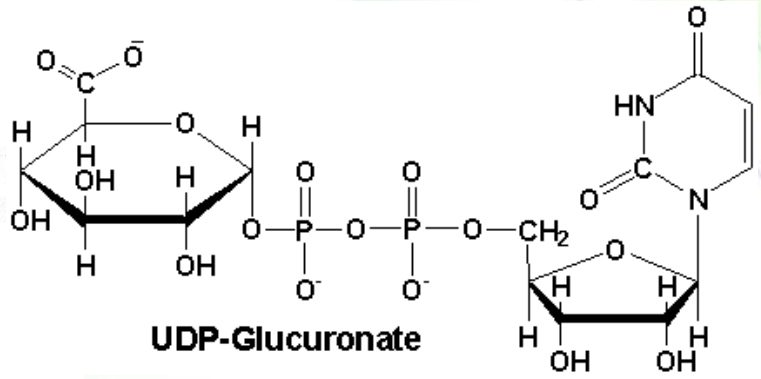




An example of alternative splicing:

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.

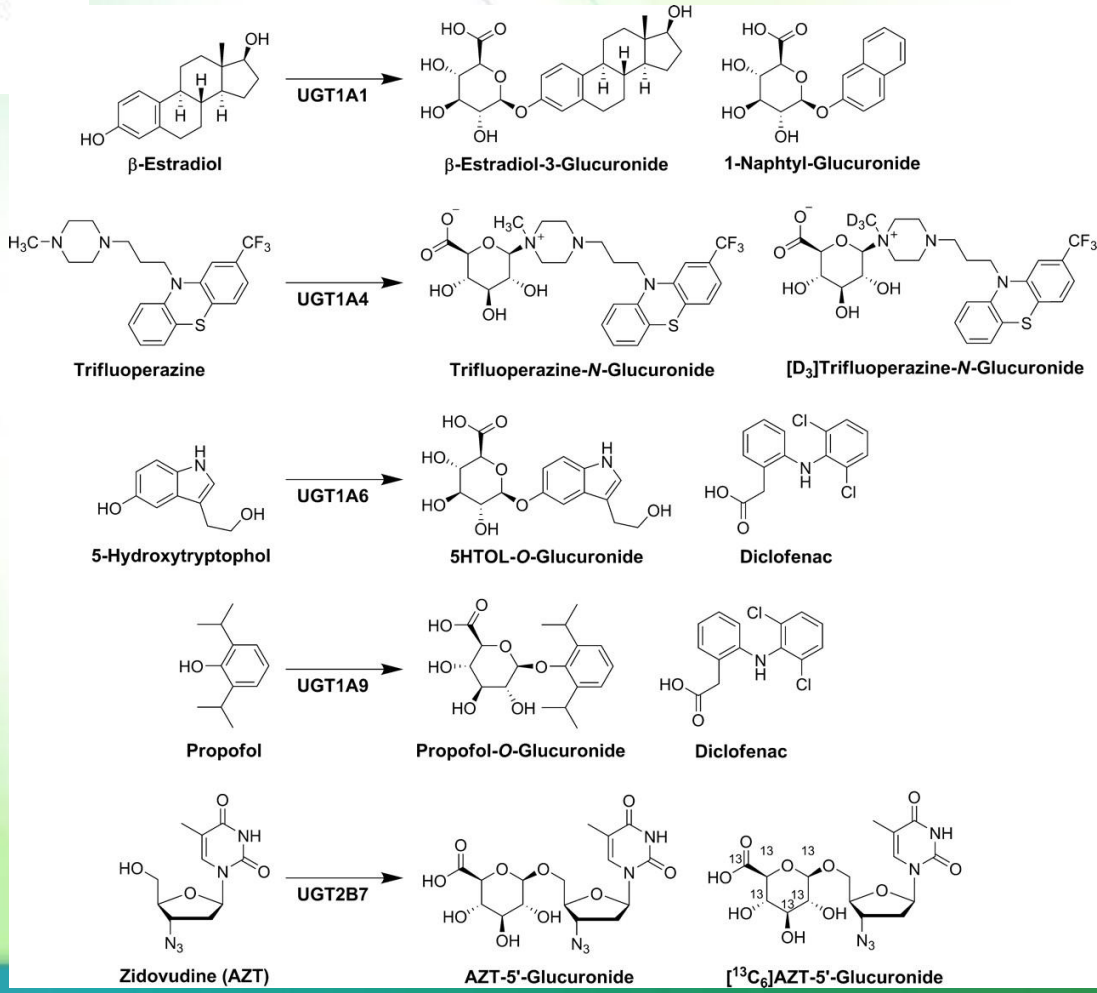
It has many substrates with different structures

Lipophilic substrate

Therapeutic drugs	Biliary acids
Carcinogens	Steroïds
Environmental toxicants	Retinoic acids
Dietary constituents	Fatty acids
Bilirubin	

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

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



and reactions are catalyzed in different tissues

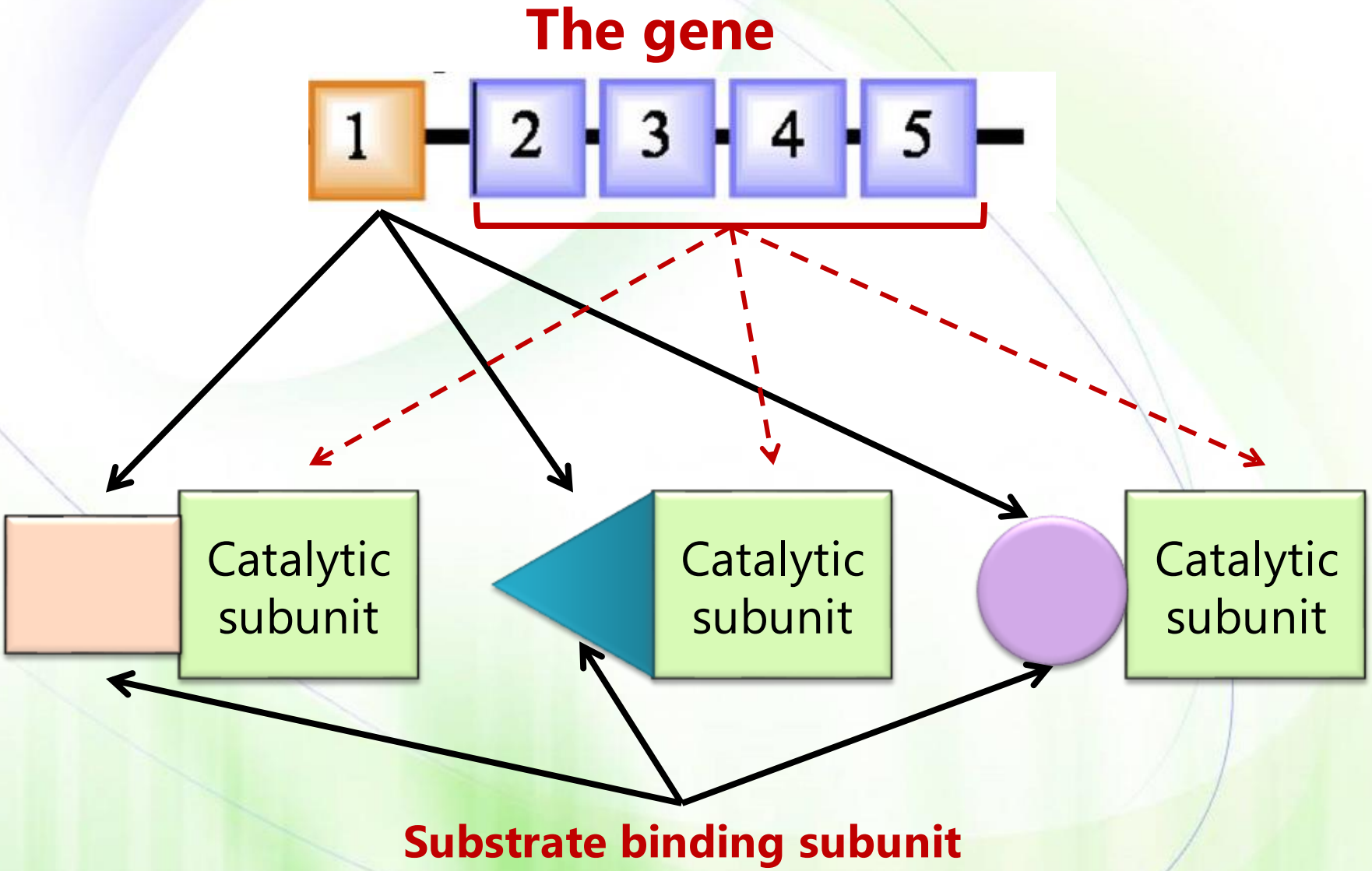
Substrates	Place of reactio
Etoposide	Biliary tissue, colon, intestine, liver, stomach
Genistein	Biliary tissue, colon, liver, stomach
Tamoxifen	Biliary tissue, colon, intestine, liver
PCBs	Biliary tissue, brain, colon, kidney, larynx, liver, lung, stomach
Heterocyclic amines	Esophagus, intestine, kidney, 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

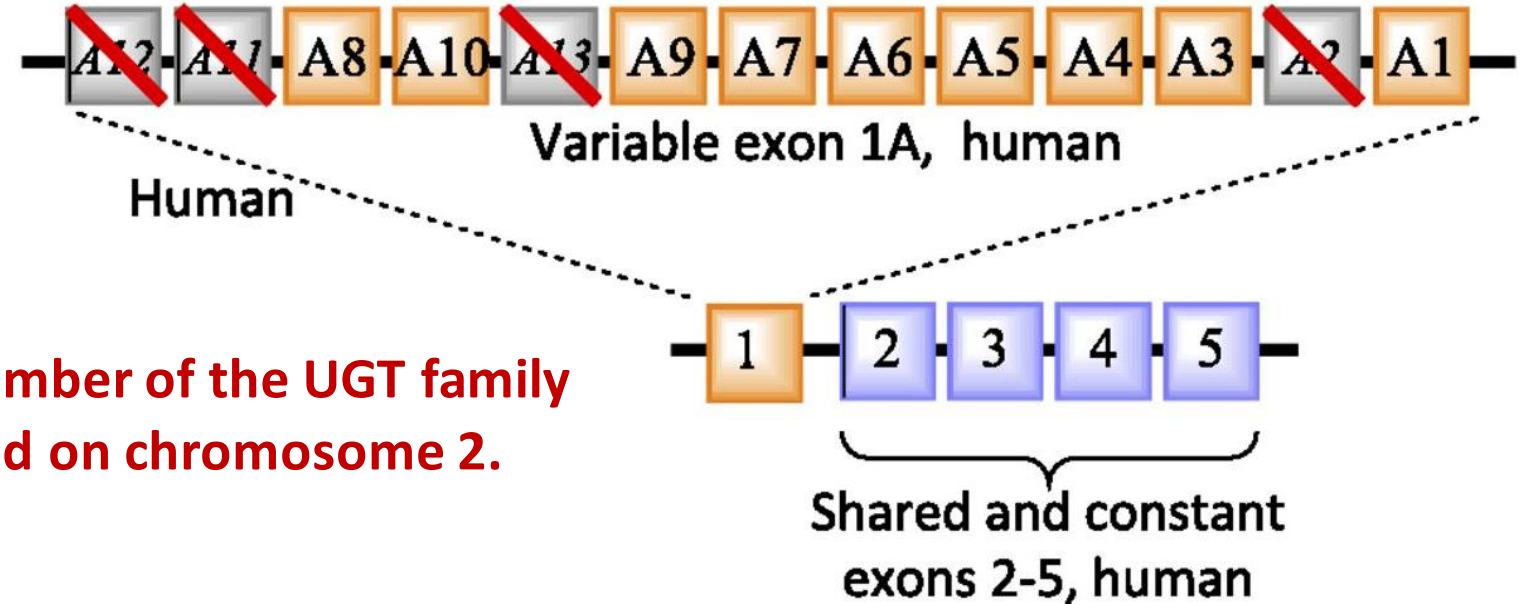
One head, many hats

Then this…


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

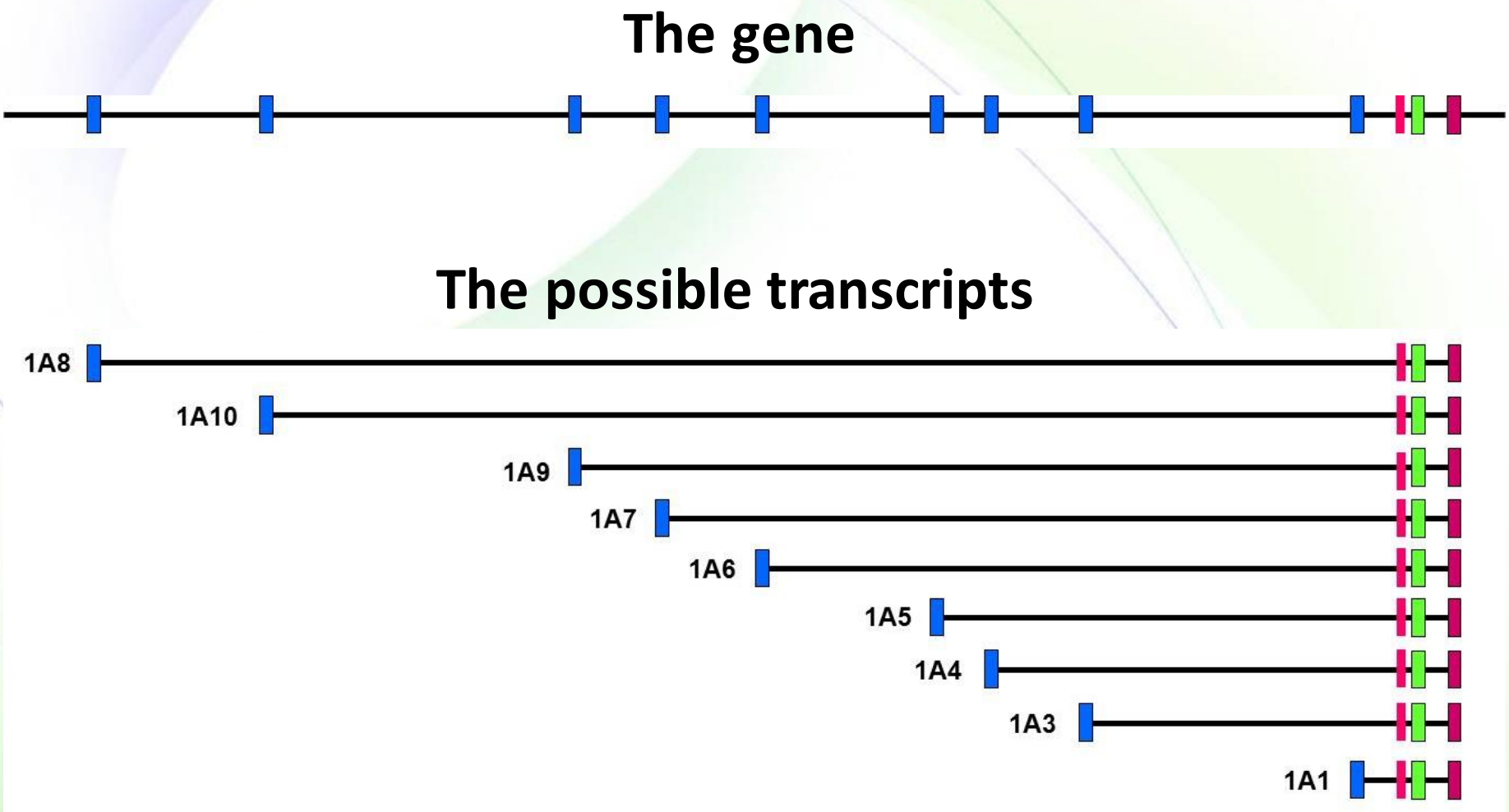
How does it do this?

- 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

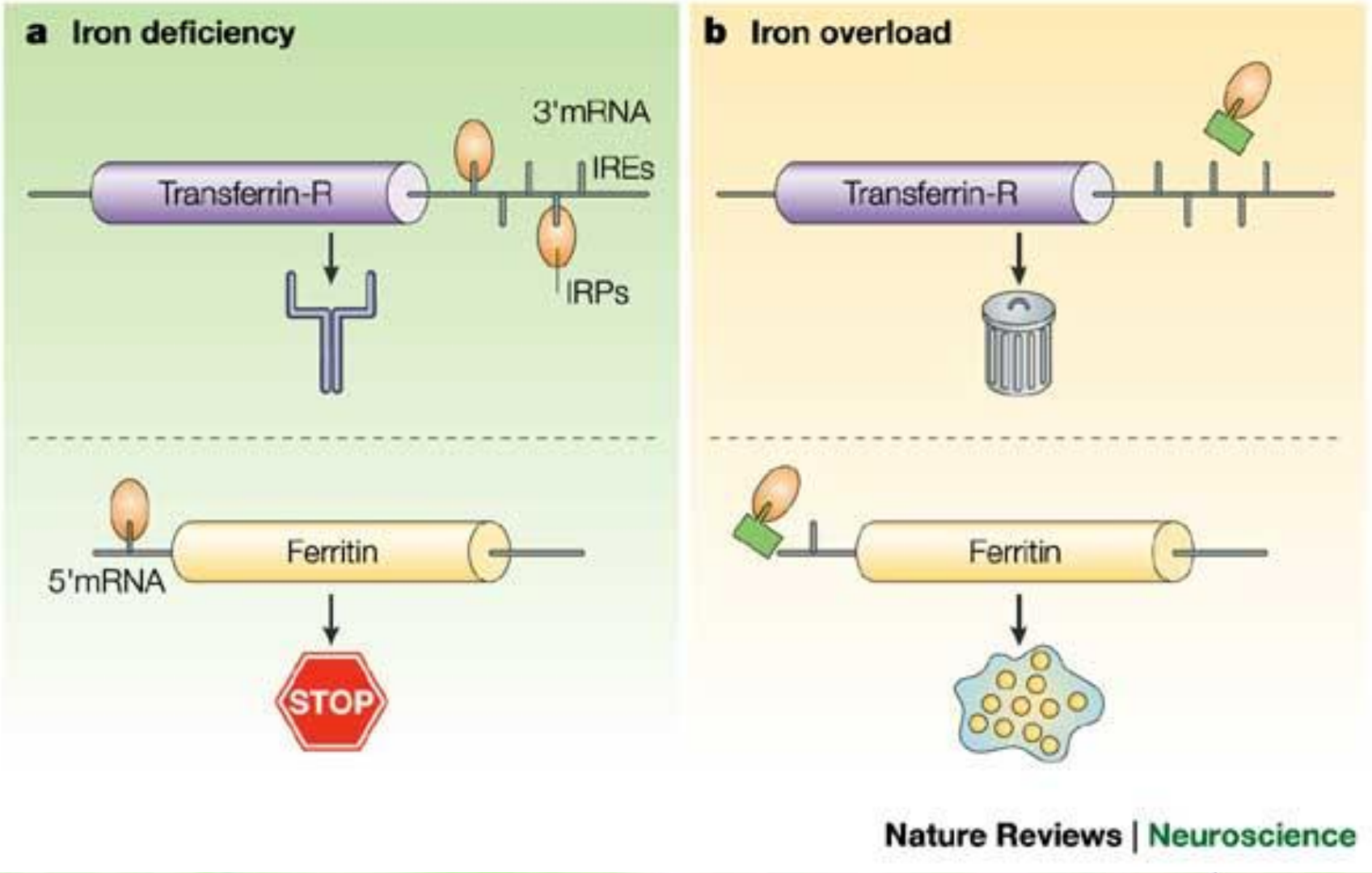
Regulation of mRNA stability

Iron-responsive elements

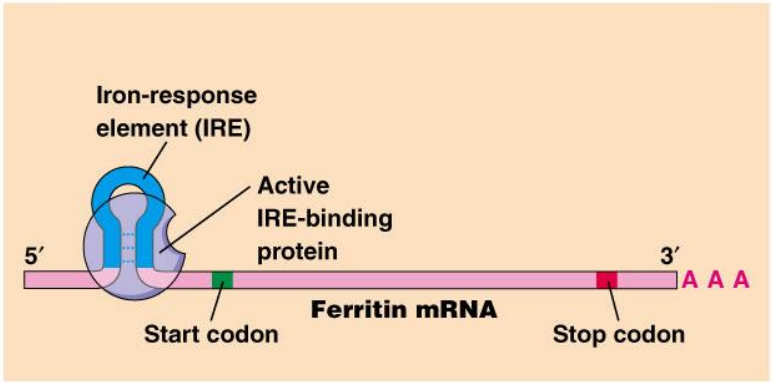
- In human cells, there are regions of mRNA called iron responsive elements (IREs).
- These regions are contained within the mRNA sequences that code for certain proteins that regulate the levels of iron.
  - Ferritin, transferrin receptor, ferroportin, and DMT1
- Iron responsive element binding protein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

Effect on expression

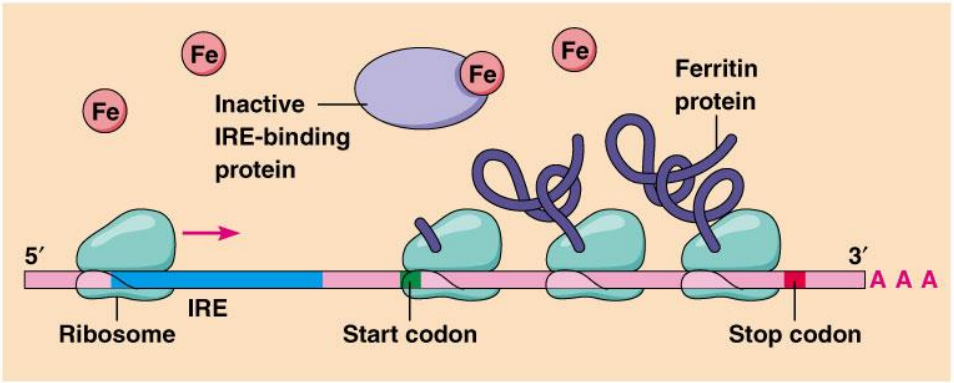
- When iron is abundant, it binds to IRE-BP, disabling the binding of IR-BP to ferritin mRNA
  - This prevents the degradation of the mRNA molecules allowing the production of more ferritin protein
  - Therefore, the iron itself causes the cell to produce more iron storage molecules
- On the other hand, 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


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

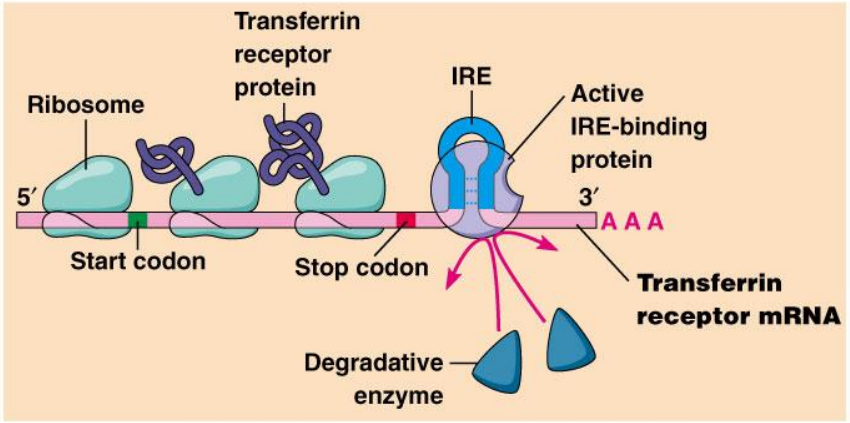
(a) Low iron concentration. IRE-binding protein binds to IRE, so translation of ferritin mRNA is inhibited.



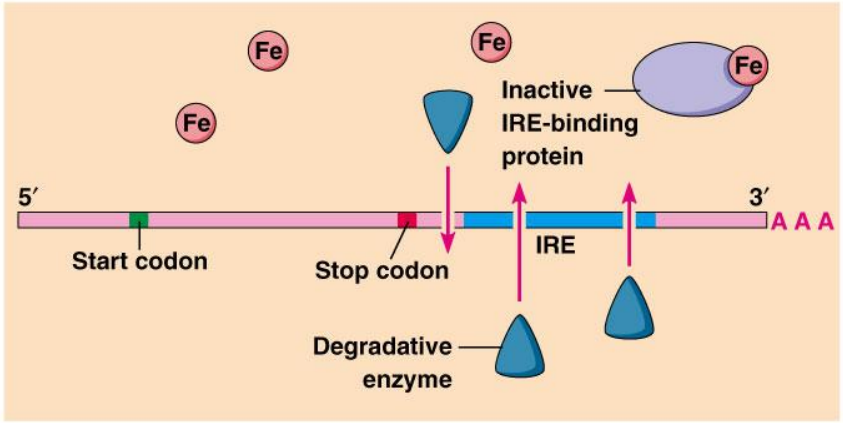
(b) High iron concentration. IRE-binding protein cannot bind to IRE, so translation of ferritin mRNA proceeds.



(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.

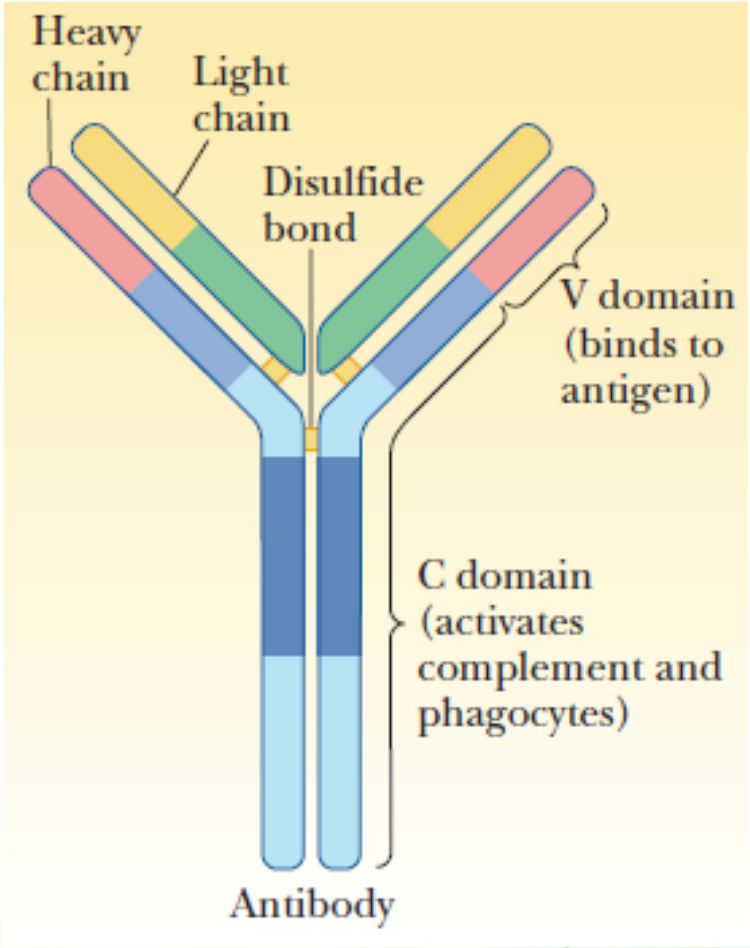


(b) High iron concentration. IRE-binding protein cannot bind to IRE, so mRNA is degraded and synthesis of transferrin receptor is thereby inhibited.



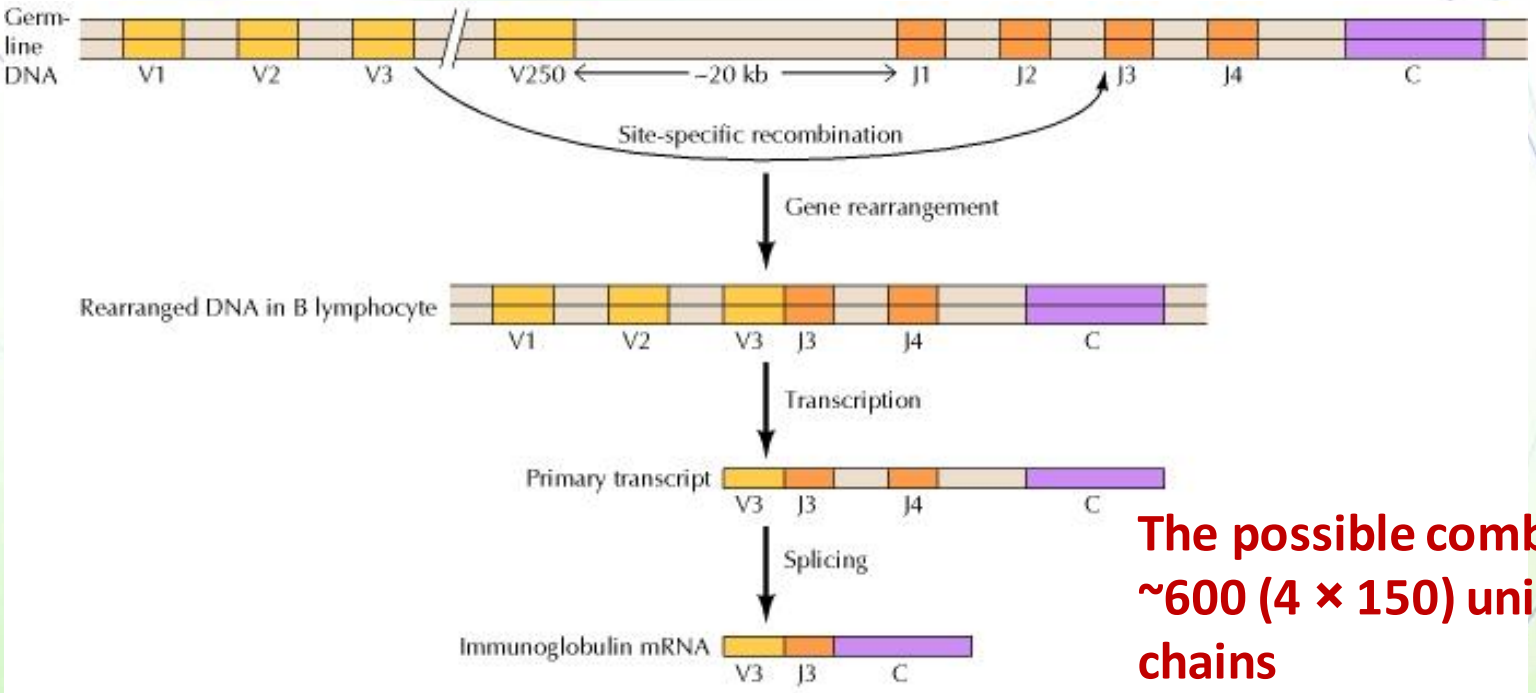
Structure of antibodies (immunoglobulins)

- Antibodies consist of two identical heavy chains and two identical light chains held together by disulfide bonds.
- Both contain constant and variable regions.
- The variable regions are responsible for recognition of antigens.
- immune system has the ability to produce about 10^{10}-10^{11} different antibodies.
- **How is diversity generated?**



Gene rearrangement of the light chain

- There are two types of immunoglobulin light chains: κ and λ
- Each is a product of at least 3 genes:
  - Variable (VL) gene; 150 genes
  - Joining region (J) gene: 4 genes

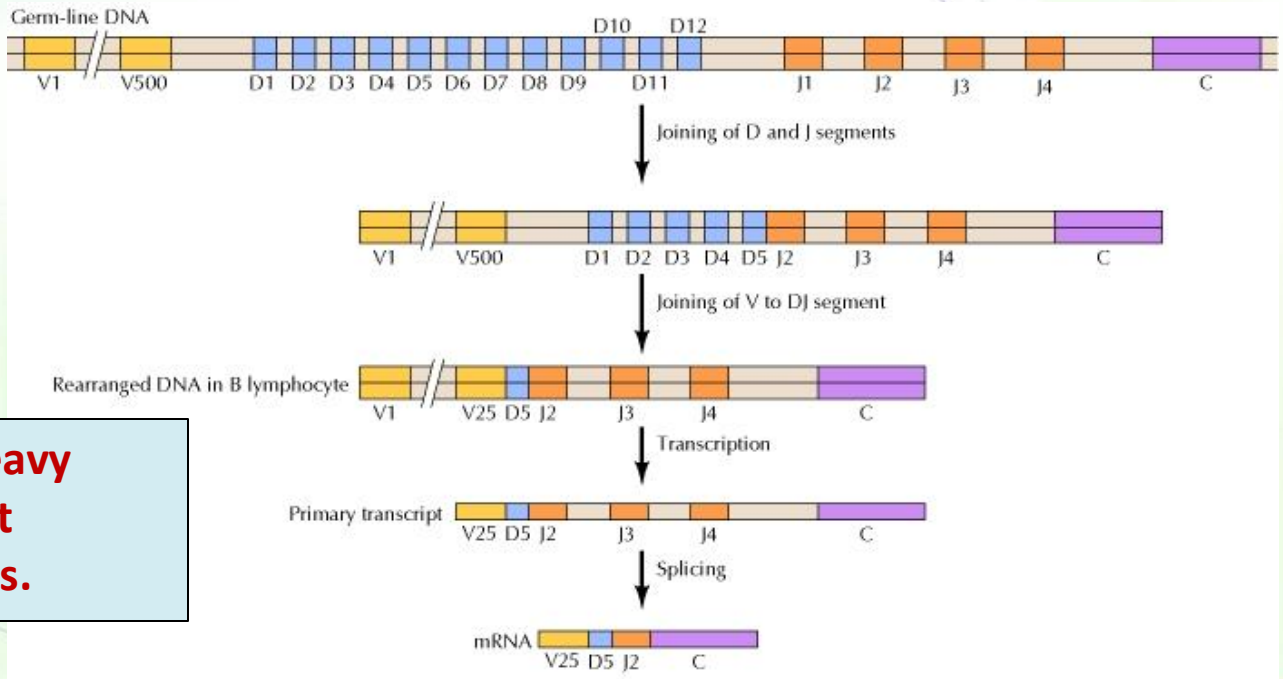


The possible combinations is ~600 (4 × 150) unique light chains

Gene rearrangement of the heavy chain

- Heavy chain is a product of at least 4 genes :
  - Variable region (VH) gene: 150 genes
  - Diversity region (D) gene: 12 genes
  - Joining region (J) gene: 4 genes

The possible combinations is ~7200 (500 x 12 x 4) unique heavy chains

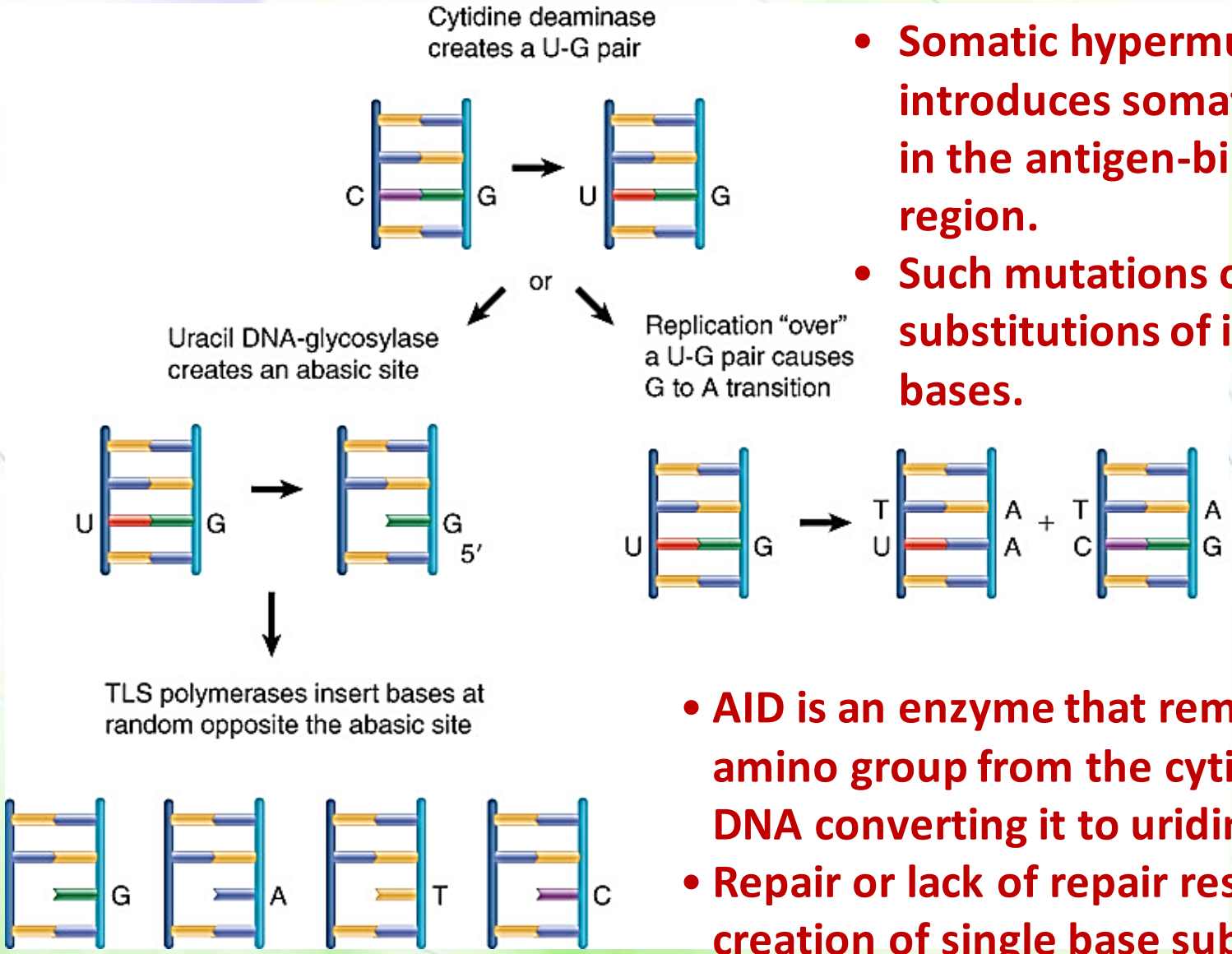


600 light chains x 7,200 heavy chains = ~ 4 × 10^6 different immunoglobulin molecules.

Additional mechanisms

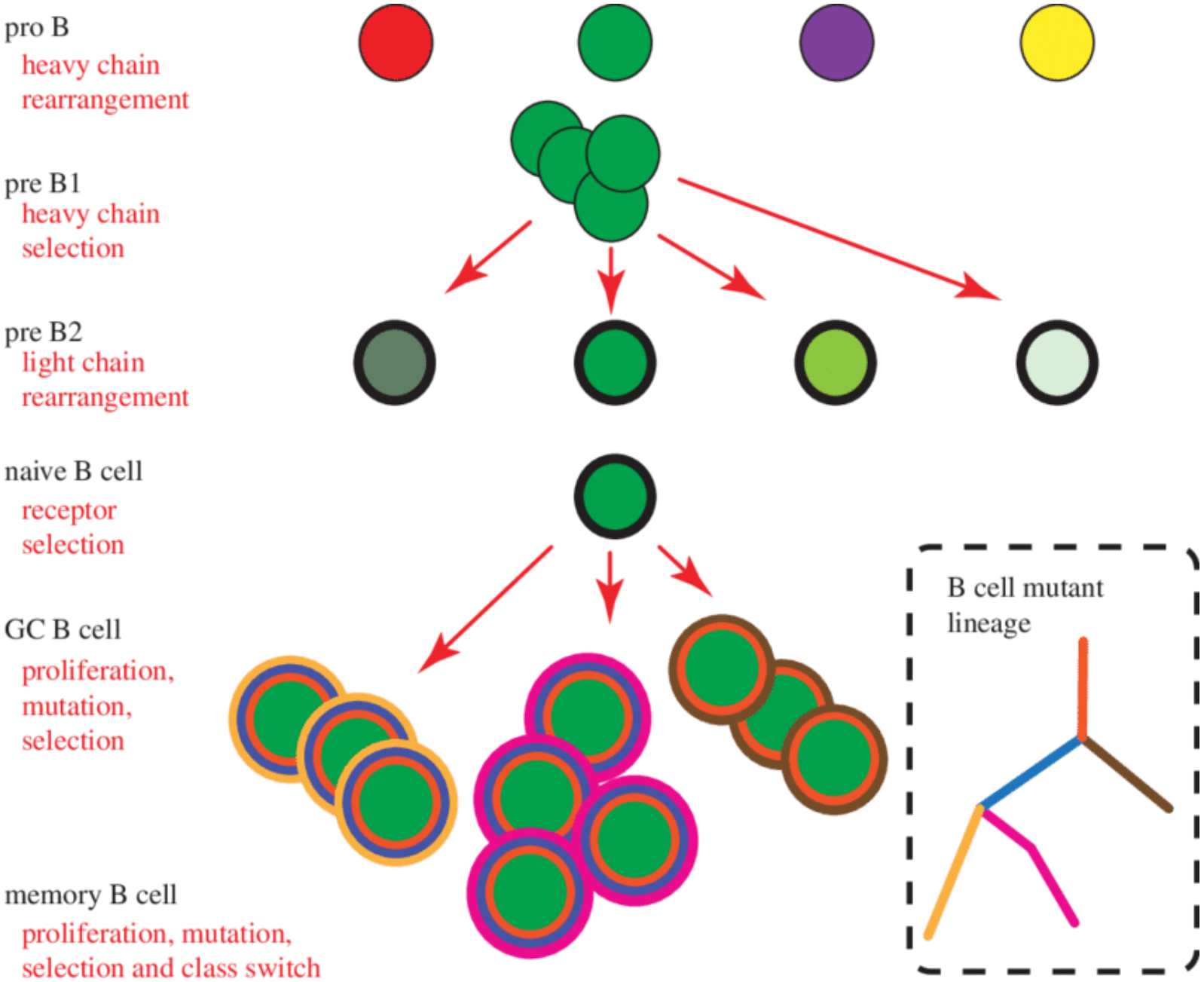
- The joining of immunoglobulin gene segments is often imprecise, resulting in the formation of ~10^5 different light chains and ~10^6 heavy chains, which can then combine to form more than 10^{11} distinct antibodies.
- During recombination, nucleotides are added or deleted.
- Further antibody diversity is generated by a process known as somatic hypermutation, which results in the introduction of frequent mutations into the variable regions of both heavy-chain and light-chain genes.

Activation-induced deaminase



- **Somatic hypermutation introduces somatic mutations in the antigen-binding variable region.**
- **Such mutations occur mostly as substitutions of individual bases.**

- **AID is an enzyme that removes the amino group from the cytidine base in DNA converting it to uridine.**
- **Repair or lack of repair results in creation of single base substitutions.**


pro B
heavy chain
rearrangement
pre B1
heavy chain
selection
pre B2
light chain
rearrangement
naive B cell
receptor
selection
GC B cell
proliferation,
mutation,
selection
memory B cell
proliferation, mutation,
selection and class switch
B cell mutant
lineage