
# Transcription

# Resources

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

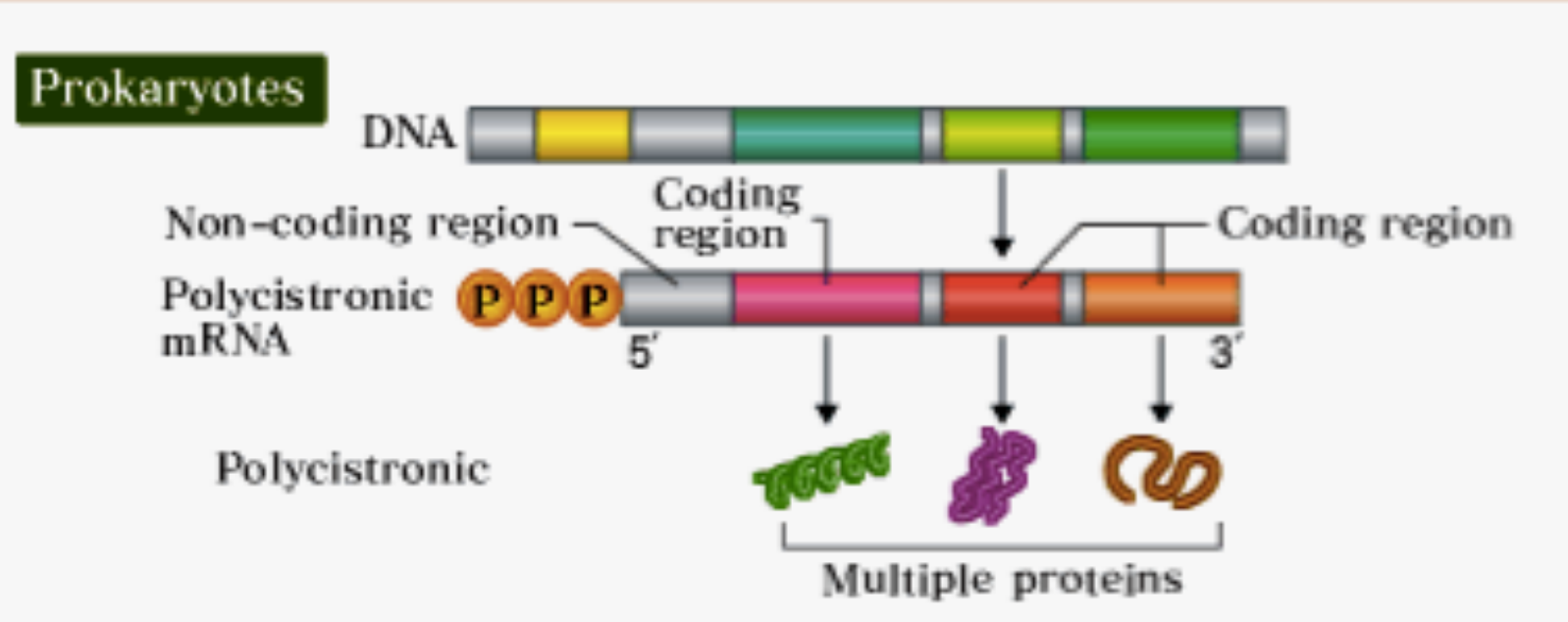
# Definition of a gene

- The entire nucleic acid sequence that is necessary for the synthesis of a functional polypeptide.

- A cistron: a genetic unit that encodes a polypeptide(s)
  - 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.

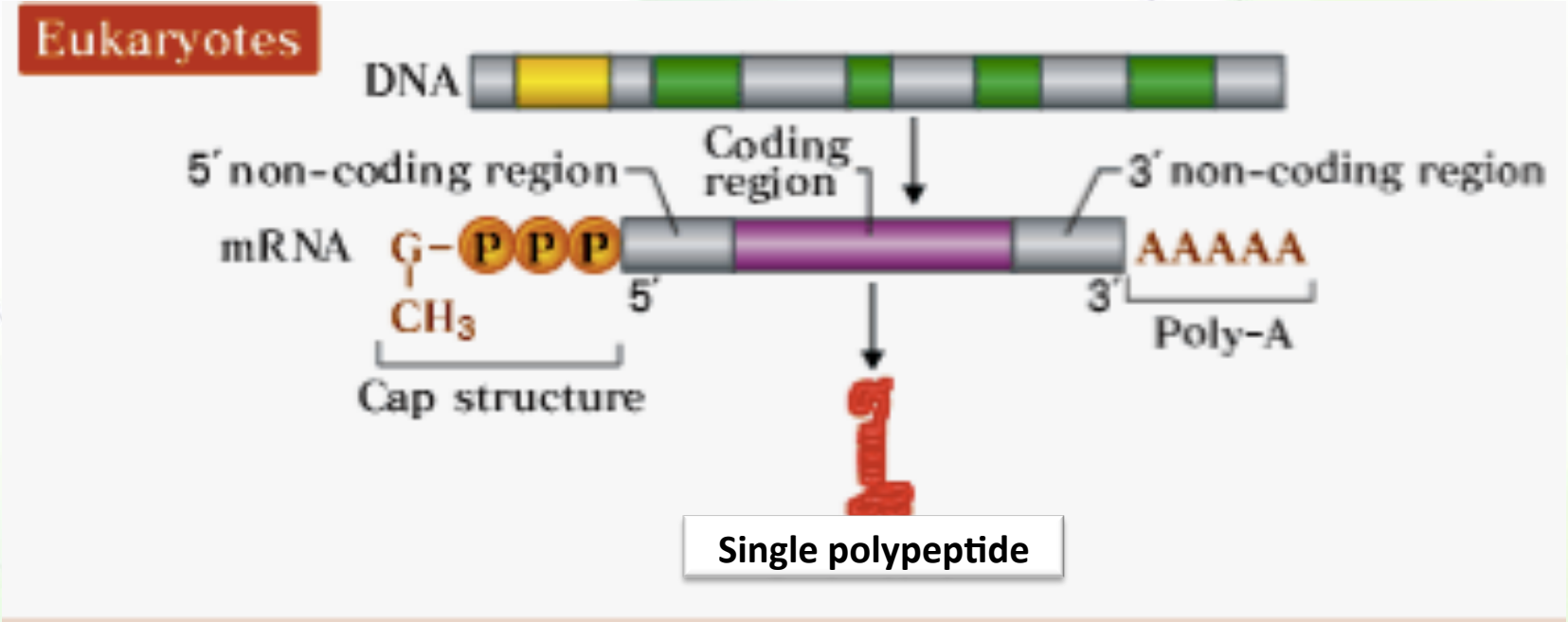
# Prokaryotic genes (operon)

- 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 the tryptophan are located in one contiguous stretch.
- This cluster of genes comprises a single transcriptional unit



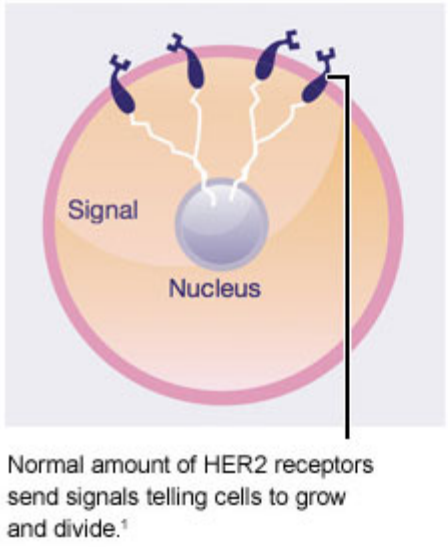
# Eukaryotic genes

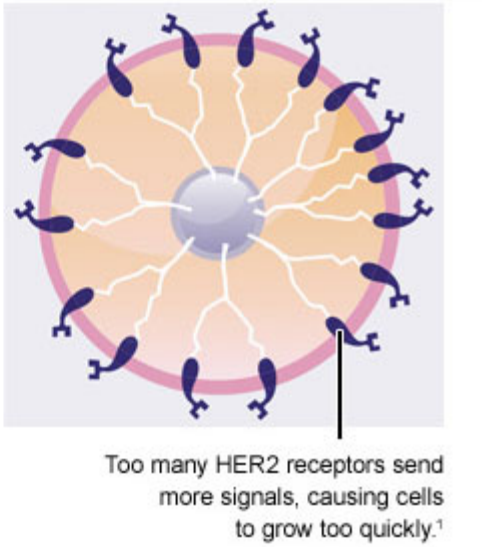
- Most eukaryotic transcription units produce mRNAs that encode only one protein, thus termed monocistronic.



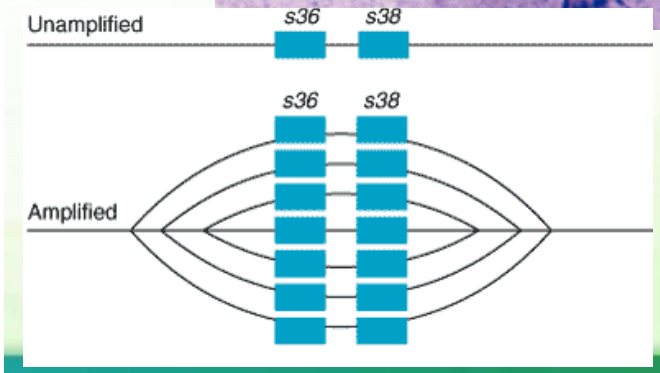
# Gene amplification

- 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.



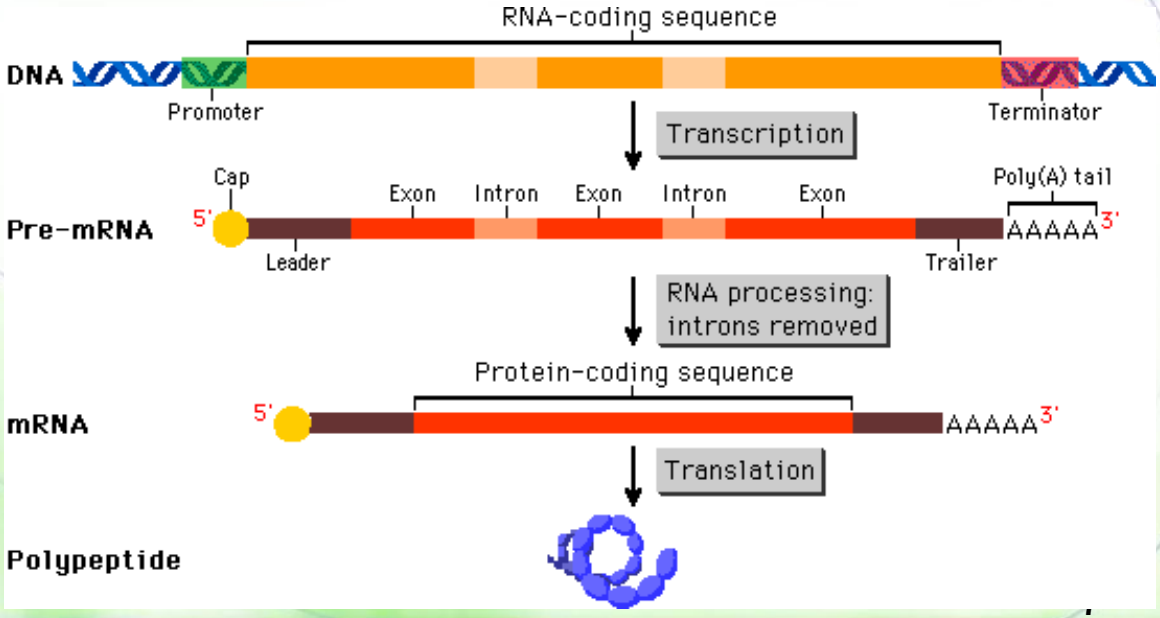







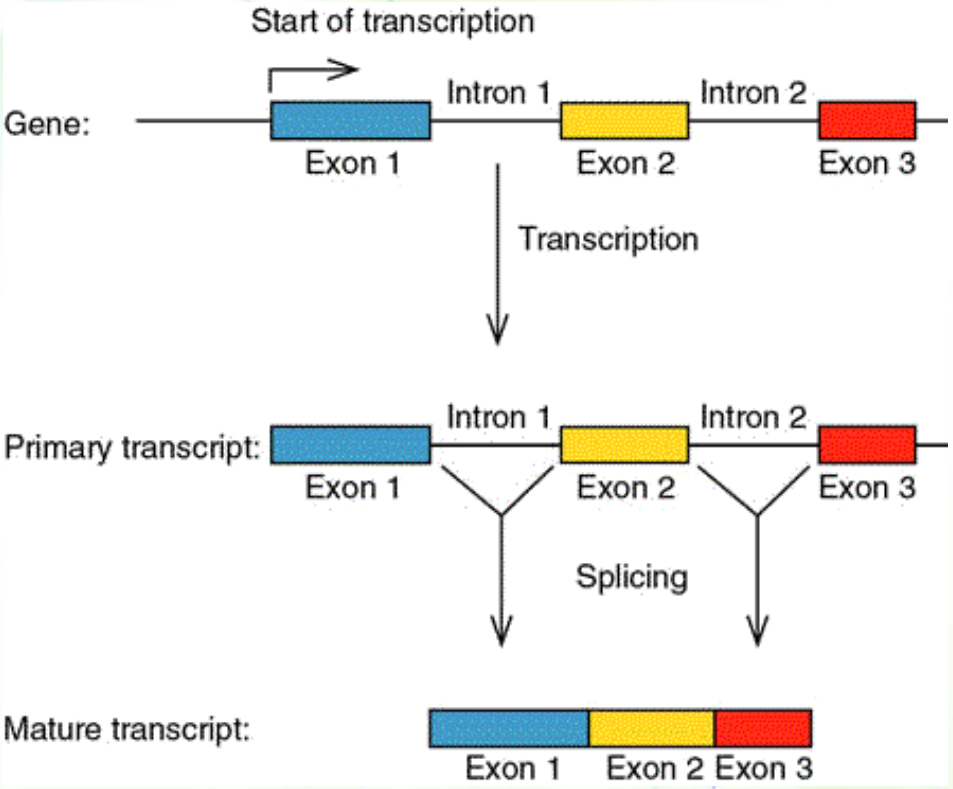
# Introns vs. exons

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



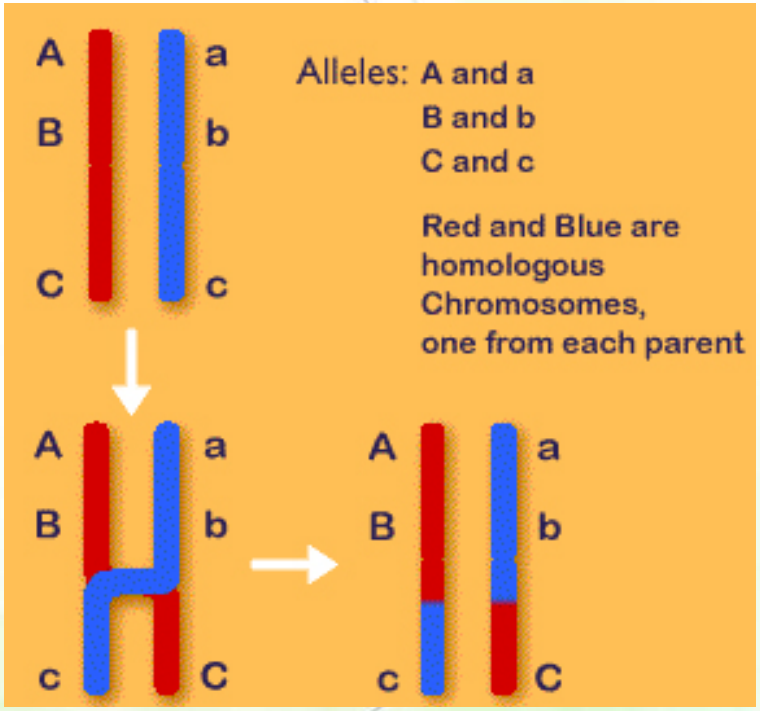
# RNA splicing

- 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).



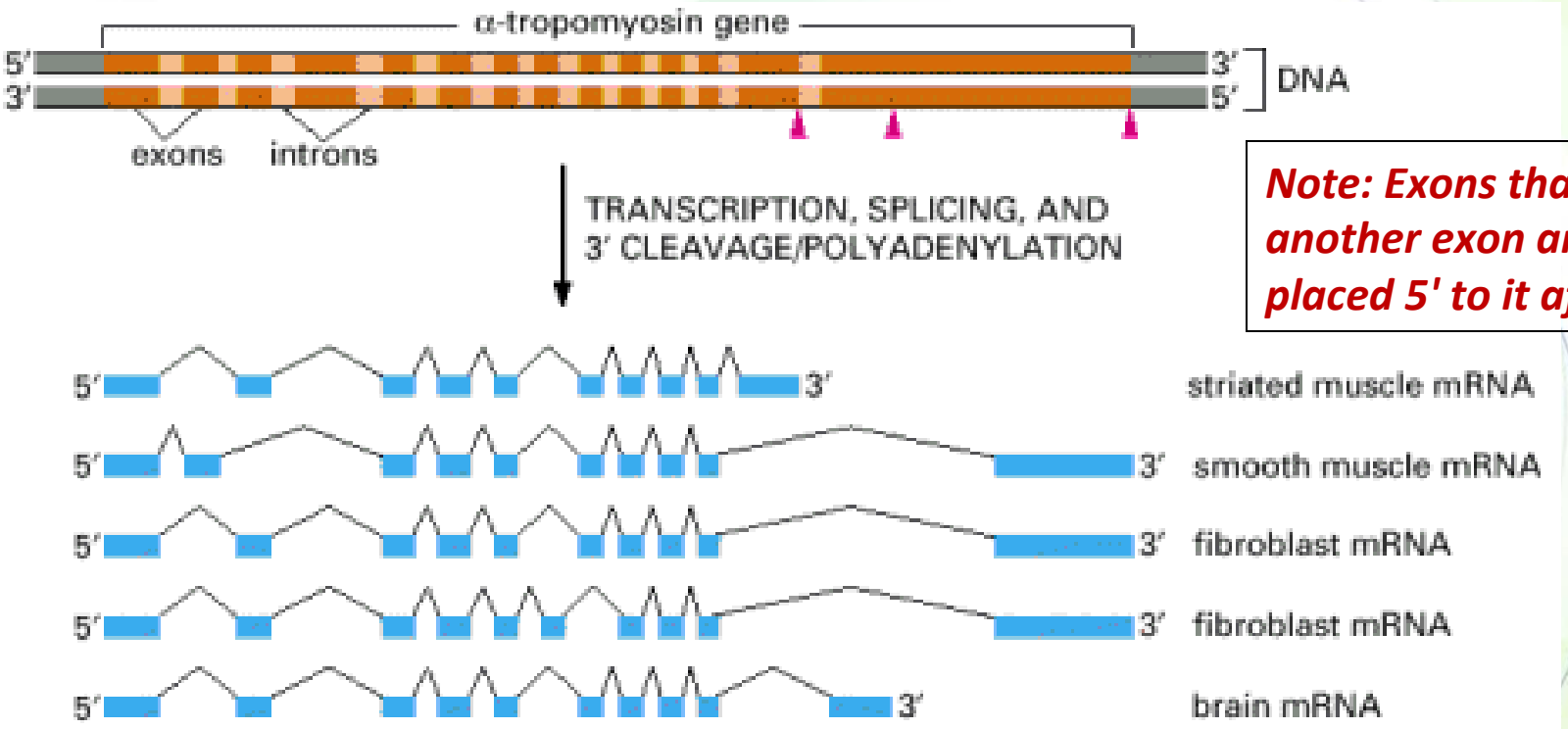
# Significance of introns

- They can encode functional RNAs such nucleolar RNA that function in ribosomal processing as well as microRNAs.
- They contain regulatory sequences of gene expression.
- The exon-intron arrangement may facilitate the emergence of new proteins via:
  - Genetic recombination (minimizing risk of error)
  - Alternative splicing



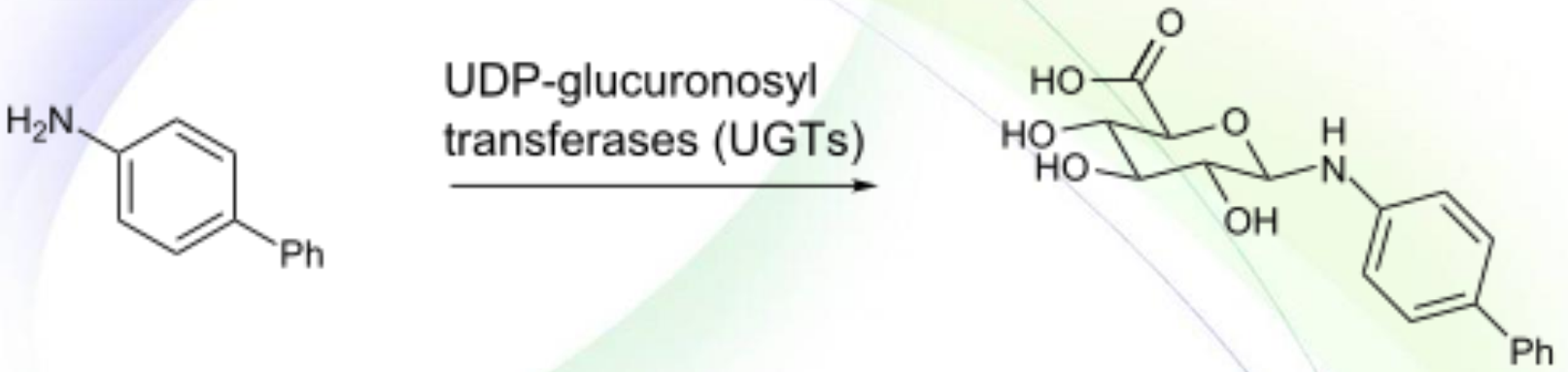
# Alternative splicing

- The transcripts are spliced in different ways to produce different mRNAs and different proteins
  - These are known as protein isoforms



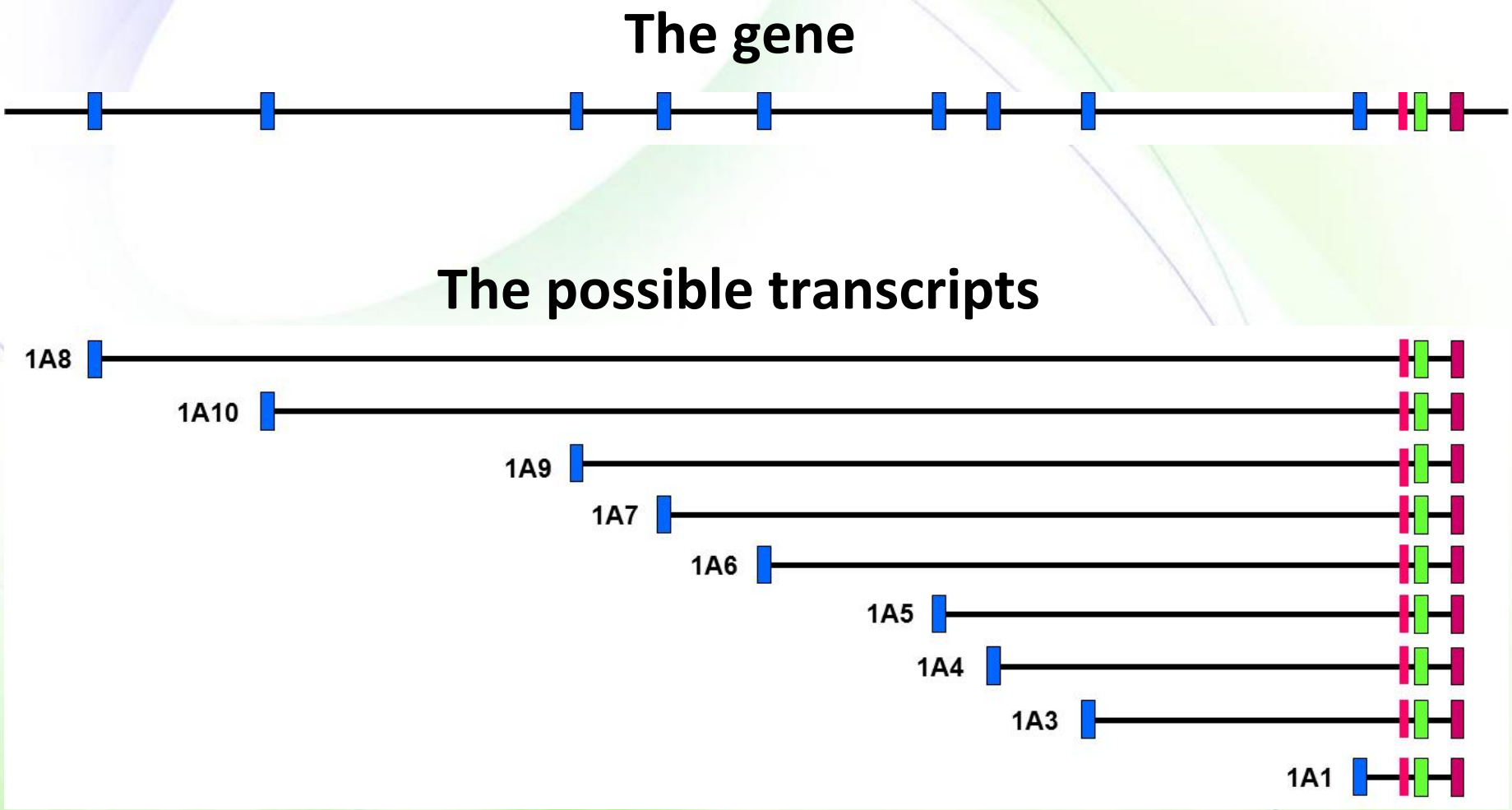

***Note: Exons that are 3' to another exon are never placed 5' to it after splicing.***

# UDP-glucuronosyltransferase gene




- The 5' region of the UGT1A complex contains 9 viable tandemly arrayed first exons
- Each first exon has its own promoter element
- Exons 2, 3, 4, and 5 encode the catalytic domain that interacts with UDP-glucuronic acid
- The 9 exons determine substrate specificity is independently and of tem at a time is spliced to exon 2 generating 9 UGT1A transcripts

# Splice variants for UDP glucuronosyltransferase 1 (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 (UGT1A4) |
| UGT1A10 | Biliary tissue, colon, esophagus, intestine, orolaryngeal tissue, stomach | Raloxifene |

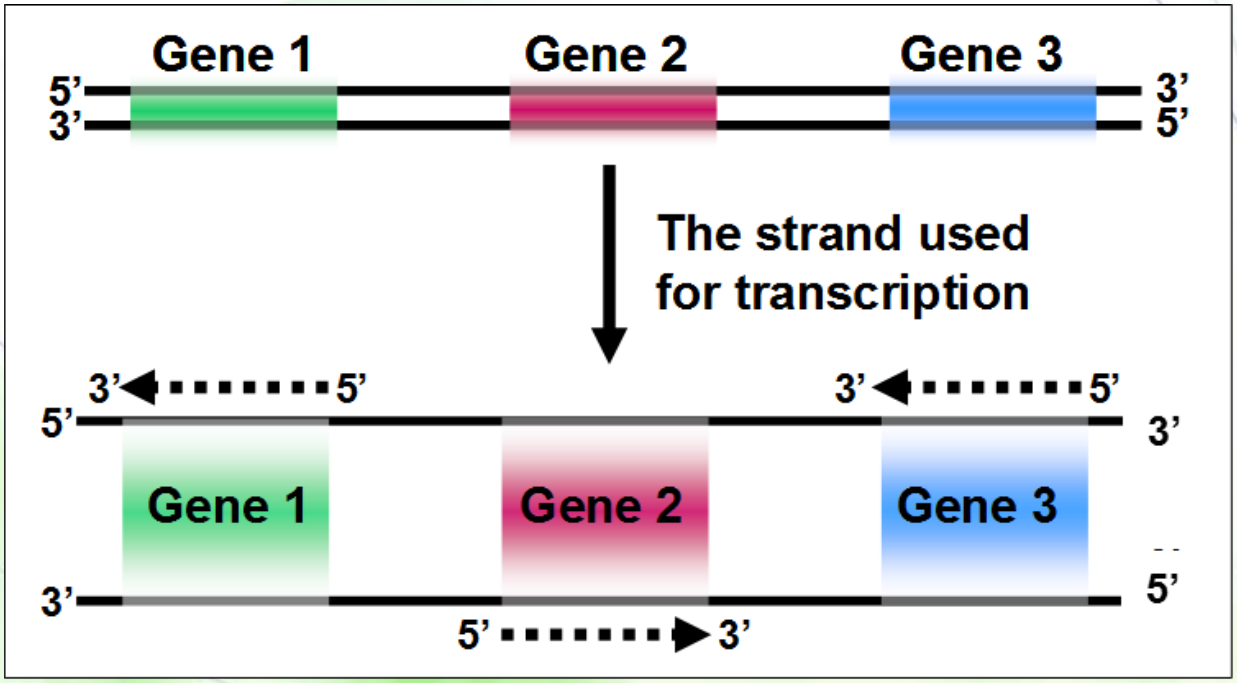
# The general mechanism of transcription

# 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.

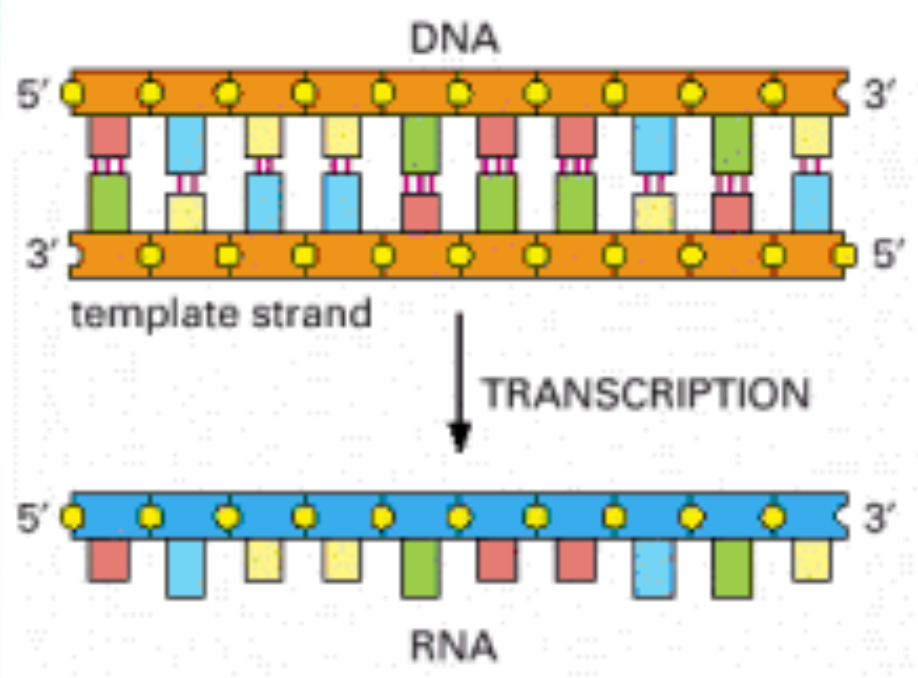
# Using DNA strands

- Although both enzymes can read both DNA strands, RNA polymerase uses one strand at a time in order to make a RNA molecule.
- The transcribed DNA strand = template, anti-sense, (-) strand
  - The other strand: sense, (+)strand, coding strand



# Complementary sequences

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



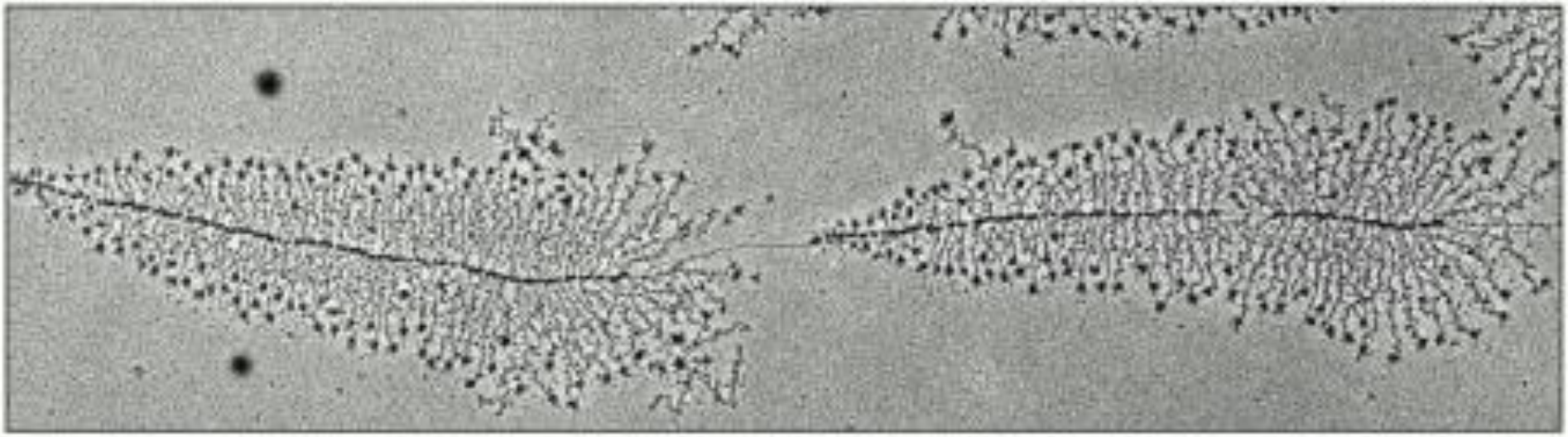
# Enzyme and substrate

- The enzymes that perform transcription are called RNA polymerases.
- RNA polymerases catalyze the formation of the phosphodiester bonds between two nucleotides.
- The growing RNA chain is extended in the 5’ to 3’ direction.
- 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

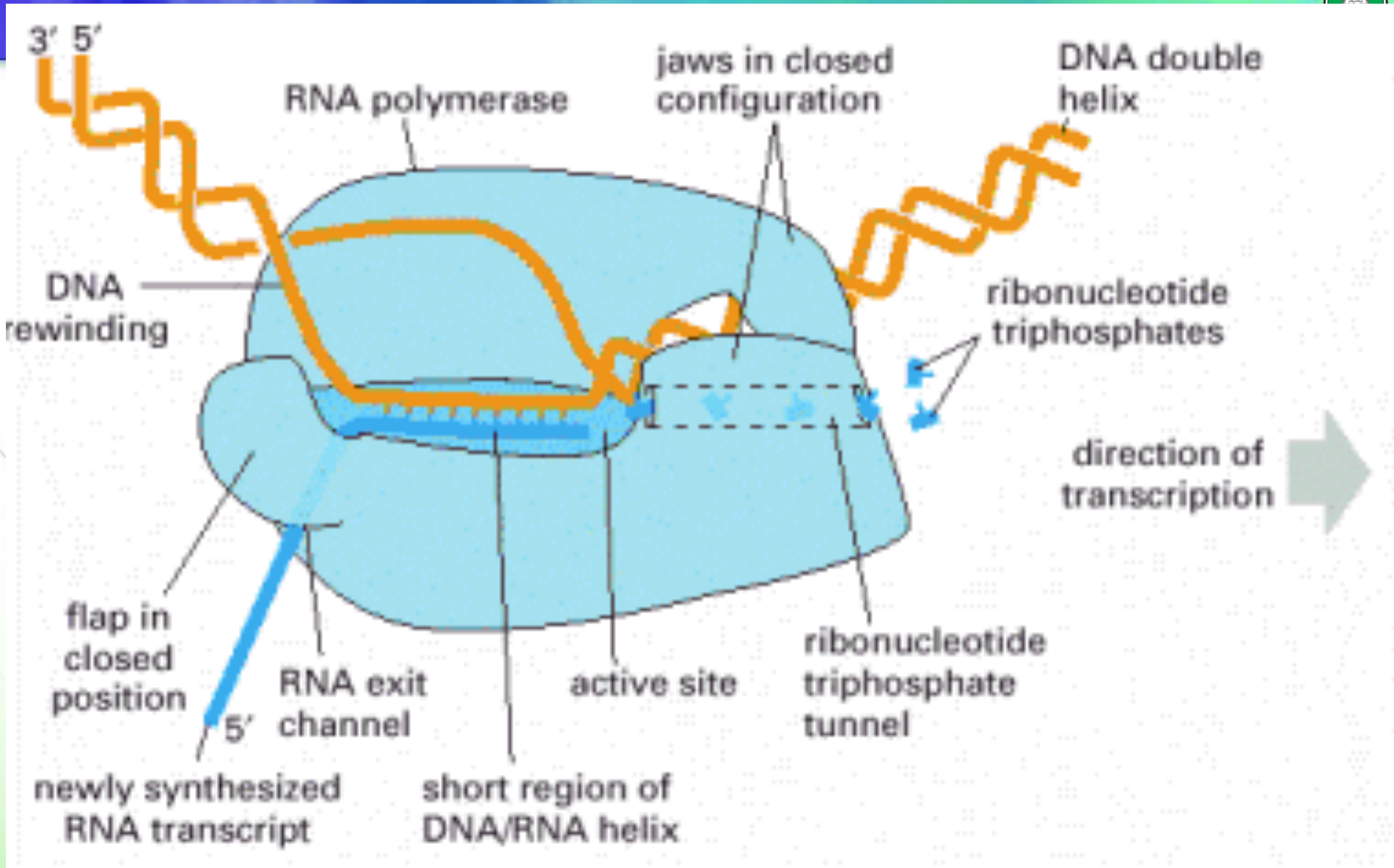


# DNA replication vs. transcription

- 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

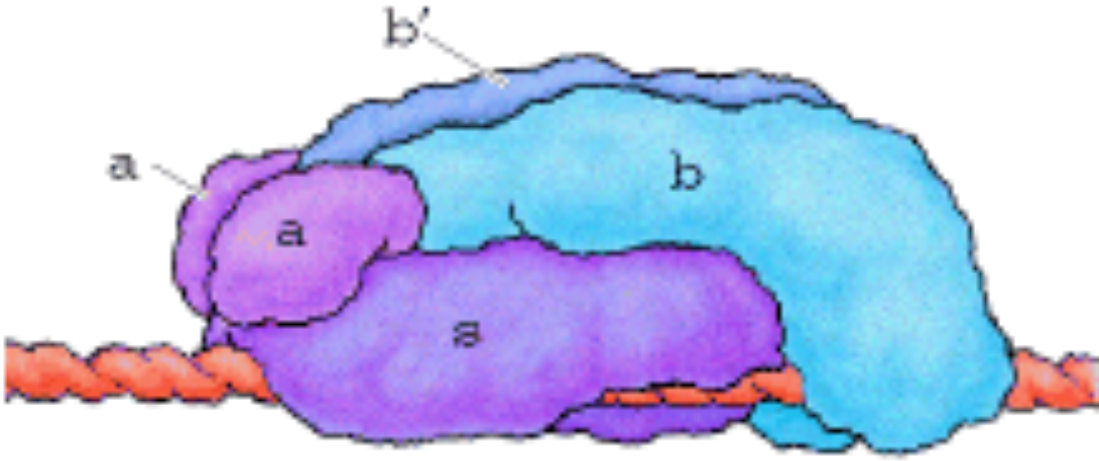
- RNA polymerase catalyzes the linkage of ribonucleotides, not deoxyribonucleotides.
- Unlike DNA polymerases, RNA polymerases can start an RNA chain without a primer.
- RNA polymerases make about one mistake for every $10^4$ nucleotides.
  - the consequences of an error in RNA transcription are much 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

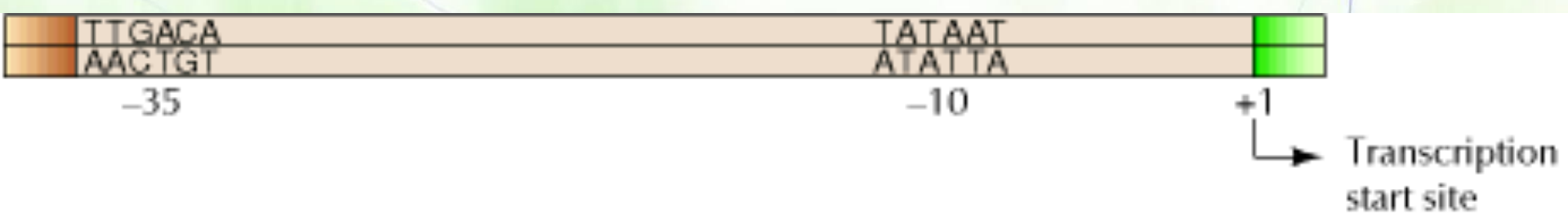
# The RNA polymerase

- E. coli RNA polymerase is made up of multiple polypeptide chains.
- The core polymerase consists of two $\alpha$, one $\beta$, and one $\beta'$ subunits.
  - The core polymerase is fully capable of catalyzing the polymerization of NTPs into RNA
- The $\sigma$ 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.



# How do we know they are important?

- Genes with promoters that differ from the consensus sequences are transcribed less efficiently than genes whose promoters match the consensus sequences.
- Mutations introduced in either the -35 or -10. consensus sequences have strong effects on promoter function.
- RNA polymerase generally binds to promoters over approximately a 60-base-pair region, extending from -40 to +20.
- The $\sigma$ subunit binds specifically to sequences in both the -35 and -10 promoter regions.

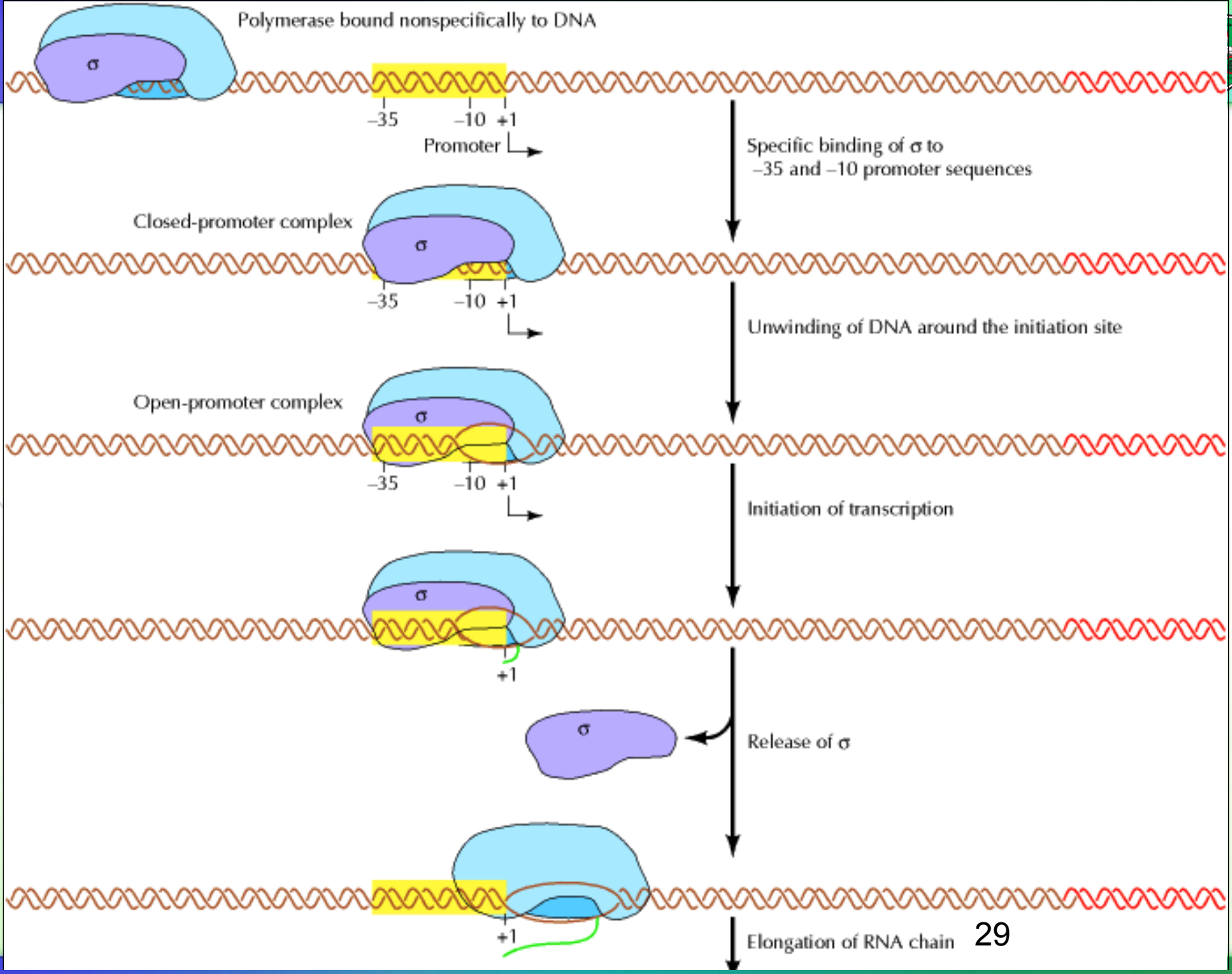
# Role of the σ subunit

- In the absence of σ , RNA polymerase binds to DNA with low affinity and nonspecifically.

- The role of σ is to identify the correct sites for transcription initiation and direct the polymerase to promoters by binding specifically to both the -35 and -10 sequences.

# Mechanism of transcription (initiation)

- The binding between the polymerase and a promoter is referred to as a closed-promoter complex.
- The polymerase unwinds approximately 15 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, $\sigma$ is released from the polymerase.

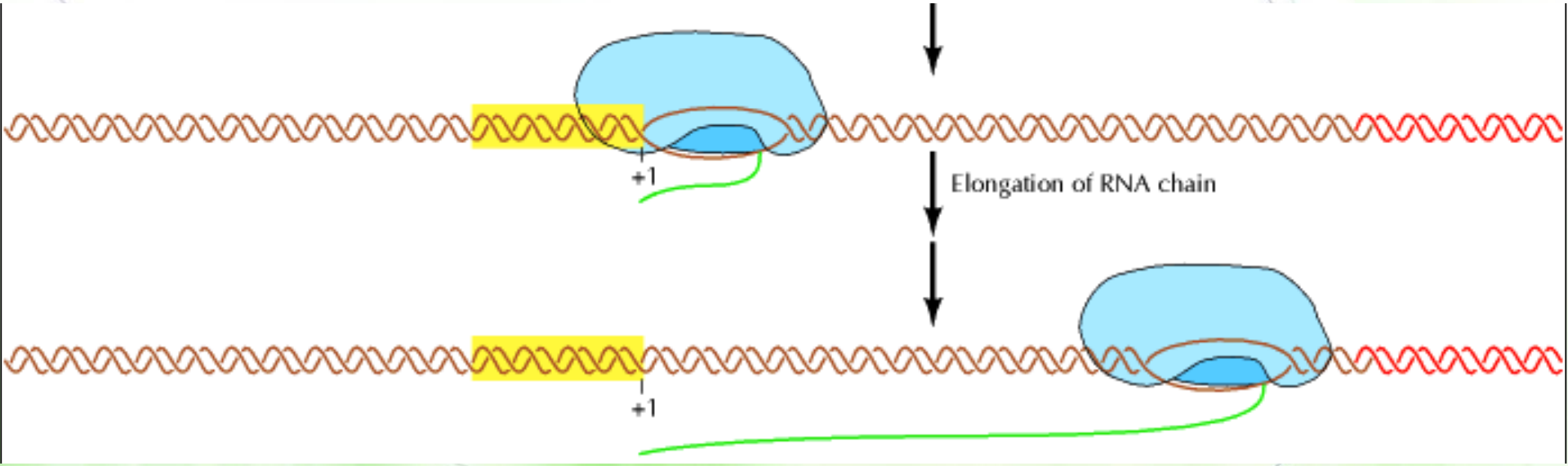

Polymerase bound nonspecifically to DNA
σ
–35
–10
+1
Promoter
Specific binding of σ to
–35 and –10 promoter sequences
Closed-promoter complex
σ
–35
–10
+1
Unwinding of DNA around the initiation site
Open-promoter complex
σ
–35
–10
+1
Initiation of transcription
σ
+1
σ
Release of σ
+1
Elongation of RNA chain

# 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

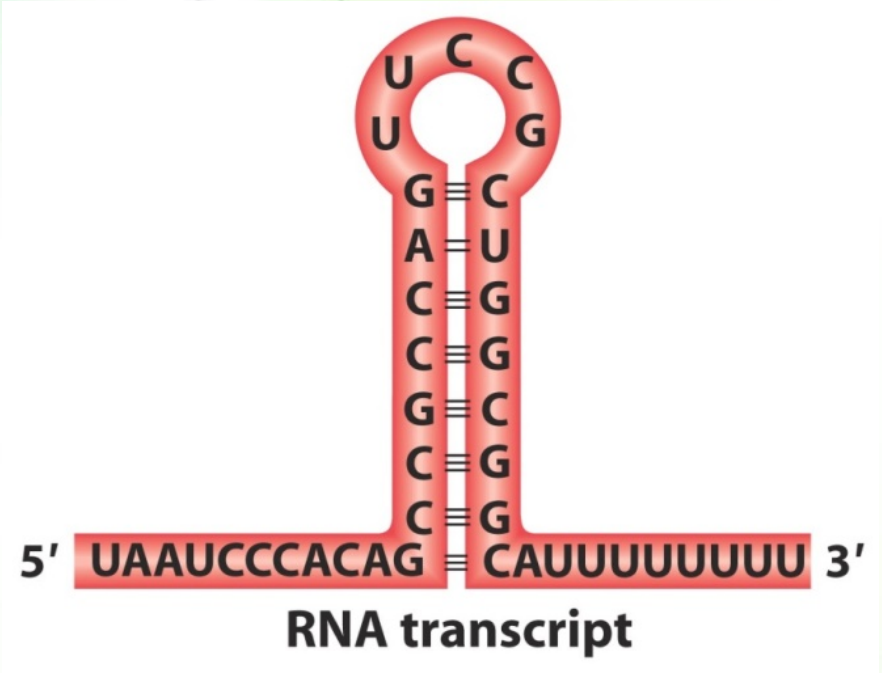
# Mechanism of transcription (termination)

- 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.



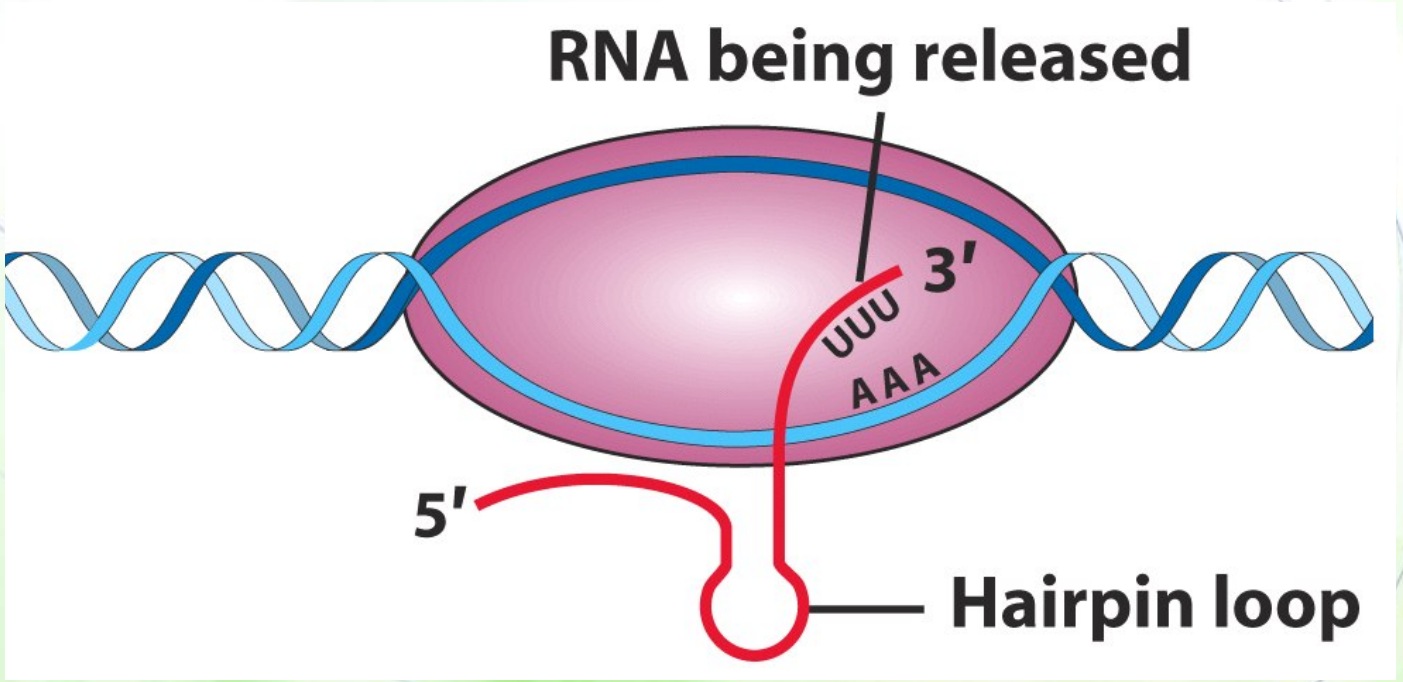
# Termination sequences

- 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.



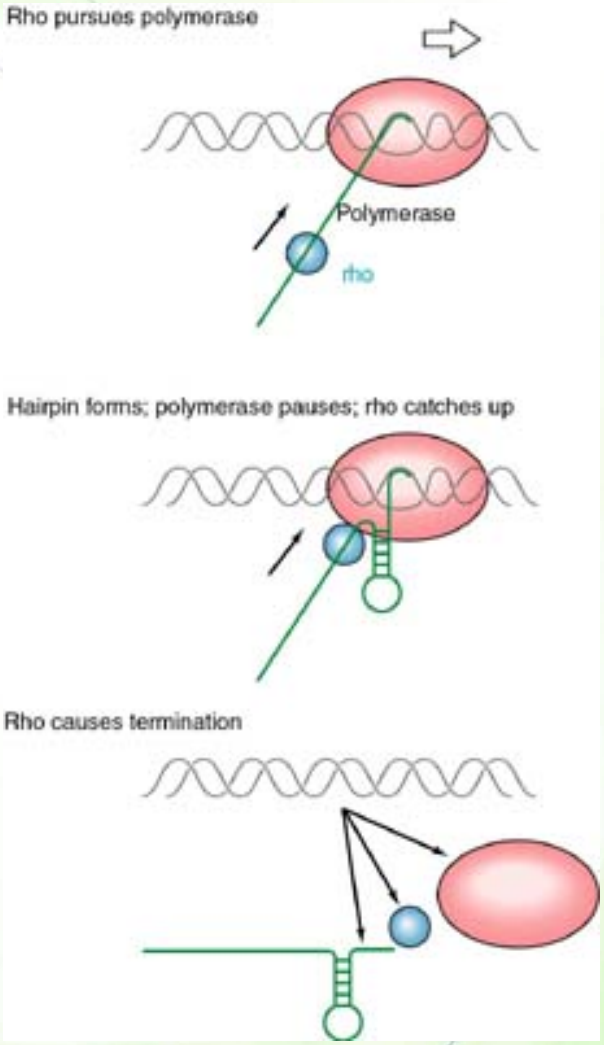
# The effect of the stem loop structure

- 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 is a helicase that 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

# 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
  - 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

# General transcription factors

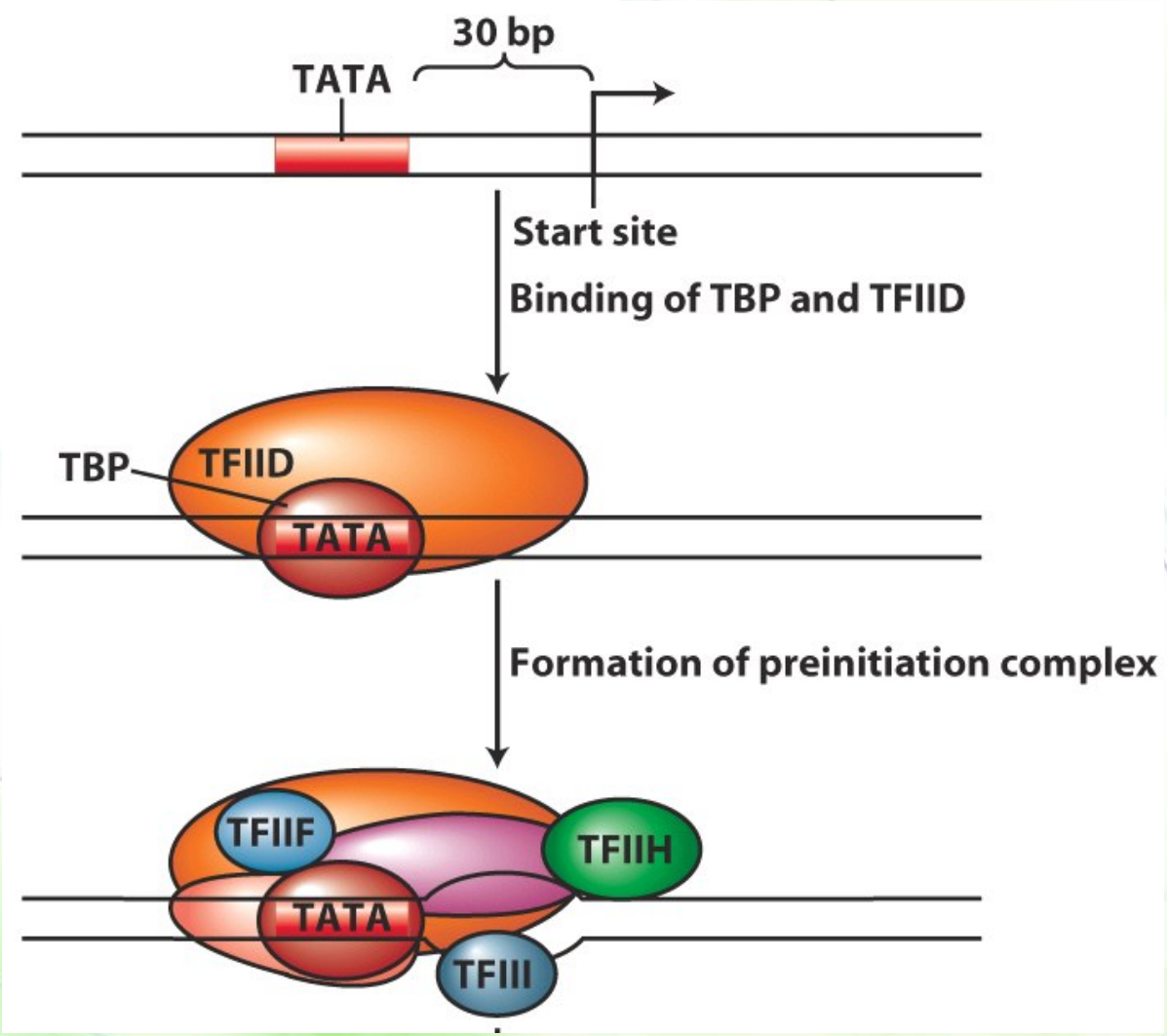
- These general transcription factors
  - help position the RNA polymerase correctly at the promoter
  - aid in pulling apart the two strands of DNA to allow transcription to begin
  - push the RNA polymerase forward to begin transcription

# Why are they general?

- 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

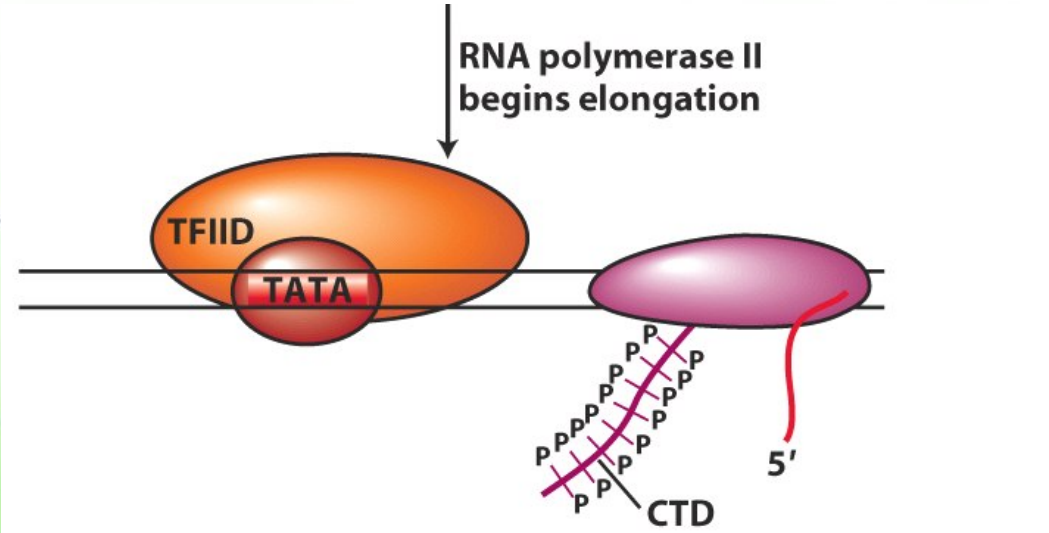
# Mechanism of transcription (initiation)

- TFIID binds to a TATA box located upstream from the transcription start site.
  - The binding of TFIID causes a bend in the DNA of the TATA box.
  - This bend attracts other proteins to assemble on the promoter.
  - Along with RNA polymerase II, these protein factors form a transcription initiation complex.
- One of them is TFIIH, which contains a DNA helicase.
  - TFIIH creates an open promoter complex exposing the DNA template to the RNA polyemerase.


30 bp
TATA
Start site
Binding of TBP and TFIID
TBP
TFIID
TATA
Formation of preinitiation complex
TFIIF
TFIIH
TATA
TFIII

# 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

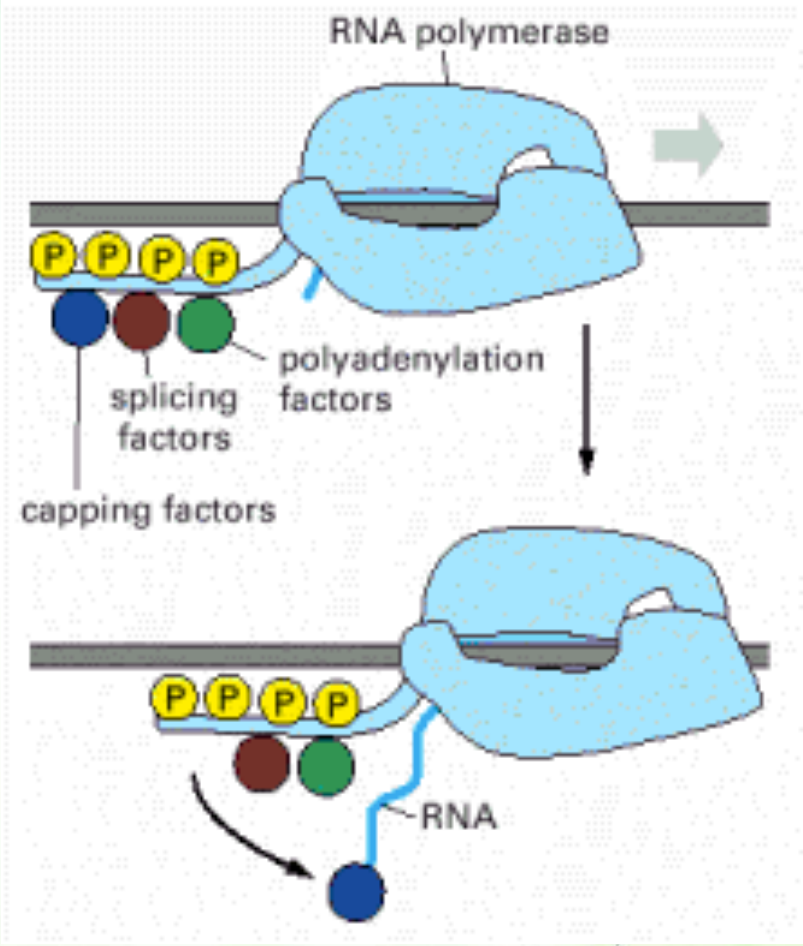


# Mechanism of transcription (termination)

- Termination begins by stopping the RNA polymerase. There is a eukaryotic consensus sequence for termination, which is AAUAAA.
- After termination occurs, the transcript is released, and the phosphorylated Pol II is released from the DNA.
- The phosphates are removed by phosphatases, and Pol II is recycled for another round of transcription.
- Termination is coupled to the process that cleaves and polyadenylates the 3 end of a transcript.

# Phosphorylation of RNA polymerase II

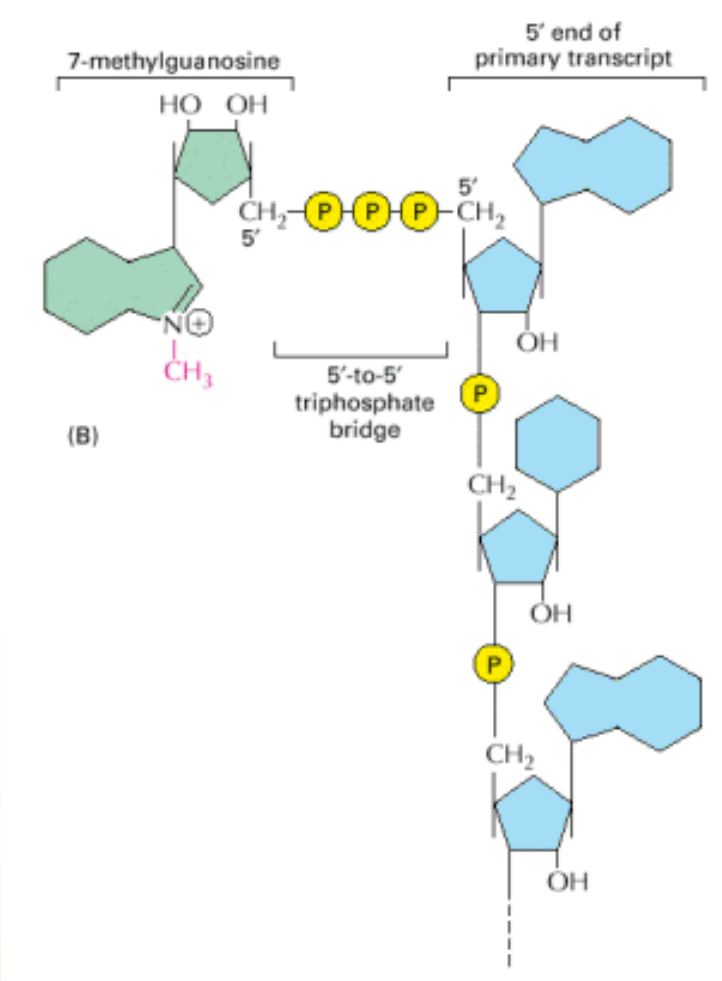
- RNA is processed and modified extensively
  - Capping
  - Splicing
  - Polyadenylation
- Some of these processing proteins are associated with the tail of RNA polymerase II.
- These proteins jump from the polymerase tail onto the RNA molecule as it appear.



# 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 a modified guanine nucleotide

- The guanine is added in a reverse linkage (5’ to 5’ instead of 5’ to 3’)

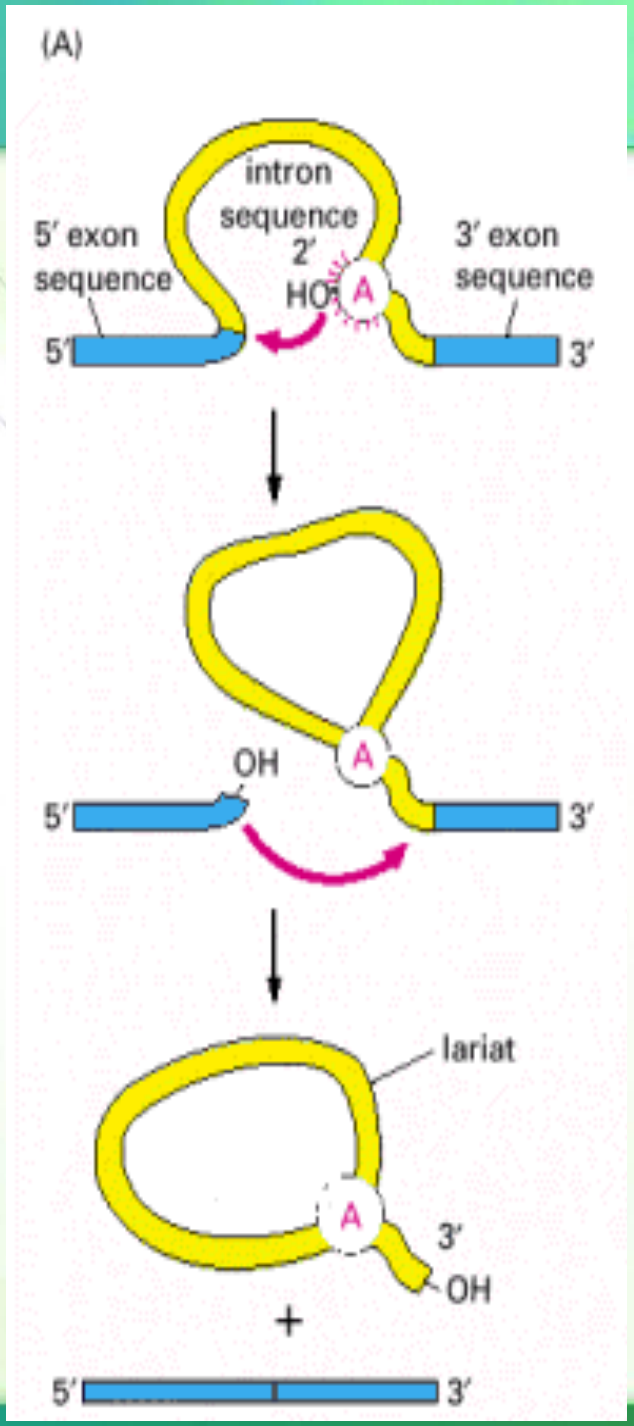


# Importance of capping

- The 5’-methyl cap 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

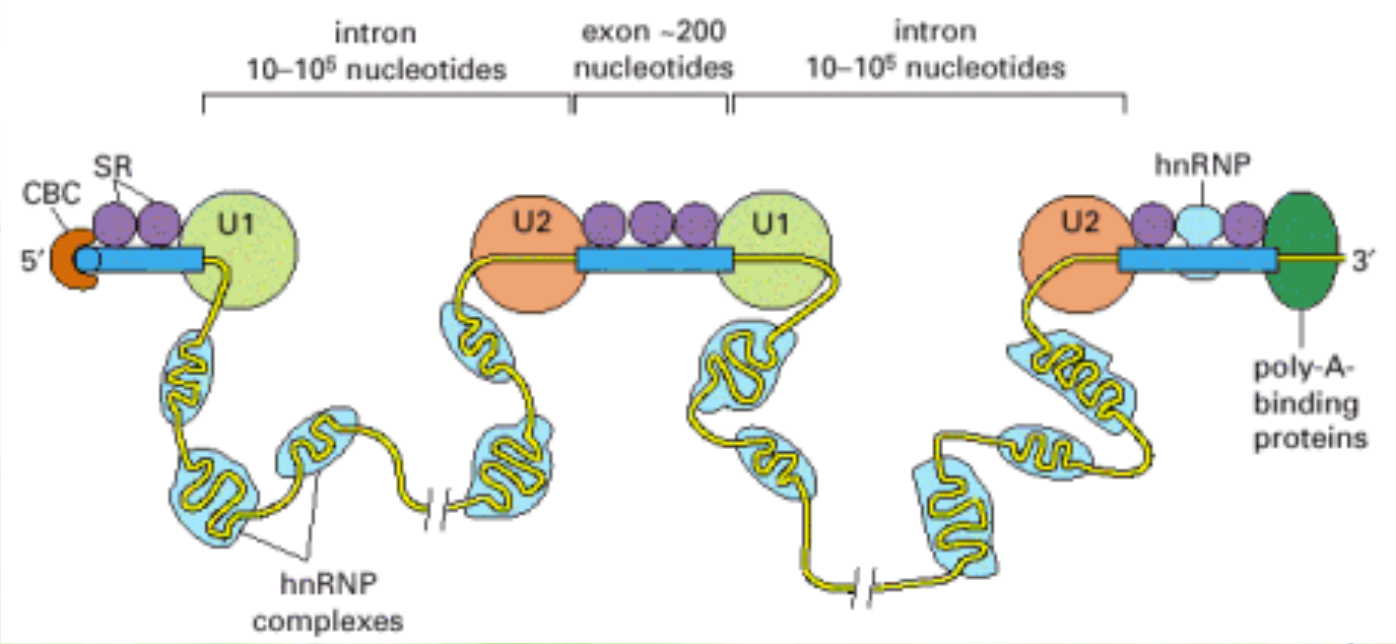
# RNA splicing

- The machinery that catalyzes pre-mRNA splicing consists of 5 RNA molecules and over 50 proteins.
  - The RNA molecules are known as snRNAs (small nuclear RNAs)
  - Each one of them is complexed with protein subunits to form a snRNP (small nuclear ribonucleoprotein)
  - These snRNPs form the core of the spliceosome, the assembly of RNA and proteins that perform pre-mRNA splicing
  - The catalytic site itself is largely formed by RNA molecules instead of proteins



# hnRNP

- Another class of proteins that assemble on pre-mRNA is hnRNPS (heterogeneous nuclear ribonuclear proteins)
  - hnRNP particles bind to introns
  - They have different functions

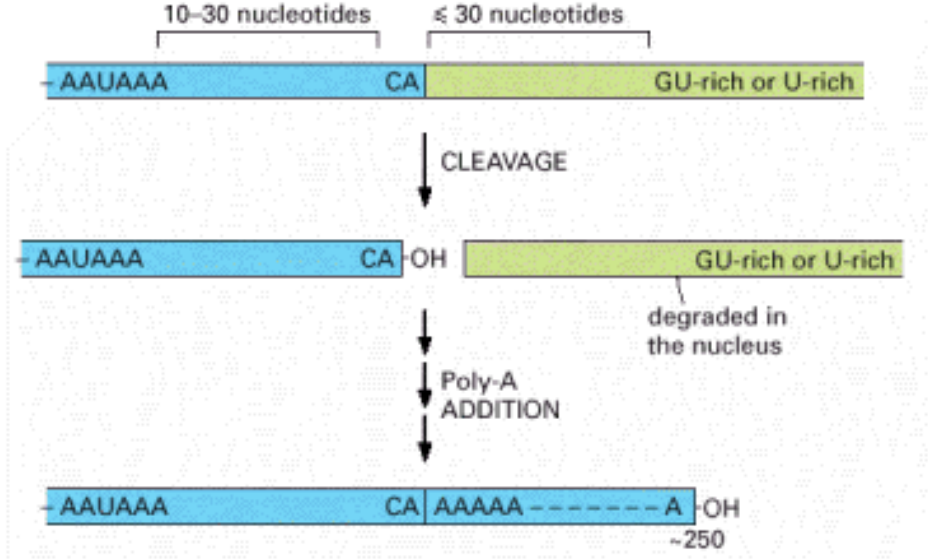


# Accuracy of splicing

- The consistent exon size (more uniform than introns)
- The assembly of the spliceosome occurs as the pre-mRNA emerges from the RNA polymerase II
- As RNA synthesis proceeds, spliceosome components, called the SR proteins, mark the 3’ and 5’ splice site
- hnRNPs define introns
- Spliceosome assembly is co-transcriptional, but splicing occurs post-transcriptionally

# Polyadenylation

- A certain sequence in the mRNA (AATAA) 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

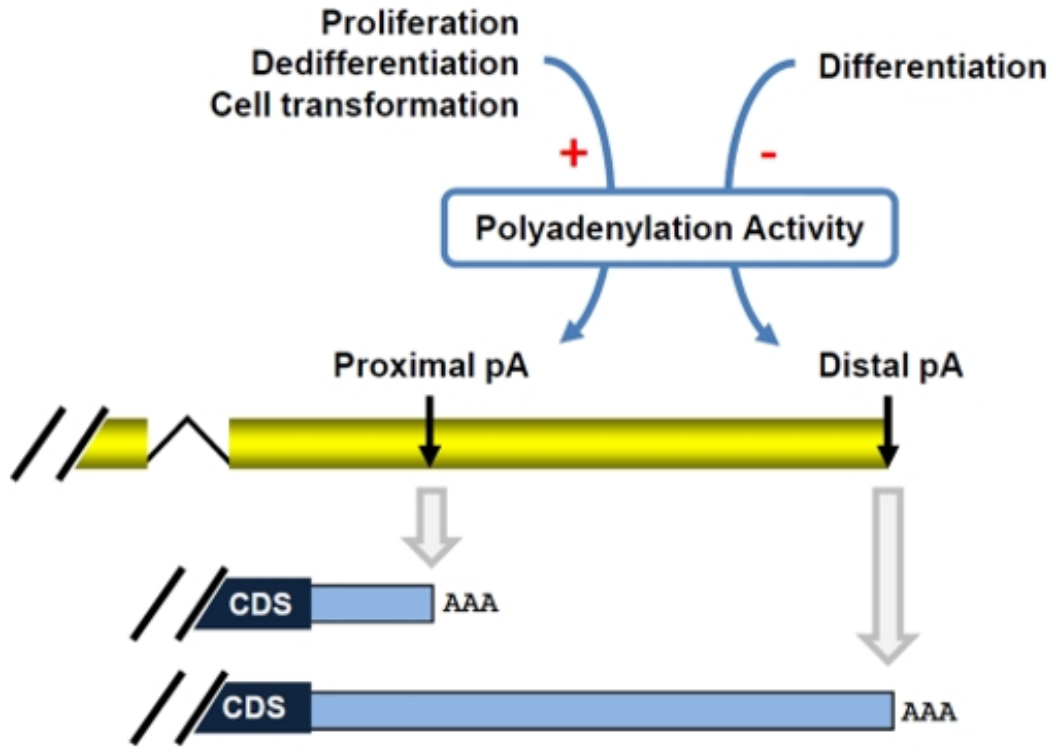
- Poly-A polymerase does not require a template
- hence the poly-A tail of eukaryotic mRNAs is not directly encoded in the genome

# Poly-A-binding proteins

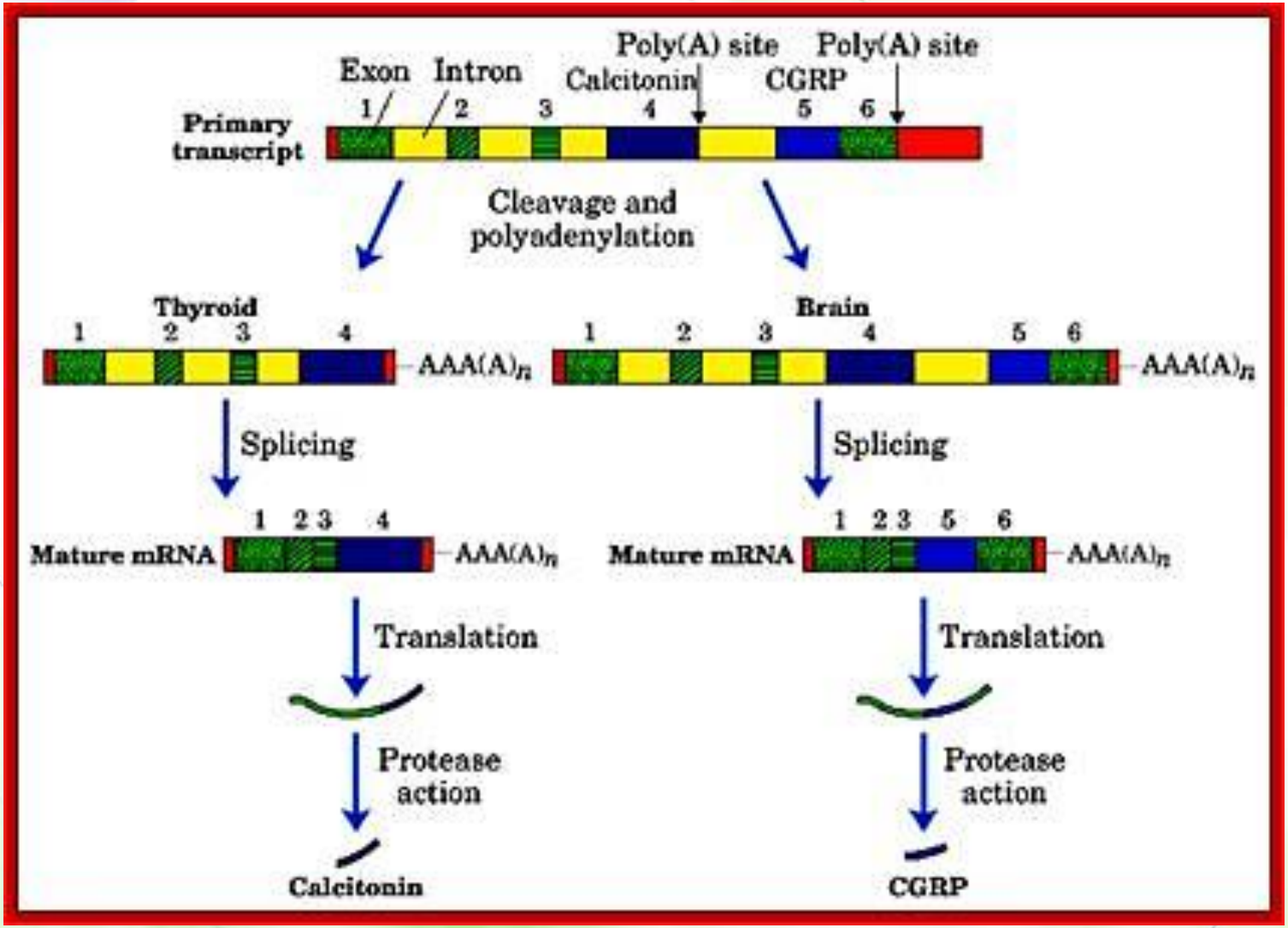
- Poly-A-binding proteins bind to the poly-A tail
  - Help in transporting mRNA from the nucleus to the cytosol
  - Help in protein synthesis
  - Stabilize mRNA

# Alternative polyadenylation

- Many 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).



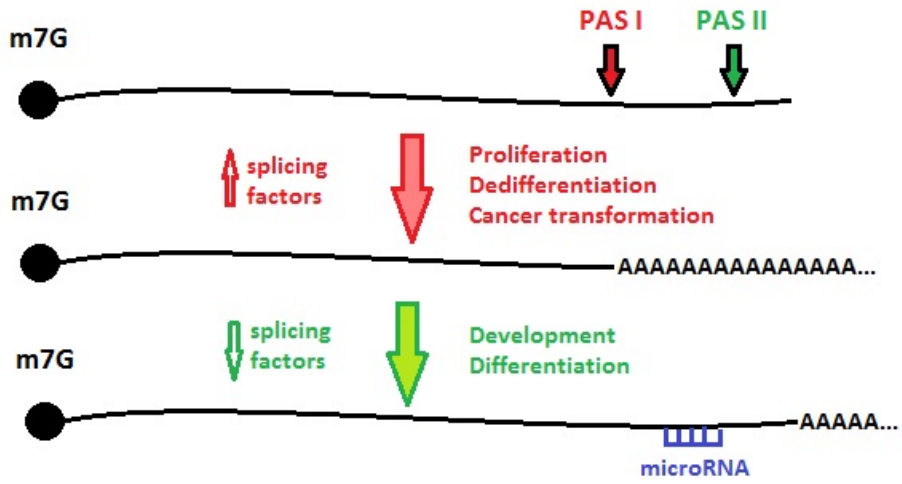
# Alternative polyadenylation



# Alternative polyadenylation

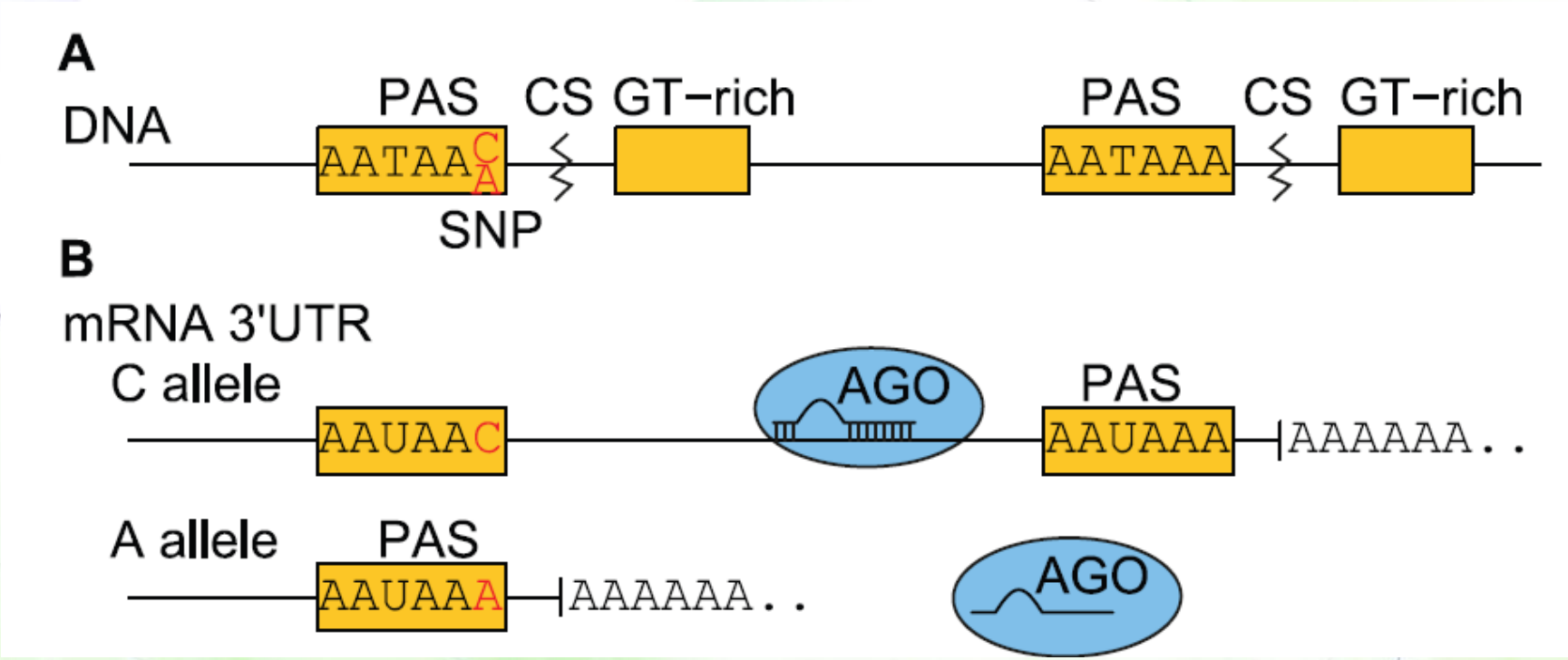
- The choice of poly(A) site can be influenced by extracellular stimuli that regulate the expression of the proteins that take part in polyadenylation.
- Having a shorter transcripts would remove regulatory elements in the 3 ‘-UTR and influence the half-lives of mRNA and, hence the amount of generated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



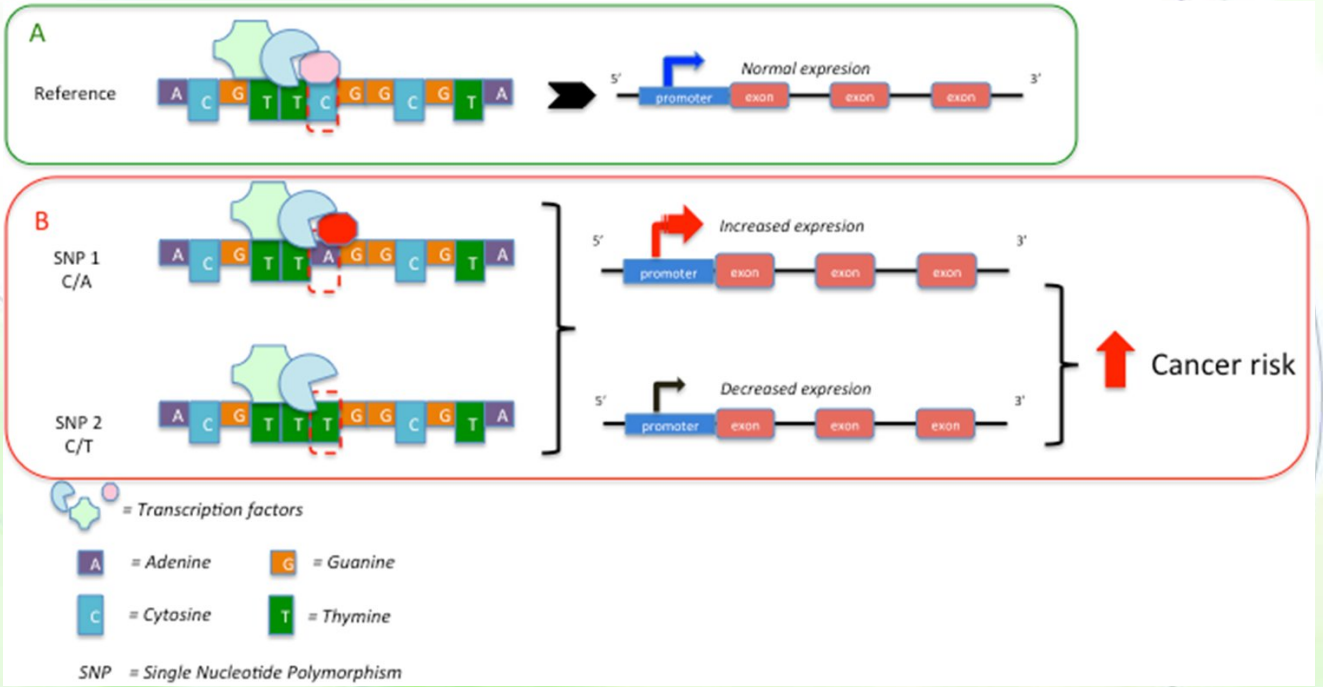
# SNPs and alternative polyadenylation

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



# 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.



# 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)
- Exonucleases are responsible for degradation

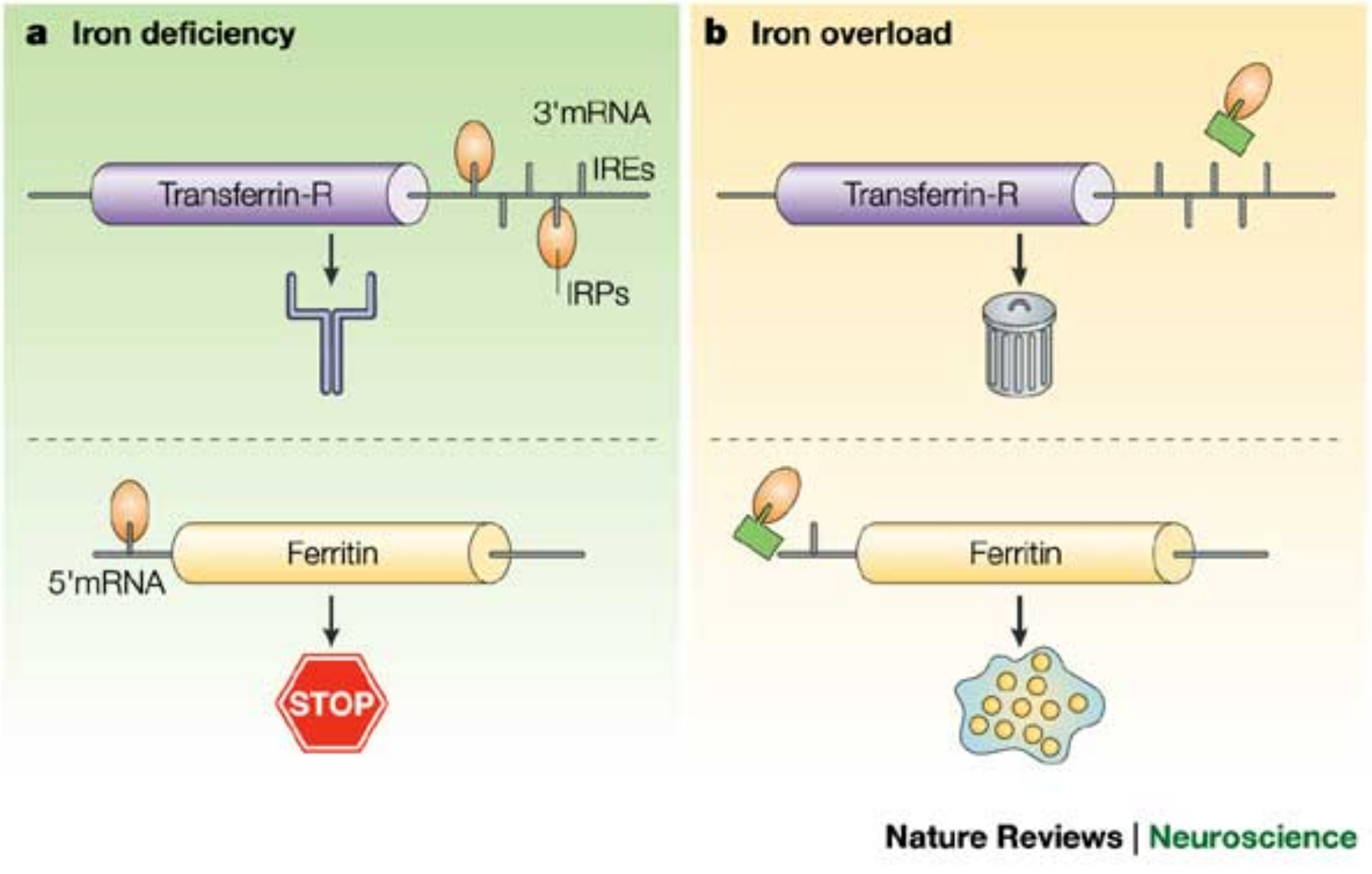
# Regulation of mRNA stability

# Iron-responsive elements

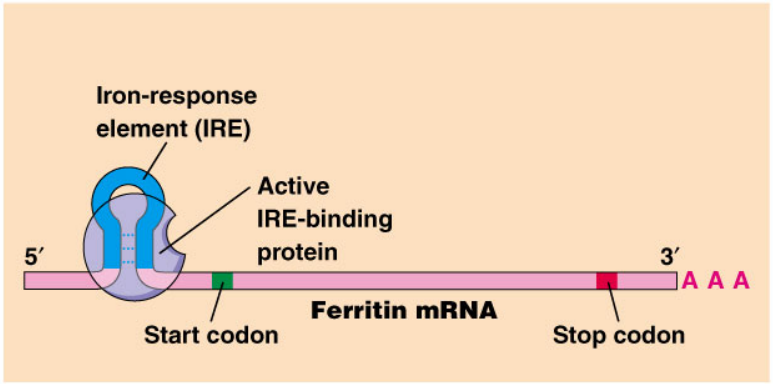
- In human cells, there are regions of mRNA called iron responsive elements (IREs)
- These regions are contained within the mRNA sequences that code for certain proteins that regulate the levels of iron
  - Ferritin, transferrin receptor, ferroportin, and DMT1
- Iron responsive element binding protein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

# Effect on expression

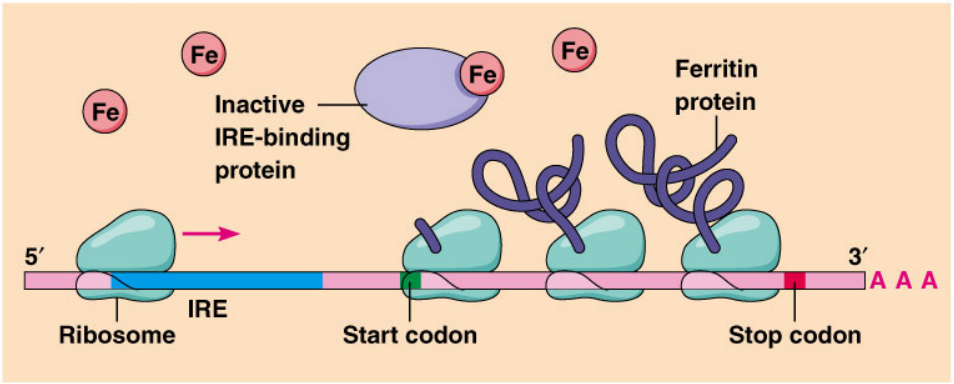
- When iron is abundant, it binds to IRE-BP, disabling the binding of IR-BP to ferritin mRNA
  - This prevents the degradation of the mRNA molecules allowing the production of more ferritin protein
  - Therefore, the iron itself causes the cell to produce more iron storage molecules
- On the other hand,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

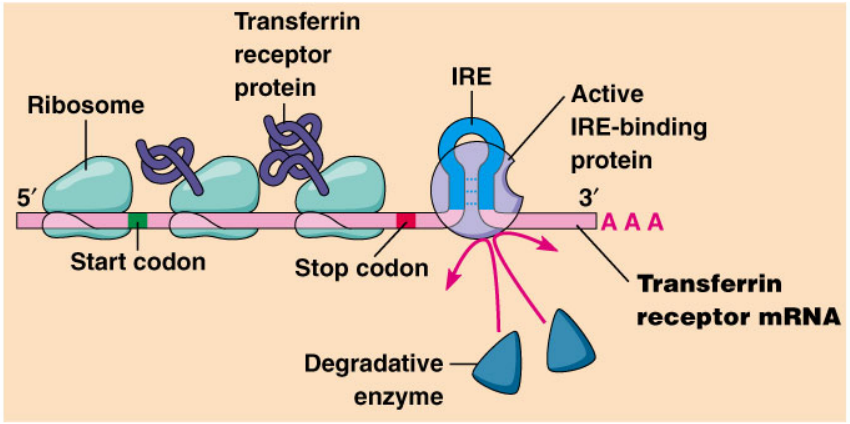



**(b) High iron concentration.** IRE-binding protein cannot bind to IRE, so translation of ferritin mRNA proceeds.

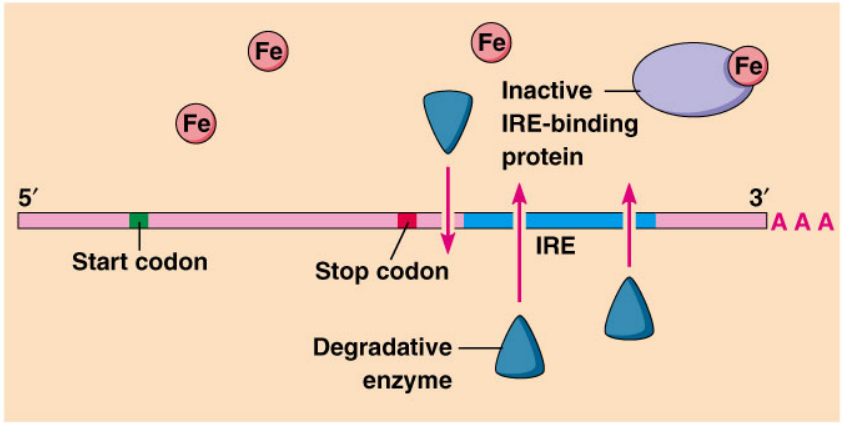





**(a) Low iron concentration.** IRE-binding protein binds to the IRE of transferrin receptor mRNA, thereby protecting the mRNA from degradation. Synthesis of transferrin receptor therefore proceeds.

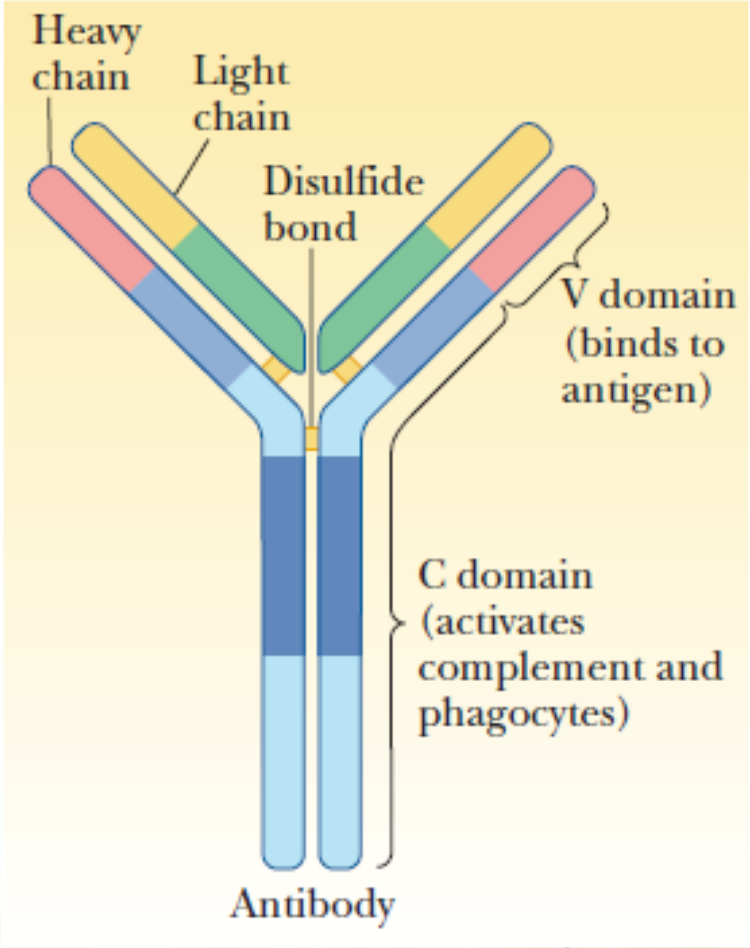



**(b) High iron concentration.** IRE-binding protein cannot bind to IRE, so mRNA is degraded and synthesis of transferrin receptor is thereby inhibited.



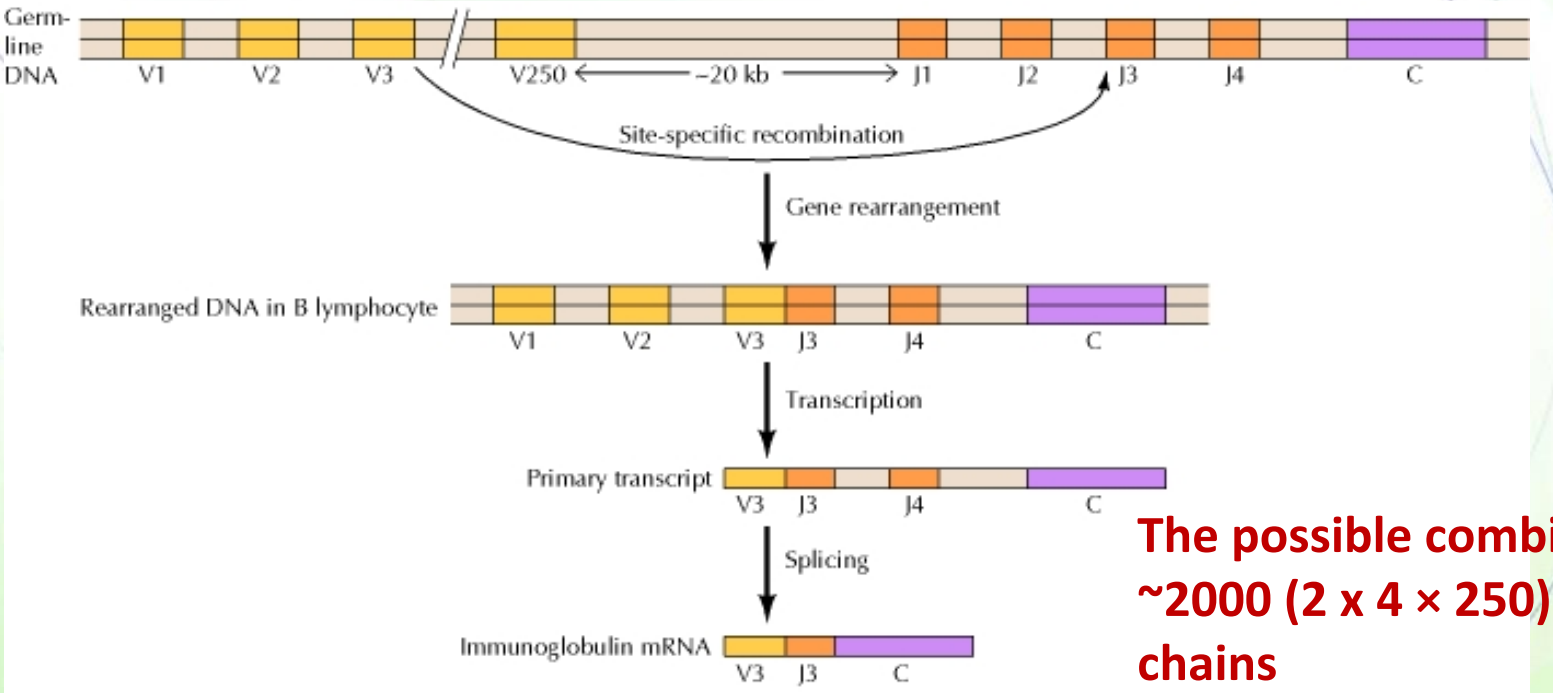
# Structure of antibodies

- 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; 250 genes
  - Joining region (J) gene: 4 genes
  - Constant region (CL) gene: 1 gene

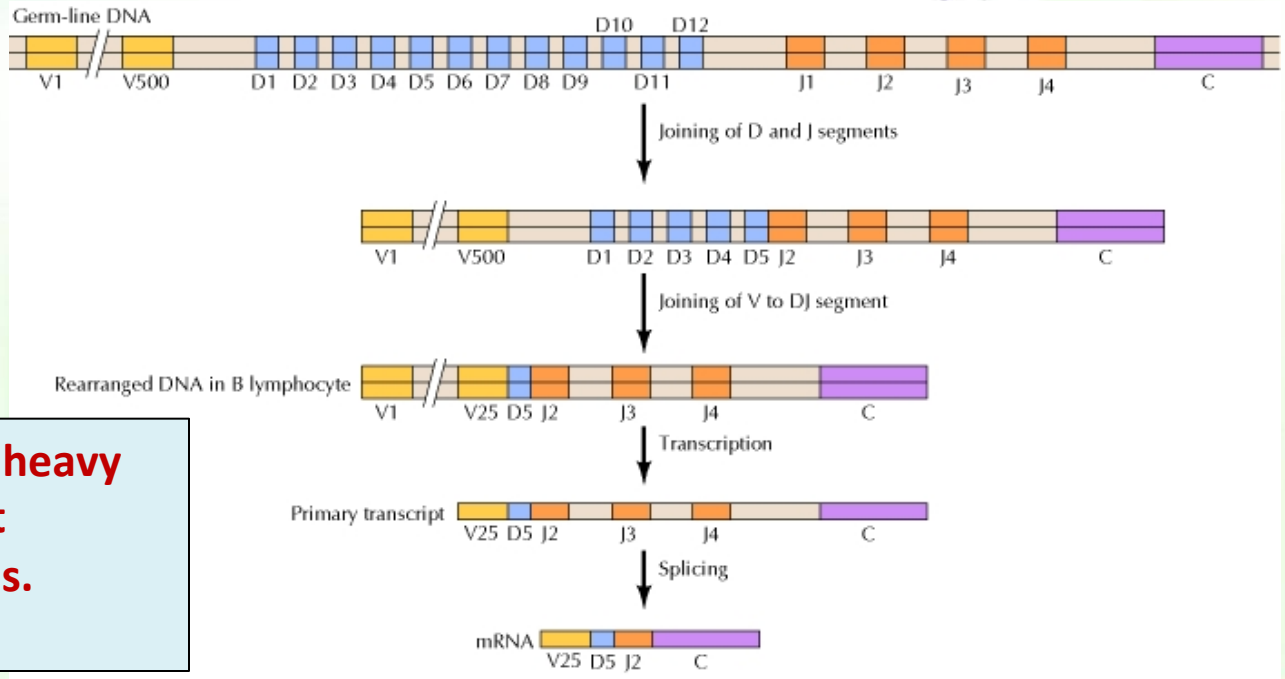



**The possible combinations is ~2000 (2 x 4 × 250) unique light chains**

# Gene rearrangement of the heavy chain

- Heavy chain is a product of at least 4 genes :
  - Variable region (VH) gene: 500 genes
  - Diversity region (D) gene: 12 genes
  - Joining region (J) gene: 4 genes
  - Constant region (CH) gene: 1 gene

**The possible combinations is ~24000 (500 x 12 x 4) unique heavy chains**

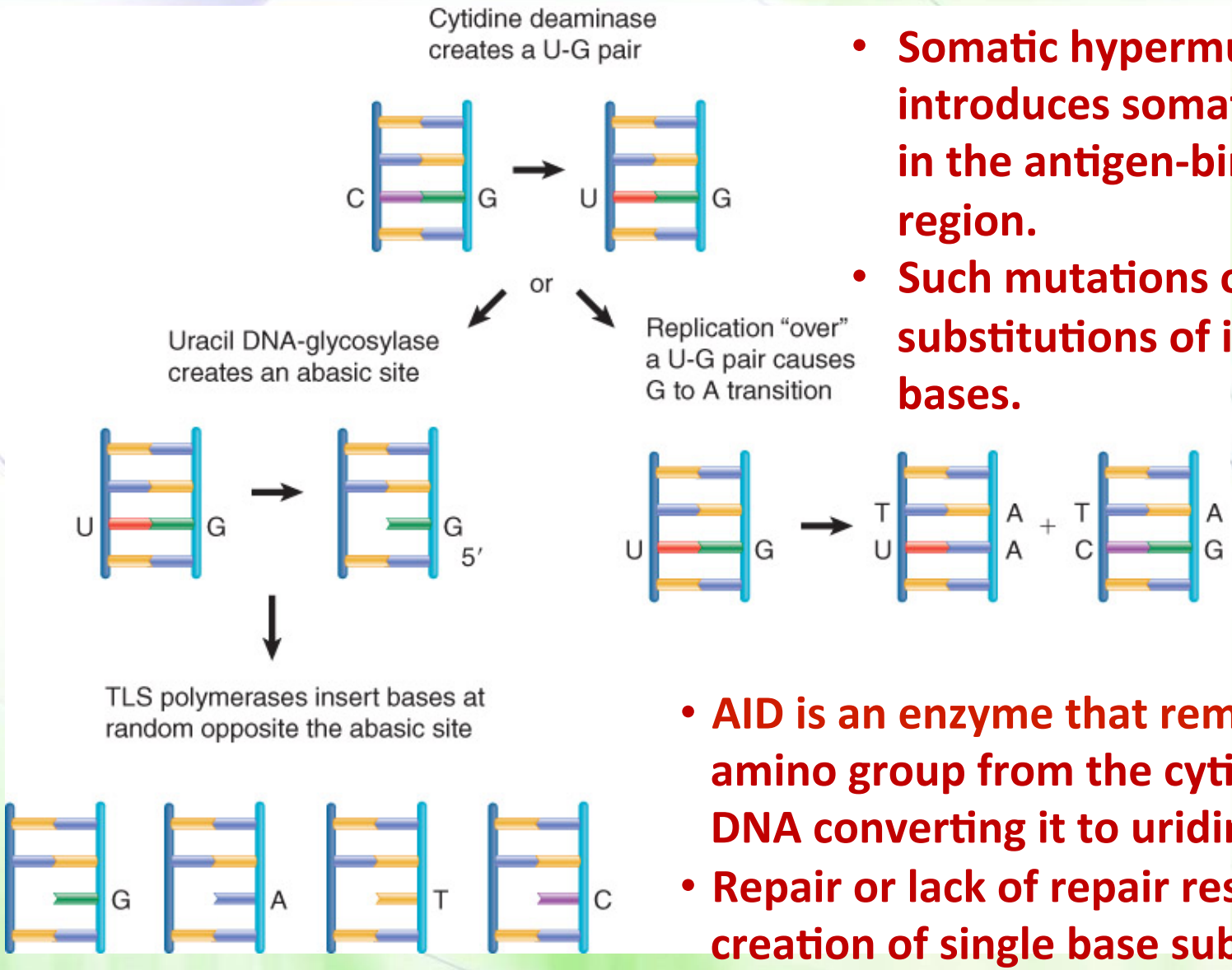



**2000 light chains x 24,000 heavy chains = ~ 5 × $10^7$ different immunoglobulin molecules.**

# Additional mechanisms

- The joining of immunoglobulin gene segments is often imprecise, resulting in the formation of ~$10^5$ different light chains and ~ $2 \times 10^6$ 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.**