University of Jordan
Faculty of Medicine

GENETICS & Molecular Biology

Number: 10

Done By: Anas Mourad

Corrected By: Amer Sawalha

Subject: DNA Transcription

Doctor: Mamoun Ahram

Price:

Date: 22.2.2016

In the previous lecture, we talked about:

1. **Genes in different types of organisms:**
   a) Prokaryotic genes can be classified as operons or polycistronic genes meaning that they produce one mRNA from which multiple polypeptides are generated.
   b) Eukaryotic genes are classified as monocistronic, meaning that a single gene produces one mRNA which generates a single polypeptide.

2. **Human genome:**
   a) Only 3% of the genome codes for proteins.
   b) 97% of the genome is considered "junk" DNA because scientists thought that "junk" DNA has no function, but it turns out that some of this DNA has a certain function.

The main difference between eukaryotic and prokaryotic genomes is:

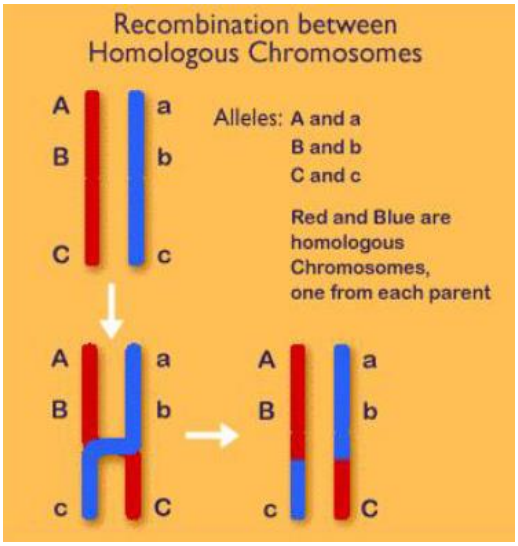
Eukaryotic genomes have exons and introns:

a) Exons are parts of the genome that code for proteins.
b) Introns are parts of the genome that don't code for proteins.

Therefore, the mRNA is synthesized as a "pre-transcript" precursor-mRNA (or pre-mRNA) which is processed by removing the introns and connecting the exons to each other forming the mature mRNA (the transcript).

Please refer to the diagram in slide 215 to visualize how introns are removed from the newly synthesized RNA through the process of **RNA splicing** (primary transcript to mature transcript).

Importance of introns



1. **Genetic recombination:**

In the process of mitosis and meiosis in Biology, we learned that chromosomes line up and genetic material is exchanged. One of the **homologous chromosomes** (alleles) is from the mother, but part of it will come from the father due to exchange of genetic material resulting in **hyperchromosomes**. Part of the **hyperchromosomes** is from the mother while the other part is from the father.

Usually this exchange occurs outside of the exons (remember 97% of the genome doesn't code for proteins) so this genetic recombination occurs in different regions as well as introns.

Thus, part of the gene will have two exons from the father and the rest of the exons are from the mother. For that reason, different combinations of exons of the same gene arise and that's why siblings look similar. However, there are differences between them and that's because of these small changes in exons.

If recombination occurs within exons in some cases, it is fine, but in other cases it can cause a mutation because this recombination is not 100% accurate and as a result, one nucleotide will be missed by **addition or deletion** and that will cause a frame shift mutation so different proteins will be produced.

Note: The exchange in the introns may also cause mutations and that's why scientists are not labeling the 97% of the genome as "junk" DNA anymore.

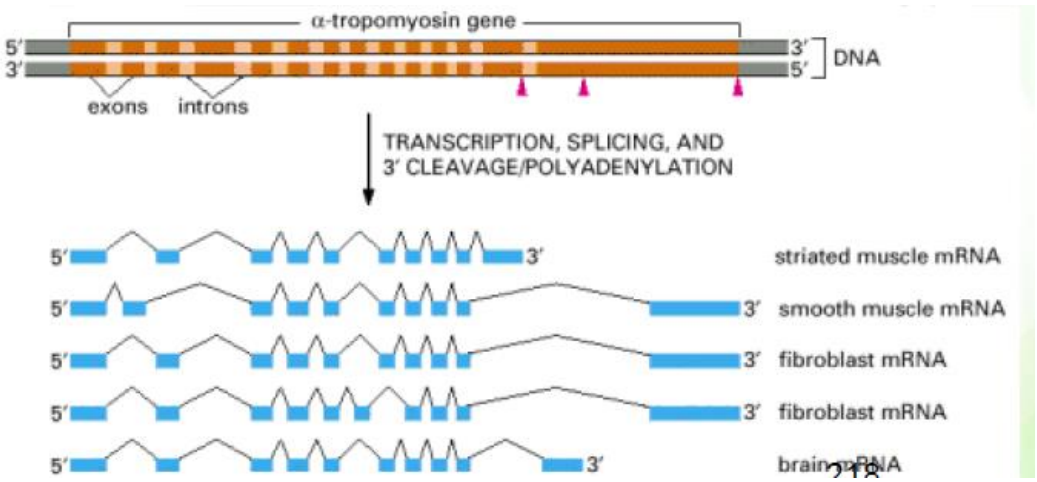
Remember:

- ✓ The removal of introns occurs at the level of RNA and not DNA.
- ✓ The DNA is the same in all cells and during transcription the removal of introns occurs.

2. **Alternative splicing:**

Tropomyosin: is an actin binding protein, which is important for muscle contraction.

The tropomyosin gene is composed of many exons (1, 2, 3, etc...). The pre-mRNA is produced, which contains the whole DNA sequence that is composed of different exons and introns. Nevertheless, the mRNAs of the same gene are processed differently in each cell type.



We can deduce from the figure that:

a) Striated muscles mRNA contains exon 1 which is connected to exon 3 and there is no exon 2.
b) Smooth muscle mRNA contains exon 1 which is connected to exon 2 and there is no exon 3.

As a result of this alternative splicing, different protein isoforms are produced.

Although it is the same protein that binds to actin, it has different locations (smooth muscles, striated muscles, fibroblast, epithelial cells, brain cells, and neurons). It progresses differently producing different isoforms. This is important because, for example, the way the protein is regulated in smooth muscle cells is different from the way it is regulated in the brain cells due to the fact that exon 10 is needed to bind to a certain regulatory protein. However, this regulatory protein is not present in fibroblast in which exon 11 is needed instead.

In slide 219, there is another example on the same process about calcitonin gene that's composed of 4 exons so either the calcitonin hormone or the calcitonin gene relating protein "CGRP" are produced in this case.

In slide 220, **UDP-glucuronosyltransferase** gene which codes for an important enzyme that modifies chemicals by conjugating glucose to make them more hydrophilic so they can be easily removed from the body and it does the same exact thing in different cells. However, this enzyme has different specificity for different chemicals, so it can recognize them because it is produced from the gene that's composed from exon 1, 2, 3 and 4. Exons 2, 3 & 4 code for the catalytic domain of the enzyme (the part that adds glucose to the chemical).

Exon 1 is 9 different exons (1A, 2A, 3A,... 9A), so they can form different enzymes from one gene. Each one of the nine exon 1 forms binds to the rest of exons (2, 3 & 4).

All 9 different forms of exon 1 code for the chemical recognition part (domain) of the enzyme so that each one of the enzymes recognizes a specific chemical.

- Here the doctor added two slides before the lecture and I'll try to bring them from the doctor and post them on Facebook but for now I'll try to mention every single word.

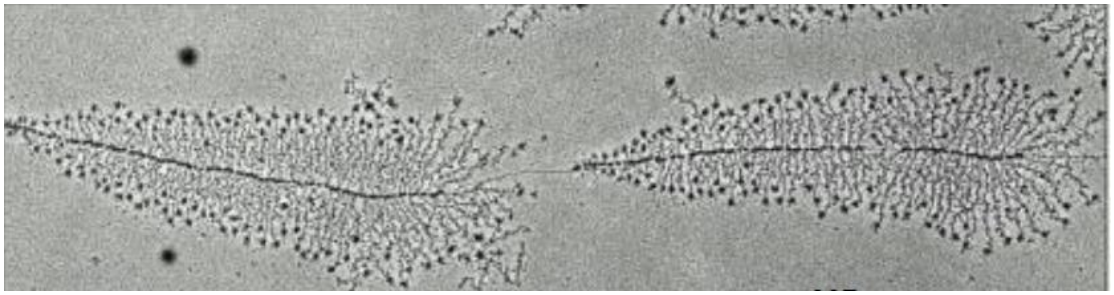
Isoforms have same function, different specificity, different regulation and different expression in tissues (like the tropomyosin).

Alternative splicing doesn't always give us the exact same functions but they must be the same.

The General Mechanism of Transcription

1. **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 “while in replication we use both strands”.
2. **Complementary sequences:**
   - ❖ The mRNA produced is complementary to the DNA (the strand that is read).
   - ❖ The RNA chain produced by transcription is also known as the transcript.
3. **Enzyme and substrate:**
   - ❖ The enzymes that perform transcription are called RNA polymerases (there are different RNA polymerases).
   - ❖ RNA polymerases catalyze the formation of the phosphodiester bonds between two nucleotides; they use the substrate that is the ribonucleoside triphosphate connecting them to the 3’ end of every nucleotide.
   - ❖ The growing RNA chain is extended in the 5-to-3 direction (all the time).
   - ❖ The substrates are nucleoside triphosphates (ATP, CTP, UTP, and GTP).
4. **Energy:**
   - ❖ The hydrolysis of high-energy bonds in NTP by cleavage of the triphosphate into a monophosphate and inorganic diphosphate (PP_i) provides the energy needed to drive the reaction forward. This is somewhat analogous to DNA polymerase that uses the substrates themselves as sources of energy.

No single RNA polymerase starts transcription of the DNA and releases the mRNA when it is done (stopping the RNA polymerase and recycling it to read the DNA again). In fact, there are multiple RNA polymerases reading the same DNA simultaneously where the first one starts, then the second one comes in after it and then the third one and so on until they form a structure known as a polysome.



So here we have two different genes; each one is surrounded by these new mRNA molecules making 2 separate polysomes.

The beginning of each gene is to the left; because mRNA is still short indicating the beginning of transcription on that side.

Please refer to slide 228 to see the differences and similarities between "DNA replication and DNA transcription" & "DNA polymerase vs. RNA polymerase".

The RNA polymerase has an exonuclease activity (proofreading activity) but it is not as efficient as the DNA polymerase. Why?

The consequences of an error in transcription are not as dramatic as those in the DNA since the DNA stores genetic information.

If one faulty RNA is produced, it is not a problem given that it is going to be degraded. And if a faulty protein is produced from the faulty RNA, it can be degraded as well.

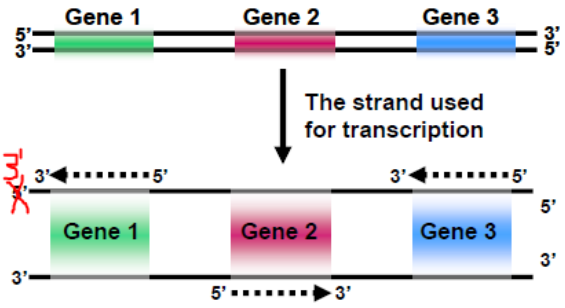
On the other hand, faulty DNA is transmitted from one generation to the next.

Transcription in prokaryotes

It is catalyzed by RNA polymerase which has 5 subunits
(2 α, β, β', and σ).
The polypeptide (co-polymerase) that's needed for polymerase activity exists in subunits α, β and β' but the σ (sigma) subunit is not necessary for the catalytic activity. Therefore, if a δ subunit is removed from RNA polymerase, the catalytic activity of the RNA polymerase will not be affected.

Using DNA strands:

- RNA polymerase uses one strand at a time in order to make a RNA molecule.
- The RNA is synthesized 5’ to 3’ so the DNA is read from 3’ to 5’.



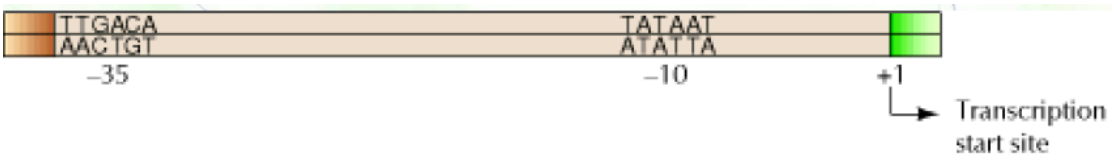
Please refer to slide 234 and replace the mistakenly mentioned 5’ in the figure (shown on the right) by 3’.

Question: Which strand is used for each gene?
The strand with the promoter sequence.

Question: Why is it different among genes?
Because it depends on the presence of a promoter.

Consensus sequences (the promoter):

The DNA sequence to which RNA polymerase binds, to initiate transcription of a gene is called the promoter “It is a binding site”.
A promoter is "upstream" of the transcription initiation site.
As a result the RNA polymerase doesn’t change sequences by which it binds to “it has to bind to this sequence or similar sequences” and we call these sequences that are common among genes Consensus sequences (sequences that are agreed “they are same or similar” among different genes and even in different bacterial species).



If you look at these sequences, they are called the (-10) and (-35) elements (sites).
The (-10) site is also called the TATA box because it is full of (T)s and (A)s and it looks like a box.

The (+1) site is the first nucleotide that’s synthesized in the mRNA and it is the same (+1) nucleotide for every mRNA that is going to be synthesized by RNA polymerase from that gene.

- The open reading frame: DNA sequence that can be transcribed into mRNA from first base (+1) to last one.
  The open reading frame is the only sequence that's going to be transcribed.

- The (-10) site (the region of TATA box) is present in the same location in different promoter regions in different genes in different bacterial species. This also applies for the (-35) site. These are called consensus sequences. Note that the DNA in the figure above is double stranded.

- All of the previously mentioned points are part of the DNA (prokaryotes DO NOT have introns), but it is not part of the open reading frame and not part of the RNA. It is not read by the RNA polymerase, it is only the region where the RNA polymerase binds to.
  It tells me where to find the gene; by finding the consensus sequences like the TATA box "the -10 region" or the -35 region. You can think about it as flag (like 9-mers and 13-mers in determining the origin of replication).

- A promoter is "upstream" of the transcription initiation site because it comes before it.
  For example, the (-35) site is upstream to the (-10) site and the (-10) site is downstream of the (-35) site.

How do we know that they are important?

a) By conducting scientific experiments to identify the common points among different genes. Comparing the upstream sites confirms that they are common.
b) By mutating the (-35) site or the (-10) site. This will affect the transcription efficiency. As a result, it is going to be reduced.
   Example: adding a G nucleotide to the TATA box and then examining the gene activity and transcription confirms that they are reduced but if you change the sequence between the (-10) and the (-35) sites, nothing will happen.
c) The RNA polymerase (σ subunit) binds specifically to the (-10) and (-35) sites on the DNA.

The role of the σ subunit is binding the RNA polymerase to the promoter. Although the RNA polymerase "jumps" on the DNA to bind to it, it doesn't bind without the σ subunit.

The RNA polymerase can scan the DNA but falls off to jump on another region to scan it. When it finds the promoter, it binds to it using the σ subunit which stabilizes the interaction between the RNA polymerase and the DNA. Thus, the RNA polymerase can start the transcription.

Nonetheless, if we remove the σ subunit there will be an initiation but the efficiency of the RNA polymerase will be reduced because it might fall easily away from the promoter without strong interactions (in some cases it can start transcription but in others it cannot). The σ subunit guides the RNA polymerase to the promoter region.

Mechanism of transcription:

It is composed of three steps:

1- Initiation:
   - ❖ The bond 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 10 nucleotides, σ is released from the polymerase and the transcription continues.
     The fate of the σ subunit is not degradation; instead it is recycled and binds to another RNA polymerase to start transcription.

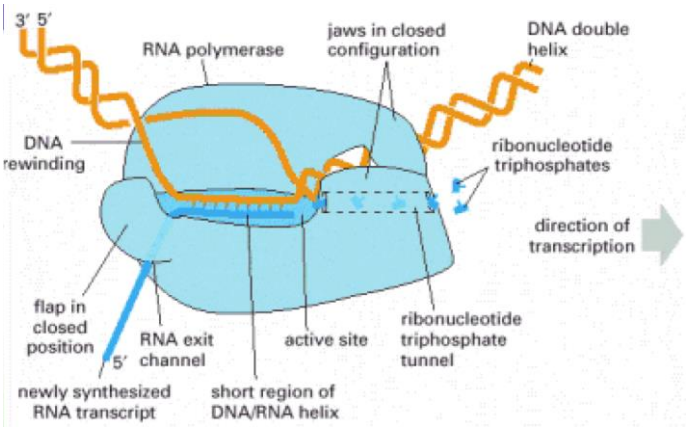
2- Elongation:
   As the polymerase moves forward it:
   - ❖ Unwinds the template DNA ahead of it.
   - ❖ Separates the two strands from each other.
   - ❖ Elongates the RNA "synthesis".
   - ❖ Reforms the double stranded DNA behind it.
   - ❖ Rewinds the DNA behind it.

3- Termination:

The complex formed by the DNA, the RNA polymerase and the mRNA which are, in this stage, separated from each other and that's how we end the transcription.

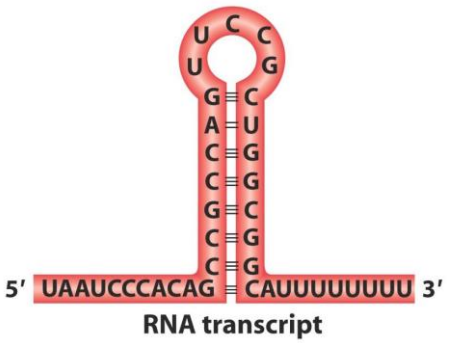
Observe the figure on the right:

As the RNA polymerase moves towards the direction of transcription, part of the mRNA is associated with the template and the old part is released, so this complex is stabilized by the interaction between the mRNA, the DNA and the RNA polymerase.



This complex dissociates as mentioned earlier through termination. This is achieved through a termination sequence that's composed of (G)s and (C)s followed by (U)s in the RNA, while in the DNA it is composed of (G)s and (C)s followed by (A)s.

Transcription of the (G) & (C) rich inverted repeat results in the formation of a stable stem-loop or "hairpin" structure (due to the fact that there are 3 hydrogen bonds between each G & C). The structure of the "hairpin" is shown in the figure on the right.

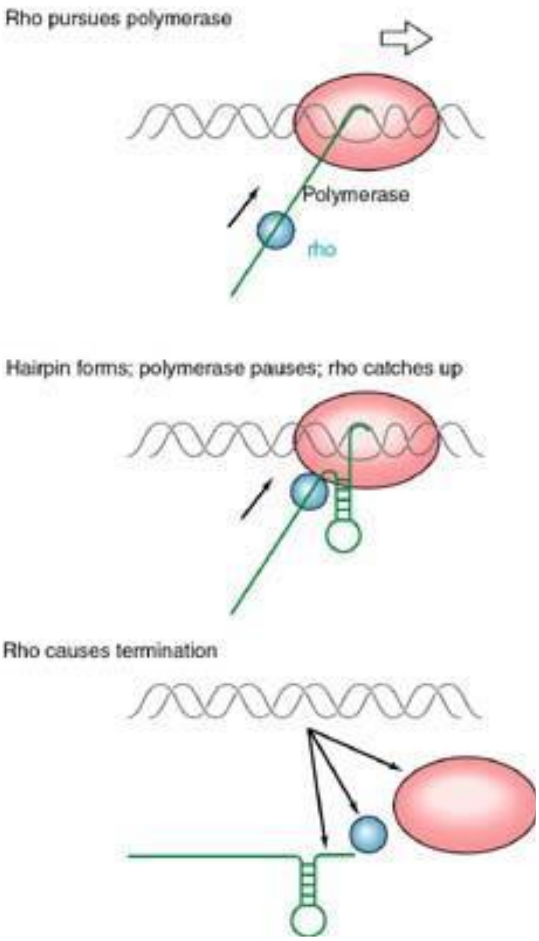


After the stem-loop structure, the (U)s are complementary to the (A)s from the DNA forming the U-A bonds but they are less stable since each bond contains two hydrogen bonds "compared to three in the G-C bonds".

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

You can think of this, as the doctor said, as a movie where the hero (mRNA) is running away and suddenly the monster (termination sequence) comes out from the ground (DNA template) creating a huge earthquake so the hero can't walk on the ground because it is unstable (weaker U-A bonds). In the end, the hero (mRNA) would rather fly away (separation).

The other way of termination is Rho-dependent terminator. It is also associated with stem loop structure but there are no (U)s or (A)s in the RNA so the RNA polymerase still binds with good stability to the DNA. However, the function of the (G)s and (C)s is slowing down the movement of the RNA polymerase by forming the stem loop structure. Right behind the RNA polymerase is a protein known as Rho which binds to the RNA and keeps following the RNA polymerase. When the RNA polymerase slows down, Rho protein catches up, binds to it and releases the RNA polymerase from the DNA. The mRNA is released as well, causing the destruction of the complex.



Sorry for any mistakes ☺

Best wishes

Revised and Edited by: Amer Sawalha