Genetics & molecular biology

◉ Sheet ○ Slide

Number:	8
Done by:	Moh Tarek
Corrected by:	Ahmad Badwan
Doctor:	Mamoun Ahram

The doctor skipped 4 slides in mutation and explained them in lecture 9.

TRANSCRIPTION

INTRODUCTION

In our courses we always talk about genes, RNA, DNA, chromosomes…. without knowing their specific definitions, now we will define some concepts that are related to these definitions.

- **DNA** is a sequence of deoxyribonucleotides attached to each other by phosphodiester bonds.
- **Chromatin** is DNA wrapped around histones (packed DNA).
- One linear unit of packed DNA is called **Chromosome**.
- **Genome** is the total collection of chromosomes in each cell.

Human cells genome = 46 linear chromosomes with histones (2 meters of DNA).

Bacterial cells genome = 1 circular chromosome without histones.

- **Genes** are regions found on the chromosome, these regions have a specific function, and each chromosome has multiple genes.

****Remember that each single cell of our body has the same genome.**

Our 46 chromosomes are numbered, between different individuals we have the same genes on the same chromosomes with the same arrangement and highly similar sequences.

Between genes we have intergenic sequences (e.g SNPs, VNTR, STR) which are random sequences but are also highly similar among all people with little variation (number of repeats, single nucleotide change…)

GENE -comprehensive definition-

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 **regulatory** (non-coding) sequences like promoter, enhancer, etc..

- This RNA can be either:

1) functional by itself e.g tRNA, rRNA (they are not used to make proteins)

2) used to make **polypeptide** (mRNA)

Peptides are either short peptides (no tertiary or secondary structures) eg oxytoxin (9 amino acids) or polypeptides which are used to make proteins.

CISTRON

An alternative term of gene (very old)

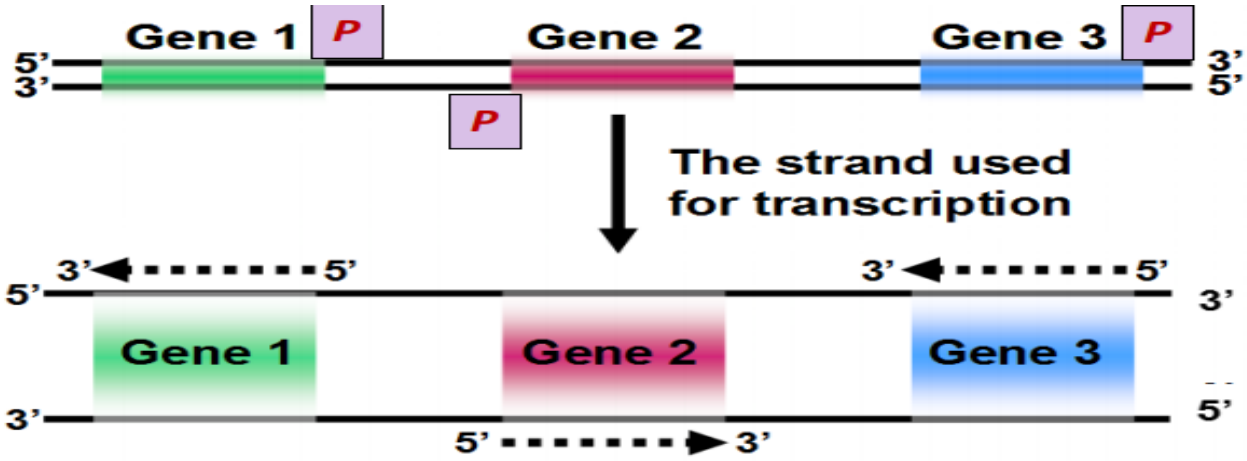
- Monocistronic produces one polypeptide from **one mRNA**, exists in both EUKARYOTES and PROKARYOTES.
- Polycistronic produces multiple different polypeptides from **one mRNA**, each will form different proteins, exists only in PROKARYOTES.

Overview of TRANSCRIPTION

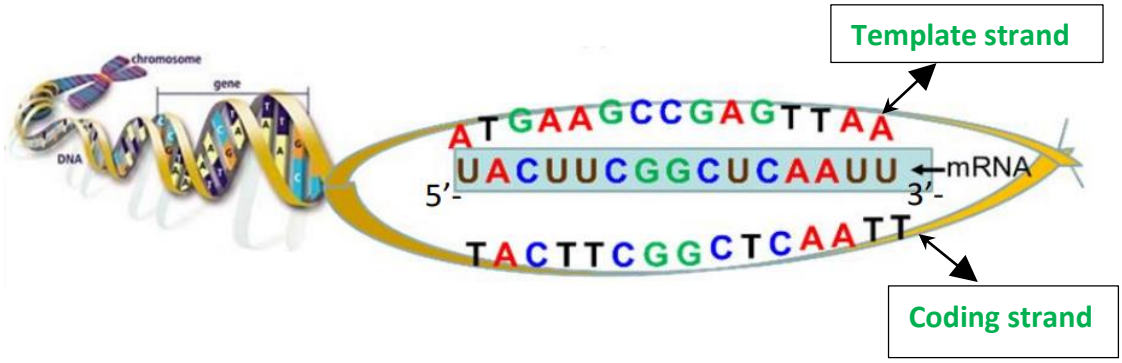
- Transcription is the process of making RNA from DNA.

Replication	Transcription
H bonds are present (2 strands)	No H bonds (1 strand only) (there are temporary H-bonds explained below)
Inserts T when it reads A	Inserts U when it reads A
Longer strands	Shorter strands (the outcome)
Stores genetic information	Doesn't store genetic information
Uses both strands as a template	Uses only one strand as a template
Needs a primer	Doesn't need a primer

- **RNA polymerase** is the enzyme that performs transcription.
- It catalyzes the formation of phosphodiester bond between two nucleotides.
- RNA chain is extended by **RNA polymerase** in the **5' to 3'** direction (remember that DNA is also replicated from 5' to 3')
- **RNA polymerase** uses nucleoside triphosphate as a substrate (ATP, UTP, GTP, CTP) and by breaking two phosphates it gets the required energy to connect the nucleoside monophosphate monomer and drive the reaction forward.
- RNA polymerase can read both DNA strands, however **the presence of a promoter** on a particular strand determines which strand will be used as a template for transcription.
- The RNA chain produced by transcription is also known as **TRANSCRIPT**



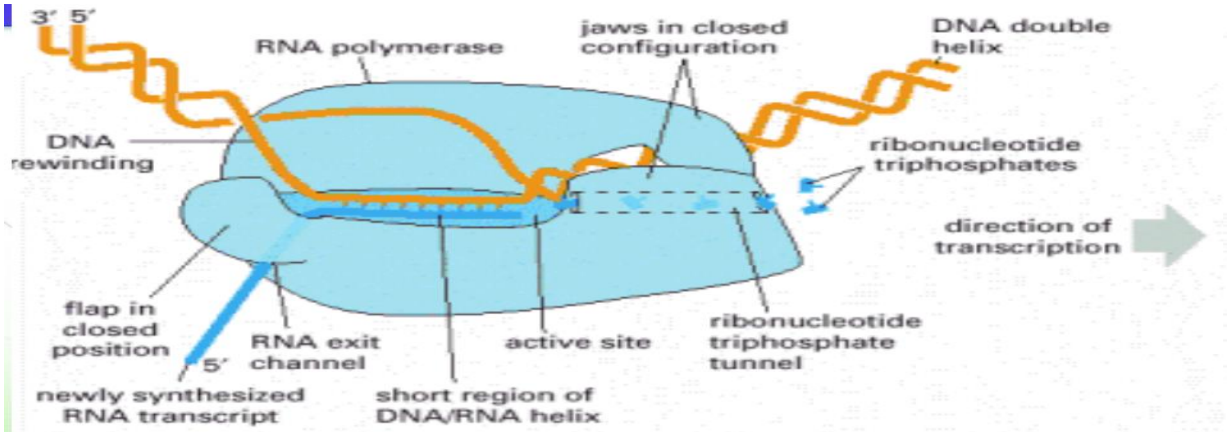
- One of the DNA strands is **the template** and the other strand is called **the coding strand**, The RNA is complementary to DNA template strand and is the same as the coding strand with the same direction but the RNA has U instead of T as in the picture below.



DNA polymerase	RNA polymerase
Deoxyribonucleotide as a substrate	Ribonucleotide as a substrate (nucleoside triphosphates like ATP, CTP, GTP and UTP)
Requires a PRIMER	Doesn't require a PRIMER
Has proof reading	Has proof reading
Less error prone (1 mistake per 100million)	More error prone (1 mistake per 10.000)

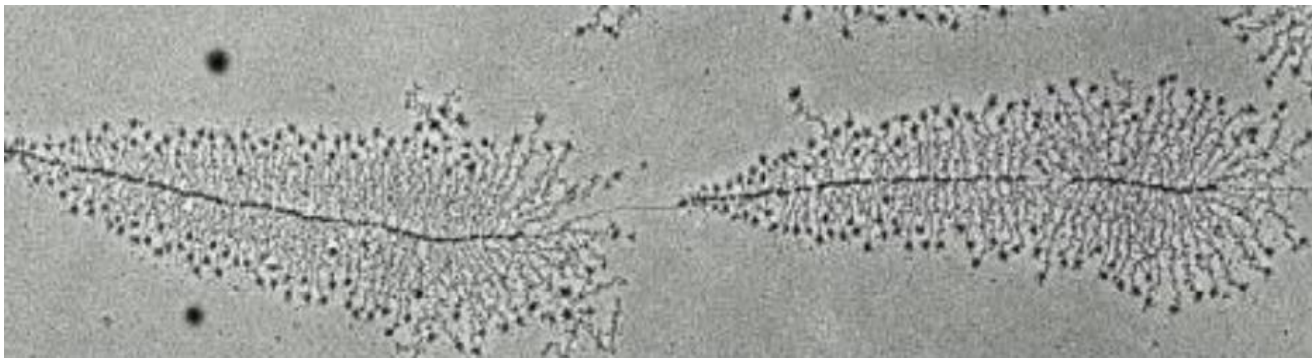
Mistakes in RNA are less harmful than the mistakes in DNA, because for example if we have a mistake in 1 mRNA, it will result in a mistake in 1 protein and the lifespan for proteins is usually short, whereas if we have a mistake in DNA it would be long lasting and will be inherited through generations.

POLYSOMES



The RNA polymerase is a large protein which makes a bubble when it binds to DNA

There are H bonds between the RNA and the DNA template which stabilize the whole complex and as the RNA polymerase is moving forward, the older polymerized RNA nucleotides are separated and the newer ones become bonded, thus allowing the simultaneous synthesis of many RNA chains from the same gene, forming structures known as **POLYSOMES.** (Multiple transcription at the same time on the same gene, which saves time)



Here we have two genes and multiple RNA polymerases at the same time on each gene

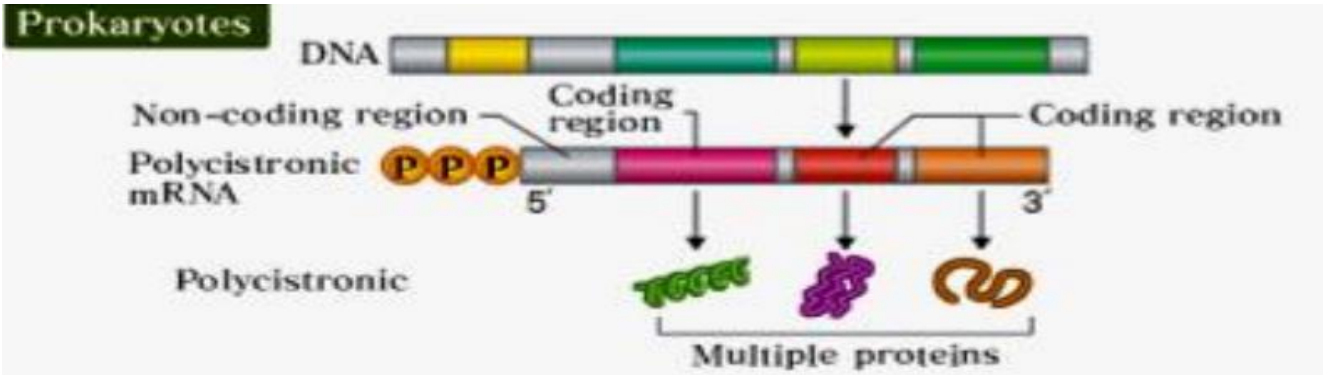
The beginning of the transcription is at the left side on each gene which shows short chains of RNA unlike the right side which shows long chains of RNA

Transcription in prokaryotes

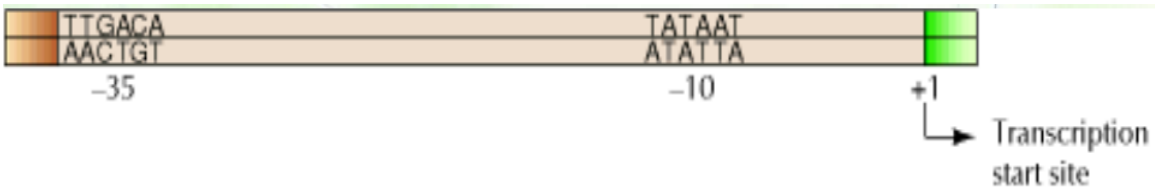
Prokaryotic genes are organized in a certain way; they're either organized as Monocistronic or Polycistronic

- **Monocistronic:** one gene makes **one mRNA** that makes one polypeptide.
- **Operon** which is a cluster of genes, all of them get transcribed together as one unit to form **one mRNA**, different parts of mRNA can code for different polypeptides that forms different proteins with totally different function **(Polycistronic)** but all proteins are related to the same pathway e.g Lac operon, Trp operon.

The operon is an efficient way to produce multiple enzymes for the same pathway from one mRNA all at the same time. (different proteins but all contributes in metabolism of lactose or tryptophan….)



How to know where to start?!



By **the promoter** (part of the gene but it doesn't code for anything), it's the DNA sequence to which the RNA polymerase binds in order to initiate transcription of a gene (**RNA polymerase binding site).**

Transcription initiation site is the site where the first nucleotide is used to make RNA, we number that nucleotide as +1 and all the other nucleotides are also numbered positively or negatively.

Open reading frame is the DNA sequence that can be transcribed into mRNA from the first base to the last base (from +1 to the end of transcription)

Important terms regarding the sites: **Upstream** → from +1 to -10 or -10 to -35 for example
Downstream → from -35 to -10 or -10 to +1 and so on

So we say that +1 site is downstream of the -10 (relatively) and -10 is upstream of the +1.

The process of transcription is always **downstream** (+1, +2, +3,…) (downstream is always easier than upstream)

Just before the transcription initiation site, we have two non-coding fixed **consensus sequences** in a variety of genes (متفق عليهم) numbered negatively (-10,-35), these consensus sequences help the RNA polymerase to recognise and start transcription (they are actually ☺ the promoter)

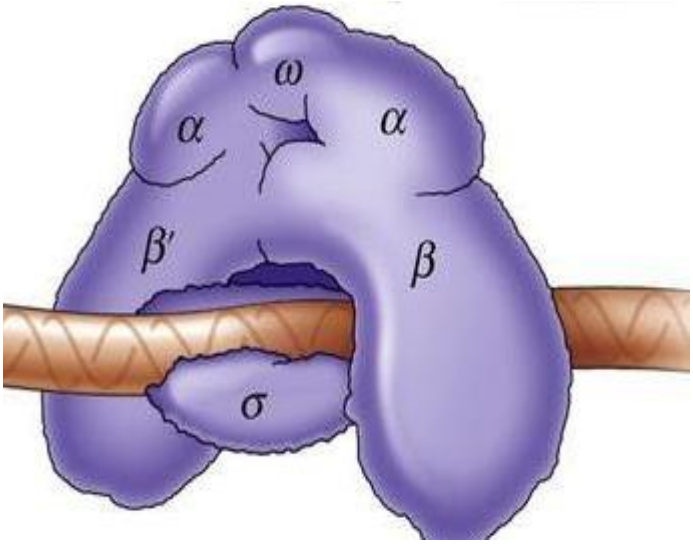
They are called the (-10) and (-35) elements because they are located approximately 10 and 35 base pairs upstream of the transcription start site.

The direction of synthesis is from (-35 → -10 → +1 ….), if we flip these two sites (-35,-10), the direction of synthesis will be flipped, and that's how the promoter determines the direction of transcription and which strand on the DNA will be the one used as a template.

The -35, -10 (promoter) sites are upstream to the +1 site (transcription initiation site).

- If we change the sequence around the consensus sequences (-10 or -35), the strength of binding is affected, therefore the efficiency of transcription is reduced.
- If the change is in the sequences between the consensus sequences, the transcription efficiency won't be affected at all.

The RNA polymerase



It's a huge protein made of multiple subunits

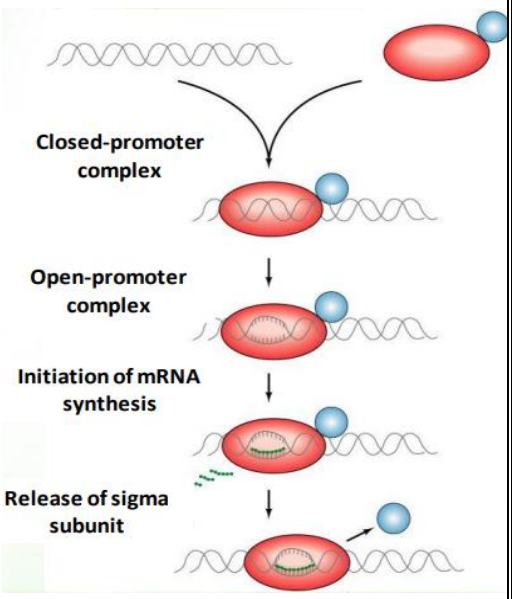
- Two α, one β, one β', and one ω (omega) subunits which form something called the **core polymerase**
- The core polymerase is fully capable of catalysing the polymerization of NTPs (NucleosideTriPhosphate) into RNA, so if any of these subunits is removed, the enzyme will become **functionless**.
- The sigma subunit (σ) is **not** a part of core enzyme, **it guides and directs the enzyme** to the -35 and -10 promoter sites so if it's removed, the enzyme can still function by making phosphodiester bonds but the ability to find the promoter and bind to it will be affected and would take much more time (low affinity and non-specific binding which results in a less efficient enzyme)

Transcription is divided into 3 stages - replication and translation have the same stages but now we will talk about transcription-

1) INITIATION:

a- RNA polymerase catches and hangs on the promoter site (-35, -10) forming the **closed promoter complex.**

b- RNA polymerase separates the strands of DNA by breaking H bonds between both strands (just like helicase enzyme in replication) **forming open promoter complex**



c- As the transcription progresses, the sigma subunit isn't required anymore as the RNA polymerase has already recognized the promoter site and is bound to it, so **the sigma subunit dissociates** (this happens after the addition of about the first 10 nucleotides, reaching the +10 site) and binds to another RNA polymerase to guide it to the promoter region again.

- When the sigma subunit dissociates, the RNA polymerase stays fixed because of the H bonds between the template and the RNA which stabilize the whole complex.

2) ELONGATION

As the polymerase moves forward, it

- Unwinds the template DNA ahead of it
- Elongates the RNA (adds more nucleotides)
- Rewinds the DNA behind it

> Imagine that the RNA polymerase has 6 hands: 2 in the front to open the strands (unwinding), 2 for adding the nucleotides (synthesis) and 2 behind for rewinding of DNA strands

3) Termination

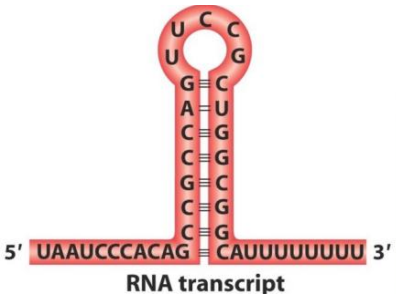
RNA synthesis continues until the polymerase reaches a termination signal (consensus sequence) where the RNA is released from the polymerase, and the enzyme dissociates from its DNA template.

There are 2 mechanisms of termination:

a-Rho independent

On the gene, we have a G-C rich region followed by a lot of AAAAA… **(talking about one strand)**

On RNA these will be transcripted to G-C rich region forming H bonds(inverted) between the single strand, which leads to **stem** formation followed by a lot of UUUUU…, this will form weak H bonds with the template (stem-loop structure).



RNA transcript

- When the stem-loop structure forms, imagine that the ground will shake, leading to the dissociation of RNA polymerase due to the weak H bonds
- Additional ☺ :- the stem is formed because of the complementarity of the bases in the RNA which will bind together.

In the loop, there is no complementarity.

b- Rho dependent

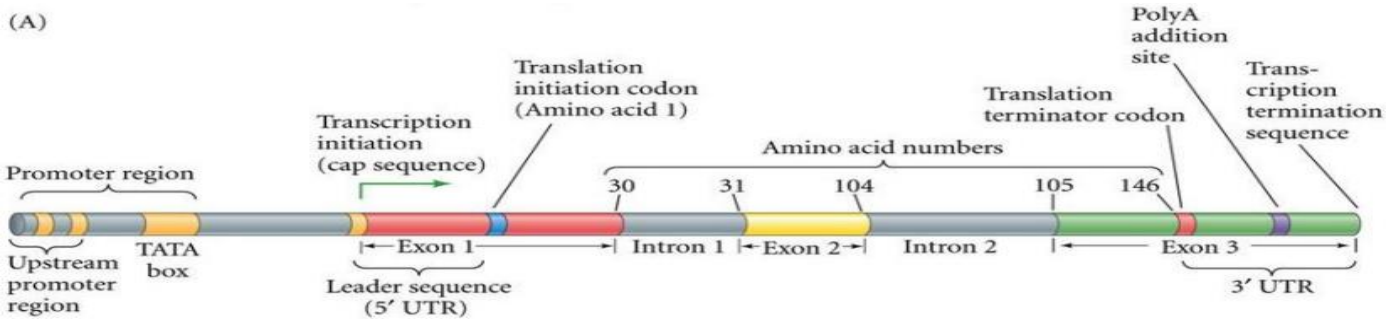
Depends on a protein known as **Rho** which recognizes a specific sequence on the RNA and binds to it, and then follows the RNA polymerase along the transcript.

When the RNA polymerase stalls at a stem-loop structure (note that there's no UUUU in the Rho-dependent termination), the RNA polymerase **slows down**, so the Rho protein finally catches up to the RNA polymerase and binds to it, breaking the RNA-DNA base pairs and releasing the transcript, ultimately causing termination of transcription.

> Note that Rho-independent termination signals have the string of U residues at the end of the RNA, whereas Rho-dependent termination signals don't.

Watch this https://www.youtube.com/watch?v=MjfdDW84A08

TRANSCRIPTION IN EUKAROYOTES



The Eukaryotic genes have **promoters with consensus sequences, transcription starting site (+1) and open reading frame just like prokaryotes**.

In addition to that, there are other sequences such as **proximal promoter element** (explained next lecture)

The resulting mRNA needs **PROCESSING** (not all translated), including removing some parts called **introns** and connecting other parts called **exons**.

The translation doesn't start from the beginning of the first exon, there is a region that won't be translated, which is called **UnTranslatedRegion (UTR 5')** , this also applies for the last exon **(UTR 3')**

Last things will be repeated later on

GOOOOOOOOOOOOOOOOOOOOOOOOD LUCK