Molecular Biology & Genetics

Dr.Said Ismail

Lecture #4

❇ Transcription means to copy the DNA into RNA.
- But why do we copy the DNA, why don't we use it every time we need it? Because the cell can`t afford getting the genetic material from nucleus to the cytoplasm each time the cell needs a protein.

❇ Plus the genetic material is preserved in the nucleus for protection and preservation, like this old map that is preserved in a library, you can`t take the original map but you can take copies of the parts you need, and after you are done with them you can dispose of them but the old map will remain safe and the same goes for DNA, the gene we need is transcripted into mRNA which goes to the cytoplasm to be translated, then it will be destroyed, mRNA is a very short lived messenger.

❇ **<u>We have 3 types of RNA:</u>**

1. mRNA (messeneger RNA):

-Messenger RNA, from its name you can illustrate the function which is delivering a message.

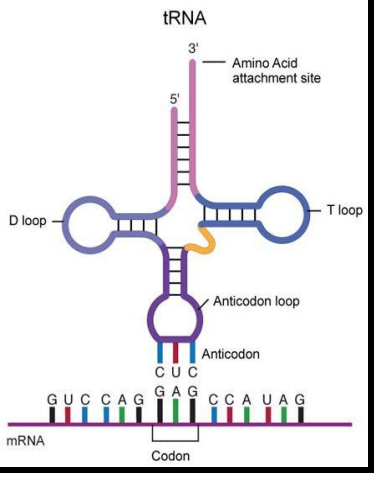
- mRNA length differs according to the length of the gene.
- mRNA has to have 3 processing events before leaving to the cytoplasm:
  1. On the 5` end we put a G-cap (GTP cap) attached to it **(capping)**.
  2. On the 3`end we put a poly. A tail **(polyadenylation).**
  3. Introns (non-coding sequences) will be cut out **(splicing).**

2. tRNA (transfer RNA):

-This is the true translator from codons (nucleic acids) into amino acids (proteins).

- It is a single stranded RNA made of 80 nucleotides (scientists thought there was no dsRNA at all until they discovered the microRNA). tRNA looks like a clover (a name of a rose).

- The 5` end will make loops till it reaches the 3` end (which is attached to the amino acid by an **ester bond)** , tRNA can`t stay straight , it has to make a 2ry structure between intra-complementary sequences (hybridization)



forming the stem of the tRNA, and when there are **NO** intra-complementarities it will form a loop.

* **Inter-complementarity**: between different molecules.

* **Intra-complementarity**: within the same molecule.

We call these loops **hairpin loops** because they look like it. The tRNA contains 3 hairpin loops according to the complementarities it contains, the most important one is the 2nd one because it will expose 3 nucleotides called the **anticodon**, which recognizes a codon on the mRNA. So, each codon on mRNA has an anticodon on a tRNA, and each specific anticodon on tRNA has a specific amino acid and matchin anticodons with their AA's is the function of enzymes called **aminoacyl tRNA synthetases**; a family that contains 50-60 enzymes, they read the anticodon on loop no.2 then they decide which amino acid to put on the tRNA.

3.rRNA (ribosomal RNA):

- it is a functional and a structural component of the ribosome .

- Structural: it is a part of the small and large subunits of the ribosome.

- Functional: one of the rRNA's in the large subunit shows an enzymatic activity, it is called **peptidyl transferase**. This enzyme joins the peptide bond between the amino acids. (By this, the old theory that said that all enzymes are proteins is proved false; this enzyme for example is an RNA not a protein).

❇ RNA polymerases:

- Direction: 5`→3`
- Substrates/precursors: "NTPs: nucleoside triphosphates": ATP, GTP, CTP and **U**TP.
- The nucleoside triphosphate that comes will be broken down into a pyrophosphate (to provide energy for the synthesis process) and a monophosphate which is the one that will be joined.

** RNA polymerases start the transcription from a place just before the start site that is called **PROMOTER**. The promoter exists before every single gene and it's the key for starting the transcription. When the promoter is working the RNA polymerase will be recruited and will copy the gene downstream the promoter. The promoter is helped by a sequence called **the enhancer** (to be discussed later in the sheet).

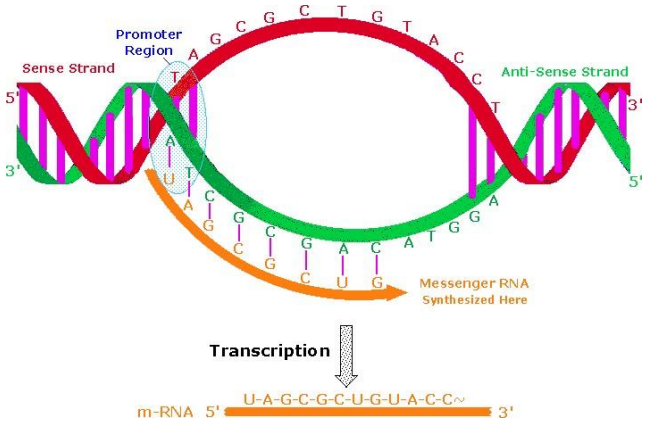
REMEMBER we have two strands of DNA: the coding and the non-coding (template). The RNA polymerase copies one strand which is the non-coding (the template) so that we get a strand that looks like the coding strand of the DNA except that there is a U instead of T. And that's actually why we have two strands for DNA.

-Types of polymerases:

1. In **prokaryotes** the polymerase is made of **5** subunits :
   #core enzyme: **4** subunits (**2** alpha & **2** Beta) this is the one that does the polymerization but it can't initiate the process by itself. It needs the help of the sigma.
   #Holoenzyme (**1** sigma factor)
   The function of the sigma, it leads the core enzyme to the correct promoter of the gene then leaves it to start copying the gene.
   ## Every group of enzymes has a different sigma factor (there are different sigma factors for different enzymes).

2. In **eukaryotes** we have 3 RNA polymerases
   - Pol. I produces most of the rRNA
   - Pol. II produces mRNA (the one important for expression)
   - Pol. III produces small RNAs such as tRNA and 5S rRNA.

❇ **The Transcription Process:**



The start:

- We already said that we copy the **template / non-coding/anti-sense** DNA strand to have an RNA strand that is similar to the coding/sense DNA strand.
- We know that we have codons on the mRNA and anti-codons on the tRNA to translate it into Amino acids.
- We said that we have a **promoter** to specify for the RNA polymerase where to start the transcription.
- What about stopping? We have a **TERMINATOR** after every gene.
  - In humans we call the terminator " the poly A signal "

  -Before every gene there is a promoter and after it a terminator.

The location of the process:

- The RNA polymerase starts the transcription AFTER the promoter, so it will **NOT** copy the promoter. That's why we say that the "transcription start point" is **Downstream** the promoter. "Downstream means in the direction of the 3`". For example if we take the start point as a reference the promoter will be upstream or 5` while the terminator will be downstream or 3`.
- To specify the site we can use the words "downstream proximal" and "downstream distal" according to the distance from the reference (here it

is the start point) for ex. The terminator is downstream distal to the starting point, while downstream proximal means close to the start point, also in the direction of the 3`.

- And we can be more and more specific by using numbers. For ex. We say that the promoter is between -10 and – 35 with reference to the start point, while after it will be positive.
- Don't mix between the start and the end points of transcription and the start and the stop points of translation.

  ***** The transcription starts from the nucleotide **directly** after the promoter (the start point (+1)) and it ends at the terminator nothing sooner or further.

  ****** But in translation the ribosome keeps searching for a start codon "AUG" to begin (it doesn't start at the first codon of mRNA); it could be -AUG- after 9 nucleotides and could be after 900! So the area that is **not** translated before the start codon is called **5`UTR "un-translated region" or leader region.**

  Then it will keep translating till it finds a stop codon, which also doesn't have to be the last nucleotide, the mRNA could have 10 or 1000 nucleotides after it! So we have a region After the stop codon that won`t be translated we call it **3`UTR or trailer.**

***Conclusion: the beginning and ending of the transcription id determined by the promoter and the terminator, while those of the translation are determined by the start and stop codons.**

The promoter:

- We call it the cap site.
- It is not just one sequence, it has promoter elements (different sequences), all of them together form the promoter region, and the most important component of it is the **TATA box;** named this way because it is rich in T and A.
- **Box**: a small sequence rich in certain nucleotides.

  1. TATA box: is the most conserved or consensus sequence in the promoters of most of the living things.

  We also have it at a certain site in almost all our genes which is upstream of the cap site / transcription start site, between -20 and -30.

  2. **GC and CAAT boxes** are other main components. The GC box is more conserved and is rich with G and C. Between GC box and the TATA box are -20 to -110 nucleotides.

  The Mechanism:

  The promoter actually isn't the one that recruits the RNA pol. Or makes it start copying because the RNA pol. Can't read the promoter, the ones

that read it are proteins called **transcription factors**, the transcription factors attach to the promoter to make it work, it can`t actually work without them, and then they together recruit the RNA pol. To start working, and they tell it where to start exactly. Then, when there are enough copies of the gene for the cell, the transcription factors leave the promoter so the RNA poly. stops working.

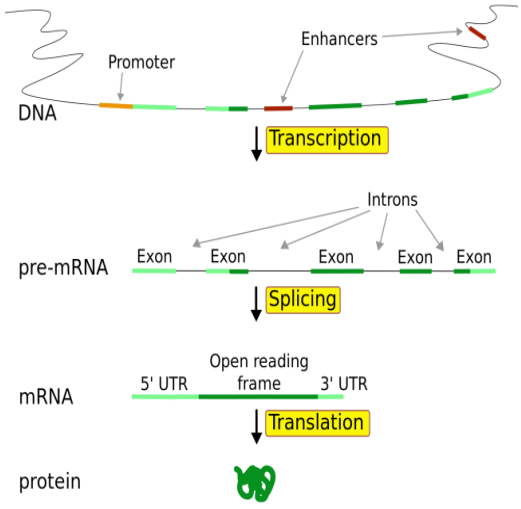
- The **terminator sequence /polyAdenylation signal/ polyA signal/ AATAAA** is the one who tells the RNA pol. where to stop. But it won't stop immediately because it was moving really fast! It transcripts a bit after it like 30 or 40 nucleotides (due to its momentum) then it falls down; so there is some transcription after the poly.A signal.

- The promoter gets help also from the **Enhancer**, it enhances the function of the promoter, which means that if the promoter was alone with its transcription factors it will transcript let`s say 10-20 copies but with the enhancer, it can transcript 20 million copies. " Referring to slide 81figure " We can see a break--//-- which means that the distance is not certain, it can be close (immediately upstream the promoter) or farther (10000 nucleotides upstream the promoter).

- The enhancer doesn't have a consensus (constant) position or sequence <u>unlike</u> the promoter elements.

- The promoter is the same among all genes while the enhancer is different among genes and that's why the cell can transcript this gene but not that gene even though they have similar promoters. And since the enhancer differs in sequence and position among genes, the transcription factors that will attach to this enhancer will also differ from gene to gene.(the transcription factors are the ones that do the work in both the enhancer and the promoter)

- We have AGGT and another AGGT on each side of the borders of an intron (intron boundaries). It helps the splicing proteins know the beginning and the ending of the intron in order to cut it. AGGT is a part of the boundary, there are other 13 sequences but AGGT is the most conserved among them.



-The first RNA that results (from the beginning of the CAP site to a little after the polyA signal) is called **the pre-mRNA/ primary RNA/ hnRNA** "heterogenous nuclear RNA", heterogenous because it is a mixture of both introns and exons, and nuclear because it is still in the nucleus.

-Then we have **mRNA (mature/messenger/final)** that is produced by splicing the hnRNA, has NO introns, it goes to the cytoplasm to be translated, it is much shorter than the hnRNA (because the introns have been removed and usually the introns are much longer than the exons). For ex. The globin gene is 10000 nucleotides and the hnRNA is 10000 (a bit less or more) while the mRNA is 1000.

NOTE that: Our single hnRNA gives a single mRNA and the mRNA gives a single protein. (monocistronic RNA), but one gene can give more than one RNA (alternative splicing) but each one of these mRNA's will only give one protein.

Bacteria`s one mRNA can give more than one protein, because it may contain more than one start and one stop codons. (polycistronic RNA). But in the bacteria a gene can only give one mRNA because there is no splicing (no introns in bacterial genes).

So, the bacteria have both; poly and monocistronic while we only have monocistronic.

- The transcription is a really complicated process, we have this promoter that is attached to 10-12 transcription factors but it won`t work alone, it needs the enhancer which is also attached to 10-12 transcription factors, and they (the promoter and the enhancer) need mediators (20-30 proteins) to communicate together, all of this is done to recruit the RNA pol. to start transcripting the needed no. of copies.

As we said before mRNA must go into 3 processes before going to the cytoplasm.

- 1.**5` capping** (putting methylated GTP)
- 2. **3` polyadenylation** (it consists of 200-250 Adenine residues)
- **3. Splicing**

RNA processing events before reaching the cytoplasm (in details):

1. **Capping**:

-on the 5` end

-GTP attaches as triphosphate plus it is methylated. But why do we add it? We add it for the protection of the RNA from degradation and to facilitate the identification of the 5` end by the ribosome when it starts translating this mRNA.

2. Adding a Poly A. tail:

- it is at the 3`end, it is copied to the mRNA as we said because the promoter is moving very fast so it will pass the polyA signal by 30-40 nucleotides before it falls off (stops).

- Then there are some cleavage enzymes that cut a no. of nucleotides that is half the extra transcrited no. after the poly.A signal, for ex. if 40 were transcripted after poly.A tail, then 20 nucleotides will be cleaved from the end.

- Then we are going to have the poly.A polymerase → this gene puts about 200-250 Adenine residues then the RNA leaves to the cytoplasm. Again why is it there?

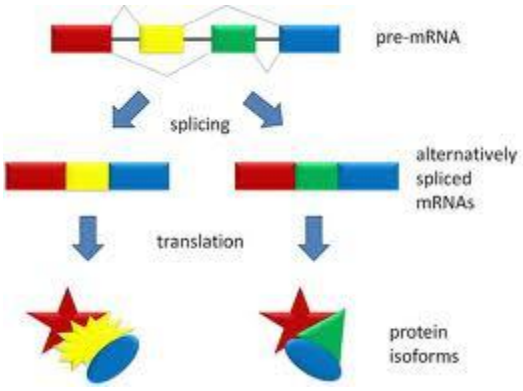
a. For protection, because we said that the RNA is a short lived material, it is broken down by RNAases once it is translated to prevent its accumulation. Most RNAases start the breaking process from the 3`end. And because we don't want the processing of breaking down to be so fast (the mRNA needs some time to be translated) we add the polyA tail which will delay the RNAases before they reach the coding sequence which is being translated into protein. (It prolongs the half-life of the mRNA in the cytoplasm).

b. It helps in the export of the mature RNA from the nucleus to the cytoplasm, this process is done by chaperons they works as shuttling molecules between the cytoplasm and the nucleus; they enter the nucleus, take the mature mRNA then deliver it to the cytoplasm, then they go back to the nucleus to bring more. But the nucleus has huge numbers of RNA's some of them are mature while others are not. So, How can the chaperons know which RNA is mature and which is not? By having the tail, the one that has the poly.A tail is ready to be exported; they actually attach to the poly.A tail and export the mRNA to the cytoplasm.

3. RNA splicing:

Introns need to be removed very accurately by its boundaries as mentioned before, like AGGU, so this is done by the splicing proteins in high accuracy. Because one mistake here (one extra or less nucleotide) will make a frame shift.

We have "**alternative splicing"**; which means that we cut the introns differently each time, to have more diversity in the proteins. (Some introns become exons at times, while some exons become introns, depending on the protein we want to produce).



Example: if we have 5 paragraphs, 1, 3 and 5 (the exons) give a meaning if we remove 2 and 4 (here 2 and 4 are the introns). If we want another meaning we can remove 3 and 4 (they are the

introns here), so that 1, 2 and 5 (the exons) will make a meaning, but it is different than the first meaning.

P.S: Sorry it`s hard to explain this lecture by writing it, here are some links that could be useful to you:

http://www.youtube.com/watch?v=WsofH466lqk

http://www.youtube.com/watch?v=SMtWvDbfHLo

Done by: Wala'a Alsalhi