Lesson 05: Protein synthesis

1- Protein synthesis tools: Gene, ribosomes and tRNA

1-1-The messenger RNA (mRNA)

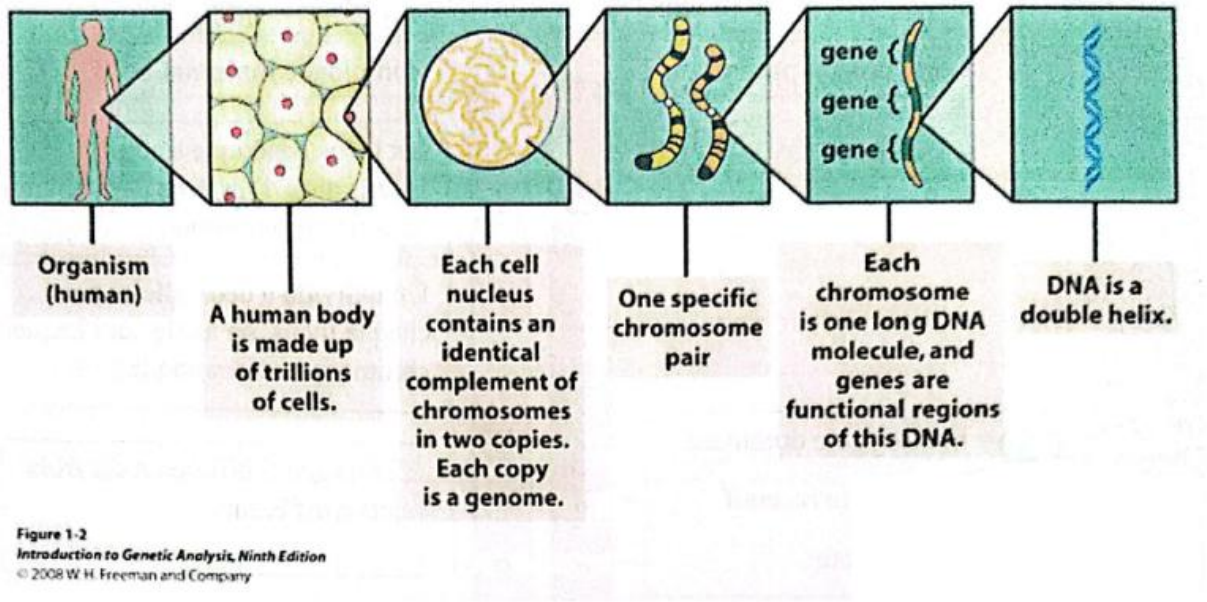


Figure: From a living organism to a gene

Genes contain the instructions that enable cells to polymerize amino acids in a specific order, thereby synthesizing specific proteins.
The precise sequence of nucleotides along the DNA molecule encodes the information required for protein synthesis.
Genes are hereditary factors responsible for the phenotypic characteristics of an individual, although a single trait may be influenced by several genes.

- **Gene structure**

Eukaryotic genes have a more complex structure than prokaryotic genes. They contain:

- **Promoter:** A regulatory sequence located upstream of the gene, adjacent to the transcription initiation site, which controls gene activation.
- **Exons:** In eukaryotes, the coding information of a gene is often divided into multiple coding sequences called exons, separated by introns.
- **Introns:** Non-coding sequences within the gene. Introns are generally longer than exons and may constitute the majority of the gene. The number and length of introns and exons are highly variable.

- **Terminator:** A sequence located at the end of the gene, adjacent to the transcription termination site, which signals the end of transcription.

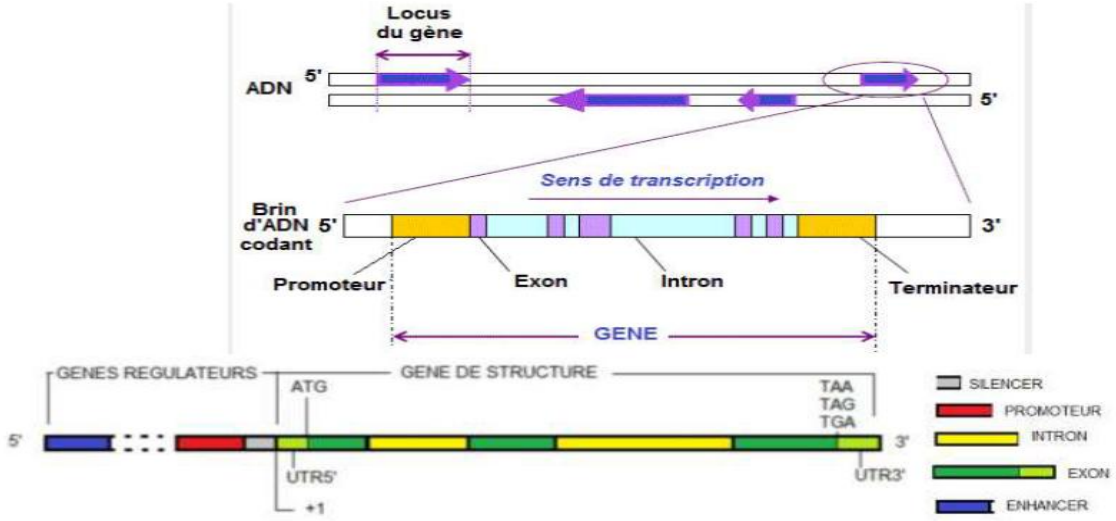


Figure : Structure of a gene in eukaryotes

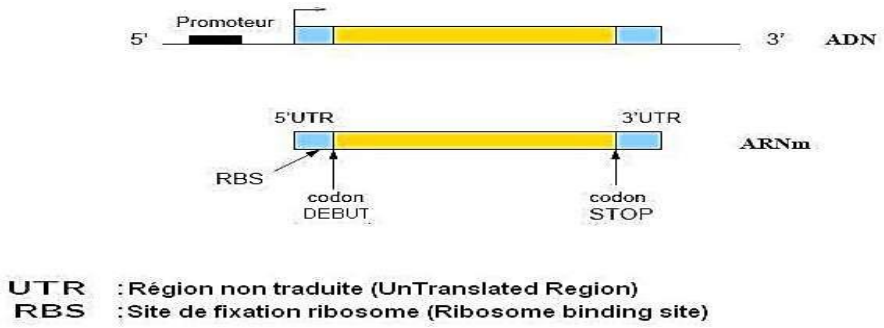


Figure : Structure of a gene in prokaryotes

1-2- The ribosome

Ribosomes are complex "organelles" composed of:

- Ribosomal proteins
- rRNA, synthesized in a specific region of the nucleoplasm called the nucleolus

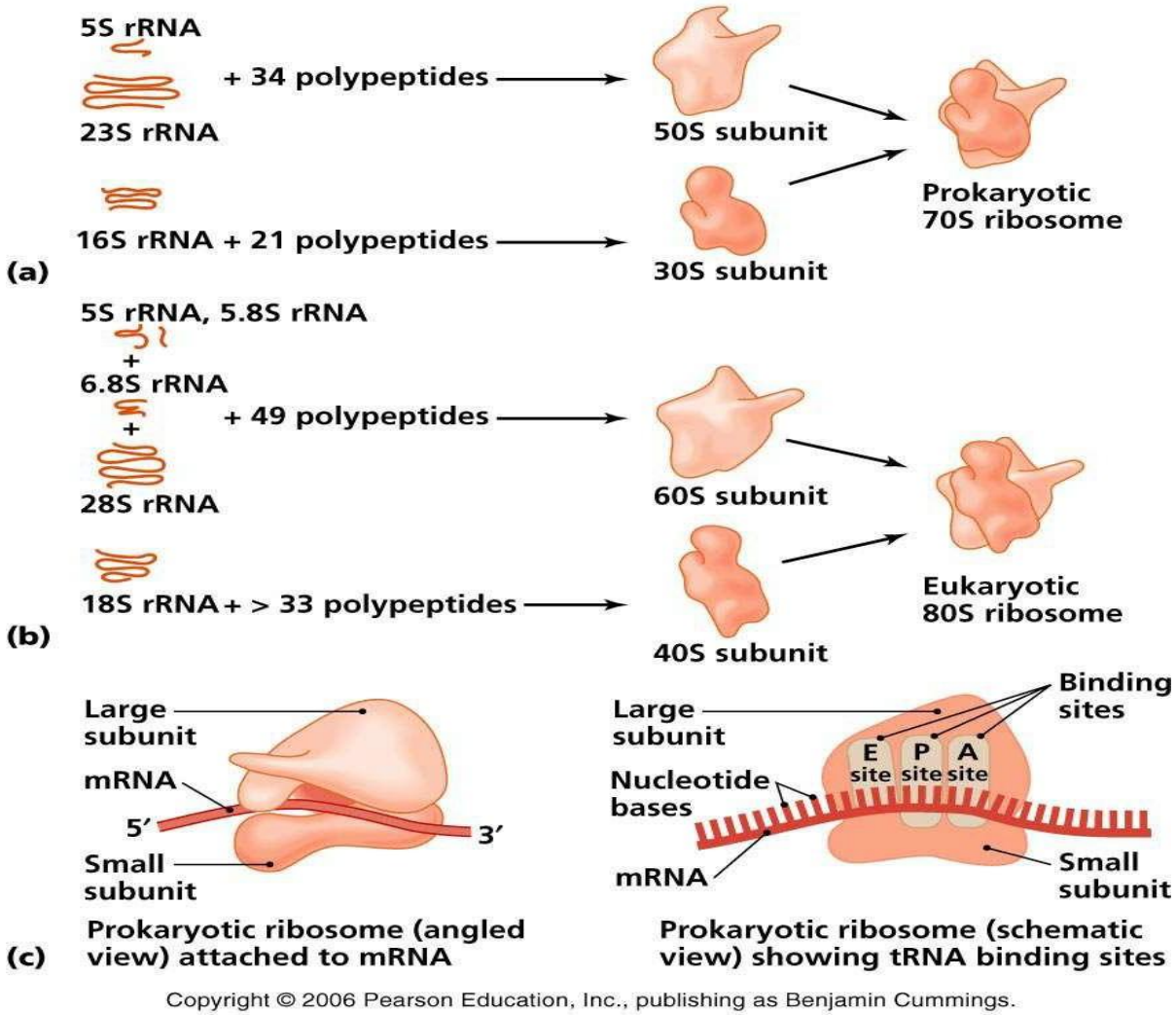
Each ribosome consists of two ribonucleoprotein subunits, generally designated by their sedimentation coefficient in Svedbergs (S):

- **The big subunit**(Large subunit) **:** 50S at the prokaryotes and 60S at the eukaryotes
- **The small subunit :** 30S at the prokaryotes and 40S at the eukaryotes

The synthesis of the different ribosomal subunits occurs in the cell nucleus. The subunits then move into the cytoplasm through nuclear pores.

In the cytoplasm, ribosomes may be found free in the cytosol or bound to the rough endoplasmic reticulum (RER). The ribosomal subunits associate with mRNA to participate in protein synthesis. Both subunits are essential for translation and are assembled during the initiation phase of this process.

The small ribosomal subunit recognizes and binds to the mRNA, ensuring proper alignment. The large subunit then joins to complete the ribosome and provides the recognition and catalytic sites for each tRNA carrying an amino acid.



- **Figure : Structure and types of ribosomes**

1-3- tRNA (RNA of transfer)

tRNAs are small molecules (about 70 nucleotides) synthesized in the nucleoplasm, with a secondary structure in the shape of a "cloverleaf," as shown in the figure below. Two characteristic regions can be distinguished:

- **Acceptor arm:** Formed by the two ends of the molecule. The 3′ end binds the amino acid, and this sequence is invariable in all tRNAs (CCA).
- **Double-helical regions with loops:** These include the **anticodon loop**, which contains the specific triplet that serves as the recognition site for the complementary codon on the mRNA bound to the ribosome.

tRNAs have two essential functions: (1) the ability to bind to a specific amino acid, and (2) the ability to recognize a specific codon through their complementary anticodon.

Anticodon–codon recognition is based on base complementarity and involves the primary structure of the tRNA. By contrast, the specific recognition of an amino acid is more complex and depends on the three-dimensional structure of the molecule.

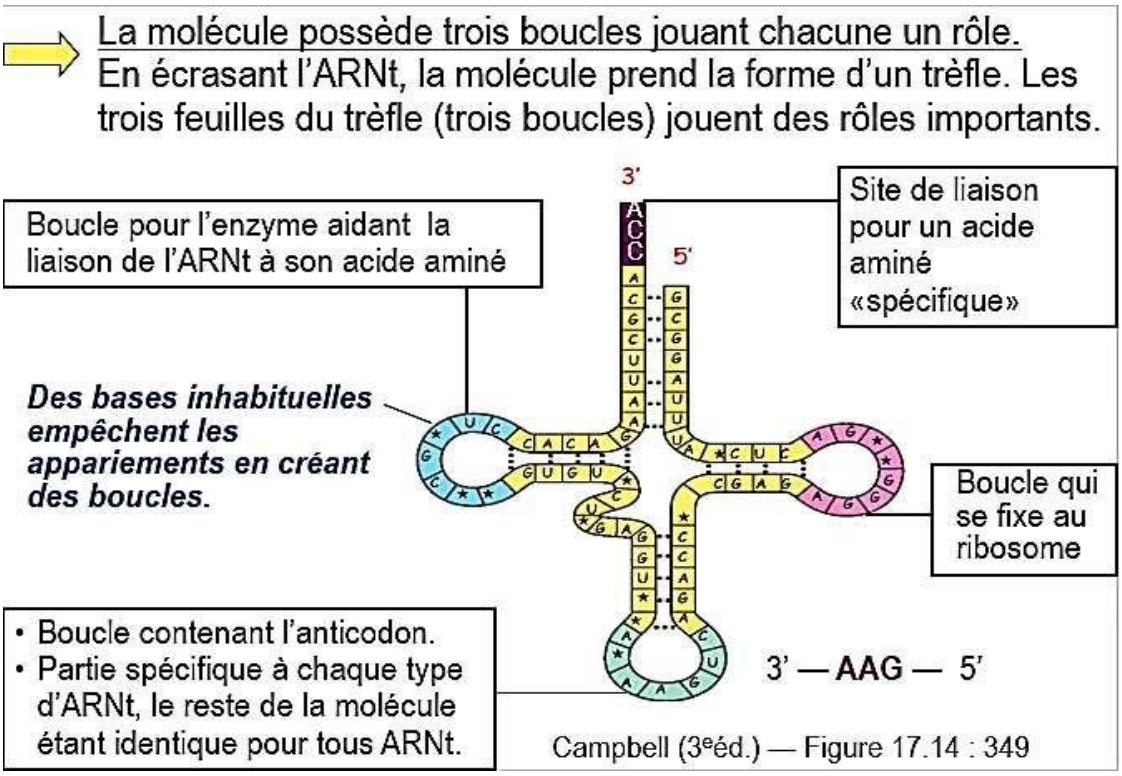


Figure : Structure of a tRNA

2- Transcription

By definition, transcription is the synthesis of an RNA molecule from a DNA template. Genes are transcribed only when their products are required by the cell. These products correspond either to a polypeptide chain or to a functional RNA.

Transcription is an RNA biosynthesis process based on base complementarity:

- Only small portions of the genome are transcribed at any given time in the life of the cell.
- Transcription begins at a precise point on the DNA and ends at another precise point; the region between these two points constitutes a transcription unit.
- Only one strand of DNA is transcribed (the 3′→5′ strand), which serves as the template for ribonucleotide polymerization. The resulting RNA molecule has a 5′→3′ orientation, corresponding to the NH_2→COOH orientation of the protein. The genetic code is read (translation) in the same direction as transcription.
- The entire transcription product does not necessarily correspond to the protein-coding sequence; at the 5′ end (upstream), a leader sequence enables the attachment of the mRNA to the ribosome.
- Transcription is carried out by RNA polymerase, which uses the single-stranded DNA template to polymerize ribonucleotides (local denaturation of the DNA molecule is required).
- Numerous protein factors are involved in ensuring the proper steps of transcription.
- A gene can be transcribed simultaneously by several RNA polymerases.
- There are important differences between transcription in prokaryotes and in eukaryotes.

- **The mechanism of transcription** (see figures below) **:**

  - **Initiation**

RNA polymerase binds to the gene promoter. The two DNA strands unwind at this site (weak hydrogen bonds are broken), and transcription begins on a single strand **(the 3′→5′ template strand).**

  - **Elongation**

Transcription proceeds in the 5′→3′ direction, catalyzed by RNA polymerases (which differ depending on the type of RNA being synthesized). In prokaryotes, a single RNA polymerase carries out transcription. In eukaryotes, RNA polymerase II synthesizes mRNA, while RNA

polymerases I and III synthesize other types of RNA.

ARN polymérase I : transcrit les gènes codant pour les deux molécules d'ARN ribosomial
ARN polymérase III : transcrit les gènes de l'ARN transfert, de l'ARN 5S, et quelques autres petites molécules d'ARN
ARN polymérase II : transcrit la plupart des gènes des eucaryotes qui codent pour des protéines, leur régulation est la plus complexe

ADN	ARN
A	U
T	A
C	G
G	C

In an antiparallel and complementary manner relative to the DNA template, RNA polymerase unwinds the DNA strand, exposing 10 to 20 bases at a time.

Ribonucleoside triphosphates (NTPs), synthesized in the cytoplasm and imported into the nucleus, pair with the exposed bases.

Once matched, the nucleotides are polymerized through the formation of phosphodiester bonds.

Behind the advancing transcription complex, the newly synthesized RNA strand detaches, and the DNA double helix rewinds.

- **Termination**

Transcription continues until the end of the termination region. At this point, the RNA transcript and the RNA polymerase are released.

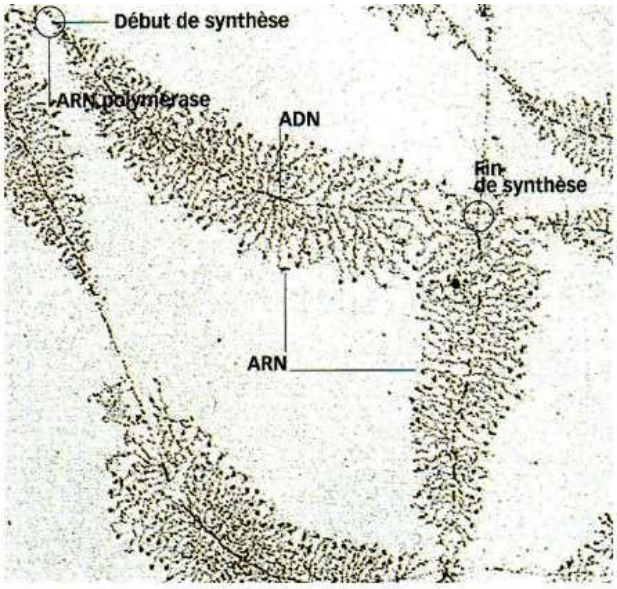


Figure: Microscopic observation of DNA transcription

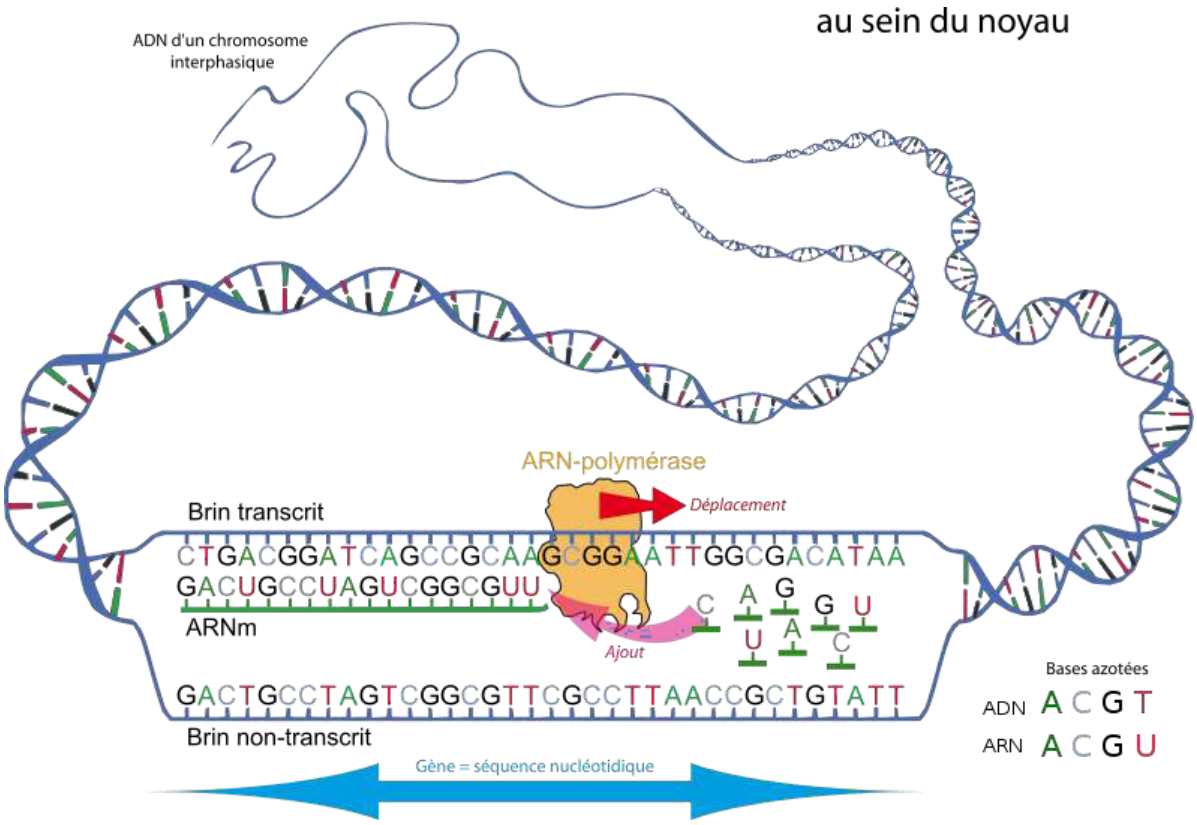


Figure : DNA Transcription

- **Prokaryotic transcription**

Transcription occurs in 3 steps:

1- Initiation

In prokaryotes, RNA polymerase binds directly to the promoter region. Once bound, the enzyme initiates local unwinding of the DNA helix to begin transcription.

2- Elongation

The region containing the RNA polymerase, the DNA template, and the nascent RNA is called the **transcription bubble**. At the 3′ end of the growing RNA strand, base pairing occurs with an incoming ribonucleoside triphosphate (NTP). As transcription progresses, the DNA that has been transcribed re-anneals and recoils into a double helix once it exits the transcription bubble.

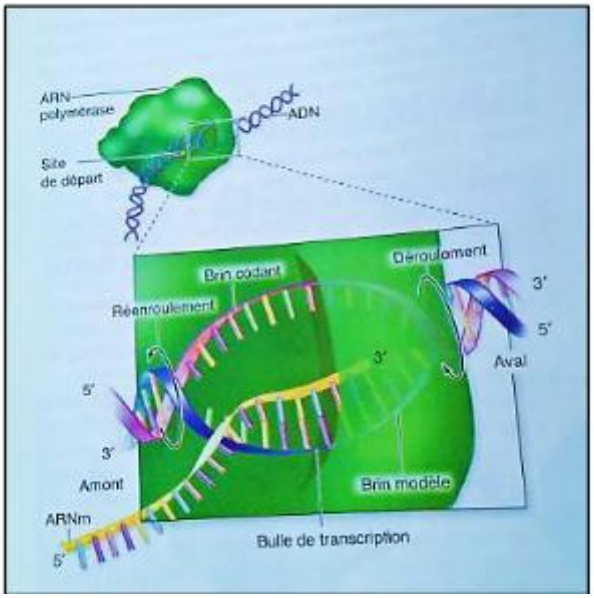


Figure: Schematic of a transcription bubble

3- Termination

The end of a bacterial transcription unit is marked by termination sequences that signal RNA polymerase to stop. In **rho-independent termination**, the RNA transcript folds into a hairpin structure followed by a stretch of at least four uracil nucleotides. This destabilizes the RNA–DNA hybrid within the transcription bubble, causing dissociation of the RNA transcript and RNA polymerase, and allowing the DNA helix to re-anneal.

- **Eukaryotic transcription**

1- Eukaryotic RNA polymerases

Most eukaryotic cells have three types of RNA polymerases:

- **RNA polymerase I:** Transcribes ribosomal RNAs (rRNAs), primarily 28S, 18S, and 5.8S rRNAs.
- **RNA polymerase II:** Transcribes precursor messenger RNAs (pre-mRNAs), small nuclear RNAs (snRNAs), some microRNAs (miRNAs), and certain small nucleolar RNAs (snoRNAs).
- **RNA polymerase III:** Transcribes small RNAs, including transfer RNAs (tRNAs), 5S ribosomal RNA, some microRNAs (miRNAs), and some small nuclear RNAs (snRNAs).

2- **Structure of the promoter**

RNA polymerase II promoters contain regulatory elements located upstream of the transcription initiation site. These DNA sequences are commonly referred to as “boxes”:

- **TATA box:** Located approximately −25 base pairs upstream of the transcription start site. It is a short sequence rich in adenine (A) and thymine (T). The consensus sequence most frequently observed is **TATAAA**.
- **GC box:** Typically located between −110 and −40. It often appears as the hexanucleotide sequence **5′-GGGCGG-3′**, and it may be present in multiple copies.
- **CCAAT box:** Usually located between −120 and −80. This sequence may occur before, after, or even between GC boxes.

3- The phases of transcription

- **Initiation complex**

In eukaryotes, RNA polymerase II binds to the promoter with the assistance of transcription factors, including several proteins (e.g., TFIIA, TFIIB, and others). Together with RNA polymerase II, these proteins form the **transcription initiation complex**, which catalyzes the formation of the first phosphodiester bond between the initial two nucleotides of the mRNA.

- **Modifications of the primary transcript (pre-mRNA)**

The primary transcript corresponds to the complete copy of a gene, including both exons and introns. The final functional form, known as **mature mRNA**, is obtained through the following modifications:

- **Addition of the 5′ cap**
  The first nucleotide of the transcript is usually an adenine (A) or a guanine (G). This nucleotide is modified by the addition of a GTP molecule to the 5′ phosphate group, forming the **5′ cap**. The cap protects the 5′ end of the mRNA from degradation by exonucleases and also plays a crucial role in the initiation of translation.

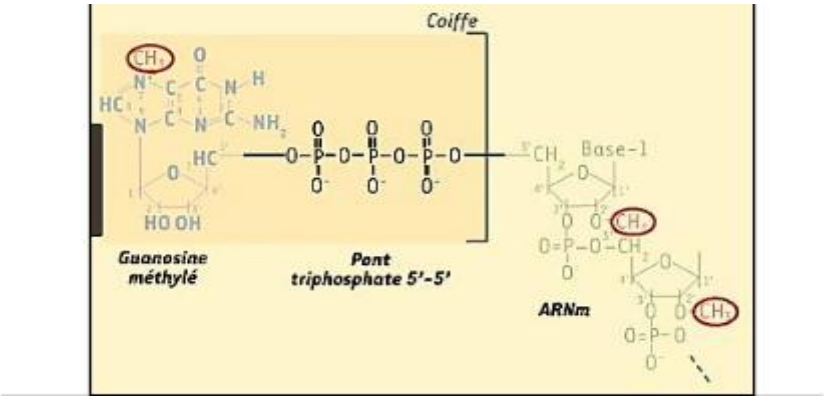


Figure: Molecular structure of the cap

- Addition of the poly(A) tail to the 3' end

After synthesis, mRNAs are cleaved by an endonuclease in their 3′ region, approximately twenty bases downstream of the specific sequence AAUAAA. Following this cleavage, the enzyme poly(A) polymerase, in the presence of ATP, adds a variable number of adenine residues. The poly(A) tail also provides a protective function for the mRNAs at the 3′ end and facilitates the attachment of ribosomes.

- Excision-splicing

Excision-splicing allows the maturation of the primary transcript into mRNA. This process involves the removal of introns by excision, followed by the splicing of exons (end-to-end joining of exons). Excision-splicing occurs within the spliceosome in the nucleus, before the mRNA is exported to the cytoplasm.

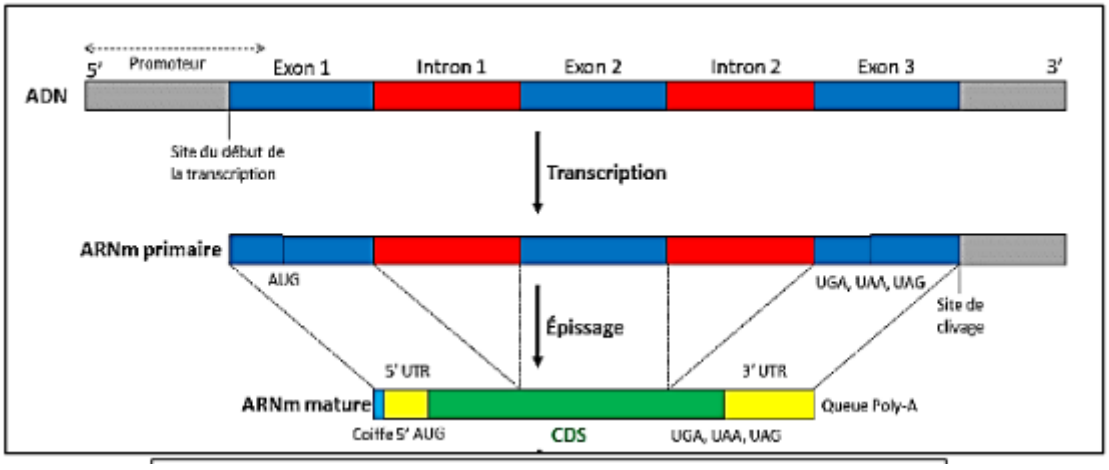


Figure: Excision-splicing process

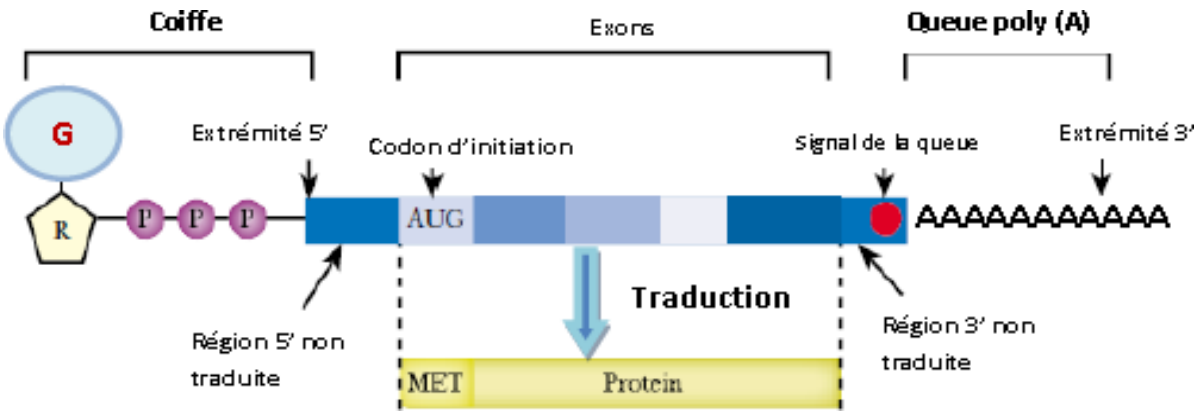


Figure : Mature mRNA

3- The translation

Translation is the mechanism by which the information encoded in nucleic acids (4-letter alphabet) is converted into protein sequences (20-letter alphabet) according to the universal genetic code. Translation involves matching the nucleotide sequence of the mRNA to the amino acid sequence of the corresponding protein.

The Genetic Code is the system that defines the correspondence between RNA nucleotide sequences and the amino acid sequences of synthesized proteins. The four nitrogenous bases of DNA constitute the genetic alphabet. The relationship between genetic information written with these four bases and the sequence of the polypeptide chain composed of 20 amino acids is determined by specific combinations of the four letters, which form the genetic code. The basic unit of the genetic code is called a codon, consisting of three consecutive bases.

.

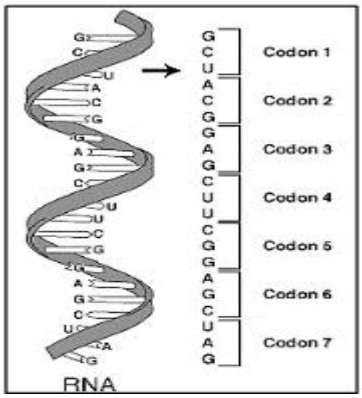
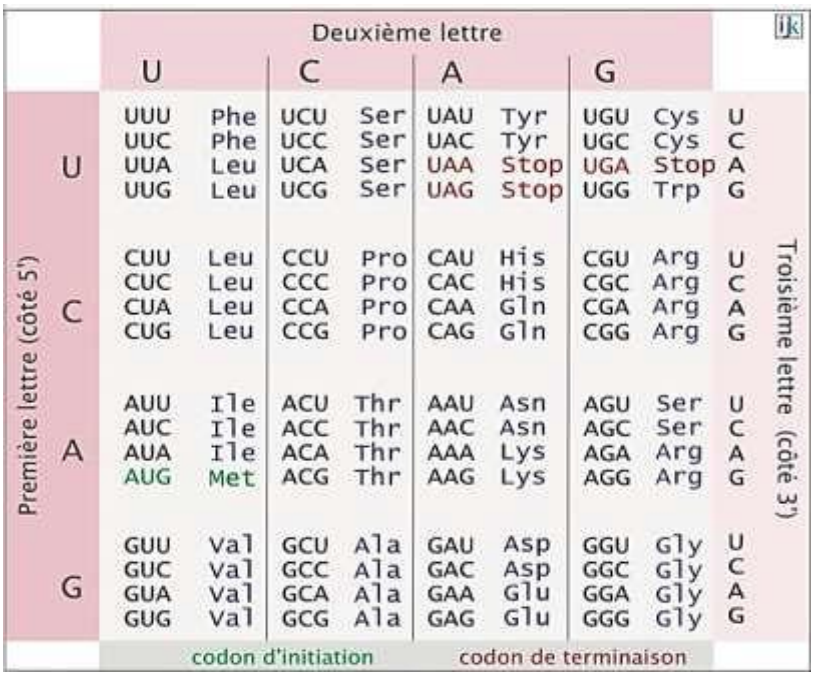


Figure: Succession of codons on an RNA molecule

With three nucleotides per codon, there are 43 = 64 different possible codons



Deuxième lettre									
	U		C		A		G		
Première lettre (côté 5')	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	U
	UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys	C
U	UUA	Leu	UCA	Ser	UAA	Stop	UGA	Stop	A
	UUG	Leu	UCG	Ser	UAG	Stop	UGG	Trp	G
	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	U
	CUC	Leu	CCC	Pro	CAC	His	CGC	Arg	C
C	CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg	A
	CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg	G
	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	U
	AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser	C
A	AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg	A
	AUG	Met	ACG	Thr	AAG	Lys	AGG	Arg	G
	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U
	GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly	C
G	GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly	A
	GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly	G
	codon d'initiation				codon de terminaison				

Troisième lettre (côté 3')

Acide aminé			Codons
Alanine	A	Ala	GCU, GCC, GCA, GCG.
Arginine	R	Arg	CGU, CGC, CGA, CGG ; AGA, AGG.
Asparagine	N	Asn	AAU, AAC.
Acide aspartique	D	Asp	GAU, GAC.
Cystéine	C	Cys	UGU, UGC.
Glutamine	Q	Gln	CAA, CAG.
Acide glutamique	E	Glu	GAA, GAG.
Glycine	G	Gly	GGU, GGC, GGA, GGG.
Histidine	H	His	CAU, CAC.
Isoleucine	I	Ile	AUU, AUC, AUA.
Leucine	L	Leu	UUA, UUG ; CUU, CUC, CUA, CUG.
Lysine	K	Lys	AAA, AAG.
Méthionine	M	Met	AUG.
Phénylalanine	F	Phe	UUU, UUC.
Proline	P	Pro	CCU, CCC, CCA, CCG.
Pyrrolysine	O	Pyl	UAG, avant élément PYLIS.
Sélénocystéine	U	Sec	UGA, avec séquence SECIS.
Sérine	S	Ser	UCU, UCC, UCA, UCG ; AGU, AGC.
Thréonine	T	Thr	ACU, ACC, ACA, ACG.
Tryptophane	W	Trp	UGG. (UGA)
Tyrosine	Y	Tyr	UAU, UAC.
Valine	V	Val	GUU, GUC, GUA, GUG.
Initiation			AUG. (UUG, CUG)
Terminaison	*		UAG, UAA ; UGA.

The genetic code has the following characteristics:

- Degeneracy

Of the 64 codons, three are stop codons, which signal the termination of translation. The remaining 61 codons, called sense codons, encode the 20 amino acids found in proteins. The code therefore contains more information than is strictly necessary to specify these 20 amino acids and is said to be degenerate. Only tryptophan (Trp) and methionine (Met) are specified by a single codon. For the other amino acids, the number of codons specific to each amino acid varies (2, 3, 4, or 6). Codons that specify the same amino acid are referred to as synonymous codons.

- The reading frame and initiation codons

The genetic code is usually non-overlapping; each nucleotide normally participates in only one codon. Each nucleotide sequence can be read in three different ways, depending on the reading frame used. The translation system must use the correct reading frame, which is determined by the initiation codon, typically AUG, which specifies methionine (Met).

- Termination codons

Three codons—UAA, UAG, and UGA—do not specify an amino acid. These codons signal the end of a protein and are called stop codons, termination codons, or nonsense codons. There is no tRNA whose anticodon pairs with a termination codon.

- The universality of the code

The genetic code is nearly universal across all living organisms, with only a few exceptions.

- **Essential Elements of translation**: From a genetic perspective, the essential components of translation are :
  - **messenger RNA (mRNA):** Provides the sequence of codons specifying each amino acid of the protein.
  - **The ribosomes:** Serve as the structural and catalytic support, ensuring the sequential connection of amino acids
  - **The RNA of transfer (tRNA):** Recognizes codons and connects them to their corresponding amino acids.
  - **Aminoacyl -tRNA synthetases** Enzymes that ensure the correct attachment of

each amino acid to its specific tRNA..

- **Initiation, elongation and termination factors,** Each step requires specific molecules, including guanosine triphosphate (GTP) for energy, proteins such as eIF2, eIF3, eIF4 (A, B, F), eIF5, and release factor RF.

➢ **Steps of translation**

1. Initiation

Initiation involves three main stages:

1. The small ribosomal subunit, assisted by initiation factors, recognizes the 5′ cap of the mRNA and binds to it. The small subunit then scans the mRNA until it locates the first AUG codon.
2. The initiator tRNA binds to the mRNA at the initiation codon, forming antiparallel base pairing with the codon.
3. The large ribosomal subunit binds to the initiation complex.

In eukaryotes, the first amino acid incorporated is methionine (Met), whereas in bacteria it is a formylated methionine (fMet) at the NH_2 terminus.

The movement of each tRNA inside the ribosome follows the pathway: **A site → P site → E site**. The initiator tRNA directly occupies the P site without passing through the A site, while all other tRNAs first enter the A site. Immediately after initiation, the ribosome is attached to the mRNA, the Met-tRNA is positioned at the start codon in the P site, and the adjacent A site is unoccupied.

2. Elongation

Elongation corresponds to protein synthesis by sequential addition of amino acids to the C-terminal end of the growing peptide chain. It is assisted by elongation factors (EF-Tu, EF-Ts, and EF-G) and proceeds via a cyclic three-step mechanism:

1. A charged tRNA, in complex with the elongation factor EF-Tu and a GTP molecule, enters the A site of the ribosome. The tRNA anticodon pairs with the next codon in the mRNA. Upon correct pairing, GTP is hydrolyzed to GDP, and the EF-Tu–GDP complex is released.

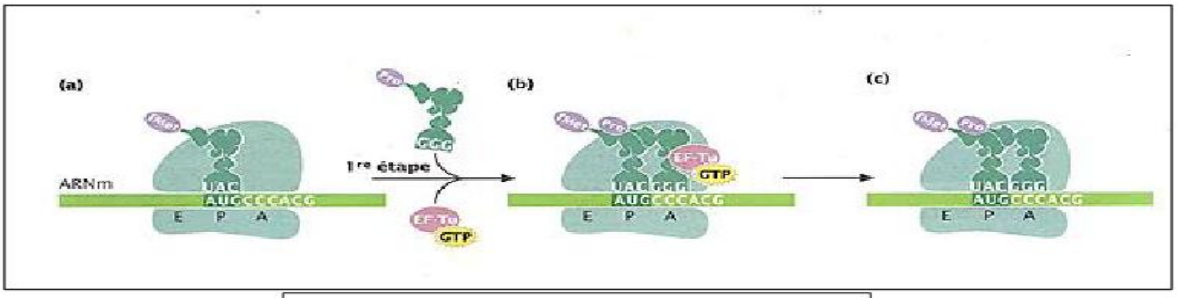


Figure: First stage of elongation

- In the second step, a peptide bond is formed between the amino acids attached to the tRNAs in the P and A sites. This reaction releases the amino acid from its tRNA at the P site.

- The third step, translocation, involves the movement of the ribosome by one codon along the mRNA. This step requires the elongation factor EF-G and the hydrolysis of GTP to GDP. Since the tRNAs in the P and A sites remain base-paired with the mRNA codons, they do not move with the ribosome. As a result, the tRNA that occupied the P site is now positioned in the E site, from which it exits to the cytoplasm to be recharged with an amino acid. Translocation also shifts the tRNA in the A site, which carries the growing polypeptide chain, into the P site, leaving the A site vacant for the next incoming aminoacyl-tRNA.

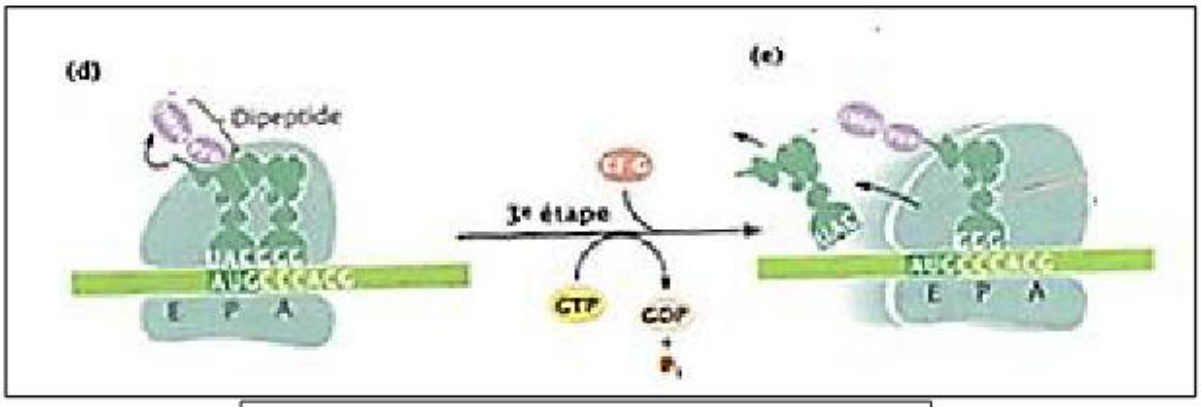


Figure: Ribosome translocation

After translocation, the A site is vacant and ready to accept the tRNA corresponding to the next codon. The elongation cycle then repeats: a tRNA carrying its amino acid enters the A site, a peptide bond forms between the amino acids in the A and P sites, and the ribosome advances to the next codon.

Throughout elongation, the polypeptide chain remains attached to the tRNA in the P site. The process of elongation in eukaryotes proceeds in a similar manner.

3. Termination

Translation terminates at the stop codons UAA, UAG, and UGA, which do not specify any amino acids. These stop codons are recognized by release factors (RFs). Because there are no tRNAs with anticodons complementary to the stop codons, no tRNAs enter the A site. The polypeptide chain is released from the tRNA in the P site, and the ribosome detaches from the mRNA and dissociates into its subunits.

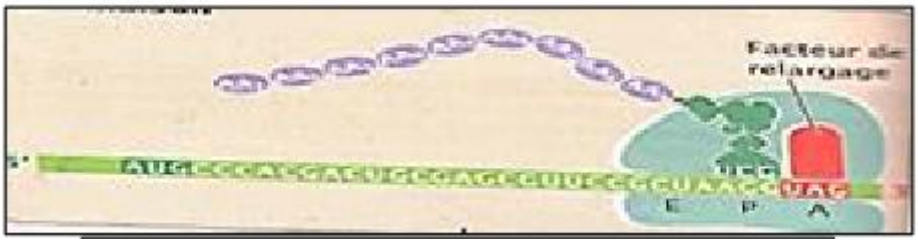


Figure: Termination of the translation

- **Post-translational modifications of proteins**

Polypeptide chains undergo post-translational modifications to become fully functional:

- Some proteins are synthesized as larger precursor molecules that must be cleaved and processed by specific enzymes to acquire their active form.
- For others, glycosylation—the addition of carbohydrate chains—may be required for proper activation.
- The function of many proteins depends on correct folding. While some proteins fold spontaneously into their functional conformations, others require assistance from molecular chaperones.

2- Post-translation

modifications occur after the translation stage (polymerization of the acids amino acids by rhe ribosome), For there production of the proteins by there cell. We quote:

- Elimination of the amino acid (Methionine): 50% of the proteins of prokaryotic cells And eukaryotes have there Methionine1 removed by the enzyme Met- aminopeptidase .
- Cleavage of the connections peptides : exp . Insulin
- Modification of the residues : glycosylation
- Association has others chains polypeptides identical Or No Or has of the ligands : exp . Hemoglobin.

Post-translational modifications occur after the translation stage (after the polymerization of amino acids by the ribosome) and are essential for the production of functional proteins. Key examples include:

- **Removal of the initial methionine:** Approximately 50% of proteins in both prokaryotic and eukaryotic cells have their initial methionine removed by the enzyme methionine aminopeptidase.
- **Cleavage of peptide bonds:** Example, processing of insulin from its precursor form.
- **Residue modifications:** Such as glycosylation.
- **Association with other polypeptide chains or ligands:** For instance, hemoglobin associates with multiple subunits and ligands to function properly.