Plant molecular biology

I. Nucleic acids:

Nucleic acids are biological macromolecules that contain genetic information (molecular masses up to several million Daltons103 to 109 Da). They are involved in the conservation, transmission and expression of genetic information in all prokaryotic or eukaryotic cells and viruses.

There are two types of nucleic acids:

- Deoxyribonucleic acid (DNA): located in the nucleus of eukaryotic organisms and the cytoplasm of prokaryotes. It is the support of heredity responsible for the conservation and transfer of genetic information.

- Ribonucleic acid (RNA) is found mainly in the cytoplasm, it transcribes and translates genetic information during the synthesis of the various proteins of a cell.

Nucleic acids are formed by polymerization as long chains of simple units called "nucleotides"

II. Chemical Composition and Structure of Nucleic Acids

II.1. Chemical composition of nucleic acids:

II.1.1 The constituents

Nucleic acids are composed of 3 elements:

- Nitrogenous bases: define the sequence, whether at the DNA or RNA level
- 5-carbon sugars (pentoses).
- Phosphoric acid: links nitrogenous bases together.

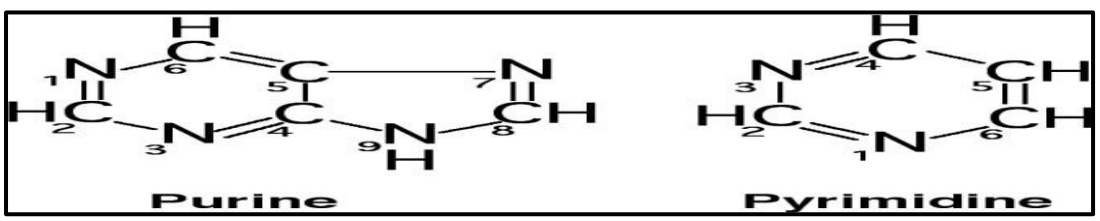
1. Nitrogenous bases

-The nitrogenous bases of nucleic acids belong to two types: aromatic heterocycles of pyrimidine type and purine type (fig 1).

-Purines have a double aromatic nucleus comprising on the left a hexagonal cycle of 4 carbons and 2 nitrogens and on the right a pentagonal cycle of 3 carbons (including 2 common with the previous one) and 2 nitrogens (We number so as to see the nitrogens N as 1, 3, 7, 9).

Pyrimidines have a six-membered aromatic ring of 4 carbons and 2 nitrogens (N in positions 1 and 3.)

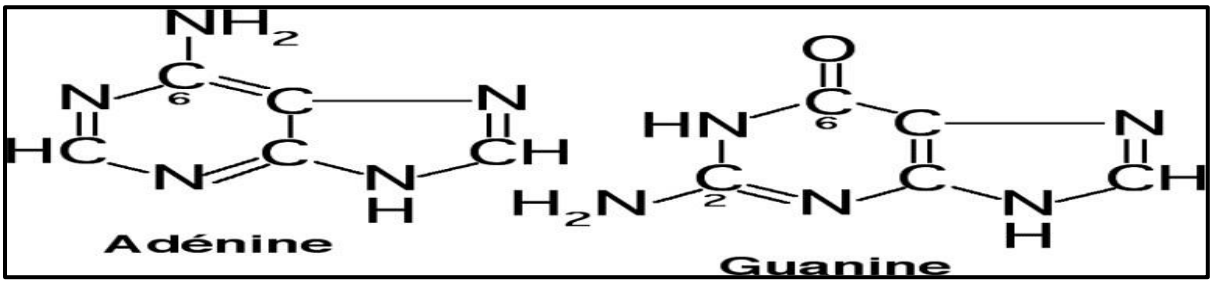
1.1. Purine bases (fig 2)



Purine bases are derived from the purine nucleus:

Among the purines there are Adenine and Guanine which are found in both DNA and RNA.

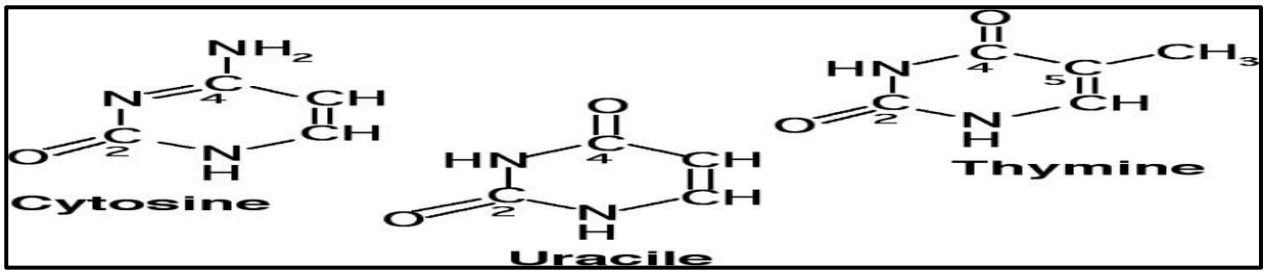
- **Adenine**"A" (6-Amino-purine) (an amine group in position 6) carbon 6 is substituted by an amine function.
- **Guanine "G"**(2 amino 6 - oxy-purine) carbon 2 is substituted by an amine function and carbon 6 by a ketone function.



1.2. pyrimidine bases(fig 3)

There are 3 types of pyrimidine bases:

- **Cytosine "C"**(2-oxy-4-amino-pyrimidine) carbon 4 is substituted by an amine function and carbon 2 by a ketone function.
- **Uracil "U"**(2,4-oxypyrimidine) carbons 2 and 4 carry ketone functions
- **Thymine "T"** (2,4-oxy-5-methyl-pyrimidine) carbons 2 and 4 carry ketone functions, but carbon 5 is substituted by a methyl



Thymine and uracil have a very similar structure. Uracil is specific to RNA and thymine is specific to DNA.

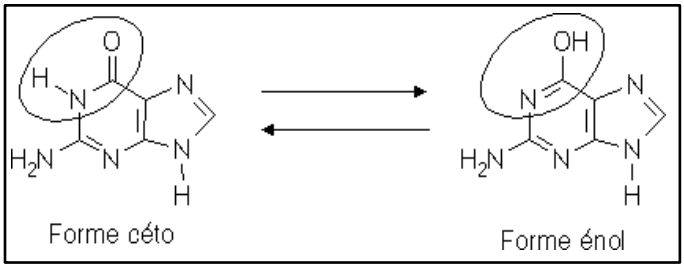
All other bases are found in DNA and RNA, except for the Uracil/Thymine pair.

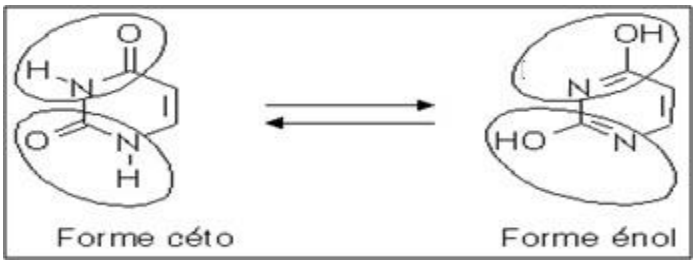
- **Tautomerization:**

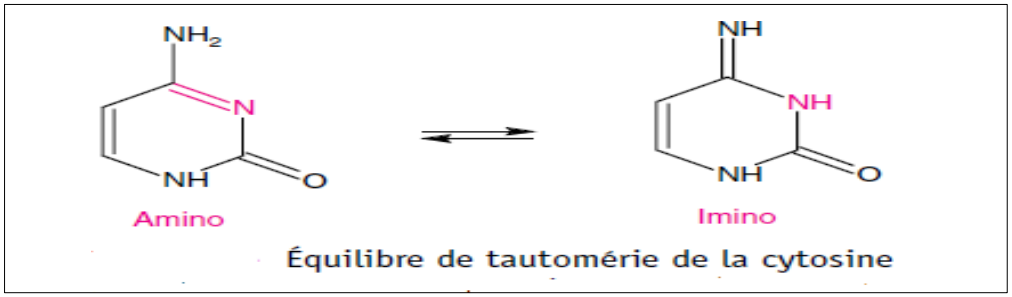
The formulas of nitrogenous bases can switch to other form(s) called tautomeric form (or tautomeric equilibrium). This equilibrium corresponds to the equilibrium between the ketone form and the enol form and is therefore called the keto-enolic equilibrium.

We also say, when the base carries ketone functions, that it is in the lactam form, and that when it carries enol functions that it is in the lactim form. If the tautomerism equilibrium affects an amine function, as with cytosine, we say that the equilibrium is established between the amino form (NH2) and the imino form (=NH)

The figures above are in the keto form which is more present at physiological pH than the enol form (eg Guanine and Uracil).







The basic derivatives:

There are derived bases found in RNA and DNA, these particular bases allow the RNA to adopt conformations adapted to the function of these RNAs.

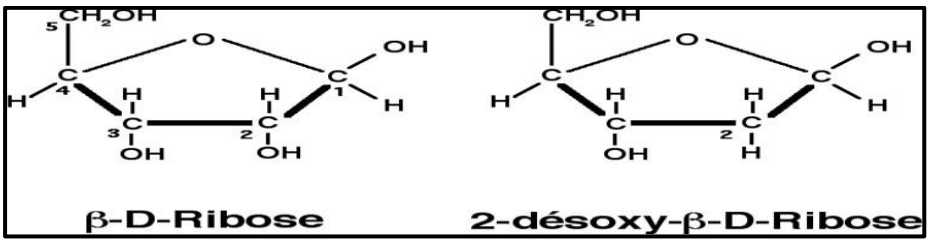
Puric acid:•Hypoxanthine•1-methyl adenine•7-methyl guanine

Pyrimidines:•5-methyl cytosine

2. **Sugars (pentoses)**(fig 4):

   Ribose is a D-series pentose, with all hydroxyls oriented to the right (Fisher representation). In ribonucleic acids (RNA), it is cyclized to ribofuranose: β-anomer.

   Deoxyribose, a component of deoxyribonucleic acids (DNA), is derived from ribose by a reduction of the OH function of C'2. Deoxyribose gives this nucleic acid greater stability, specific to its function of preserving genetic information.

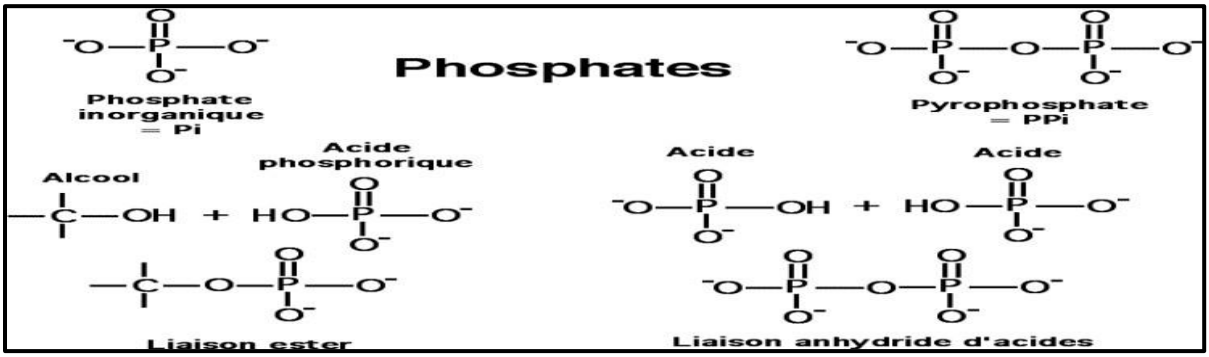


3. **Phosphoric acid (H_3PO_4)**(fig 5)

Inorganic phosphate is a stable ion formed from phosphoric acid PO4H3. It is often written Pi.

Phosphate esters can form between a phosphate and a free hydroxyl group (alcohol). The condensation of a phosphate and another acid, for example another phosphate, gives an anhydride.

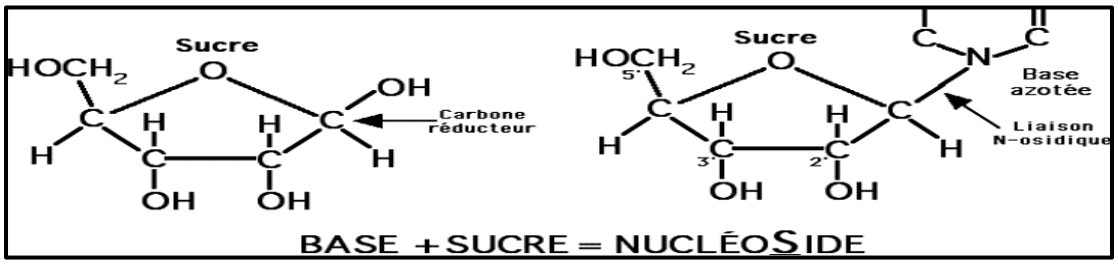
Pyrophosphate is an ion derived from pyrophosphoric acid (P2O7H4), which itself is an anhydride of phosphoric acid.

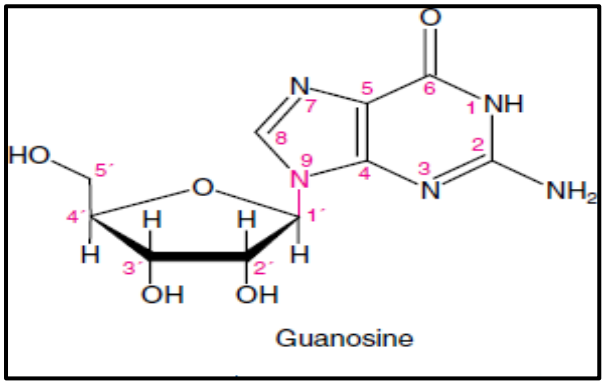


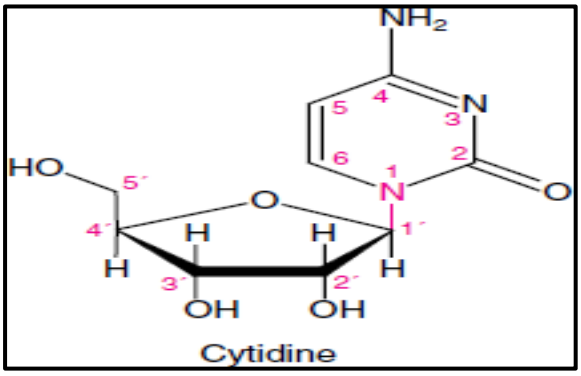
II.1.2. Constituents asssemblage:

1. **Nucleosides (**fig 6)

A nucleoside is composed of a base (pyrimidine or purine) and a sugar (ribose or deoxyribose) always associated by a covalent bond: the β N-osidic bond. The latter is made between C 1' of the ose and N 9 for the case of purine bases, and between C 1' of the ose and N 1 for pyrimidine bases.







The nomenclature of nucleosides comes from the combination of the name of the base with that of the sugar (ose). We use the suffix osine (purines) and suffix idine (pyrimidines)

The two types of nucleosides

1.1.Ribonucleosides:

Adenine + ribose = adenosine

Guanine + ribose = guanosine

Cytosine + ribose = cytidine

Uracil + ribose = uridine

1.2. Deoxyribonucleosides:

Adenine + deoxyribose = deoxyadenosine

Guanine + deoxyribose = deoxyguanosine

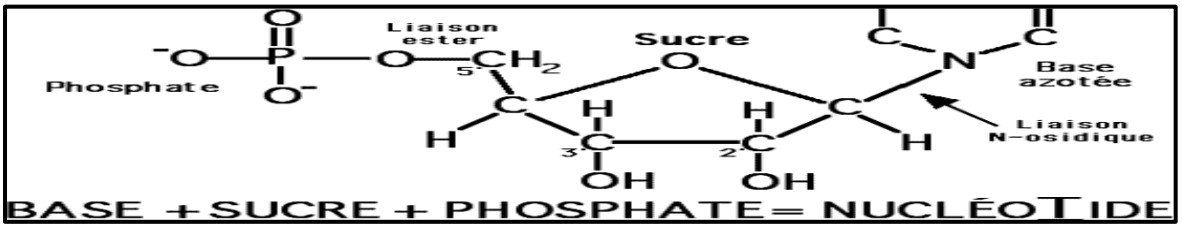
Cytosine + deoxyribose = deoxycytidine

Thymine + deoxyribose = deoxythymidine

2. Nucleotides(fig 7)

A nucleotide is a molecule composed of a nucleoside linked by an ester bond to a phosphoric acid, therefore, nucleotides are phosphoric esters of nucleosides.

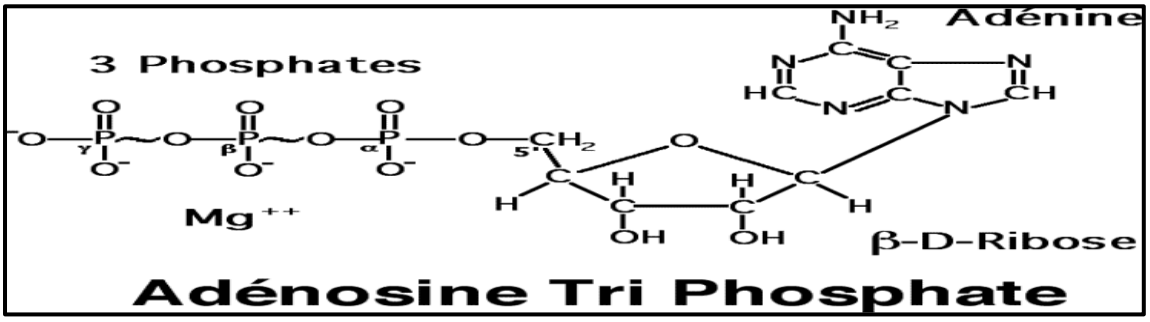
A nucleotide is therefore formed from a nitrogenous base (pyrimidine or purine), linked by an osidic bond with a sugar (ribose or deoxyribose), itself linked by a covalent ester bond with a phosphate. We will therefore have ribonucleotides and deoxyribonucleotides.



-The bonding of a nucleoside with a phosphate is done by an esterification of the OH of (C5') of the sugar and an acid function of the phosphate.

-Each nucleoside can be linked to one, two or three phosphates (fig 8) in the form of 5' monophosphate, 5' diphosphate and 5' triphosphate respectively. They are designated by conventional acronyms: GMP for guanosine monophosphate, CDP for cytidine diphosphate, ATP for adenosine triphosphate, etc.

-All 5' nucleoside triphosphates (5' NTPs) and 5' deoxynucleoside triphosphates (5' dNTPs) are precursors in nucleic acid synthesis and exist free in the cell.



3. Polynucleotides (fig 9):

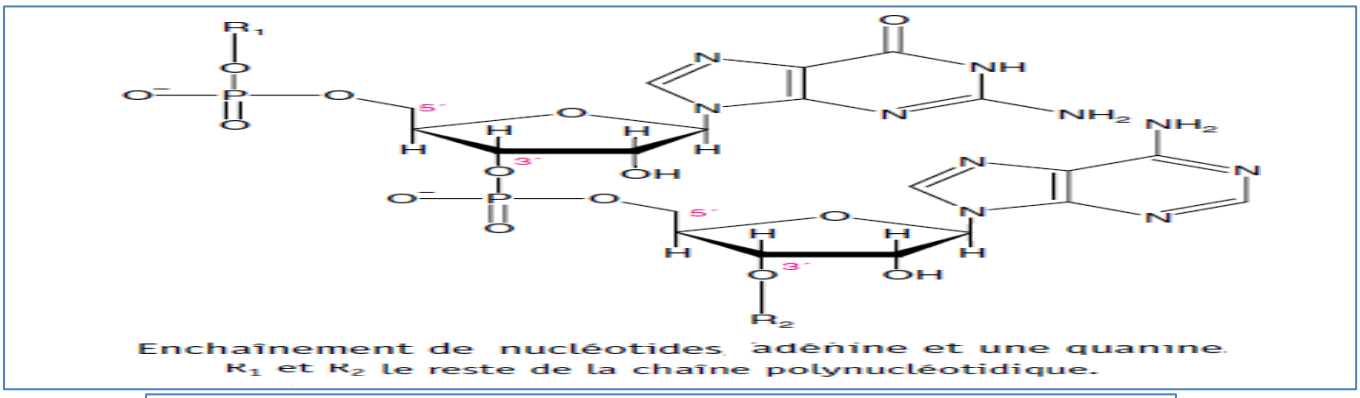
-Nucleic acids are formed by a polycondensation of nucleotides AMP, CMP, GMP and UMP for ribonucleic acids, dAMP, dCMP, dGMP and dTMP for deoxyribonucleic acids.

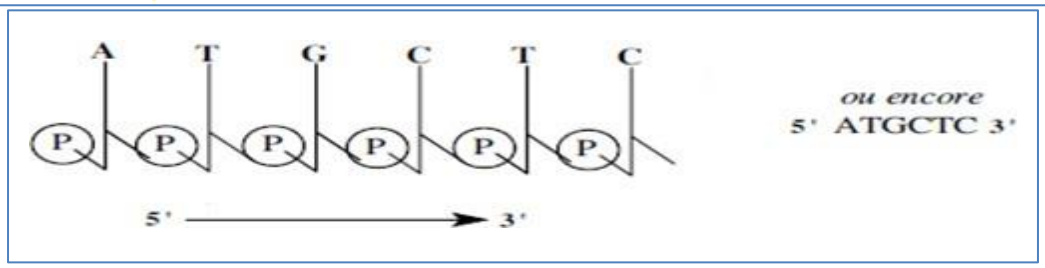
-The bond between two nucleotides is made by the formation of an ester bond between the phosphoric acid of one nucleotide and the alcohol function carried by the 3' carbon of the ose of the other nucleotide: 3'-5' phosphodiester bond.

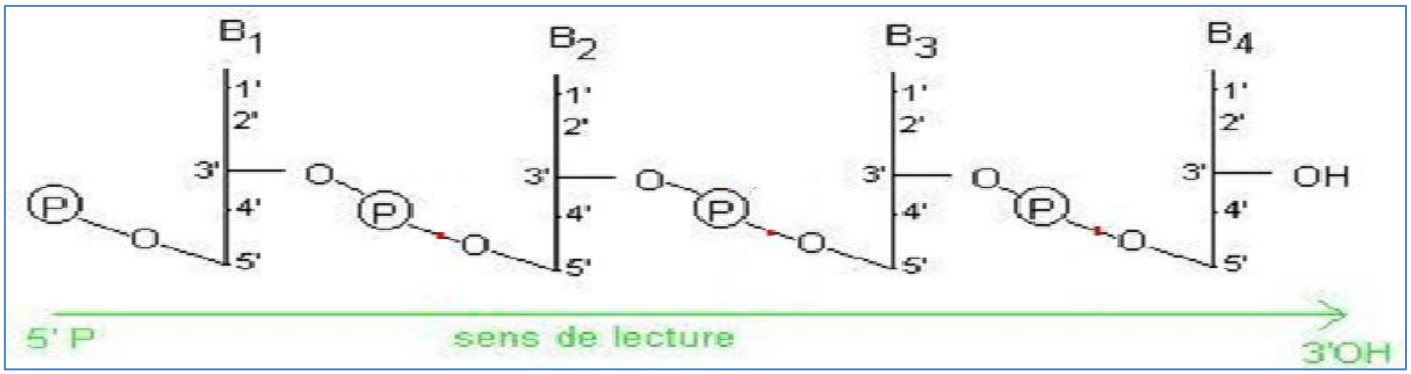
- Phosphoric acid therefore forms two bonds, one with the ose of its nucleotide, the other with the ose of a second nucleotide. There therefore remains an acid function on each phosphoric acid residue.

-If we consider the two ends of a nucleic acid, one contains a phosphoric acid residue carrying two free acid functions (5'P ends), the other contains a free hydroxyl on the 3' carbon of the ose (3'OH end).

-By convention, polynucleotide chains are oriented from 5' to 3'.



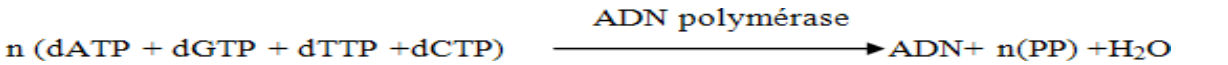


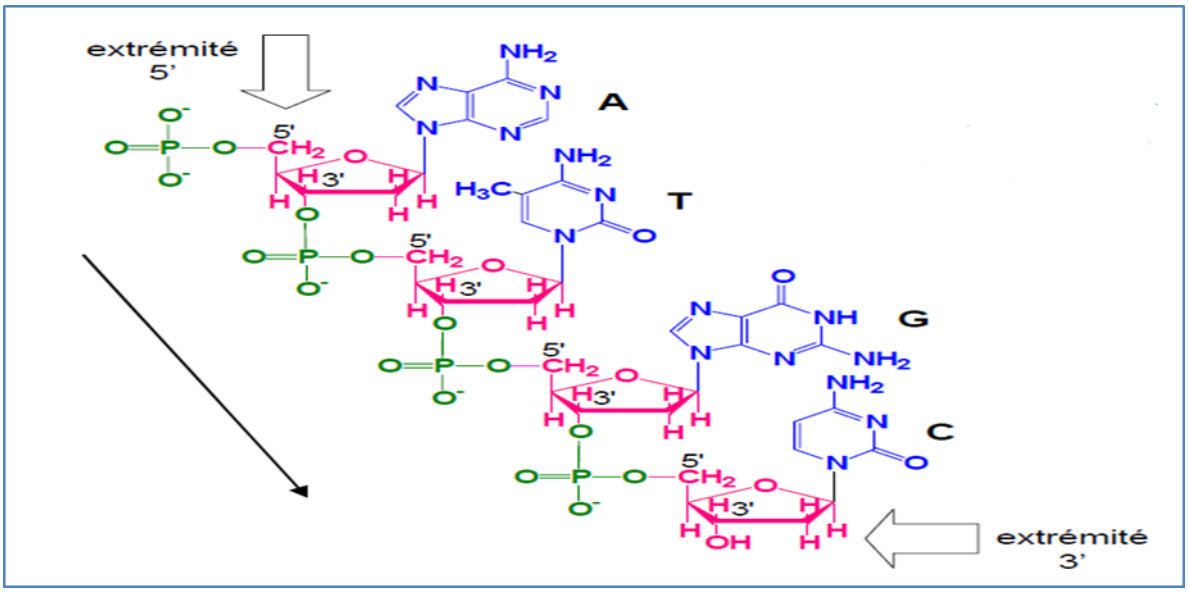


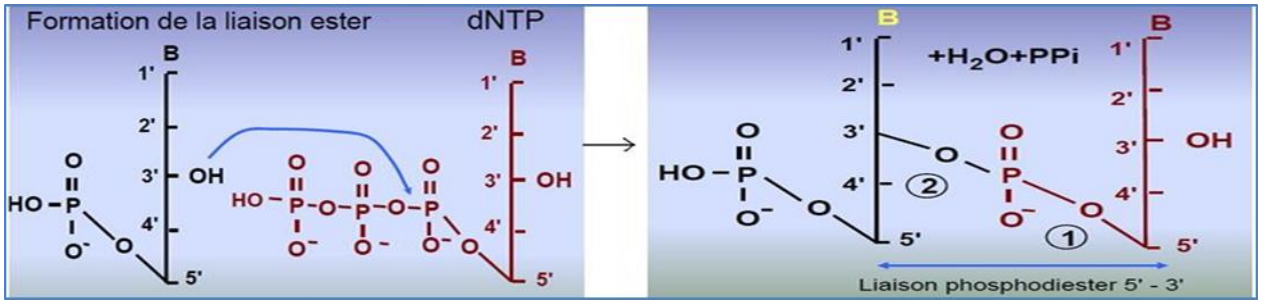
The structure of DNA

1. The primary structure of DNA

The primary structure of DNAs is made up of the linear chain of the four main deoxyribonucleotides: dAMP, dCMP, dTMP, dGMP which are linked by a 3' - 5' phosphodiester bond (fig.).

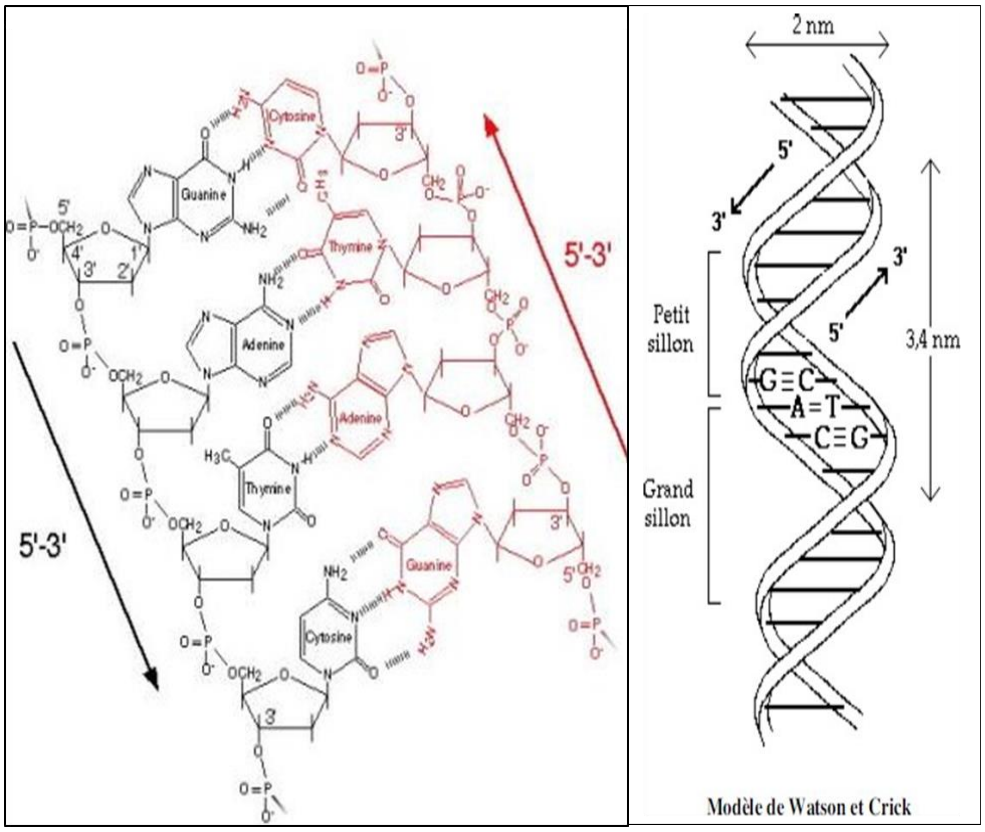






2. The secondary or three-dimensional structure of DNA (Watson and Crick double helix model, 1953):

- The secondary structure of DNA is made up of two strands (chains) which wrap around each other in a right double helix (not right-handed). This shape is also called a helical double-stranded helix.



The two DNA chains have several characteristics:

- **Antiparallelism:**

The two chains are antiparallel: Each of the two strands is oriented (5'→3') in the opposite direction to that of the other strand (3'→5'), that is to say that the 5' end of one is on the side of the 3' end of the other.

- **Complementarity:**

The bases of each of the strands are complementary: this complementarity is due to the conjunction of two phenomena:

- *Steric constraints*: opposite a purine (made up of two cycles), we find a pyrimidine (made up of a single cycle). The purine-pyrimidine complementarity means that each base pair having the same dimension, the structure of the DNA helix is very regular (fig).
- *The creation of hydrogen bonds:* The nitrogenous bases are turned towards the inside of the double helix so that each hybridizes with a base of the other strand by hydrogen bonds (hybridization A with T and C with G). The guanine-cytosine hybridization is more stable (3 hydrogen bonds) than that between adenine and thymine (2 hydrogen bonds).
- **Chargaff's rule:**

  The chemist Chargaff demonstrates that the ratio between the quantity of adenine and thymine (A/T) on the one hand, and the quantity of guanine and cytosine (G/C)

on the other hand, is constant and close to 1 but that the base composition of nucleic acids varies according to the species. This characteristic is designated under the name of Chargaff's rule (Chargaff coefficient). So whatever the origin of the DNA, the number of purines is always equal to the number of pyrimidines:

[Pur] = [Pyr] or [A] + [G] = [T] + [C]

- in addition, the molar fractions of the bases are such that: the quantity of A always = quantity of T (A = T) and the quantity of G always = quantity of C (G = C). The bases A and T are said to be complementary, the same is true for G and C. But of course the proportions ([A] + [T]) and ([G] + [C]) are not equal and vary from 35 to 75% depending on the species, but constant for all members of a given species. Example in humans: A = 30.4% and T = 30.1%; C = 19.6% and G = 19.9%.

- When we represent the double helix along its axis, we highlight certain particularities.

➢ The double helix has a "pitch" of 3.4 nm (34 A°) meaning that there are approximately 10 pairs of nucleotides for each turn of the helix. for a diameter of 2 nm (20 A°). We therefore have 0.34 nm (3.4 A°) between 2 consecutive base pairs (fig).

➢ The double helix structure allows for two grooves of unequal size, one large and one small; the large groove 12 A° allows the accessibility of the DNA to the enzymes regulating transcription, and the small groove 6 A° allows binding to histones.

2. Types of double-stranded DNA conformations:

The stability of the secondary structure of these different DNA double helices is essentially due to:

- hydrogen bonds between the complementary bases of each strand

- hydrophobic and electrostatic interactions of successive bases stacked in the helix structure.

Conformation B:

This is the one in the model described by Watson and Crick and is found as the main and most important native form in physiological conditions. It is found when the DNA fibers are at 90% humidity (90% water) and in a solution of low ionic strength.

Conformation A:

Is a more compact right helix. When the water content of a solution containing a DNA molecule is decreased (70% water), for example during crystallization, the molecule changes conformation and adopts an A conformation. This change is reversible.

The A conformation of DNA is specific to transcription. During transcription, RNA stimulates a transfer of DNA from type B to A. At the end of transcription, the DNA returns to its B conformation. This conformation is found in vivo in: - the DNA of certain bacterial

spores, formed in response to the desiccation of the medium. - DNA-RNA hybrids that form transiently at the beginning of replication and during transcription.

Conformation Z:

Is a left-handed helix (zigzag shape) or the DNA molecule is more stretched in this conformation. This conformation is present in vivo in short regions of DNA. These specific regions are generally segments of alternating Pur/Pyr sequence (GCGC) whose bases are often methylated.

DNA shape	DNA B	DNA A	Z DNA (zigzag)
Helix	right (right-handed)	right (right-handed)	left (levorotatory)
General appearance	Long and thin	Short and stocky	Stretched and thin
Helix tour	34Å	29Å	46 Å
Number of base pairs per spire tower	10	11	12
Axial distance between pairs basics	3.4 Å	2.3 Å	3.8 Å
Twist angle	36°	32.7°	-30°
Presence	very presence frequent	relatively presence frequent	rare

3. Tertiary structure:

3.1. Definition of topoisomers

Two circular DNAs are called "topoisomers", having exactly the same number of nucleotides and the same sequence of bases, but differing from each other by the number of twists. That is, the number of turns that one of the two strands makes around the other.

It mainly concerns the different cases of circular DNA which are small sized DNA such as:

- Prokaryotic DNAs and plasmid DNAs.
- in eukaryotes; In addition to the DNAs of mitochondria and chloroplasts which are circular, the DNAs which normally have a linear structure are also found in circular form where each of the ends will be linked to an anchor point in the cells

3.2. The different states of topoisomers:

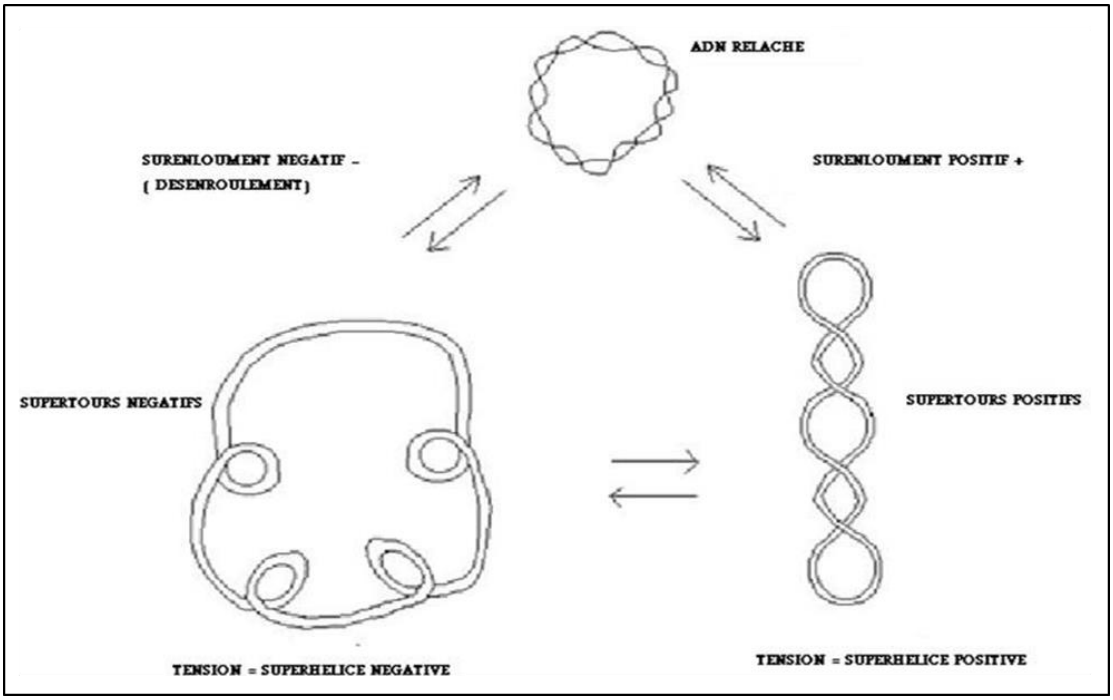
- **The relaxed state:**

  The tension on the double helix is minimal (without constraint), it is the most stable configuration of the DNA molecule, and that we find when the DNA is in a B form. It is observed for 10 pb per turn

- **The supercoiled state:**

The axis of the double helix can be wound on itself to form a superhelix. Two forms of supercoiling are theoretically possible:

- ***Positive overcoiling***: the winding of the right double helix occurs in the same direction (right super helix), and leads to the formation of a positive superhelix and the number of intertwinings increases.
- ***Negative overcoiling***: the winding of the double helix is carried out in the opposite direction to that of the initial helix in a negative way forming a negative superhelix (left superhelix). This promotes the unwinding also called "unwinding", in this case, the number of interlacings decreases.



3.3. Topoisomerases:

The passage of the different forms is controlled by enzymes called topo isomerases which are enzymes that modify the number of interlacements. To modify the number of superturns, it is necessary to cut 1 or 2 strands.

- **Topoisomerase 1:**

It cuts only 1 of the 2 strands of DNA in a positively supercoiled form. This allows access to the enzyme to attach to the DNA.

These topoisomerases allow the release of superhelices under tension. They exist in prokaryotes and eukaryotes.

- **Topo isomerase 2:**

They are dimeric and cut the 2 strands of double-stranded DNA and re-weld after inducing negative supercoils from the relaxed form which allows molecular compaction, or after removing positive supercoils. This type of topoisomerase includes bacterial gyrase.

3.4. Interest of topoisomerases

Most DNA molecules found in nature form a negative supercoil which has the dual advantage of

being more compact, and they facilitate accessibility to replication enzymes, transcription and DNA repair. In eukaryotes, the advantage of the superhelix is that the DNA forms a superhelix while wrapping itself around histones (proteins).

4. Higher order structures:

4.1. DNA in living organisms:

The amount of DNA in living cells varies greatly from one species to another: - virus 104 to 105 bp, - bacteria 106 bp,- human 109 bp. LDNA is characterized by its number of nucleotides, expressed in (bp).

- **In Prokaryotes:**

There are 2 types of DNA found in bacteria:

- *Chromosomal DNA*: There is only one DNA molecule per cell. It is a circular DNA of about 1Mbp. The DNA makes loops rotating in the opposite direction to the helix, called "supercoils". This allows it to have a degree of compaction.
- *Plasmid DNA***:** in most bacteria small circular DNAs that constitute extrachromosmic DNA or plasmids that replicate independently of the chromosome.

- **In DNA Viruses:**

Are particles formed from a single nucleic acid and proteins, constituting the nucleocapsid.

- **In Eukaryotes:**

DNA is contained in the nucleus of the cell. It is organized in a linear double helix, according to the W and C model. The total DNA, in a man, is compacted, thanks to a system of proteins, histones (basic proteins) as well as other non-basic proteins.

4.2. Types of human DNA organization:

DNA is associated with histone proteins. It is located in the nucleus of the cell, folded, condensed and compacted in an orderly manner that ensures the storage of a large quantity of DNA. We distinguish: chromatin (cells in interphase) and chromosomes (cells in division or mitosis).

DNA compaction depends on the cell cycle; but also on the need for accessibility to certain areas of the genome. This compaction is not by several levels of DNA organization.

1. **10nm fiber (pearl necklace) (nucleofilament):**

Also called a "pearl necklace" structure. This first level of organization is formed by the winding of DNA around a protein core to produce a "pearl" called the nucleosome. The DNA is wound around a small disk 11 nm in diameter and 6 nm thick, formed of an octamer of proteins, histones.

- **Nature of histones**

There are five families of histones: H1, H2A, H2B, H3 and H4.

- The five types of histones have the following properties in common: 1) high richness in basic amino acids lysine and arginine (20%) positively charged, 2) high evolutionary conservation.

  The four histones H2A, H2B, H3 and H4 associate in duplicate (an octamer); H2A,

H2B, H3 and H4 are called "nucleosomal", H1 is called "internucleosomal".

- The N-terminus of these histones undergoes post-translational modifications (acetylation, methylation,

  phosphorylation…) which will modify the interactions with the DNA to control its condensation or decondensation. These modifications constitute modes of regulation of gene expression; since the genes will or will not be accessible to the different enzymes or proteins, and therefore will or will not be expressed.
- Basic amino acids have (+) charges that will allow ionic bonds to be established with the (-) charges of the DNA phosphates. This allows the curvature of the DNA around the histones.

➢ **The nucleosome**

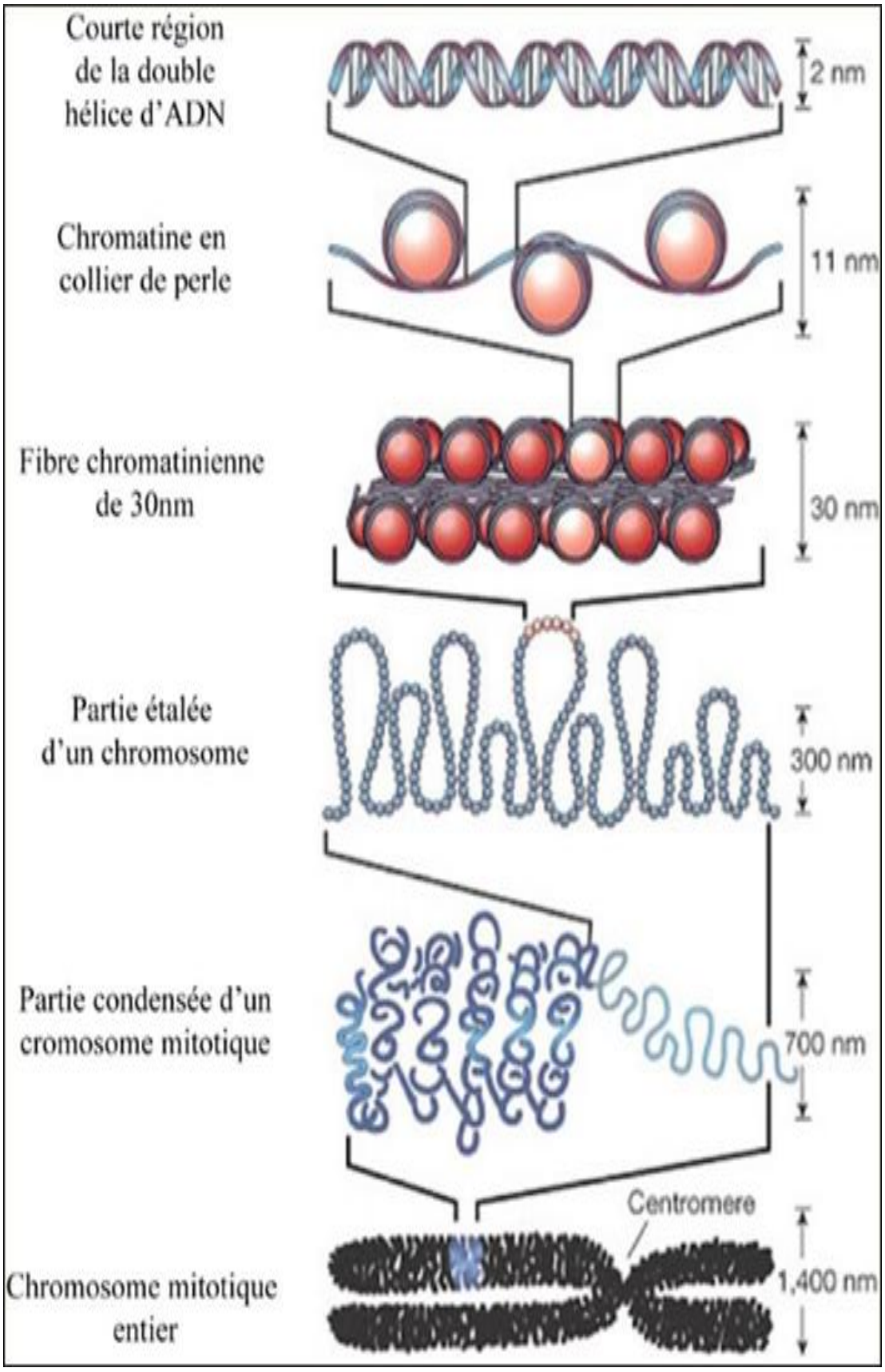
The DNA forms a ¾ turn around the histone octamer, about 200bp, but only 146bp make two turns around the histone core and the rest of the bases connect this nucleosome to the next one (fig).
Interactions take place at the minor groove of DNA.

2. 30nm fiber:

Six nucleosomes will assemble into a helix with a diameter of approximately 30 nm and a length of 11 nm. This helix therefore has 6 * 200 bp, or 1200 bp. Since the turns are joined, we rather speak of a solenoid. Therefore, the 30 nm fiber is a structure that resembles a solenoid with 6 nucleosomes per turn. The stability of this structure is ensured by a histone H1. Due to the contribution of phosphorylation-dephosphorylation system, this histone H1 can either promote the formation of this 30 nm fiber (chromatin compaction), or discourage it (decompaction).
This form of DNA constitutes interphase chromatin, and is transcriptionally inactive.

3. Higher levels of chromatin organization:

Maximum condensation is achieved when the fiber is organized into a loop, with both ends of the loop held together by proteins forming the central axis of the chromosome skeleton. Each loop contains 20,000 to 100,000 bp, with a size ranging from 150 to 800 nm.
This structure will form a helix with a diameter of approximately 700 nm containing around sixty loops. This is the maximum compaction rate of DNA in the chromosome.


Courte région de la double hélice d'ADN
2 nm
Chromatine en collier de perle
11 nm
Fibre chromatinienne de 30nm
30 nm
Partie étalée d'un chromosome
300 nm
Partie condensée d'un cromosome mitotique
700 nm
Centromere
Chromosome mitotique entier
1,400 nm

Structure of RNA

1.Primary structure

RNA is called to be "single-stranded from 5' P to 3' OH chaining of the 4 ribonucleotides AMP, GMP, CMP, UMP, by a 3' 5' phosphodiester bond. The sugar β DRibose

2.Secondary structure

Three-dimensional conformation

Internucleotide pairings

In space, there are double helix folds that occur in regions between two distant segments of the same RNA strand. This type of conformation involves hairpin-like secondary structures.

Pseudo-helical conformation:

Under the action of stacking forces between successive bases, RNA tends to spontaneously form an irregular right single helix.

These secondary structure motifs have been found in ribosomal RNAs and transfer RNAs:

-The stems are helices whose conformation is close to that of DNA. A

-loops: those participating in a hairpin pattern stabilize this structure.

3.Tertiary structure

foldings which will give this structure in space essentially due to hydrogen bonds ofdouble helix and pseudohelix A very precise conformation in the RNA space.

2) The different classes of RNA in prokaryotic and eukaryotic cells

The classification of RNA according to their functions. 5 types of RNA:

Messenger RNA: mRNA 5% ,Transfer RNA: tRNA 15% ; Ribosomal RNA: rRNA 80% and Nuclear RNA: Small nuclear RNA: snRNA ; Heterogeneous RNA: nhRNA

a)Ribosomal RNAs

1-Definition

rRNAs are located at the level of ribosomes. These are ribonucleoprotein particles.high molecular weight based on 35% protein and 65% RNA.

rRNAs are synthesized in the nucleus from DNA and assemble at the nucleoli with proteins, then pass into the cytoplasm.

Ribosomes are grouped into functional units called "polysomes." A polysome is a strand of messenger RNA to which arefixed the ribosomes.

2-Composition of ribosomes

Ribosomal subunits:

A ribosome consists of two ribosomal subunits. They are designated by their sedimentation velocity S (S for Svedberg) it depends not only on its mass, but also on its shape and rigidity.

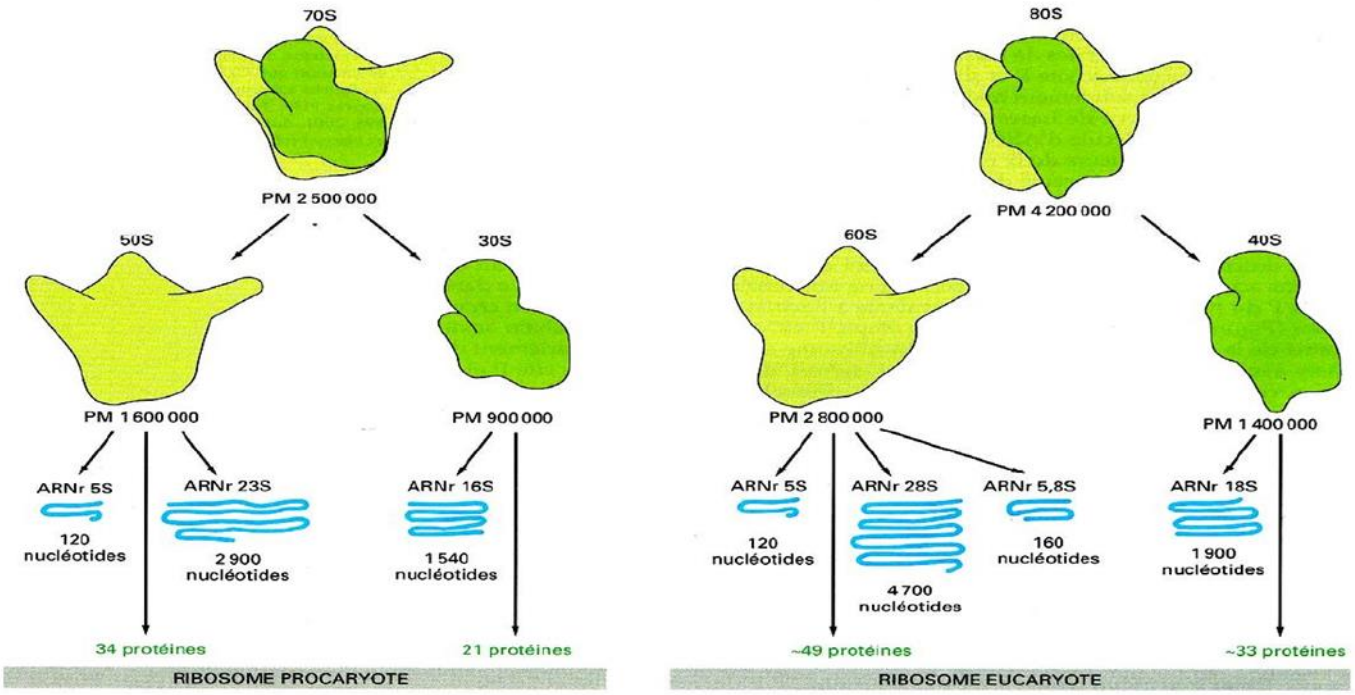
In prokaryotes:

A ribosome has a sedimentation coefficient of 70s of 2 units.

The two subunits: one of 50s and one of 30s. The large 50s subunit is formed of two types of rRNA: (5s and 23s) + 34 proteins. The small rRNA subunit of 16s and 21 proteins.

In eukaryotes:

The 80s ribosome. two subunits: one of 60s and 40s. The 60s is formed of 3 types of rRNA of 5s, 5.8s and 28s and 40 proteins. The 40s of a subunit of 18s rRNA and 30 proteins.

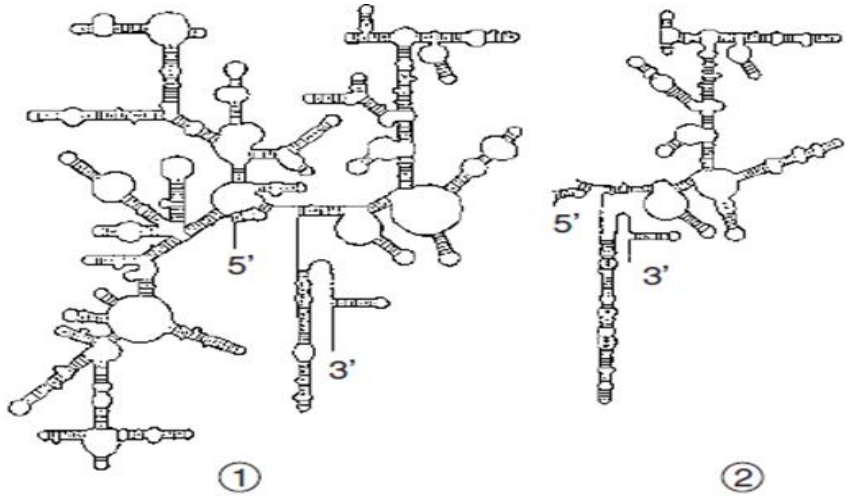


3-Structure

rRNAs are rich in G and C, and have only a few rare bases.

The secondary structure is characterized by flexible single-stranded pairing and mispairing regions.

The tertiary structure: the three-dimensional organization of the ribosome. To be functional, the two subunits of the ribosome must be associated.



Organisation comparée de la structure secondaire de l'ARNr 16 S de *Escherichia coli* **(1)** et de celle d'une partie de l'ARNr 18 S de la levure de bière **(2)**

b)Transfer RNAs (tRNAs)

1-Definition

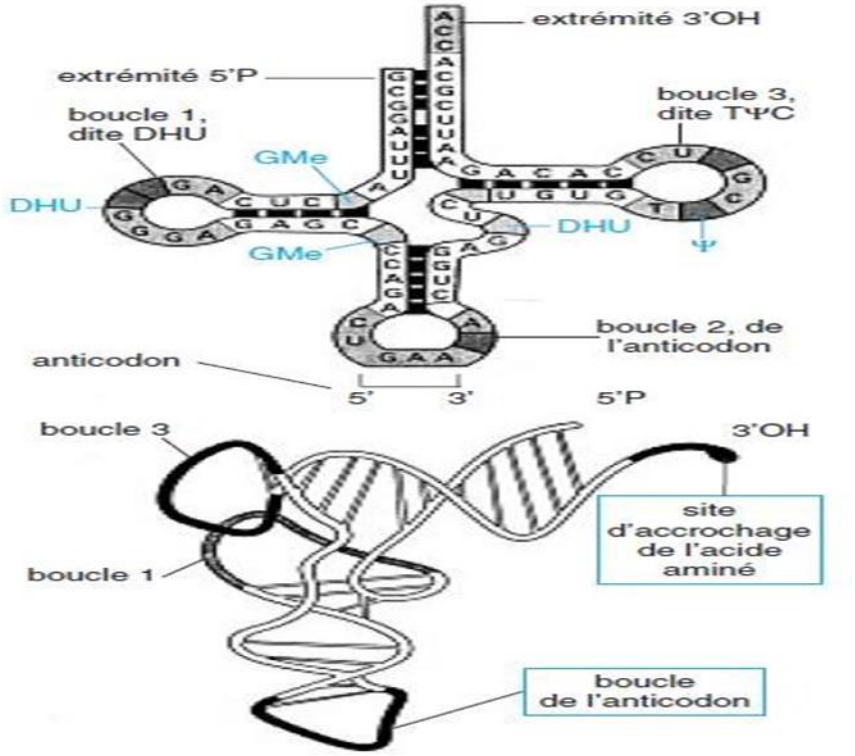
They are essential for the transfer of amino acids found in the cytoplasm to the ribosomes. For a given amino acid, there will be a tRNA capable of loading this amino acid and positioning it in the peptide chain. For 20 amino acids, there are about a hundred transfer RNAs.

The composition of tRNAs presents a large number of rare or minor bases resulting from the chemical modification of the four classical nucleotides after transcription. Example: case of adenosine: we can find 20 variants which are different methylations on Adenosine, case of uracil: pseudo-uridine (ψ), dihydro-uridine (DHU) methylguanosine (GMe), etc.,

2-Structure

tRNAs consist of 75 to 90 nucleotides and their sequence is completely known. For the secondary structure, which is due to hydrogen bonds. the structure of tRNA was made of five regions (cloverleaf structure)

A tertiary structure is seen during additional pairings which give a more compact structure.



c) RNA messengers (mRNA)

The sequence of bases is complementary to that of DNA. It carries the information contained in the genome to the ribosomes. Its chain length varies depending on the message of the protein to be encoded.

2-Life Duration

The lifespan of messenger RNAs is very short; in bacteria, a few minutes, and in eukaryotes, a few minutes to a few hours. Messenger RNAs are renewed very quickly. They are rapidly produced and then degraded. They only last the time of one message, however each one can be read several times at the ribosome level.

3-Structure

Sequences start with AUG and always end with triplets UAA, UGA UAG (stop codon). When we have a coding sequence, we have an exon and when it is non-coding, an intron.

d) RNAnuclear

Heterogeneous RNAs (nhRNAs)

They are of very variable sizes since their sedimentation constant is between 10s and 100s or 600 to 6000 nucleotides.These heterogeneous RNAs are the precursors of mRNAs, are located in the nucleus, they are destined to leave the nucleus and will undergo modifications.

Replication

Replication occurs by specific base pairing from a template which is one of the two strands.

I) General aspect of replication

1. bidirectional replication

Replication begins with an origin of replication, present between small repeating sequences recognized by proteins. It measures approximately 250 bp for prokaryotes and 2000 bp for eukaryotes.
At each origin of replication, a replication eye is formed which grows larger throughout the progression at the replication forks. There are thus two replication systems which evolve in opposite directions.

2. Unidirectional polymerization and semi-conservative replication

The polymerization is unidirectional, in fact the strands being polarized the polymerization will always take place in the same direction: 5'→**3'**. A phosphodiester bond is formed between the 3'OH end of the elongating strand and the 5'phosphate end of the added nucleotide.
Replication occurs by copying the template DNA. Each DNA molecule will separate into two strands of DNA that will serve as a template for the new strands, so there will be a parent strand and a daughter strand in the new DNA molecule. Replication is thus semi-conservative and universal in prokaryotes and eukaryotes.

3. Semi-discontinuous replication

DNA synthesis is bidirectional and both daughter strands are synthesized simultaneously. But DNA synthesis always occurs in the 5' direction.→**3'**, this therefore requires the presence of an early strand which is called the continuous (or primary) strand, which is the strand read in the direction of the fork, and a late strand which is called the discontinuous (or secondary) strand which is the strand read in the opposite direction of the fork. We thus speak of semi-discontinuous replication and the fork is called an asymmetrical fork.

II) DNA polymerases

1) General

Are the enzymes responsible for the polymerization of nucleotides during replication. They are DNA dependent, they need a DNA template to produce the newly synthesized strand. They read the 3' template (parent) strand→**5'**to synthesize DNA (strands) in the 5' direction→**3'**.
DNA polymerases in prokaryotes are 3 types (I, II and III) and eukaryotes 5 types (α, β, δ, ε and γ).
DNA polymerases require:

- The 4 deoxyribonucleosides 5' tri-phosphate (dATP, dTTP, dCTP and dGTP)
- Ions (Mg^{2+}) that stabilize DNA and proteins.
- A DNA template (single or double stranded).
- An RNA primer having a free 3'OH end.

2) <u>Activities of DNA polymerases</u>

When the phosphodiester bond is formed between a dNTP and the elongating strand, there is hydrolysis of the triphosphate function and formation of pyrophosphate (PPi).

DNA polymerases have very specific activities:

- A 5' polymerase activity→**3'**which is their main activity.
- An exonuclease activity which is the degradation of one of the ends of the new strand. 2 types:
    - From 3'→5', which corresponds to degradation from the 3'OH end. The 3' to 5' exonuclease activity allows what is called "proofreading" or "editing = editing function" which corresponds to the correction of a bad base pairing by breaking the phosphodiester bond and replacing the mispaired nucleotide.
    - From 5'→3', which corresponds to degradation from the 5'P end.

Replication in Prokaryotes

1)Prokaryotic DNA polymerases in replication

Prokaryotic DNA polymerases are of 3 types, the most studied of which are DNA polymerases I and III.

- DNA polymerases I are the most numerous (400 molecules/cell). They exhibit 5' polymerase activity→3' as well as 5' exonuclease activities→3' and 3'→5'. The synthesis rate of DNA polymerase I is low (20 nt/s); this characteristic does not allow them to do the majority of prokaryotic DNA replication.

  DNA polymerase I has 3 functions:

  - 5' → 3' polymerization function for replacement of RNA primers by a DNA strand and filling of gaps during DNA repair.
  - 5' → 3' exonuclease function which will eliminate the RNA primers
  - 3' → 5' exonuclease editing function.
- DNA polymerases III are heterogeneous multimers of large molecular weight that are responsible for the synthesis of long fragments of DNA, having a rapid synthesis rate (approximately 1000 nt incorporated/s).

  DNA polymerase III has 2 functions:

  - 5' → 3' polymerization function by addition of nucleotides to the 3'OH end of a nucleotide chain. It is this enzyme that functions at replication forks.
  - 3' → 5' exonuclease editing function like DNA poly I. but not 5' → 3' exonuclease

2) Mechanisms of prokaryotic replication in E-Coli

The entire enzymatic activity located within a replication fork constitutes a replisome.

2.1) The different proteins involved

- Dna A recognition proteins recognize initiation sites at the origin of replication and enable initiation of replication by helping to open the DNA.

- Helicases (or DNA B): one on each strand, they gradually separate the two strands of the double helix by breaking the H bonds between the two strands of DNA.
- Single stranded binding proteins (SSB) have a high affinity for single-stranded DNA and thus prevent it from rewinding during the migration of replicative forks.
- Primase is a DNA-dependent RNA polymerase that synthesizes an RNA primer.
- Topoisomerases (DNA gyrase): Type II topoisomerase releases torsional constraints on DNA formed by the progression of the replication fork.
- DNA ligases (or DNA G) form a phosphodiester bond between two Okasaki fragments.
- Tus protein binds to the Ter termination site and terminates replication.

2.2) The steps of prokaryotic replication

a) Initiation:

The opening of the double helix is allowed by recognition of the origin of replication by DNA A. The origin of replication in E.coli is called Ori C. The Ori C locus is 245 bpd. This sequence contains in tandem (a repeated sequence):

- three nearly identical nucleotide sequences of 13 bpd each rich in thymine (T).
- four binding sites comprising a 9 bpd motif for the DNA A protein.

Several Dna A proteins will bind to the 9-bp repeat sequences. The DNA will then wrap around the protein complex of Dna A. This change of DNA conformation causes the DNA double helix to open at the 13-bp A-rich repeat sequences and Twhich allows the enzymes and other factors to bind to the dissociated strands and start the replication.

The opening of the DNA leads to the formation of the replication eye and the two replication forks. The helicases (Dna B) then attach and are put in place to allow the separation of the two strands.

- DNA unwinding is ensured by a topoisomerase II, which introduces negative superturns. Topoisomerases II are present downstream (in front) of the fork, allowing the removal of topological constraints applied to the double helix by its opening by the helicase.
- SSB proteins attach to untwisted but not yet copied sites on the helix. They prevent the pairing and rewinding of single-stranded DNA and would make the template accessible for replication. They dissociate easily to allow polymerase to pass through.
- A bi-enzymatic complex called a primosome is rapidly formed between the helicase and the primase.

b) Elongation:

After synthesis of the RNA primer, DNA polymerase III inserts at the replication fork and begins synthesis of the complementary strand from the primer.

- ***Elongation of the early strand in the direction of movement of the fork:***

The strand that will serve as a template for the early strand is read in the same direction as the

advancement of the fork, that is to say from 3' to 5'.

· Need for RNA primers:

At the origin of replication, DNA polymerase does not know how to start a chain, it only knows how to lengthen a chain of nucleotides. An RNA polymerase or primase intervenes, which will start a chain of nucleic acids with an RNA fragment of about 10 nucleotides. This RNA fragment is called a "primer".

The RNA primers synthesized by primase are then elongated in the 5' → 3' direction by DNA polymerase III, but this time DNA is added. DNA polymerase III will be responsible for the initiation and elongation of the leading strand.

At the replication fork, the leading or leading strand elongates continuously from a single primer in the 5' → 3' direction that coincides with the direction of fork movement.

- ***Elongation of the late strand in the opposite direction of the fork movement:***

The synthesis of the lagging strand is more complicated since the parent strand that will serve as a template for the lagging strand must also be read in the 3' to 5' direction and the growth must follow the direction of movement of the fork, i.e. 3' to 5', but the fork moves in the opposite direction while DNA polymerase III, which will also be responsible for the elongation of the lagging strand, only attaches the deoxyribonucleotides in the 5' → 3' direction. This was resolved by synthesizing the lagging strand discontinuously in the form of small fragments grafted onto as many "primers". In this way its synthesis will be segmented into fragments of relatively constant size each time the template strand is "uncovered" enough and thus the 5' to 3' elongation direction will be respected. The RNA-DNA fragments thus formed on the lagging strand are called Okasaki fragments. Eukaryotic Okasaki fragments measure 100 to 200 bpd and prokaryotic 1000 to 2000 bpd. At each segment there is recruitment of a primase for the synthesis of an RNA primer.

· <u>Hydrolysis and replacement of RNA primers</u>

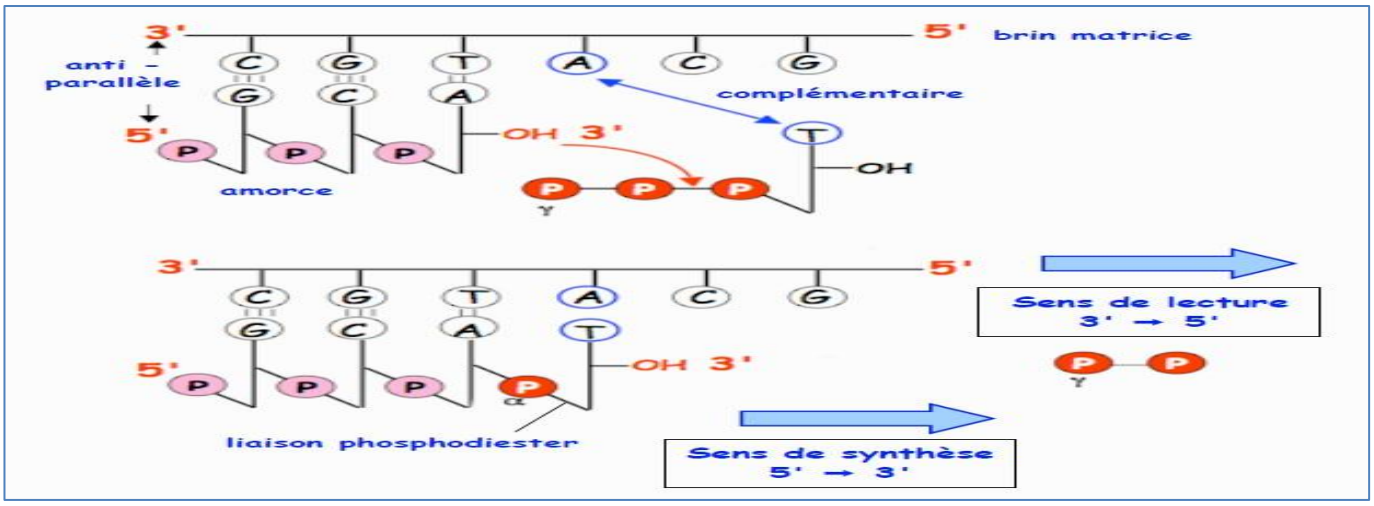
At the lagging strand, as soon as the lagging strand segment has lengthened to reach the 5' end of the adjacent fragment, DNA polymerase I takes over: it eliminates the RNA primers with its 5'→3' exonuclease activity and will complete the gap entirely by synthesizing DNA in its place.

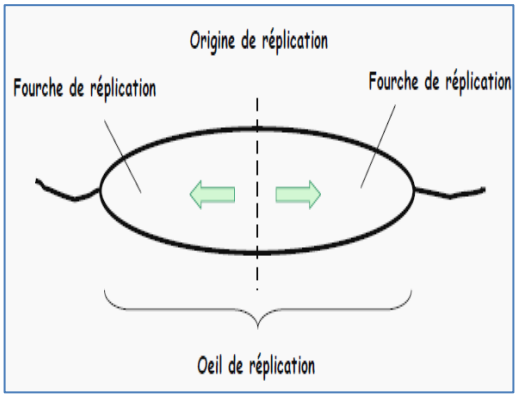
· Finally, the DNA fragments, freed from their RNA primer and having become contiguous, will be welded by a ligase (Dna G) by phosphodiester bond between the 5' end of the first fragment and the 3' end of the second fragment.

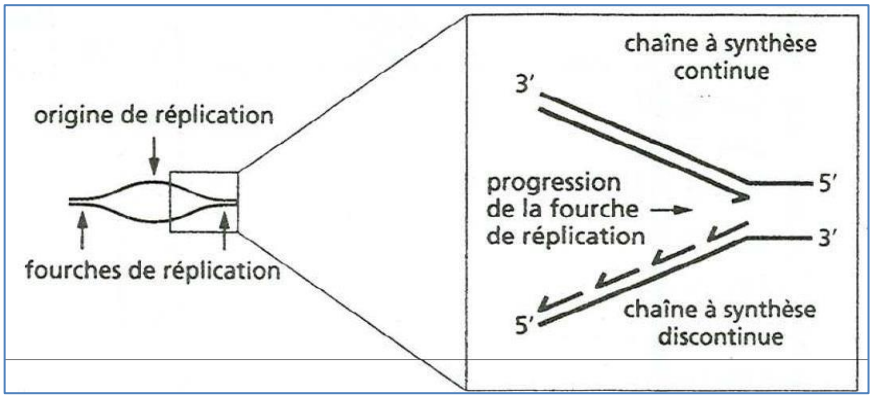
c) Termination:

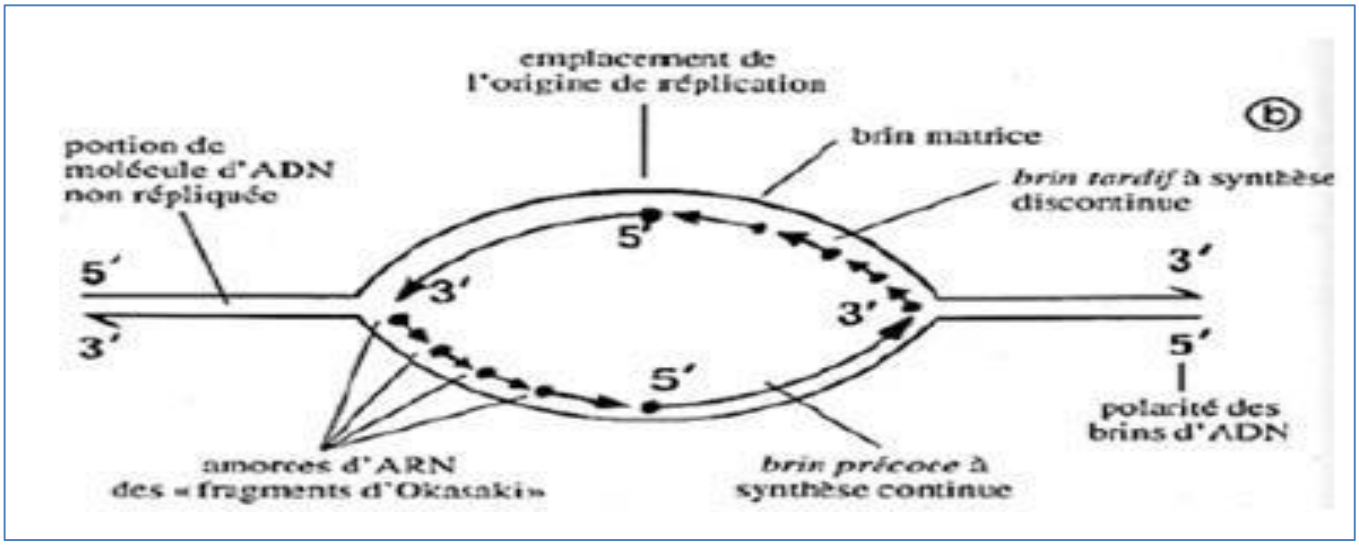
The endings are double for one origin and are rich in A-T base pairs.

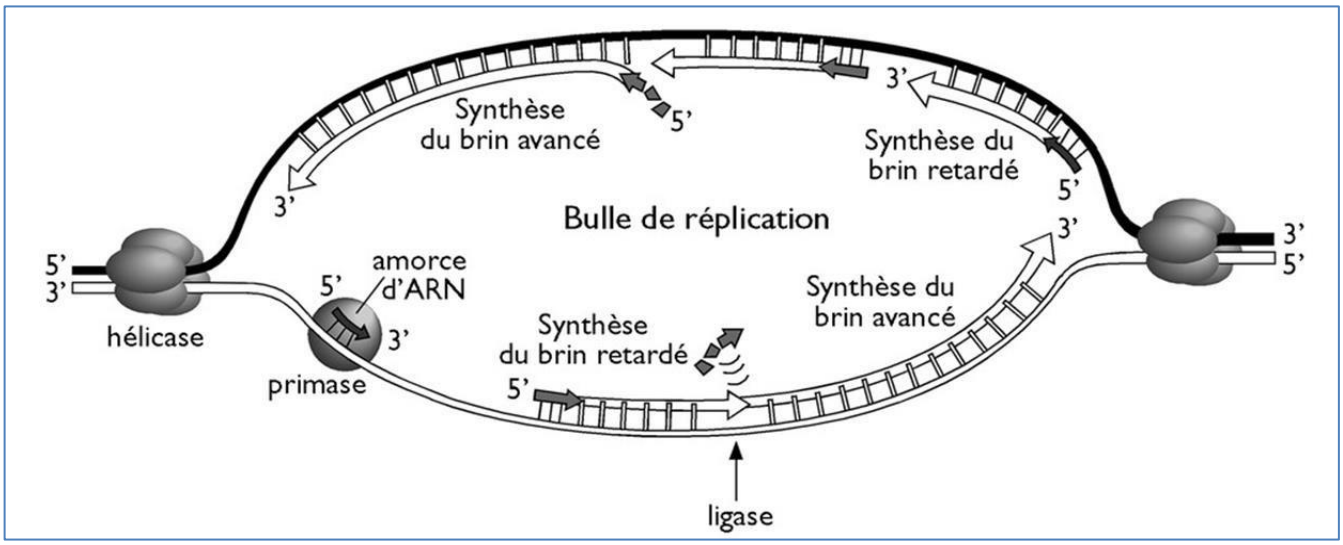
The terminator is the protein binding site "Tus" that recognizes Ter regions. In E-Coli, the part between the two terminators is not initially replicated, the two circular DNAs are thus associated, topoisomerase II is then used to dissociate them. DNA polymerase I will then complete the unreplicated parts.


3'
5'
brin matrice
anti - parallèle
complémentaire
OH 3'
amorce
OH
liaison phosphodiester
Sens de lecture 3' → 5'
Sens de synthèse 5' → 3'


Origine de réplication
Fourche de réplication
Fourche de réplication
Oeil de réplication


origine de réplication
fourches de réplication
chaîne à synthèse continue
3'
progression de la fourche de réplication
5'
3'
5'
chaîne à synthèse discontinue


emplacement de l'origine de réplication
b
portion de molécule d'ADN non répliquée
brin matrice
brin tardif à synthèse discontinue
5'
3'
3'
5'
3'
5'
5'
3'
polarité des brins d'ADN
amorces d'ARN des « fragments d'Okasaki »
brin précoce à synthèse continue


Synthèse du brin avancé
3'
5'
Synthèse du brin retardé
3'
5'
Bulle de réplication
5'
3'
amorce d'ARN
5'
3'
hélicase
primase
Synthèse du brin retardé
5'
Synthèse du brin avancé
3'
5'
ligase

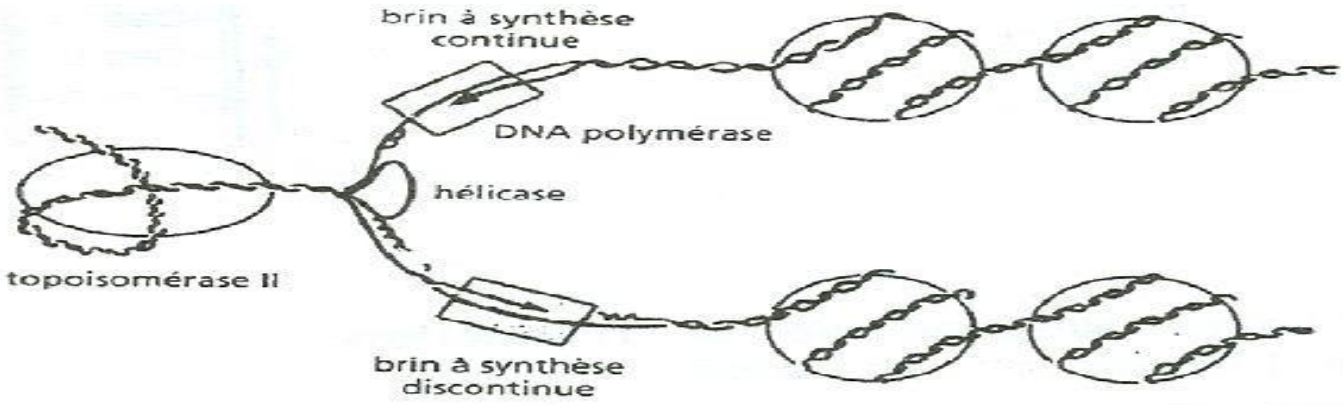
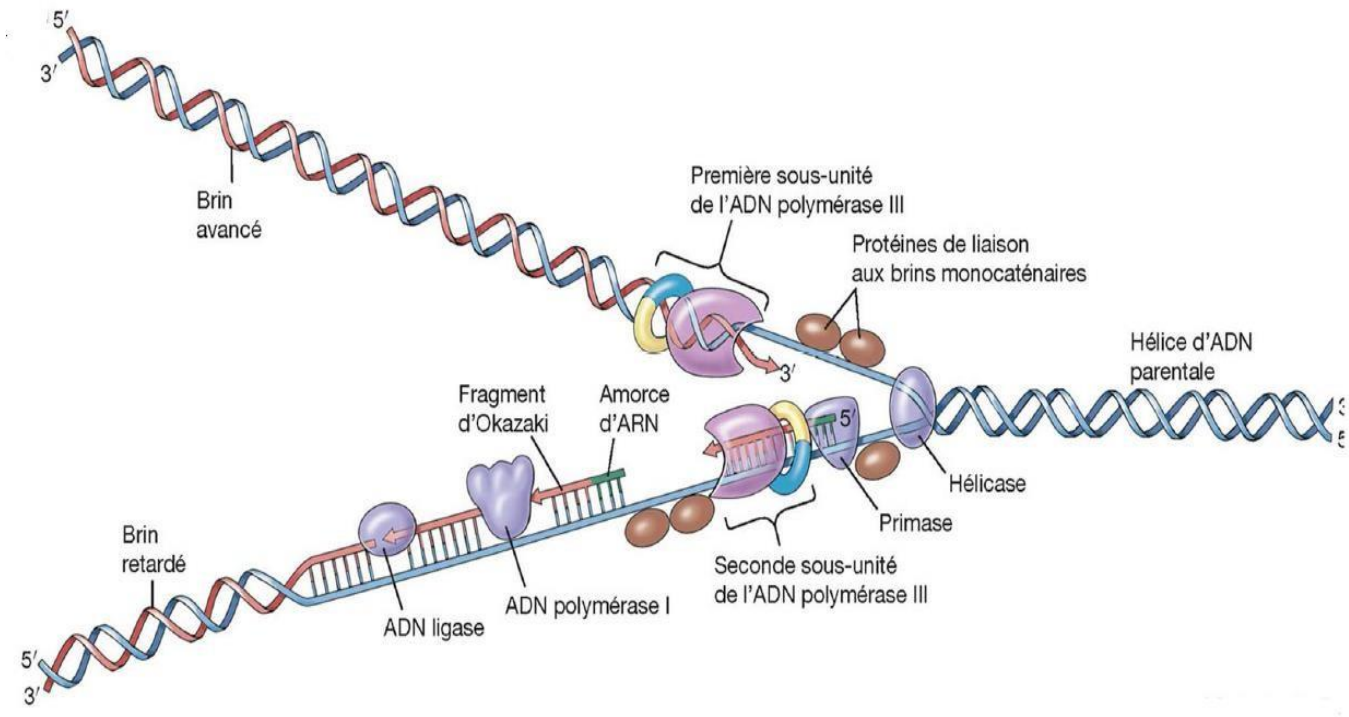
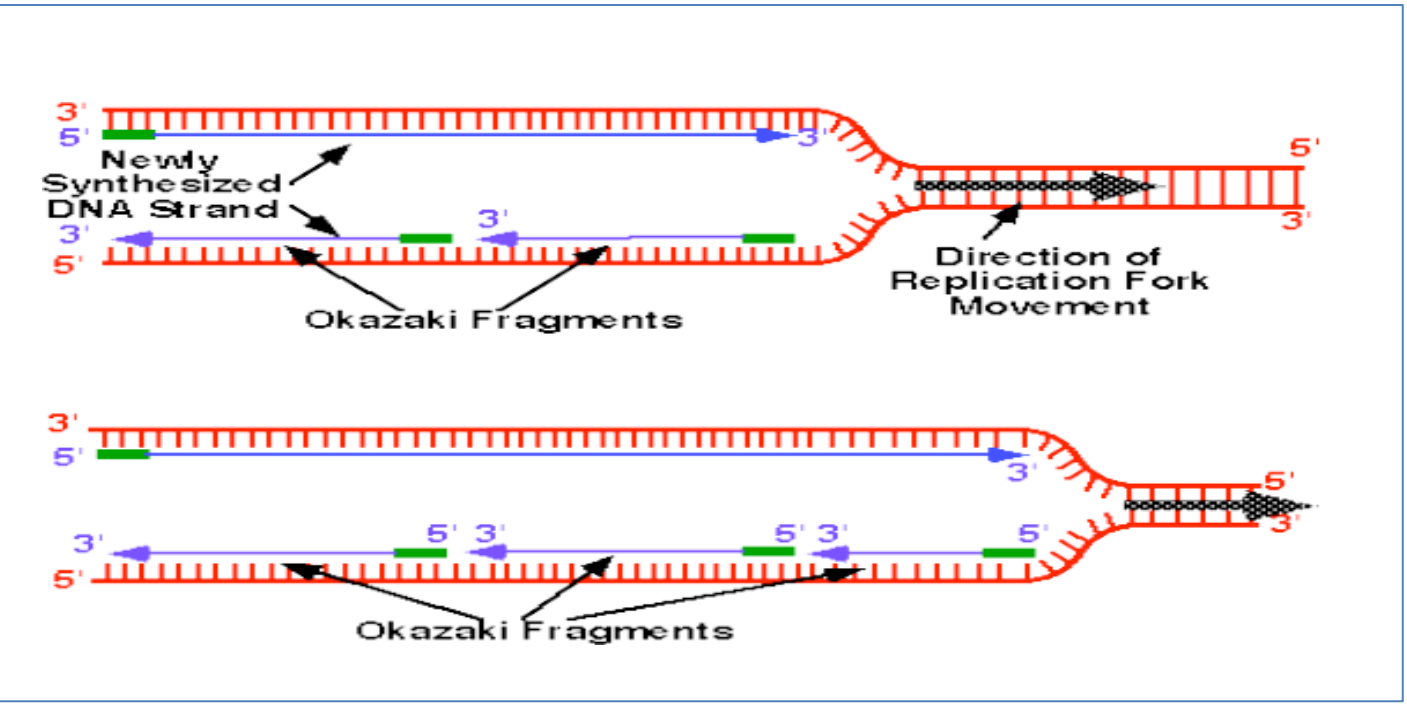


Schéma montrant l'enroulement de l'ADN au niveau des nucléosomes





Replication in Eukaryotes

- Replication is rare in eukaryotes. It occurs in the "S" phase.
- DNA is longer. The replication speed is 50 nucleotides per second; this is compensated by multiple origins of replication (about 10000). (complete replication in about 8H)
- Both forks start from a given origin until they meet forks from an adjacent origin. Each region copied by an origin of replication is called a replicon.

1. DNA polymerases:

- **DNA polymerase alpha α**: has primase activity (named RNA Polymerase). synthesizes short primers. does not have a 3'→5' exonuclease function.
- **DNA polymerase beta β**: is involved in DNA repair processes.
- **DNA polymerase gamma γ**: present in mitochondria for mitochondrial DNA replication
- **DNA polymerase delta δ**: It is the main polymerase of DNA replication in eukaryotes with DNA polymerase epsilon ε in the synthesis of the leading and lagging strands. It has a 3'→5' exonuclease activity involved in error correction and repair processes. This polymerase corresponds to bacterial Poly III.
- **DNA polymerase epsilon ε**: It has a 5' →3' polymerase activity and a 3' → 5' exonuclease activity and is involved in DNA replication and repair.

  It reads in the 3' ⇒ 5' direction and synthesizes 5' ⇒ 3' the telomere region that cannot be synthesized by DNA polymerase delta δ

2. Replication mechanism

- At the replication origin level: "ARS" (Autonomous Replicating Sequence) nucleotide sequences are recognized by "ORC" (Origin Recognition Complex) proteins.
- As the replication fork progresses: helicases unwind the DNA double helix, topoisomerases reduce the twists in the DNA downstream of these forks.
- The unwinding of the DNA and the separation of the two strands reveal regions of single-stranded DNA, stabilized by RPA proteins (Replication Proteins A).
- **DNA polymerase delta δ**. reads the template strand in the 3' → 5' direction and synthesizes DNA in the 5' → 3' direction. It depends on a primer because it can only act if it finds a free 3'OH on which it fixes a new nucleotide.
- The primer is synthesized by DNA polymerase α which: has RNA polymerase activity, it synthesizes a short RNA (10 bp), then synthesizes a short segment of DNA (20 bp).
- At the origin of replication, DNA synthesis begins simultaneously: on the right and on the left and on each of the 2 strands (the 3' → 5' DNA strand: the continuous or leading strand and the 5' → 3' DNA strand: the lagging

or discontinuous strand).

- On the discontinuous 5' → 3' DNA strand, DNA polymerase delta δ synthesizes multiple small DNA fragments from the primers, these are the "Okazaki fragments" (approximately 1000 to 2000 nucleotides long).
  - the enzyme **RNAse H** will remove the RNA primers.
  - **DNA polymerase delta δ**replaces the RNA sequences of the primers with DNA fragments.
  - **DNA ligase** joins DNA fragments by establishing a phosphodiester bond.

After replication, DNA is immediately restructured into nucleosoma using newly synthesized histones.

The transcription

1. Definition

Transcription is the process that transcribes the genetic information contained in double-stranded DNA into information carried by a segment of single-stranded RNA whose base sequence is complementary to that of one of the DNA strands.

All RNA molecules, with the exception of some viruses, are synthesized by complex enzymatic systems, DNA-dependent RNA polymerases.

Three types of RNA are thus produced: messenger RNA, transfer RNA and finally ribosomal RNA.

Transcription is reminiscent of DNA replication (template, polarity, etc.) and differs from DNA replication in two ways; it does not require a primer and only involves limited segments of DNA.

- **Structure of a gene**

Transcription begins at a specific point in DNA and ends at an equally specific point, the space between the two constitutes a transcription unit. A gene can be defined as a region of DNA that is transcribed into RNA.

The gene (or cistron) is a segment of DNA that constitutes the unit of expression leading to the formation of a functional product which can be in the form of RNA or polypeptide.

In prokaryotes, several genes can be part of the same transcription unit, this is called a polycistronic unit.

In eukaryotes, transcription units are monocistronic.

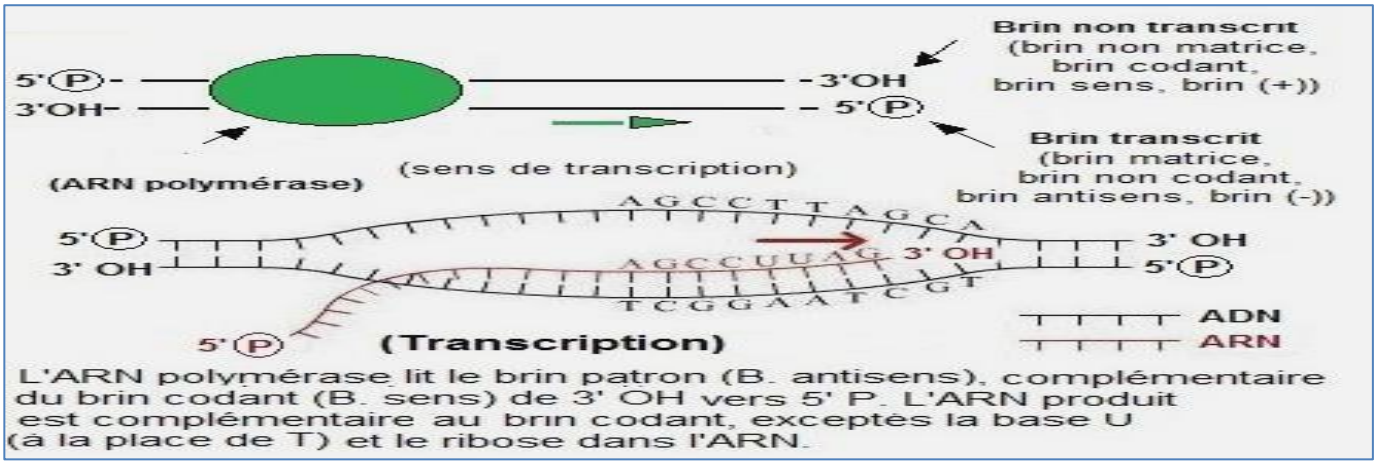
2. General mechanism of transcription

Transcription involves: a template strand, RNA polymerase, a promoter, initiation and elongation factors, and termination signals.

The DNA molecule is composed of:

- of a non-matrix strand (or coding strand or sense (+)) directed by definition from 5' to 3' and having a sequence identical to the transcribed RNA apart from the fact that thymine is changed by uridine.

- of a matrix strand (or anti-coding strand or antisense (–)) directed by definition from 3' to 5' and serving as its name indicates as a template for RNA polymerase for the polymerization of ribonucleotides to



generate a transcript (RNA). RNA is single-stranded and directed from 5' to 3'

ADN	5' TTACCTG 3' (brin codant) 3' AATGGAC 5' (brin matriciel)	➡	ARN 5' UUACCUG 3'

RNA polymerase polymerizes the complementary sequence of the template strand of the gene. Nucleotide triphosphates (NTPs) are added to the 3' end of the chain being synthesized by complementarity of the DNA template.

RNA polymerase binds to a specific site on the template strand, the promoter, which is a sequence of about a hundred nucleotides located in the regulatory region and designating the start of transcription. It is located upstream of the initiation site and carries sequence elements recognized by RNA polymerase and determining the direction of transcription.

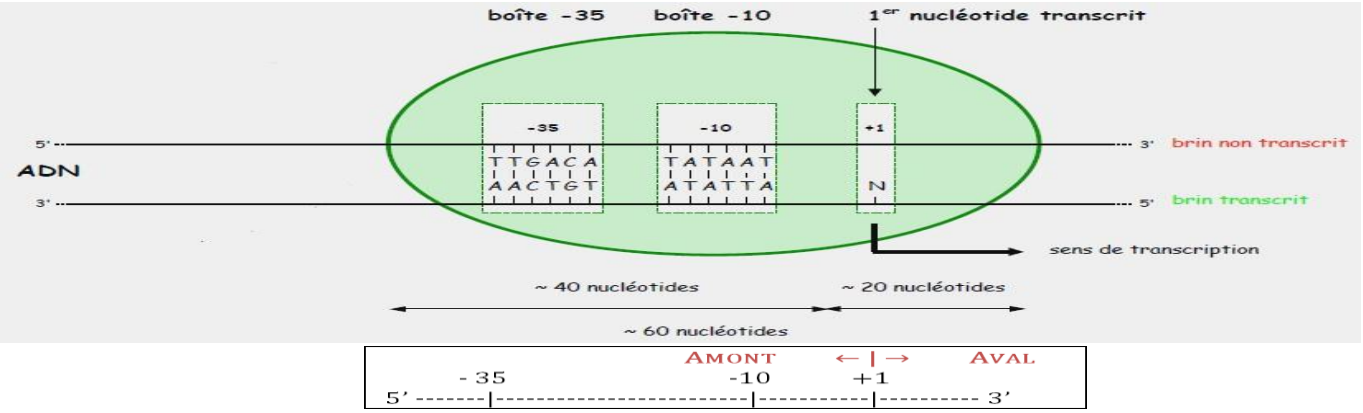
3. Transcription of prokaryotic DNA

In Prokaryotes, transcription takes place in the cytoplasm. A single RNA polymerase catalyzes the synthesis of all RNAs in the cell.

RNA polymerase is a DNA-dependent, large multimeric protein (> 500 kDa) possessing the α, β, β' and σ subunits. It is present in two forms: the core enzyme (α2ββ') and the holoenzyme (α2ββ'σ).

The promoter is made up of short sequences conserved from one transcription unit to another and called consensus sequences:

- At -10 from the initiation site we find the TATA box or Pribnow box: "TATAAT": facilitates the dissociation of the two strands of DNA, because it is rich in A and T
- At -35 from the initiation site we find: “TTGACA”



The affinity of RNA polymerase for DNA depends on the form of the enzyme: the core enzyme has a weak and nonspecific affinity, the holoenzyme has a very strong and specific affinity for the promoter.

The sigma σ subunit therefore allows specific recognition of the promoter by RNA polymerase and reduces the affinity of the enzyme for non-promoter regions. It acts cyclically, in fact after initiation, the sigma factor is detached to be recycled and reused for other gene initiations.

RNA polymerase causes the two DNA strands to denature over approximately 17 nucleotide pairs, known as an open complex, which further increases the enzyme's affinity for the double helix.

3.1. Initiation

- Initiation of transcription requires that the sigma σ subunit RNA polymerase binds to the core enzyme α2ββ' to form the holoenzyme α2ββ'σ.
- Initiation corresponds to the synthesis of the first phosphodiester bond made by the β subunit which corresponds to the catalytic subunit of RNA polymerase.

The sequence of the first steps of transcription is therefore:

1. binding of the holoenzyme and formation of a closed complex at the promoter over an area of approximately 60 nucleotides.
2. Opens the double helix by forming the open complex (transcription bubble formation) by unwinding the DNA over approximately 17 bp.
3. Placement of the first nucleotide (very often A or G)
4. Initiates polymerization by elongation of 4 to 5 nucleotides.
5. Release of the sigma factor after transcription of the first 4-5 nucleotides.
6. Moves on the template strand. The elongation of a nucleotide in the 5' to 3' direction at an average speed of 50 nucleotides/second.
7. Close the double helix at the back.

3.2. Elongation

Elongation is the movement of the transcription bubble along the DNA molecule. The unpaired region is

then 70 base pairs. During transcription, the RNA forms a short pair with the template strand of the DNA forming a hybrid DNA-RNA helix of about ten base pairs.

3.3. Termination

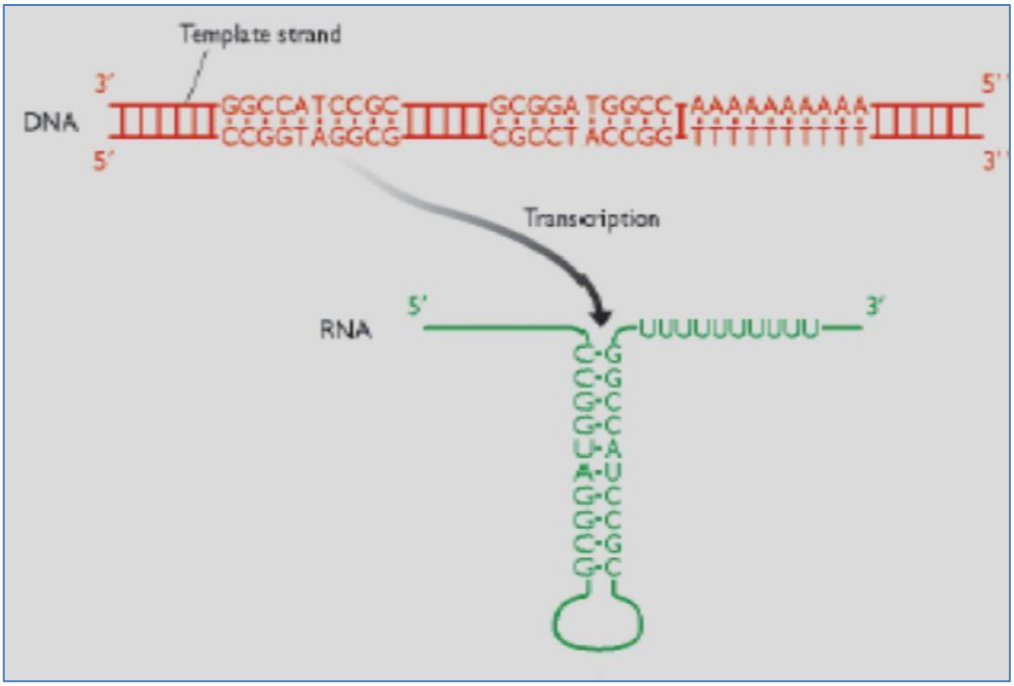
Termination occurs when the enzyme reaches a specific sequence called a terminator. The terminator is a palindrome that results in sequence complementarity at the mRNA level that allows the establishment of a hairpin (or stem-loop) structure rich in GC base pairs, followed by a poly-U sequence of about 6 nucleotides that destabilizes the RNA polymerase until the DNA-RNA hybrid dissociates.

It can be facilitated by a rho factor ρ which has an affinity for the RNAs being synthesized, traversing it from 5' to 3' until it finds the RNA polymerase. The rho factor is ATP dependent, the hydrolysis of which will allow the dissociation of the complex.

4. Maturation of primary transcripts (post-transcriptional modifications):

- The primary transcript corresponds to the immature RNA (pre-RNA) which requires maturation in the form of cleavages or base modifications to obtain the mature RNA.
- Prokaryotic mRNAs do not undergo post-transcriptional modifications.

  While primary rRNAs and tRNAs require modifications before they become functional. Exp: in Bacteria, the maturation of the primary 30S pre-rRNA transcript gives the 23S, 16S and 5S rRNAs.



1. Transcription in eukaryotes:

- **Eukaryotic RNA polymerase:**

- 3 different RNA polymerases:

RNA polymerase I: allows transcribing ribosomal RNAs (except 5S) RNA
polymerase II: allows transcribing messenger RNAs
RNA polymerase III: transcribes small RNAs such as tRNA and 5S rRNA.

- Made up of 10 to 12 sub-units

- Synthesis speed: about 20 to 30 nt/sec

In eukaryotes, RNA polymerases are unable to initiate transcription on their own. Proteins called transcription factors (TFs) are required to bind to the promoter. This assembly provides different possibilities for regulating transcription initiation. The presence of these transcription factors is necessary for the functioning of RNA polymerase.

In most cases, several transcription factors assemble on the DNA at the promoter and recruit RNA polymerase. There are a large number of transcription factors.

- **The promoter**

The eukaryotic promoter is composed of several sequences that define the start and frequency of this transcription. There is a TATA box that is very similar to that of prokaryotes, which defines where transcription begins and is located at –30 upstream of the +1 of transcription. Other sequences, such as the GC –30 and CAAT –80 boxes, define the frequency of transcription.

There is a third class of sequences regulating the transcription rate: enhancer or extinguisher element; they can be located upstream or downstream of the gene, even very far away. Hormonal regulatory elements, certain metals, thermal shocks or even certain toxins act as such.

4.1. Transcription by RNA polymerase II (transcription of cistrons expressed as proteins)

- The transcription of these cistrons is ensured by polymerase II. Most often, genes are structured in a mosaic of coding portions (exons) and portions having no protein significance (introns).
- RNA synthesis in eukaryotes generally gives rise to a "primary transcript" (or premessenger) transcription product which will have to undergo maturation to provide the mature functional cytoplasmic messenger.

Gene transcription occurs in three main phases: initiation, elongation, and termination.

4.1.1. Initiation phase

Initiation of transcription by RNA polymerase II is mediated by general transcription factors called TFII ABCDEF- and H. These factors assemble on the promoter.

- A first TFII-D complex (a multiprotein complex whose TATA box binding element is called TBP for "TATA Box Binding Protein" specifically recognizes the TATA box and binds to it), recognizes the

promoter and allows the fixation of TFII-A then protein interactions between this assembly, TFII-B and polymerase B allow the fixation of the latter, its maintenance is ensured by TFII-E. Other factors (TFII-H and J) participate in the modification of the topology of the DNA by unwinding of the helix by approximately 1/3 of a turn.

- Transcription begins 20-30 base pairs (bp) downstream of the TATA box, at the transcription initiation site (ATG), which by convention is called +1. The transcription initiation complex will catalyze the formation of the first phosphodiester bond between the first two nucleotides of the messenger RNA (mRNA). RNA polymerase II then moves along the DNA by opening a part of the DNA molecule by unwinding the double helix over a short distance, forming a transcription loop.

4.1.2. Elongation phase

The transcription loop moves in the 3'-5' direction of the template strand and the mRNA chain elongates in the 5'-3' direction.

The elongation of the mRNA molecule is done by the pairing of complementary bases and by the successive addition of nucleoside 5' triphosphates. This elongation requires additional factors called elongation factors which are necessary for the displacement of RNA polymerase II. The read DNA rewinds immediately after reading.

4.1.3. Termination phase

Transcription termination is still poor. However, there is a consensus sequence AAUAAA that apparently serves as a signal for an RNA endonuclease to cut the transcript at 3'. The 3' end thus formed is polyadenylated in the nucleoplasm.

4.2. Maturation of primary transcripts (or post-transcriptional modifications):

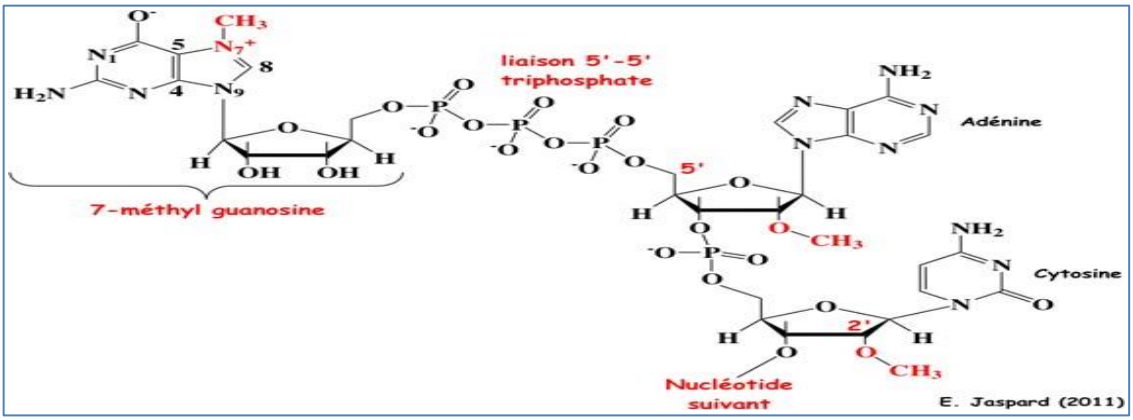
Some RNAs in eukaryotic cells, particularly so-called pre-messenger RNAs, undergo post-transcriptional modifications, catalyzed by several enzymes located in the nucleus.

- Addition of a protective cap to the 5' end
- A polyadenylation at the 3' end.
- Excision and splicing (removal of non-coding parts called introns)

1- Addition of the RNA cap

It is the addition of a 7-methylguanosine (N7) on the first nucleotide of the RNA, by a 5'-5' triphosphate bond (notation: 7mGpppN). This modification is called the cap (or "5'-cap").

The riboses of the first two nucleotides of the transcribed mRNA can also be methylated at the 2' position.



The cap:

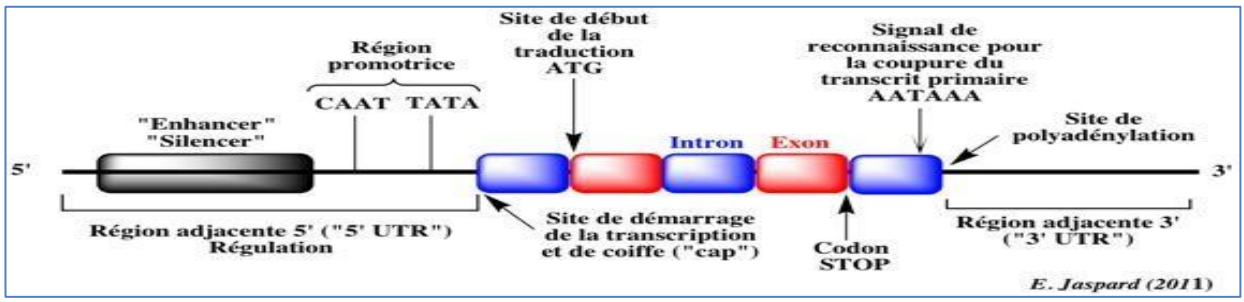
- protects the 5' end of RNAs against degradation by 5' -> 3' exonucleases, thus increasing their half-life.
- participates in the transport of RNA to the cytoplasm
- contributes to the initiation of translation

2- Polyadenylation of messenger RNA

Polyadenylation is the addition of up to 200 adenines which constitute what is called the "polyA tail" at the 3' end of the primary transcript and this without a template by poly-A-polymerase. The only RNAs that escape this maturation are those coding for histones.

Its role:

- stimulates transcription termination
- participates in the migration of messenger RNAs into the cytoplasm
- protects messenger RNA from degradation by exonucleases
- contributes to the initiation of translation



3 - Excision of introns and splicing of exons (or splicing)

In eukaryotes, a gene is not a coding sequence but a set of coding sequences (Exons) separated by non-coding sequences (Introns).

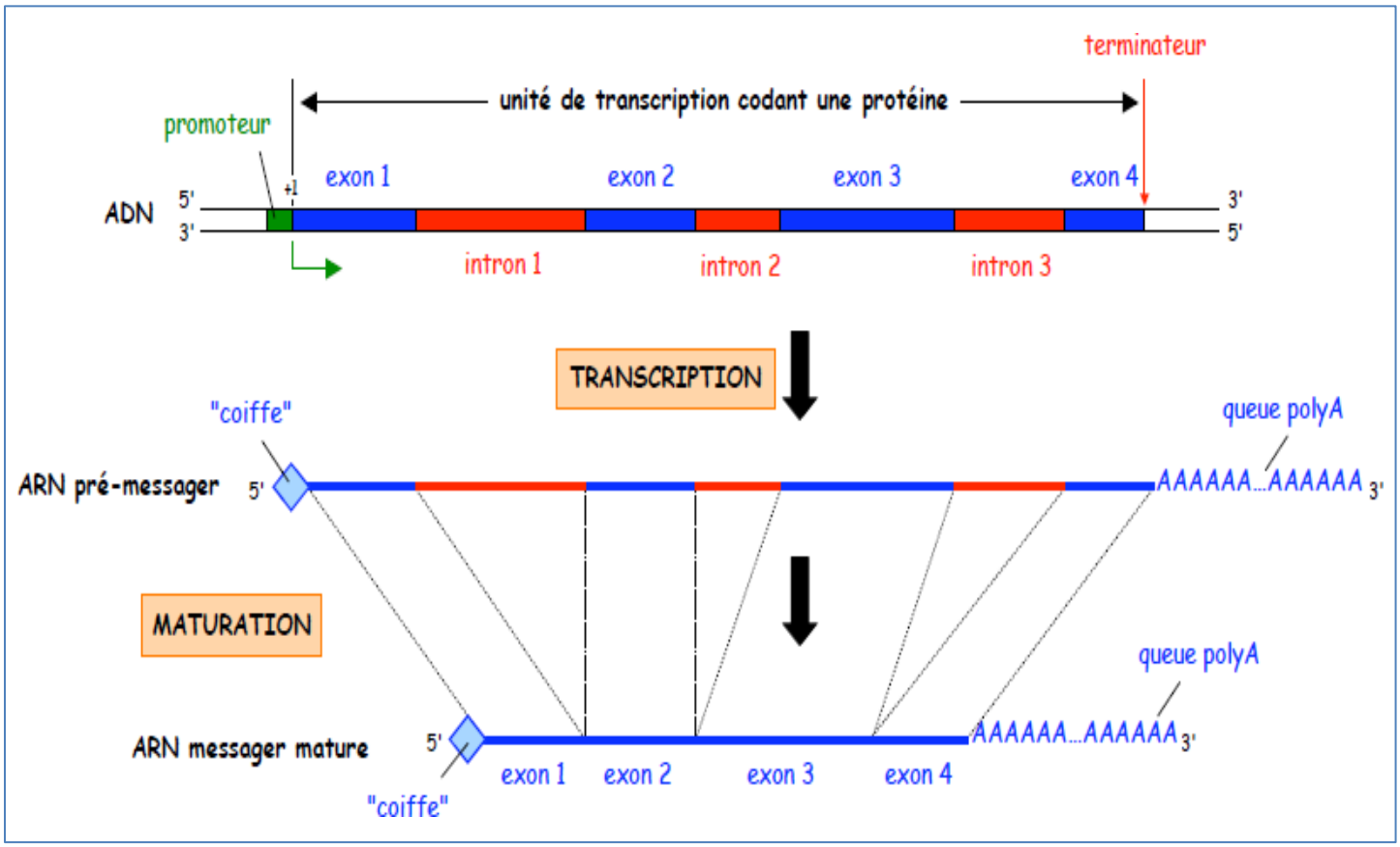
After cap addition and polyadenylation, the primary transcript is further subjected to intron excision and exon splicing; introns are thus eliminated.

Splicing is done using ribonucleoprotein complexes.

This efficiency is obviously due to the snurps (small nuclear binonucleoprotein particles) which are in fact snRNAs (small nuclear RNAs) associated with proteins. The set of snRNPs is called the spliceosome.

- **Alternative splicing**

From a primary transcript, we can have two or more mature mRNAs that will be at the origin of the formation of protein isoforms. This is possible thanks to alternative splicing which consists of the elimination of certain exons. Indeed, some exons are constant at the level of the different mature mRNAs and others are variable and specific to the tissue in which the protein isoform is found.



DNA translation

Translation is when mRNA is translated into protein: the transition from nucleotide sequences to amino acid sequences in accordance with the genetic code. Translation takes place in the cytoplasm of the cell.

I) The genetic code

The genetic code is a code that allows the conversion of a sequence of nucleotides (DNA then RNA) into a sequence of amino acids (proteins). The code involves the bases A, C, U and G as well as the 20 amino acids. The genetic code has different characteristics (Fig):

- **Codons are triplets of nucleotides**and they code for an amino acid.
- **The genetic code is universal**. In fact, each amino acid has one or more codons and this is true for a multitude of living prokaryotic and eukaryotic organisms.
- **The genetic code is redundant**(or degenerate). Several codons code for the same amino acid
  : there are 64 codons and 20 amino acids. Often it is the first two nucleotides of the codon that define the amino acid, so the redundancy is due to the third nucleotide of the codon.
- **The genetic code is non-overlapping**. The nucleotides of a codon only participate in the code for a single amino acid, so the next amino acid will be coded by the next codon present on the mRNA.
- **The code has**an initiation codon which is the AUG codon and the termination codons are the UAA, UAG and UGA codons.

II) The actors of translation

The actors in translation are messenger RNA (mRNA), transfer RNAs (tRNAs), ribosomes, amino acids, amino-acyl tRNA synthetases, Mg2+, GTP and ATP.

1) Ribosomes

Ribosomes consist of ribosomal RNA (rRNA) and proteins and are structured as two subunits in both prokaryotes and eukaryotes. Their size is defined in Svedberg units.

Schematic topography of the bacterial ribosome:

The bacterial ribosome has specific sites:

- **Site A:**(= Amino Acid or Acceptor site) attachment of amino acids.
- **Site P:**(= Peptide or Donor site) fixation of f-Met.
- **Site E:**(= Exit site) exit of the transfer RNA.

The chaining of ribosomes on the mRNA forms the polysome, it allows to increase the efficiency of translation. The minimum distance which separates two ribosomes is 100 nucleotides.

2) tRNAs

a) <u>Structure of tRNA and iso-acceptor tRNA:</u>

tRNAs have a three-leaf clover-shaped secondary structure and an upside-down L-shaped tertiary structure. During the translation mechanism there is an antiparallel pairing between mRNA and tRNA: codon-anticodon recognition at the anticodon loop (Fig.).

tRNAs also have an arm of the amino acid that attaches it in 3' (CCA) on the ribose, this is a covalent bond: an energy-rich ester bond. The amino acids will thus not arrive free on the ribosome but associated with their respective tRNAs. There are 40 to 60 different tRNAs per cell, so there are several different tRNAs for an amino acid, they are called iso-acceptor tRNAs.

b) <u>Loading the amino acid onto the tRNA:</u>

Formation of the amino-acyl-tRNA (aa-tRNA) complex requires an amino acid-specific amino-acyl-tRNA synthetase, which must recognize all codon forms of that amino acid. Correct loading of the tRNA is an important element in translation fidelity.

The amino acid (aa) is first activated and this activation requires energy in the form of **ATP** to allow the formation of aa-AMP (mixed anhydride bond). The bond formed between the tRNA and the amino acid is a covalent bond of the carboxy-ester type (Fig).

$$\text{Aminoacide} + \text{tRNA} + \text{ATP} \xrightarrow{Mg^{2+}} \text{aminoacyl-tRNA} + \text{AMP} + 2P_i$$

There are 20 amino-acyl-tRNA synthetases in the cell, as many as there are amino acids involved in translation. The complexed amino acid can thus associate with the chain.

III) The different stages of prokaryotic translation

1) <u>Initiation</u>

A ribosome recognizes the start of the coding sequence, it uses upstream addressing signals between -8 and-13 of the initiation codon (AUG) which corresponds to the Shine-Dalgarno sequence (AGGAGG). There is antiparallel base pairing between the mRNA and the small subunit (30S) of the ribosome, due to sequence complementarity between the mRNA and the 16S rRNA.

Procayotes require a particular amino acid for initiation; methionine which requires formylation on the NH2 end (addition of a formyl) to form f-Met, this is a pre-translational phenomenon. This formylation is carried out by a cofactor, vitamin B9 which recognizes the characteristic tRNA responsible for the transport of f-Met-tRNA. The conformational peculiarity of this tRNA allows it to be placed directly in the P site and not in the A site.

Initiation is enabled by the presence of initiation factors (IF):

- **IF 1**is the 70S ribosome dissociation factor.
- **IF 2**is a factor ensuring the fidelity of recognition between tRNA and the amino acid. It also has GTPasic activity (i.e. GTP hydrolysis).
- **IF 3**is a factor required for the specific binding of 30S to mRNA and for controlling the balance between

the associated and dissociated form of the ribosome (anti-reassociation factor).

Subsequently the 70S complex is reformed: when the tRNA attached to formyl-methionine is attached to the small 30S subunit there is hydrolysis of GTP and the large 50S subunit attaches to the complex.

2) Elongation

Elongation corresponds to a protein synthesis by addition of amino acids to the C-terminal end of the nascent peptide chain, a reaction catalyzed by the peptidyl-transferase activity of the large SU of the ribosomes. The reading of the mRNA by the ribosome is done from 5' to 3'. There is formation of a peptide bond between two functions (NH_2 and COOH) of two different amino acids with the elimination of H_2O. Elongation is also allowed by the presence of elongation factors (EF for Elongation Factor): EF-Tu; EF-Ts and EF-G.

For each peptide bond formed, three steps can be characterized: the coupling reaction, the formation of the peptide bond and the translocation.

a) Coupling reaction:

The coupling step corresponds to the transfer of the amino acid complexed with the tRNA onto the protein chain in the process of elongation. It can be noted that during translation it is the N-terminal end (amine function) that comes out of the ribosome first. Thus it is the C-terminal end (carboxyl function) of the first amino acid that will allow the formation of the peptide bond with the amine function of the second amino acid. Thus the second aa2-tRNA complex arrives in the A site, the f-Met being positioned in the P site.

b) Formation of the peptide bond and release of the first tRNA:

The high-energy bond between the first tRNA and f-Met breaks to allow the formation of the peptide bond, due to the attack of NH2 of the aa2-tRNA on COOH of the 1st Aa. This is possible because the COOH function is engaged in the binding to the tRNA, leading to the formation of a dipeptidyl-tRNA located in the A site and the deacylated f-Met-tRNA in the P site. The reaction is catalyzed by the peptidyl-transferase activity of the large subunit.

c) Translocation:

The reading of the mRNA is done from 5' to 3', so the ribosome moves 3 nucleotides (or one codon) in this direction so that the tRNA (dipeptidyl-tRNA) carrying the first two Aa moves towards the P site, there is translocation. Whereas the deacylated tRNA goes to another site, the E site which will be expelled into the cytoplasm where it will be recycled.The tRNA carrying the third can then take its place in the once again free A site.

This three-step cycle will therefore be repeated as many times as necessary until the stop codon.

3)Termination

Translation termination occurs at the stop codons UAA, UAG and UGA which do not code for any amino acid.

- the A site of the ribosome is positioned on the stop codon of the mRNA.

- Stop codons are recognized by RF termination factors (RF for Releasing Factor):

  RF 1(recognizes UAA and UAG); RF 2 (recognizes UAA and UGA) and RF 3 stimulates the activity of the other 2 factors.
- These GTP-activated RF factors associate with the ribosome and cleave the bond between the peptide and the tRNA
- GTP hydrolysis allows: release of RF; dissociation of ribosome subunits; release of protein, tRNA and mRNA.

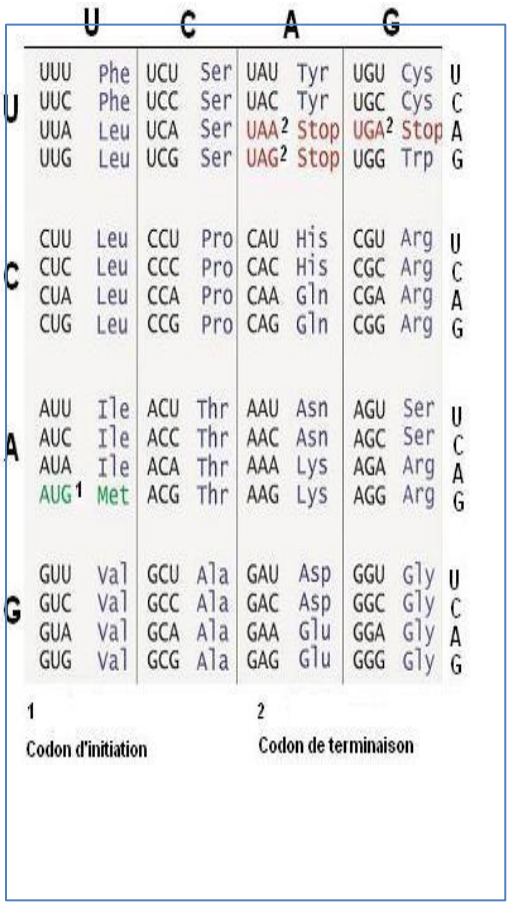
IV) The specificities of eukaryotic translation

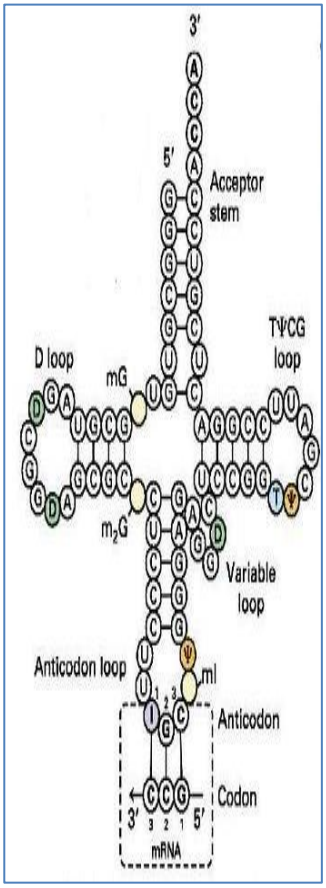
The ribosome is different in size and composed of different ribosomal RNAs although the general structure and activity is comparable.

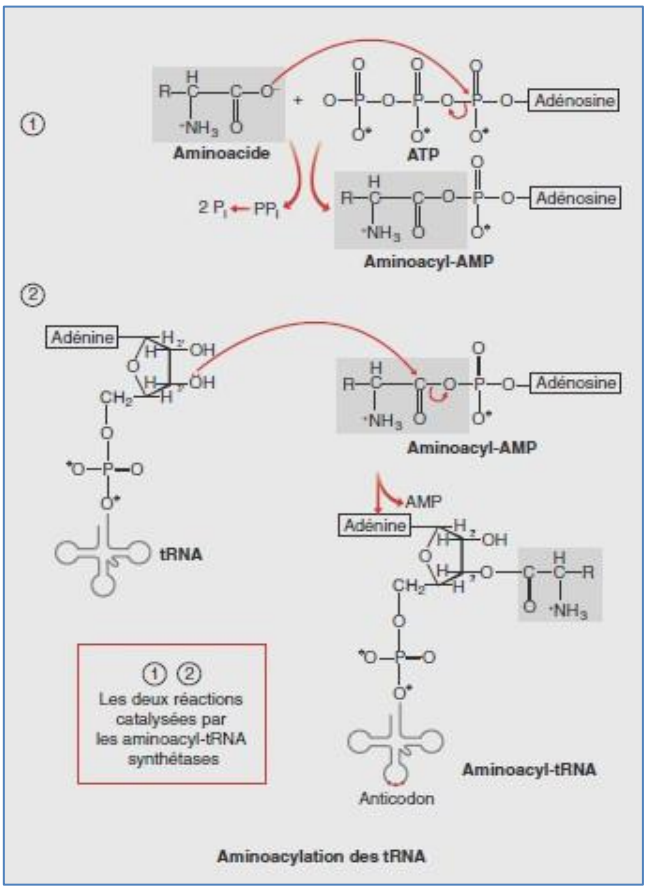
The initiator codon is also AUG and it is usually the first AUG present on the mRNA. In eukaryotes the first amino acid is methionine and not f-Met present in prokaryotes. Methionine will most often be removed just after the synthesis of the peptide chain.

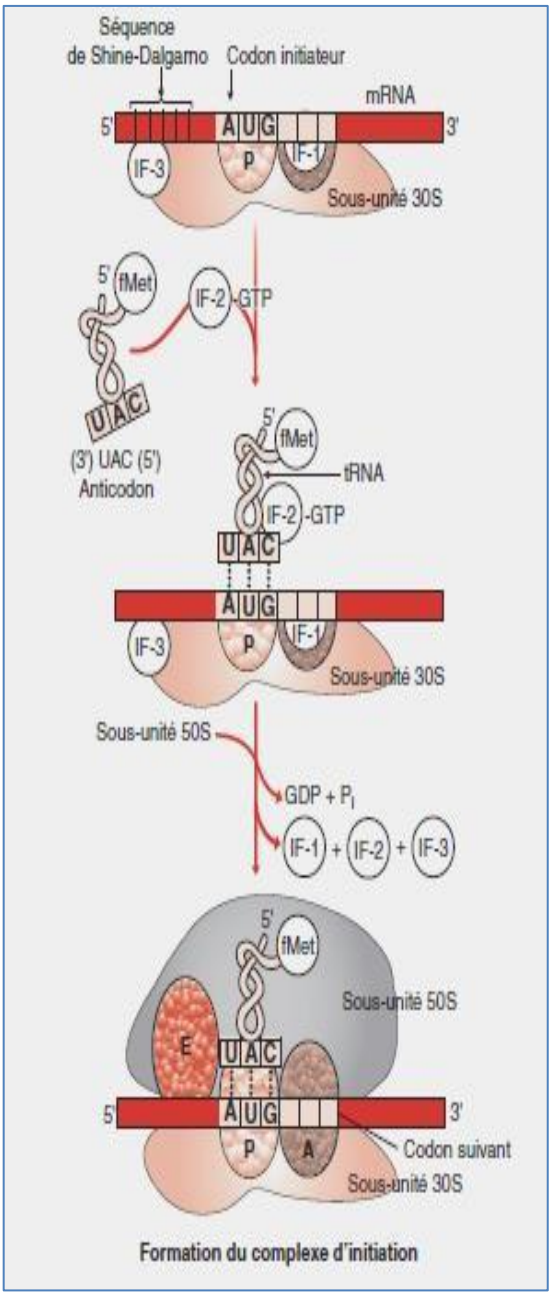
The initiation factors are of the eIF (eukaryotic initiation factor) type, from eIF1 to eIF6. The elongation factors are also of the EF type (EF1α, EF1β and EF2).

The termination factors are of the eRF type (for eukaryotic Releasing Factor).

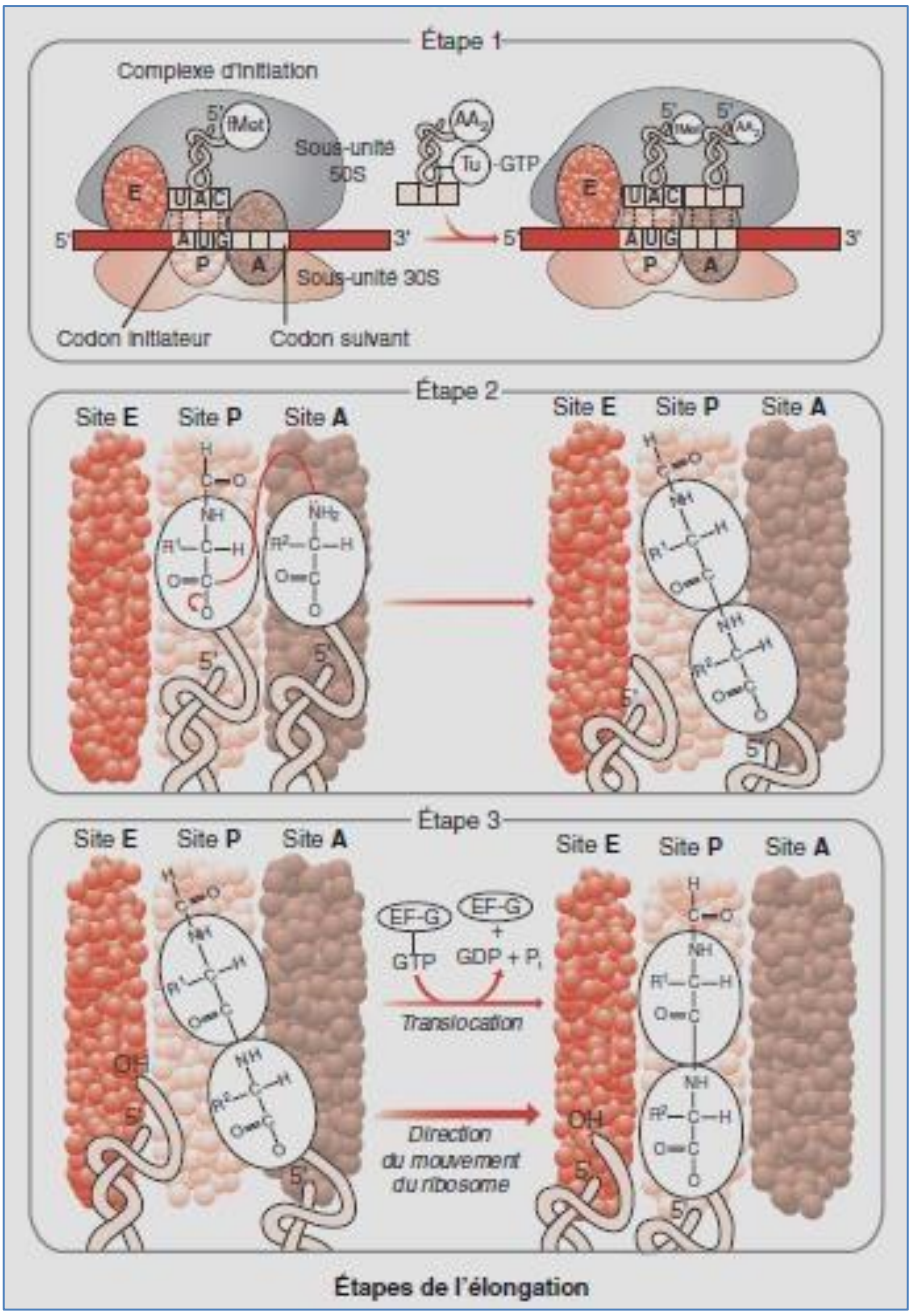








Formation du complexe d'initiation



Étapes de l'élongation