

Matière : Biologie moléculaire
Semestre : 1 (U E F1)
Nombre de crédits : .6
Coefficient de la Matière : 3
Pr. Ouldjaoui Abdallah

CHAPITRE 1 : STRUCTURES DES ACIDES NUCLÉIQUES

- 1-1- Structure des acides nucléiques
 - Les nucléotides : structure et propriétés
 - La structure primaire des acides nucléiques
 - La structure tridimensionnelle de l'ADN :
 - caractéristiques de la double hélice
 - compactage intranucléaire et chromosomes
 - Les différents types d'ARN : principales caractéristiques structurales et fonctionnelles
- 1-2- Organisation des gènes et des génomes

CHAPITRE 2 : LA REPLICATION DE L'ADN

- 1- La réplication de l'ADN chez les procaryotes
- 2- La réplication de l'ADN chez les eucaryotes

CHAPITRE 3 : LES MUTATIONS ET MECANISMES DE REPARATION DE L'ADN

- 1- Les mutations
 - les causes des mutations
 - Les agents mutagènes
 - Les différents types des mutations
- 2- La réparation de l'ADN
 - les principaux systèmes de réparation

CHPITRE 4 : EXPRESSION DES GENES

- 1- LA TRANSCRIPTION
 - Structure des gènes (chez les procaryotes et les gènes des eucaryotes)
 - La transcription chez les procaryotes
 - La transcription chez les eucaryotes
- 2- LA TRADUCTION
 - Le code génétique
 - Les étapes La traduction chez les procaryotes
 - Les étapes La traduction chez les eucaryotes

CHAPITRE 5 : LA REGULATION DE L'EXPRESSION GENIQUE

- 1- La régulation de l'expression génique chez les procaryotes
 - ❖ L'opéron lactose
 - ❖ L'opéron tryptophane
- 2- La régulation de l'expression génique chez les eucaryotes
 - ❖ Régulation poste transcriptionnelle
 - ❖ Régulation poste traductionnelle

CHAPITRE 6 : techniques utilisées en biologie moléculaire

CHAPITRE 1 : STRUCTURES DES ACIDES NUCLÉIQUES

1-1- Historique

1953 : Découverte de la double hélice (Watson et Crick).

Analyse de clichés de diffraction de rayon X par des cristaux d'ADN

1953 : un modèle du code génétique (Gamow).

1956: l'ARN est découvert.

1958 : La démonstration du modèle semi-conservatif de la réplication de l'ADN (Meselson et Stahl).

1960 : Découverte de l'ARN messenger (Monod).

1961->1966 : le décryptage du code génétique (Nirenberg et Matthaei) grâce à la synthèse de protéines in vitro à partir d'un ARN poly-U.

Affinage du "dogme central" de la biologie moléculaire.

1965 : L'opéron lactose (Jacob et Monod) Etude de l'activité de la Beta-galactosidase Modèle de la régulation de l'expression des gènes

2- Caractéristiques des acides nucléiques

Les acides nucléiques sont de deux types, l'ADN et les ARN. L'ADN ou acide désoxyribonucléique est le support de l'information génétique.

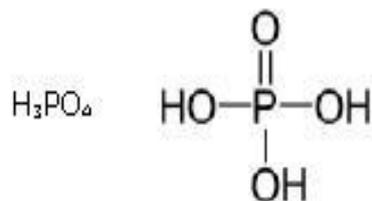
Les ARN ont soit un rôle de support de l'information afin d'être traduit en protéines (ARN messenger), ou bien un rôle structurale (ARN ribosomiques, ARN de transferts et autres petits ARN).

L'ADN est situé dans le noyau mais il est également mis en évidence dans la mitochondrie, les ARN eux sont situés dans le noyau et dans le cytoplasme (également dans la mitochondrie). En effet l'ADN est transcrit en ARN dans le noyau, lui-même traduit en protéines dans le cytoplasme pour les ARN messenger.

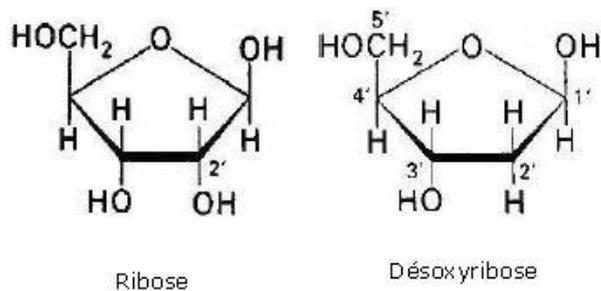
3-Structure des acides nucléiques :

Les acides nucléiques sont constitués d'acides phosphoriques, de pentoses et de bases azotés :

L'**acide phosphorique** est un tri acide dont une fonction acide est dissociée permettant de donner une charge négative à l'ADN, et dont les deux autres peuvent former des liaisons phosphodiester.

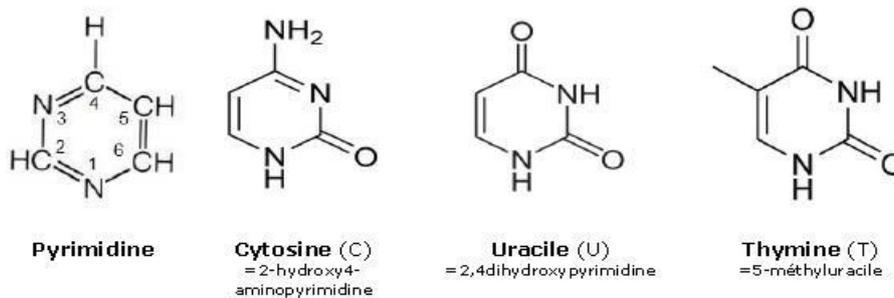


Les **pentoses** sont des glucides cycliques à 5 atomes de carbone qui sont sous forme de β -D-ribofurane et sous forme « désoxy » pour l'ADN.

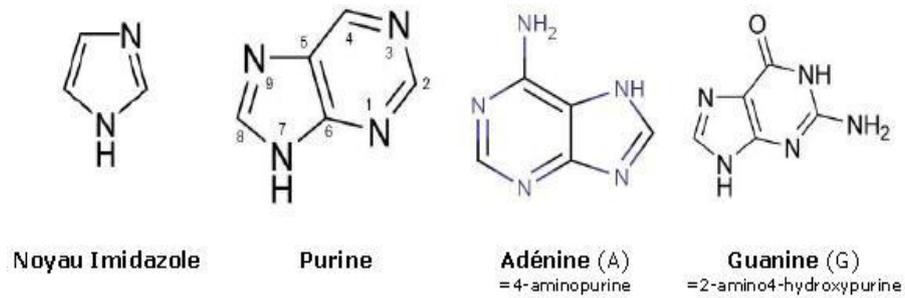


Les **bases** sont de deux types :

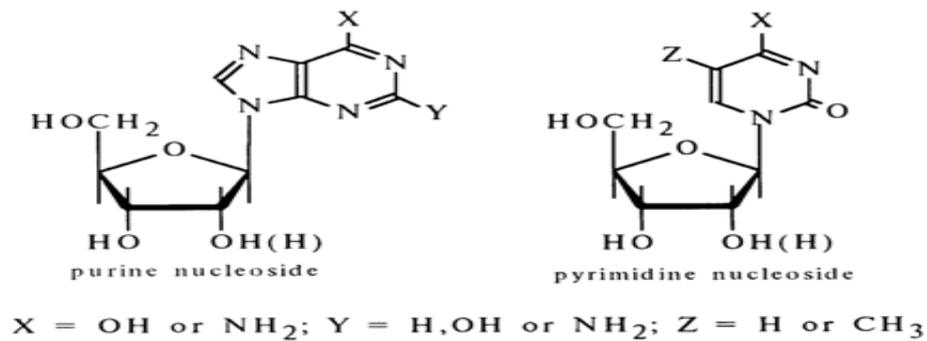
Les **bases pyrimidiques** : la cytosine, la thymine (ADN) et l'uracile (ARN).



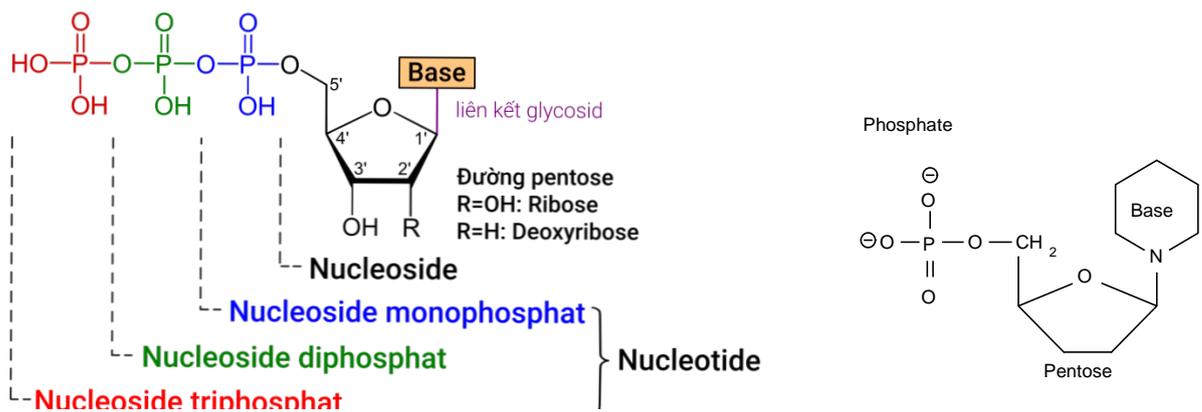
Les **bases puriques** : la guanine, l'adénine et l'hypoxanthine (précurseur des bases puriques et présent au niveau des ARNt.



L'association d'une base et d'un pentose par une liaison N-glycosidique est appelée un **nucléoside**

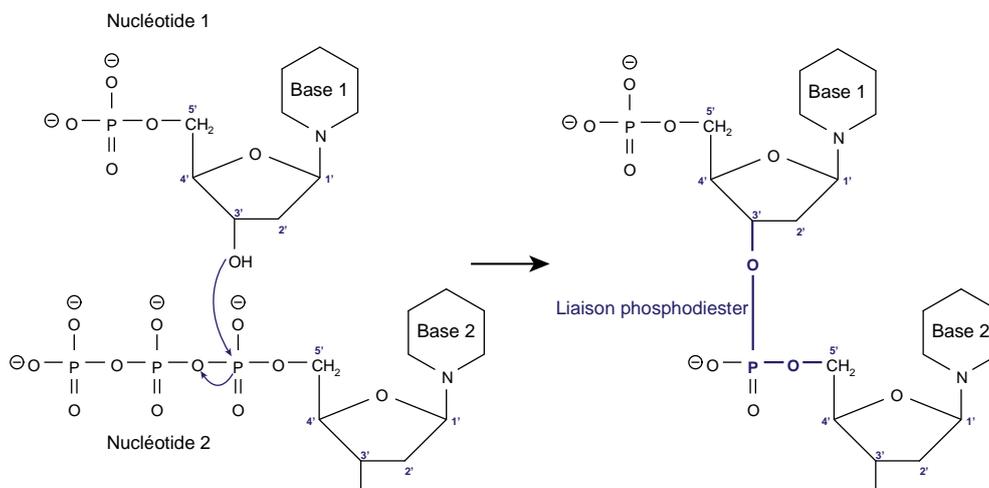


L'association d'un nucléoside et de l'acide phosphorique par une liaison phosphodiester en 5' du pentose est appelée un **nucléotide**.



4- Structure de l'ADN : L'ADN est une molécule **bi caténaire** étant constituée de deux brins dirigés de manière **antiparallèle** et associés en doubles hélices de **type B**. Il est constitué d'autant de bases puriques que de bases pyrimidiques, en effet il y a autant d'adénine que de thymine, et autant de guanine que de cytosine.

Les deux brins sont liés grâce à des liaisons **hydrogènes** présentes entre toutes les bases de l'ADN : deux liaisons hydrogènes entre les bases A et T, et trois liaisons hydrogènes entre les bases G et C. la double hélice de l'ADN a un diamètre de **20 Å**, un pas de **34 Å** qui correspond à **10** nucléotides.



Polymérisation de deux nucléotides et liaison phosphodiester

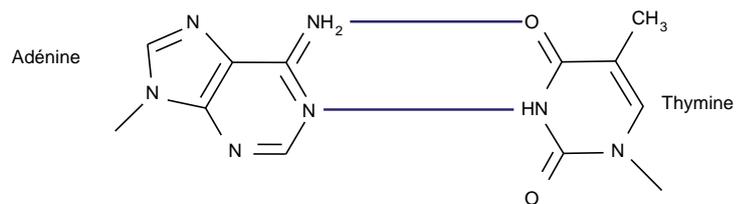
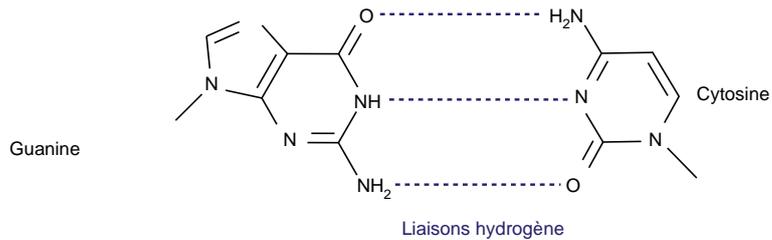
La double hélice d'ADN : La découverte de la structure de l'ADN bouleverse l'étude des phénomènes biologiques en introduisant la dimension moléculaire. Proposée par **Watson et Crick en 1953**, elle est obtenue non seulement à partir de l'interprétation de clichés de diffraction des rayons X réalisées par **Franklin**, mais aussi des travaux d'**Erwin Chargaff** (qui avaient montré que pour toute molécule d'ADN, le nombre de molécules d'adénine est égal au nombre de molécules de thymine, et que celui de cytosine est égal à celui de guanine) et enfin d'analyses en microscopie électronique, qui avaient montré que le diamètre de la molécule d'ADN est de 20Å,

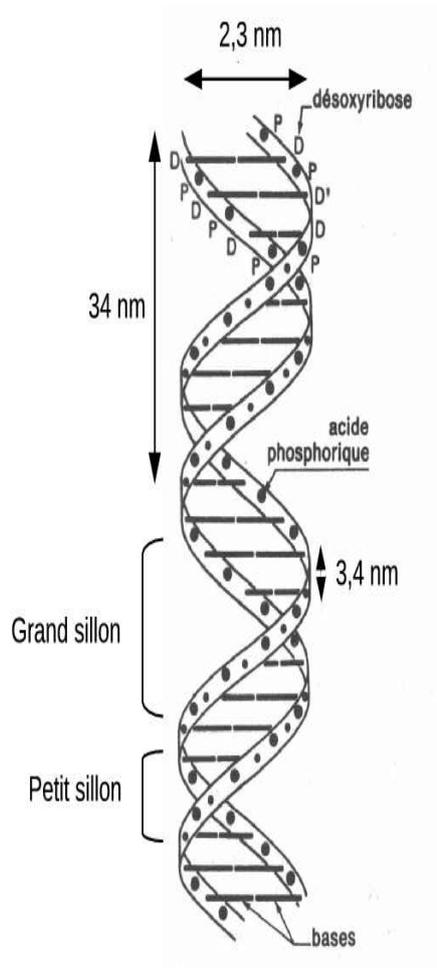
Ce qui suggérait que cette molécule comportait deux chaînes de désoxyribose-phosphate. Les deux points fondamentaux de la structure sont les suivants :

les bases ou nucléotides (A,T,C,G) s'organisent en paires **A-T** et **G-C**

Cet appariement permet un enroulement quasi-parfait en hélice droite des deux chaînes sucre-phosphate qui portent ces nucléotides ; L'hélice possède un pas d'environ 10 paires de bases, qui sont espacées de 3.4 Å, donc le pas vaut environ 3.4nm.

La structure est stabilisée par l'interaction (liaisons hydrogène) entre les bases et l'empilement successif des paires de nucléotides le long de la double-hélice ("stacking" du à des interactions hydrophobes).





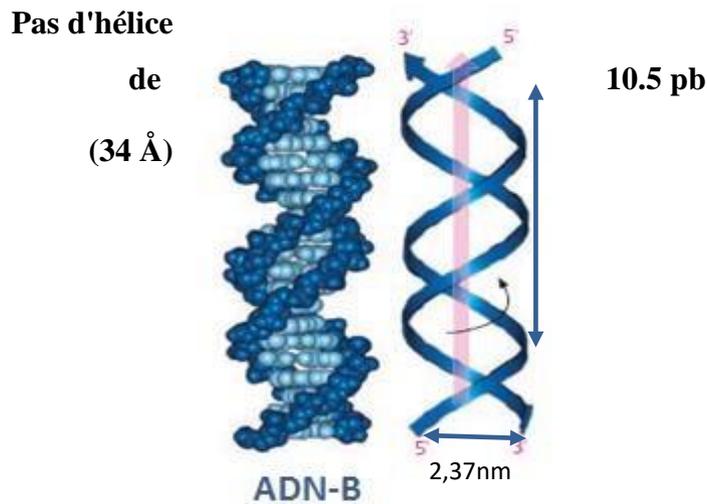
Formes de l'ADN : Dans l'espace les deux chaînes présentent une configuration hélicoïdale. Elles s'enroulent autour d'un axe imaginaire pour constituer une double hélice à rotation droite (Ex. Forme A et B) ou bien à rotation gauche (Ex. Forme Z).

Il existe plusieurs structures hélicoïdale de l'ADN jusqu'à maintenant présent, six formes ont été décrites (A → E et Z), mais la plupart d'entre elles ont été trouvées dans des conditions expérimentales contrôlées.

Forme B

La forme B de l'ADN est la forme biologique la plus importante, elle correspond à la forme décrite par **Watson et Crick en 1953**.

Cette forme est caractérisée par un pas (tour) d'hélice de 10,5 pb (34 Å) et un diamètre de 2,37 nm. La forme B de l'ADN est caractérisée aussi par la présence de deux sillons, le petit sillon et le grand sillon.



Les protéines font des liaisons avec l'ADN au fond des sillons (le grand, le petit ou les deux), ces liaisons sont spécifiques (liaisons hydrogène) ou non spécifiques (interactions de Vander Waals, et interactions électrostatiques générales).

Les différentes formes de l'ADN se distinguent par :

- 1- Nombre de paires de bases dans chaque tour d'hélice.
- 2- Pas de l'hélice (longueur)
- 3- Diamètre hélicoïdal de la molécule
- 4- Sens de la double hélice (droit, gauche).

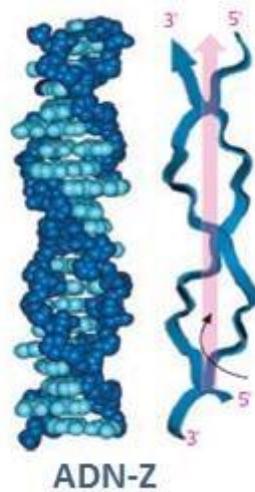
Forme Z

Forme une double hélice à rotation gauche (fig. 10); nombre de paires de bases par tour d'hélice est de 12, le pas d'hélice est de 4.56 nm, et un diamètre de 1.8 nm. Cette forme présente les caractéristiques suivantes:

Le squelette présente une conformation en zigzag favorisée par un taux de G C élevé (moins lisse que l'ADN-B).

- Il y a un seul sillon, qui ressemble au petit sillon de l'ADN-B
- Les phosphates sont plus proches les uns des autres que dans l'ADN-B.
- L'ADN-Z ne peut pas former de nucléosomes.

- La transformation de l'ADN-B en ADN-Z se fait lors de la transcription des gènes .



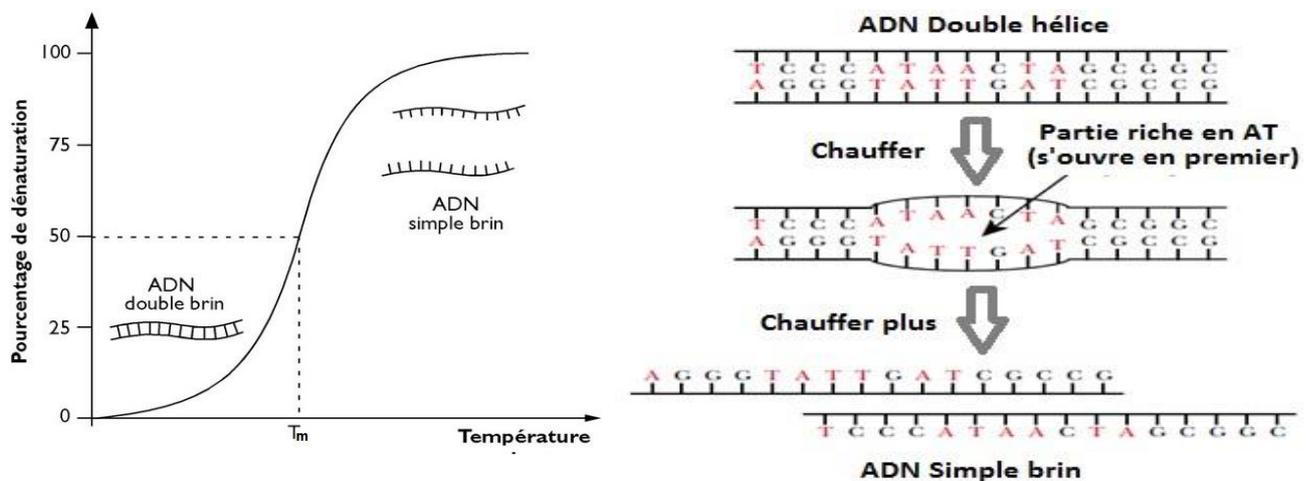
5- Propriétés physico-chimiques de l'ADN

Les liaisons hydrogène et les interactions hydrophobes qui maintiennent la structure en double hélice sont des forces faibles et des quantités relativement petites d'énergie peuvent séparer les deux brins, un processus appelé dénaturation.

1-Température de fusion : Si une solution d'ADN est chauffée, à une certaine température, les liaisons hydrogène qui assurent la cohésion des 2 brins appariés se rompent.

On parle de fusion de l'ADN caractérisée par la température de fusion (T_m).

La dénaturation et la renaturation des brins d'ADN en solution sont des reconstitutions critiques pour diverses fonctions biologiques normales (réplication; transcription ...etc.).



La séparation des brins d'ADN en fonction de la température.

T_m est la température pour parvenir à 50% de séparation des deux brins.

Facteurs influençant la T_m

La température de fusion est influencée par plusieurs facteurs :

- C + G % : la T_m augmente avec l'augmentation du % CG.
- Taille de la séquence d'ADN.
- **Absorption de la lumière ultraviolette** : La propriété d'absorption des purines et pyrimidines dans l'UV à 260nm, et les protéines à 280nm permet de doser les acides nucléiques (C: concentration), et aussi bien d'estimer la contamination par les protéines lors de la purification des acides nucléiques.
- **P (puretés) = A₂₆₀ / A₂₈₀ (Une solution d'ADN est considérée pure si 1.7 ≤ P ≤ 2)**

Solubilité : L'ADN devient un sel d'acide en milieu aqueux et est ainsi soluble. Il précipite en présence d'éthanol et d'une forte concentration saline. Cette propriété permet sa purification.

6- Les formes de l'ADN :

L'ADN de forme circulaire se trouve chez les bactéries et chez certains virus. Cet ADN est souvent double brin.

Chez les eucaryotes, l'enroulement de l'ADN autour de protéines histones et son organisation, contribuent à circulariser l'ADN, dans ces organismes (la circularité de l'ADN peut être donc considérée comme un fait universel et un attribut essentiel d'un ADN actif).

Les topoisomères: 2 ADN ayant exactement le même nombre de nucléotides et la même séquence de bases, mais différents entre eux par le nombre d'enlacement L (le nombre de tours que fait l'un des brins autour de l'autre brin) sont appelées topoisomères.

- **L'état relâché:** la tension sur la double hélice est minimale.

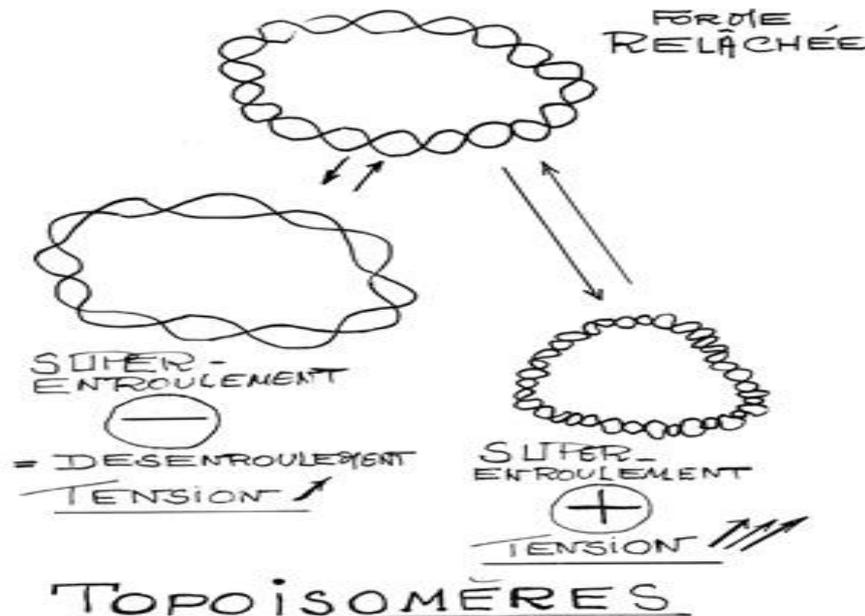
- **L'état surenroulé:** l'axe de la double hélice peut s'enrouler sur lui-même en formant une super hélice. 2 formes de surenroulements sont théoriquement possibles:

- Surenroulement positif : le nombre d'enlacement a été augmenté
- Surenroulement négatif : le nombre d'enlacement a été diminué.

La plupart des molécules d'ADN rencontrées dans la nature forment des surenroulements négatifs.

- Les formes de DNA surenroulées négativement présentent le double avantage d'être à la fois plus compactes, mais aussi plus accessibles aux enzymes de la **transcription et de la réplication**.

- **Les topoisomérases**: Sont des enzymes qui modifient le nb d'enlacement, elles sont donc capables d'introduire ou d'éliminer des super tours dans une double hélice d'ADN.



7- ARN : Structure, Différents Types et Propriétés

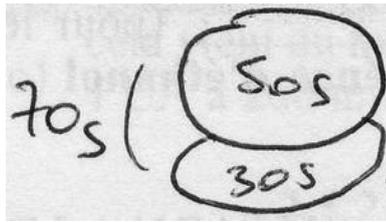
Les ARN sont des polymères de Ribo Nucléotides liés par des liaisons phosphodiester 5'-3'.

Les bases azotées sont A-U, C-G. Le sucre est le Ribose. L'ARN est une molécule monocaténaire, c'est-à-dire simple brin. Cependant il existe des structures où il y a un repliement en «doubles brins».

Classification des ARN : La classification des ARN est liée à leurs fonctions. Chaque type d'ARN possède des éléments structuraux propres.

ARN ribosomiaux (ARNr) : Représente 80% de l'ARN. Les ARNr sont localisés au niveau des ribosomes : en effet, ces derniers sont en fait des particules à base de 35% de protéines et 65% d'ARN. On peut les isoler facilement à partir de broyats cellulaires par ultracentrifugation. On les désigne pour cette raison par leur constante de sédimentation, exprimée en Svedberg (S).

Chez les bactéries : Les ribosomes ont une constante de sédimentation de **70S**. On peut les dissocier en 2 sous-unités de 50S et 30S (les constantes de sédimentation ne sont pas additives...).

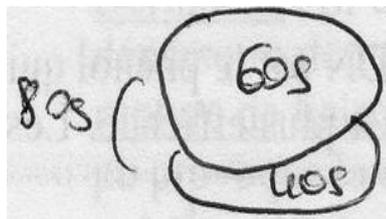


50S: 2 types d'ARN; 31 protéines

30S : 1 ARN; 21 protéines

Ainsi : 50S : 2 types d'ARN ; 31 protéines 30S : 1 ARN ; 21 protéines.

Chez les Eucaryotes : Les ribosomes ont une constante de sédimentation de **80S** \neq 60S + 40S. **60S : 3 types d'ARN ; 45 protéines 40S : 1 type d'ARN ; 31 protéines.**



60S: 3 types d'ARN; 45 protéines

40S: 1 type d'ARN; 31 protéines

Localisation des ribosomes : Les ribosomes sont présents :

Dans le cytoplasme à l'état libre (synthèse des protéines cytoplasmiques).

Dans le cytoplasme, lié au réticulum endoplasmique (protéines à destination membranaire ou intra-organite).

Dans les mitochondries et les chloroplastes : toutefois. Noter que les ARNr sont synthétisés dans le noyau à partir de l'ADN et s'assemblent au niveau des nucléoles avec les protéines, puis passent dans le cytoplasme.

Fonction des ribosomes : les ribosomes interviennent dans la synthèse des protéines (appelée aussi "traduction"), à partir des ARNm.

ARN messagers (ARNm) : 5% de l'ARN, Les ARNm sont monocaténaire et ont une durée de vie très courte. Ce sont des copies en ARN de l'ADN contenu dans les gènes : leur séquence est complémentaire de certaines séquences de l'ADN.

Localisation des ARNm : Ces copies d'ADN sont réalisées dans le noyau, puis exportées dans le cytoplasme, où elles sont ensuite traduites en protéines grâce aux ribosomes.

Fonction des ARNm : Ils sont les intermédiaires entre l'ADN et les protéines.

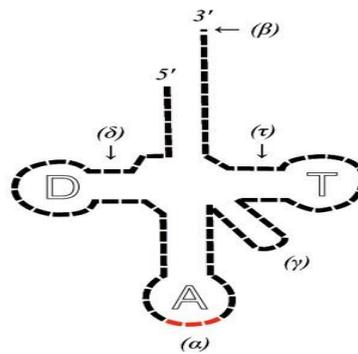
ARN de transfert (ARNt) : Autour de 15% de l'ARN, Ils présentent plusieurs particularités structurales liées à leur fonction :

Petite taille : de 73 à 93 nucléotides.

Structure typique en trèfle : avec de nombreuses régions en structure II (appariement de bases).

Existence d'un anticodon : triplet de nucléotides, complémentaire des codons de l'ARNm- l'extrémité 3'OH fixe l'acide aminé correspondant au codon de l'ARNm. Elle possède la séquence CCA (constante pour tous les ARNt).

Localisation et fonction des ARNt : Les ARNt sont synthétisés, à partir de l'ADN, dans le noyau des cellules Eucaryotes. Ils passent ensuite dans le cytoplasme, où ils jouent un rôle fondamental dans la traduction. En effet, ils apportent les acides aminés aux ribosomes pour construire les protéines. Ils sont qualifiés d'adaptateurs.



α) :anticodon (3 nucléotides), (β) : site de fixation de l'acide aminé, (γ) : boucle variable, (δ) : branche D, (τ) : branche T, A :Boucle de l'anticodon, D: Boucle D, T: Boucle T

CHAPITRE 02 : Réplication des acides nucléiques

Réplication de L'ADN

L'ADN est le matériel génétique qui définit chaque cellule. Avant une cellule en double et est divisé en de nouvelles cellules filles soit par mitose ou la méiose , biomolécules et organites doivent être copiés à répartir entre les cellules. ADN, trouvé dans le noyau , doit être reproduit afin de faire en sorte que chaque nouvelle cellule reçoit le bon nombre de chromosomes .

Le processus de duplication d'ADN est appelée **réplication de l'ADN**. Réplication suit plusieurs étapes qui impliquent de multiples protéines appelées enzymes de réplication et l'ARN . Dans les cellules eucaryotes, tels quelles et les cellules végétales , la réplication de l'ADN se produit dans la **phase S** du interphase pendant le cycle cellulaire . Le processus de réplication de l'ADN est vital pour la croissance cellulaire, la réparation et la reproduction dans les organismes.

La réplication de l'ADN doit respecter deux principes :

- 1- L'ensemble du génome doit être répliqué à chaque division cellulaire
- 2- Chaque molécule d'ADN n'est répliquée qu'une seule fois par cycle cellulaire.

Le réplicon est l'unité de réplication de l'ADN eucaryote, il contient une origine et une terminaison.

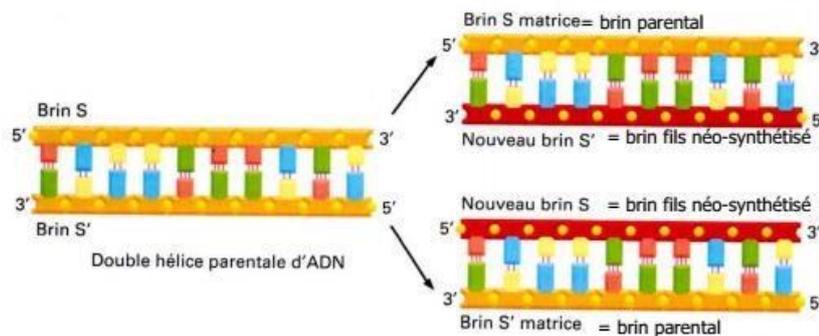
Réplication bidirectionnelle : A chaque origine de réplication, il y a formation d'un œil de réplication qui s'agrandit tout le long de l'avancement au niveau des fourches de réplication. Il y a ainsi deux systèmes de réplication qui évoluent en sens opposés. On dit que la réplication est **bidirectionnelle**.

La réplication de l'ADN est un mécanisme très important elle permet la **duplication** du patrimoine génétique pour obtenir 2 exemplaires, chacun transmis à une des deux cellules filles lors de la division cellulaire. La réplication a donc lieu **avant la division cellulaire**.

La duplication est un évènement crucial où peuvent se produire des erreurs, il y a donc des mécanismes de contrôle, pour limiter ces erreurs. Ces mécanismes peuvent être complétés par des mécanismes de réparation de l'ADN.

La réplication de l'ADN est un mécanisme cellulaire **semi conservatif** : sur la double hélice d'ADN de la cellule fille, un des deux bras provient du brin parental et l'autre est synthétisé lors de la réplication par polymérisation (brin fils).

Chaque brin parental est lu dans le sens 3' → 5' et sert de matrice à la synthèse d'un brin complémentaire et anti-parallèle, appelé brin fils ou brin néosynthétisé, synthétisé dans le sens 5' → 3'. Pour cela les deux brins parentaux doivent se séparer, une rupture des liaisons hydrogènes est donc nécessaire.



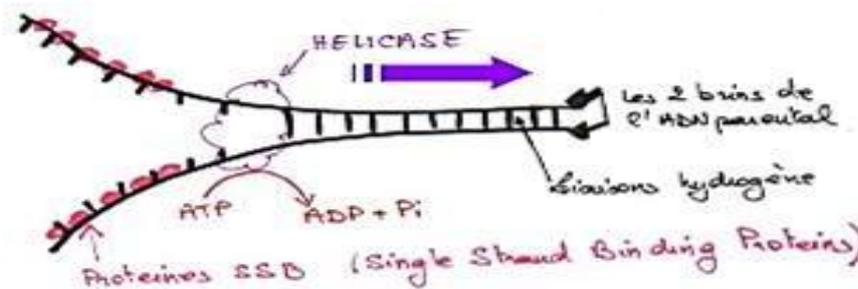
REPLICATION CHEZ LES PROCARYOTES :

ELEMENTS NECESSAIRES

- 1- **L'ADN parental**, où chacun des deux bras sert de matrice pour synthétiser le brin anti-parallèle. Séparation des deux brins (dénaturation) : pour passer à un état simple brin et le rendre accessible aux enzymes de la polymérisation. Il est aussi appelé ADN matrice.
- 2- **Les dNTP** (dATP, dCTP, dTTP et dGTP) rentrent dans la réaction sous forme triphosphate, et sont intégrés dans le brin sous forme monophosphate : le relargage du pyrophosphate apporte l'énergie nécessaire à la réplication.
- 3- Les **ions Mg²⁺** stabilisent les dNTP en les protégeant d'une hydrolyse enzymatique et sont également importants pour l'ADN polymérase.
- 4- Les **enzymes** suivantes :
 - Les hélicases, à l'initiation de la réplication
 - Les topoisomérases
 - Les ADN polymérase et Les primases.
 -

Les enzymes utilisées :

Les hélicases : Elles permettent aux deux brins de l'ADN parental de se séparer en supprimant les liaisons hydrogène entre les bases qui se font face. Le déplacement le long de l'ADN parental nécessite une énergie (hydrolyse des molécules d'ATP en ADP et Pi). Après elles, des protéines se positionnent : **les protéines SSB**, qui se lient à des simples brins, pour éviter la renaturation spontanée de l'ADN.



Les topo isomérase : modifient l'état d'enroulement (+ ou -) de la molécule d'ADN. Chez la bactérie, la topoisomérase est la **gyrase bactérienne**.

Il existe deux types de topoisomérase :

- **Les topoisomérase I** n'agissent que sur un seul brin de l'ADN et ne nécessitent donc pas d'ATP
- **Les topoisomérase II** agissent sur les deux brins de l'ADN, et nécessitent de l'ATP pour maintenir les deux fragments d'ADN. Chez les procaryotes, la gyrase bactérienne est une topoisomérase II.

Les topoisomérase agissent derrière l'action de l'hélicase, qui induit des surenroulements positifs (en avançant) qui vont super vriller la double hélice d'ADN. La gyrase va supprimer ces surenroulements positifs et introduire des surenroulements négatifs. Cette action est simultanée aux hélicases et permet une bonne accessibilité aux enzymes de la réplication à la molécule d'ADN.

Les ADN polymérase : Elles permettent la synthèse du brin fils par **polymérisation** en utilisant un brin de l'ADN parental comme matrice. L'activité ADN-polymérasique est la synthèse de l'ADN dans le **sens 5'→3'** dans la chaîne du brin en cours de synthèse.

Cette activité nécessite une initiation : démarrage par une **amorce** de nucléotides. Pour la réplication, l'amorce est un ARN synthétisée par polymérisation par une **primase (ARN polymérase)**.

L'ADN-polymérase se positionne sur l'extrémité 3'OH pour polymériser de l'ADN dans le sens 5' → 3' après incorporation successive de nucléotides. La synthèse se fait de manière complémentaire et anti parallèle à l'ADN parental. Lors de la formation de la liaison phosphodiester, il y a libération de deux pyrophosphates qui s'hydrolysent en deux phosphates inorganiques.

Les primases : Les primases sont des ARN polymérases ADN-dépendante : elles synthétisent les amorces d'ARN et sont dites ADN-dépendantes car elles se servent de l'ADN comme matrice. L'amorce d'ARN se fait dans le sens 5' → 3' du sens de la chaîne en cours de synthèse.

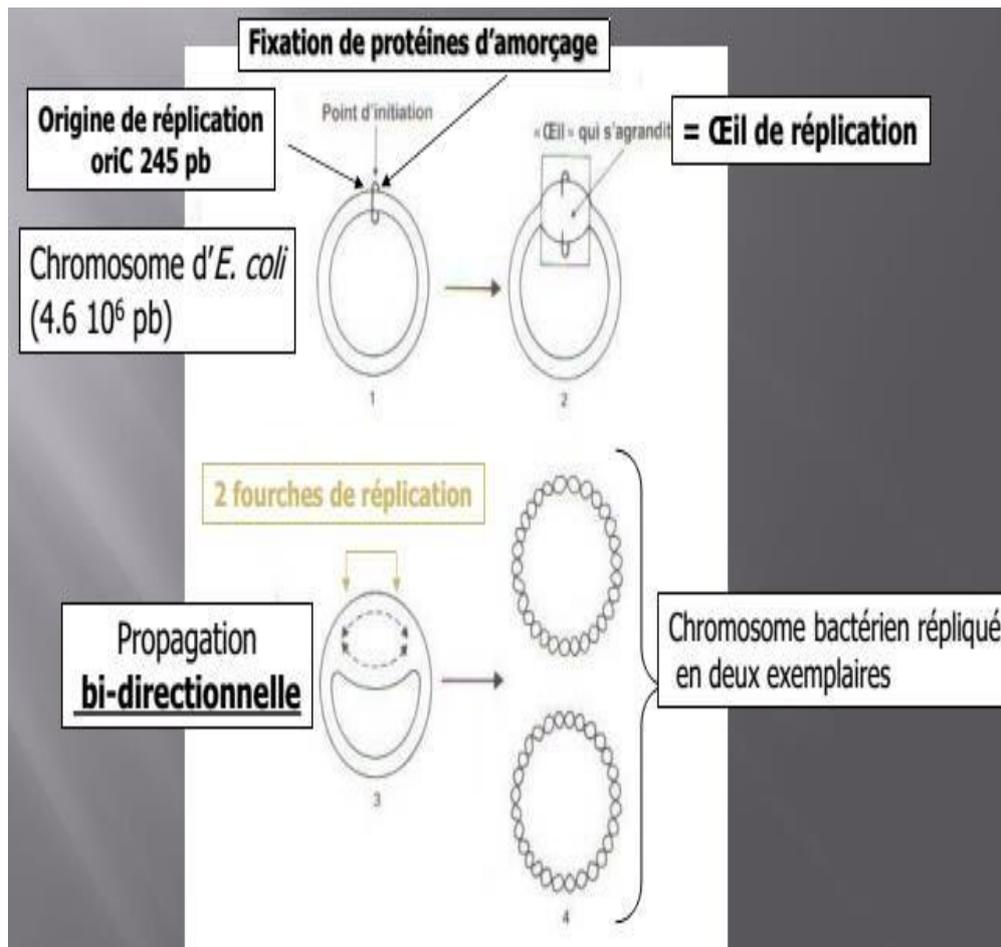
Les primases sont capables de synthétiser de courtes séquences d'ARN (de 4 à 12 ribonucléotides) qui sont ensuite utilisées comme amorces par l'ADN polymérase répliquative. Chez E. Coli, la primase et l'hélicase forment un seul complexe appelé le primosome.

MECANISME DE REPLICATION :

1. Initiation de la réplication :

Chez E. Coli, il existe une seule origine de réplication (appelée ori C : environ 245 pb), contenant des séquences répétées riches en A et T. C'est ici que se fixent les protéines d'amorçage, qui initient la réplication par l'ouverture locale de la double hélice d'ADN.

C'est une propagation bidirectionnelle. On a deux fourches de répliquations : une à droite et l'autre à gauche du point d'initiation. Les deux fourches progressent jusqu'à avoir deux molécules d'ADN double brin. On obtient donc un chromosome pour chaque cellule fille (donc deux au total), le plus fidèle possible au chromosome bactérien de la cellule mère. Chacune des deux cellules filles possède un chromosome à deux brins, qui correspondent à un brin synthétisé et à un brin parental.



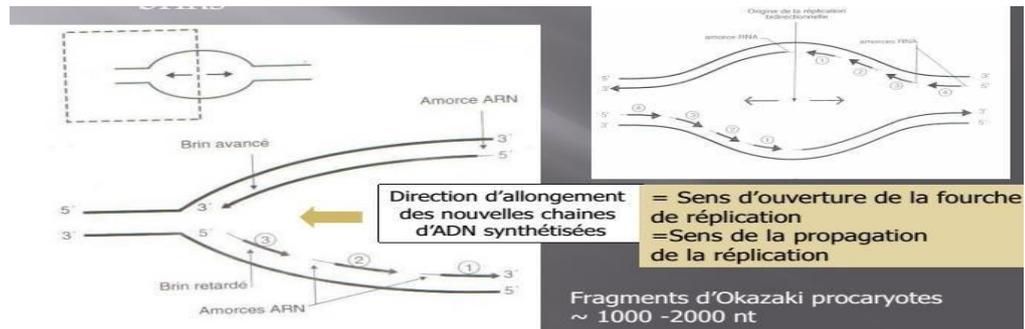
2- Réplication des 2 brins d'ADN parentaux (élongation) : La synthèse des brins fils se fait par l'ADN polymérase III. L'activité enzymatique se fait toujours dans le sens 5'→3'.

Le sens de propagation de la réplication doit s'effectuer dans le sens d'ouverture de la boucle d'ADN, c'est-à-dire dans le sens 5' vers 3'. La réplication est donc discontinue sur un des deux brins. On distingue un brin dit avancé et un brin dit retardé :

Pour le **brin avancé**, la synthèse de l'ADN se fait dans le sens d'ouverture de la molécule d'ADN, donc dans le sens de propagation de la fourche de réplication. C'est le brin **continu**.

Pour le **brin retardé** ou **discontinu**, la synthèse d'ADN se fait dans le sens contraire de l'ouverture la molécule d'ADN : l'ADN polymérase n'exerce son activité polymérasique seulement dans le sens 5'→3' de la chaîne en cours de synthèse, or sur ce brin d'ADN, le sens de propagation de la fourche de réplication est 3'→5'. La synthèse de ce brin est donc **discontinue**. Elle va s'effectuer de manière **rétrograde**, à partir de multiples

fragments d'ARN. On obtient donc de multiples fragments d'ADN (fragments de 1000 à 2000 nt), appelés **fragments d'Okazaki**.



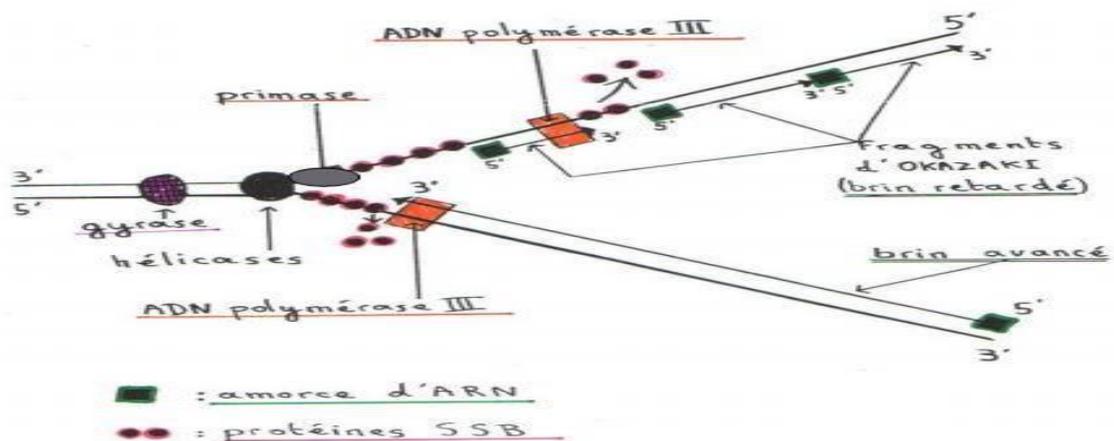
Pour finir, il faut :

- Eliminer l'amorce ARN
- Remplacer les amorces d'ARN par de l'ADN
- Transformer le brin discontinu en un brin continu

Pour avoir un fragment d'ADN continu, l'ADN POL I se positionne au niveau 3'OH d'un fragment d'Okasaki et comble la lacune qui le sépare de l'amorce d'ARN suivante.

Dès qu'elle va rencontrer l'amorce d'ARN, de par son activité exonucléasique 5'-3', elle hydrolyse cette amorce d'ARN et continue, de par son activité ADNpolymérasique, à synthétiser de l'ADN à partir du fragment d'Okasaki précédent. Elle remplace ainsi l'amorce d'ARN en ADN.

La dernière étape consiste à relier ces fragments d'ADN grâce à une enzyme, appelée ADN ligase, qui permet la création d'une liaison phosphodiester entre 2 nucléotides adjacents. On obtient alors un fragment continu. Quand la synthèse d'ADN est terminée, les 2 fourches de réplication vont se rencontrer (car l'ADN est circulaire. Il y a alors séparation des 2 doubles hélices par la gyrase.



REPLICATION CHEZ LES EUCARYOTES

La réplication chez les eucaryotes est grande majorité semblable à celle de la bactérie E. Coli mais il y a quelques spécificités.

Spécificités du génome eucaryote :

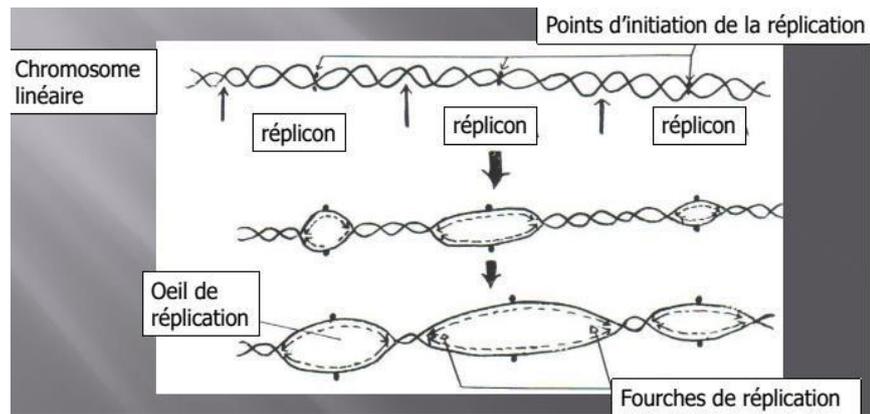
- **Sa localisation** : il se trouve dans le **noyau** cellulaire, alors que la bactérie ne possède pas de noyau.
- **Sa taille** : le génome eucaryote est **plus grand** que le génome procaryote. Par exemple, le génome humain a environ (3.10^9) paires de bases) .
- **Son organisation** : le génome eucaryote est organisé sous forme de chromosomes, formés de molécules d'**ADN linéaire et double brin**. Contrairement à la bactérie, le génome eucaryote est organisé sous forme de chromatine.
- **Sa structure** : structure en **nucléosomes** pour faciliter la condensation de l'ADN afin que le génome puisse rentrer dans le noyau cellulaire. On a donc un ADN qui n'est pas nu.

La vitesse de réplication chez les eucaryotes est **plus faible**. En effet, elle est de **50 nt.s⁻¹** : elle est donc **10 fois plus faible** que celle de la bactérie.

On a donc un génome beaucoup plus grand, qui se réplique à une vitesse plus faible et qui est associé à des structures. Cependant on a quand même une réplication active.

Chez les eucaryotes, il y a **plusieurs origines de réplication** : de 20 000 à 100 000 origines de réplication par cellule. Ces origines de réplication ne sont pas toutes activées en même temps, mais par petits groupes de 20 à 80 OR. Ces groupes constituent une **unité de réplication**.

Sur un chromosome, on retrouve donc plusieurs milliers d'OR, à partir desquelles il y a les deux fourches de réplication. A partir de chaque point de réplication, quand il y a activation, il va donc y avoir un œil de réplication qui va se former, avec les deux fourches qui vont progresser dans des sens opposés (bidirectionnel). Ces portions de chromosome répliquées s'appellent les **réplicons**.



Sur un même chromosome, il y a différents yeux de réplication qui s'agrandissent, et qui progressent jusqu'à ce que la fourche rencontre une autre fourche qui progresse en sens inverse ou atteigne l'extrémité du chromosome.

Le cycle cellulaire est une autre spécificité des cellules eucaryotes. Le cycle est subdivisé en plusieurs phases : la phase de mitose et les interphases G1, S et G2. Les événements qui aboutissent à la réplication ont lieu juste avant que la cellule ne se divise en deux cellules filles. La réplication intervient **pendant la phase S** du cycle cellulaire.

Suivant l'état de condensation de la chromatine, la réplication s'effectue après une activation plus ou moins tardive : si la chromatine est décondensée, elle est répliquée plus précocement, car elle offre une meilleure accessibilité aux enzymes et est donc plus apte à être répliquée. En revanche, si la chromatine est fortement condensée, alors la réplication aura lieu beaucoup plus tard.

LES ENZYMES ET PROTEINES EUCARYOTES

Les différentes enzymes et protéines intervenant dans la réplication eucaryote sont :

- **Les hélicases** : pour séparer les brins parentaux
- **Les protéines RPA** : pour maintenir les 2 brins séparés (ce sont les équivalents des protéines SSB chez la bactérie)
- **Les topoisomérases** : pour éliminer les surenroulements positifs et introduire des négatifs.
- **La primase** (qui est une ARN polymérase ADN dépendante) : pour synthétiser les amorces d'ARN.
- **L'ADN polymérase eucaryote** : il y en a 5 : α , β , γ , δ , ϵ .

ADN POL δ :

C'est une ADN polymérase ADN-dépendante. C'est l'enzyme principale de la réplication des deux brins fils :

- Elle réalise la réplication totale du brin avancé en 5' vers 3' (initiation et élongation)
- Elle réalise la majorité de la réplication du brin retardé (élongation)
- Elle a une activité polymérasique en 5' vers 3' et une fonction d'édition en 3' vers 5'.
- Sa processivité (capacité de l'enzyme à rester associée sur le brin d'ADN parental) est régulée par un anneau ou clamp, qui est une protéine particulière nommée la PCNA (Proliferating Cellular Nuclear Antigen).
- Elle intervient aussi sur la finition des brins d'ADN (pour compléter les lacunes).

ADN POL α :

- Elle intervient uniquement sur le brin retardé pour initier la réplication.
- Elle est associée à la primase : elle prend le relai direct au niveau des fragments d'Okazaki eucaryote (100-200 nt), sur le brin retardé.

ADN POL γ : Elle est impliquée dans la réplication de l'ADN mitochondrial.

ADN POL ϵ , ADN POL β : Elles sont impliquées dans la réparation de l'ADN (cf cours Réplication de l'ADN).

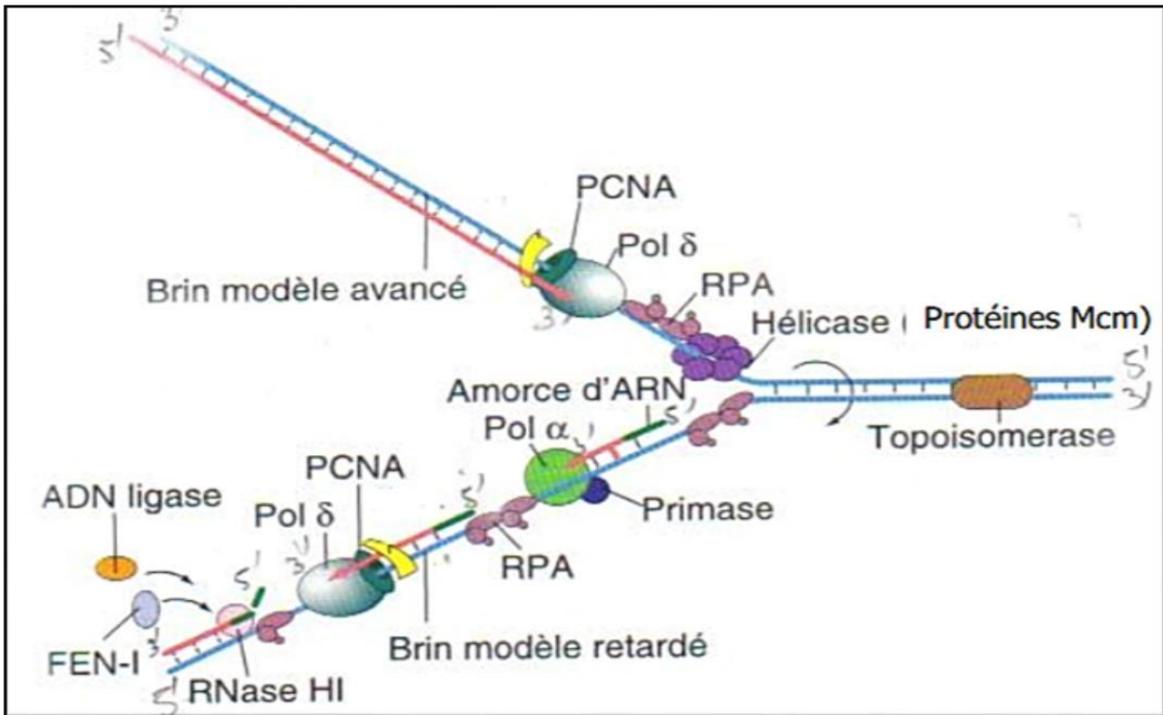
Terminaison :

Une enzyme de la classe des RNases (H-1 et FEN-1) élimine les amorces d'ARN. Le comblement est réalisé par l'ADN POL δ , puis l'ADN ligase lie les morceaux entre eux pour former un seul brin.

L'ADN POL α réalise le début de la synthèse du brin retardé, en initiant la polymérisation, puis elle est remplacée par l'ADN POL δ , qui continue la synthèse dans le sens 5' vers 3' du brin en cours de synthèse. On retrouve sur le brin retardé des fragments d'Okazaki (100 à 200 nt), comme pour le chromosome bactérien.

Pour la finition du brin retardé, l'ADN POL δ élonge le fragment en synthétisant de l'ADN, et remplit la lacune jusqu'à rencontrer le 1er nucléotide du fragment suivant.

Notion de boucles : il y a un repliement structurel de l'ADN du brin fils, de façon à ce que l'allongement se fasse dans le sens 5'-3' mais aussi dans le la fourche de réplication.



CHAPITRE 3 : LES MUTATIONS ET MECANISMES DE REPARATION DE L'ADN

Introduction : Au cours de la réplication, principalement, des erreurs d'incorporation de nucléotides (non-complémentaires, d'oubli ou d'ajout) peuvent intervenir. Afin de pallier à ces problèmes, les cellules possèdent des **systèmes enzymatiques** capables de **vérifier l'ADN** et de **corriger les erreurs** (environ à 99.9%). Cependant, quelques erreurs peuvent persister dans la molécule d'ADN répliquée.

Les modifications de la molécule d'ADN qui ont échappé aux processus de réparation sont appelées **mutations**.

Les mutations sont **aléatoires** et **rares**. En effet, elles apparaissent **spontanément** avec une **faible fréquence** (10^{-6} mutation / génération) mais certains **agents mutagènes** de l'environnement peuvent augmenter leur fréquence d'apparition :

- rayons Ultraviolets (UV)
- substances chimiques
- radioactivité (rayons X)

Les effets de ces agents mutagènes constituent aujourd'hui un problème de **santé publique** : en effet, les cellules dont l'ADN est endommagé peuvent devenir cancéreuses (ex : mélanome).

Définition : Une mutation est une modification rare, accidentelle ou provoquée, de l'information génétique (séquence d'ADN ou d'ARN) dans le génome. Selon la partie du génome touchée, les conséquences d'une mutation peuvent varier. Une mutation est dite **héréditaire** si la séquence génétique mutée est transmise à la génération suivante (voir mutations germinales).

Les mutations sont des modifications permanentes de la séquence de l'ADN. Le gène est alors défectueux. La protéine produite est modifiée. De même, nous allons voir que des modifications des lettres de la phrase entraînent des modifications du sens.

On peut distinguer plusieurs types de mutations :

Une mutation est dite **sexuelle** lorsqu'elle concerne un chromosome sexuel, par exemple X/Y chez les mammifères ou W/Z chez les oiseaux. Une mutation est dite **autosomique** lorsqu'elle touche un autre chromosome que les chromosomes sexuels.

Mutations ponctuelles : La création de nouveaux **allèles** (versions d'un gène, ex : le gène qui code pour la couleur des yeux présente différents allèles : allèle vert, allèle marron, allèle bleu) est potentiellement infinie : pour ce faire il suffit en effet qu'advienne n'importe quelle

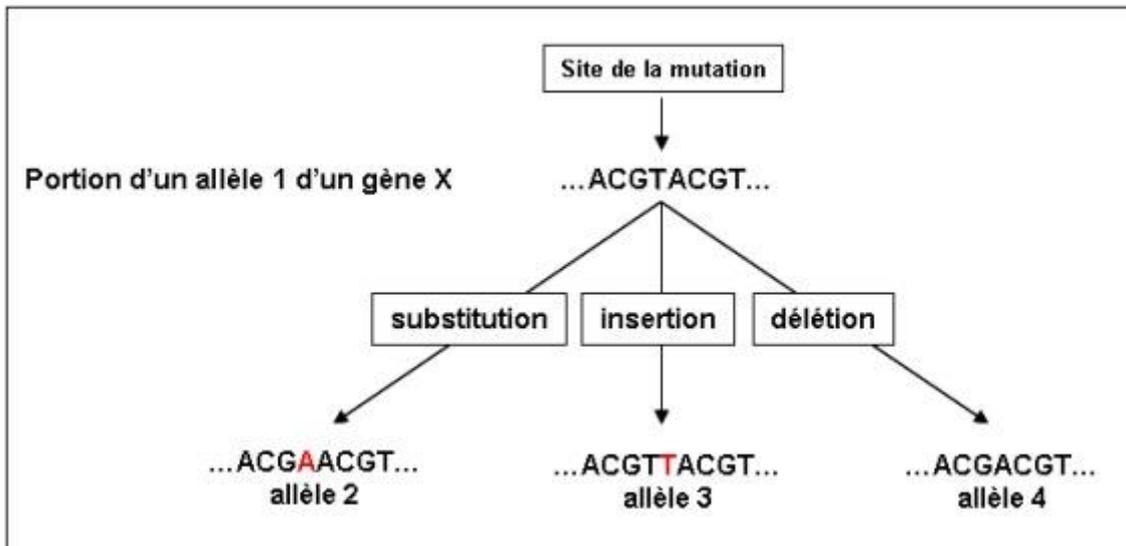
modification de la séquence d'ADN par laquelle ils sont codés. Pourtant, le nombre de types de modifications qui peuvent altérer la séquence d'un gène est très limité.

Lorsque la mutation ne concerne qu'une modification **d'une paire de nucléotides** (ou limitée à quelques nucléotides), on parle de mutations **punctuelles** parmi lesquelles on peut distinguer :

Mutation par substitution : remplacement d'un nucléotide par un autre. Cela arrive la plupart du temps à cause d'une erreur de lecture lorsque l'ADN est recopié, au cours de la division cellulaire par exemple.

Mutation par addition / insertion : gain d'un ou plusieurs nucléotide(s) au niveau de la séquence initiale.

Mutation par délétion : perte d'un ou de plusieurs nucléotide(s) au niveau de la séquence initiale.



Les mutations par insertion ou délétion d'une base azotée engendrent le décalage du cadre de lecture suivi par les ribosomes lors de la traduction de l'ARNm. Ces mutations sont dites **décalantes**.

... AAG GTA ACG ACC ... > ... Phe His Cys Trp ...

... AAG GAA CGA CC ... > ... Phe Leu Ala ...

Mutations punctuelles aux conséquences phénotypiques variées : Les mutations ont des conséquences variables sur le phénotype au sein d'une même espèce, de par leur nature, leur localisation et le rôle du gène concerné. La modification de la séquence d'acides aminés d'une protéine par certaines mutations peut altérer **la fonction de la protéine**. Ces modifications vont avoir ainsi pour conséquence une modification du phénotype.

Les mutations peuvent être répertoriées en fonction des conséquences qu'elles induisent au niveau des protéines synthétisées.

Les mutations silencieuses : une substitution peut avoir lieu sans conséquence sur le phénotype moléculaire.

La séquence de l'ADN est modifiée mais pas la séquence en acides aminés de la protéine codée par le gène, du fait de la **redondance** du code génétique. La fonction de la protéine modifiée est ainsi maintenue intacte. Ces mutations ne s'expriment pas au niveau du phénotype.

Portion du gène initial > protéine initiale : ... AAG GTA ACG ACC ... > ... Phe His Cys Trp

Portion du gène muté > protéine mutée : ... AAA GTA ACG ACC ... > ... **Phe** His Cys Trp ...

Les mutations non-sens : une substitution peut faire apparaître un codon stop qui arrête prématurément la traduction. La protéine se trouve ainsi écourtée et devient souvent non-fonctionnelle (ex : perte du site actif d'une enzyme).

... AAG GTA ACG ACC ... > ... Phe His Cys Trp ...

... AAG GTA ACT ACC ... > ... Phe His **Stop**

Les mutations faux-sens : une substitution peut modifier un acide aminé lors de la traduction. Les conséquences à l'échelle de la cellule et de l'organisme dépendront des modifications structurales et fonctionnelles subies par la protéine. Celles-ci sont très variables suivant les cas. Certaines de ces substitutions sont dites **conservatrices** (aucune modification des propriétés de la protéine) ou **non conservatrices** (changement plus ou moins important des propriétés de la protéine).

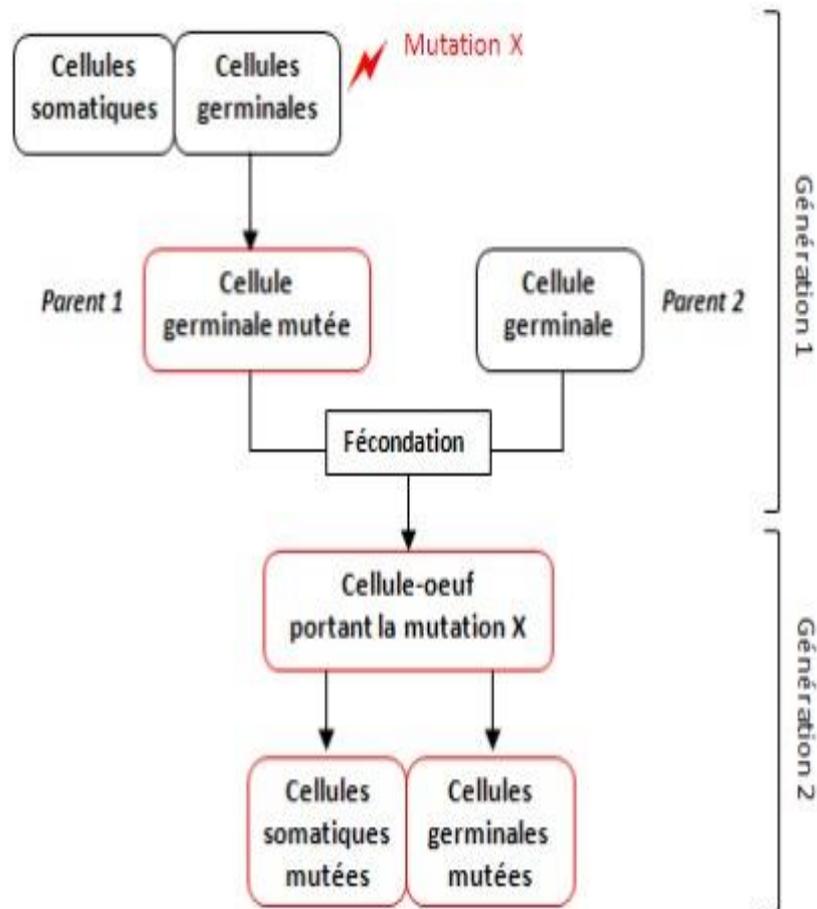
... AAG GTA ACG ACC ... > ... Phe His Cys Trp ...

... AAG TTA ATG ACC ... > ... Phe **Asn** Cys Trp ...

Transmission des mutations : Ainsi, les espèces sont capables de se modifier au cours du temps et de transmettre ces modifications à leur descendance, faisant ainsi évoluer le pool génétique des populations de l'espèce.

Les mutations affectant les cellules **somatiques** sont, bien évidemment, **sans conséquence** sur la descendance. Elles disparaîtront au plus tard avec la mort de l'individu. Les mutations somatiques sont cependant une cause importante de la **cancérisation** des cellules.

Les mutations qui affectent les cellules de la lignée **germinale** (lignée reproductrice) sont **transmissibles à la descendance**. Elles se reproduisent par mitoses à l'ensemble des cellules du nouvel organisme. La mutation devient alors héréditaire.



Transmission d'une mutation d'une génération à l'autre

CHPITRE 4 : EXPRESSION DES GENES

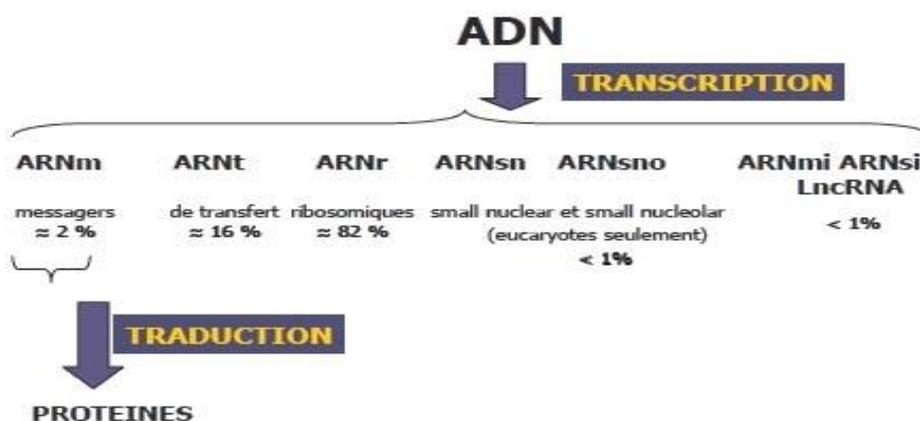
La **synthèse des protéines** est le processus de fabrication des protéines à partir de l'information portée par les gènes. En d'autres termes, il s'agit de l'acte par lequel une cellule assemble des acides aminés ensemble afin de former des protéines, selon l'information contenue dans l'ADN.

La **synthèse des protéines** comprend deux **étapes** : - la transcription permet de copier l'ADN en ARNm au niveau du noyau. Elle **est** réalisée grâce à l'ARN polymérase. - la traduction correspond au décodage de l'information portée par l'ARN m en **protéines**, grâce au code génétique.

La transcription

Est un mécanisme faisant intervenir un système enzymatique, qui convertit l'information génétique d'un segment d'ADN en un brin d'ARN complémentaire (simple brin). Cette molécule est appelée transcrite. Il y a 3 milliards de paires de bases dans le génome mais toute l'information n'est pas transcrite. Les portions qui sont transcrites sont appelées les gènes.

Les ARN ne sont pas seulement les ARN messagers (ils ne représentent que 2% des ARN totaux), il y a aussi d'autres catégories telles que les ARN ribosomiques ou les ARN de transfert. Aujourd'hui, on découvre encore des nouvelles classes d'ARN. Ces autres ARNs ont des fonctions biologiques en tant qu'ARN et ne sont pas traduits en protéine.



Les différentes catégories d'ARN :

Les ARN messagers : Ils représentent **2%** des ARN totaux. Ce sont des **messagers**, c'est à dire qu'ils portent l'information génétique qui provient de l'ADN génomique jusqu'au ribosome, où se réalise la synthèse des protéines : la **traduction**. Ce sont les seuls ARN à être traduits en protéines.

Les ARN de transfert : Ils représentent **16%** des ARN cellulaires. Ils jouent un rôle dans le **transport** et le **transfert d'acide aminés** du cytoplasme vers le ribosome, où se fait l'incorporation de l'acide aminé dans la chaîne protéique en cours de synthèse.

Les ARN ribosomiques : Ils représentent 82% des ARN totaux. Ils rentrent directement dans la **constitution des ribosomes** avec des protéines ribosomiques. Certains de ces ARN jouent un rôle important, car certains sont porteurs de **fonction enzymatique**.

On retrouve aussi de manière minoritaire des petits ARN, principalement au niveau du noyau

Les ARNsno ou ARNsn (small nuclear et small nucleolar) : Ils sont présents uniquement chez les eucaryotes et représentent < 1% des ARN totaux.

Ils ont des rôles très différents :

- Les ARNsn vont intervenir dans le mécanisme d'épissage donc maturation des ARNm.
- Les ARNsno vont intervenir dans les modifications chimiques des ARN ribosomique (maturation de ces ARNr).

Les miRNA et ARNsi : Ils représentent < 1% des ARN totaux. Ils ont un rôle dans la **stabilité des ARNm**. Ils ont la propriété d'être complémentaires au niveau de la séquence de certains ARNm.

Les ARN long non coding (LncRNA) : Ils représentent < 1% des ARN totaux.

Ce sont des ARN de **grande taille**, synthétisé par transcription (ils peuvent être constitués de plusieurs kilobases). Ils jouent un rôle très important en cancérologie (ex : dans le cancer du sein).

Ils jouent un rôle dans la **régulation de la transcription d'ARNm** en se liant aux facteurs de transcription et en régulant l'expression de ces facteurs de transcription. Ils ont aussi un rôle dans l'épissage et la traduction.

LES ELEMENTS NECESSAIRES

LES RIBONUCLEOTIDES : Pour la synthèse d'ARN, les ribonucléotides rentrent sous **forme triphosphate** (ATP, CTP, GTP, UTP), ce qui amène l'énergie nécessaire à la polymérisation par libération d'un pyrophosphate.

L'ARN POLYMERASE ADN-DEPENDANTE : L'ADN double brin sert de matrice à l'ARN polymérase qui synthétise un ARN simple brin dans le sens 5' à 3' grâce à l'énergie apportée par le relargage du phosphate.

La synthèse de l'ARN :

- Ne nécessite **pas d'amorce**
- Se fait dans le **sens 5' à 3'**
- Se fait de manière **antiparallèle** à l'ADN matrice et **complémentaire** de l'un des deux brins de l'ADN

Chez les procaryotes : Il n'existe qu'**une seule ARN polymérase**, constituée de plusieurs sous unités : α , β , β' et σ (à c'est complexe multimérique). Cette enzyme a deux dénominations : **enzyme cœur** ou **holoenzyme**, selon sa constitution.

Le complexe holoenzyme correspond à l'association des sous unité, α , β , β' et σ .

Il initie la transcription. Puis le facteur σ est relargué pour former l'enzyme cœur (constituée des sous unités α , β et β'), qui synthétise tous les ARN.

Il existe :

- **2 α** : Assemblage de l'enzyme, assure la **liaison au promoteur**.
- **1 β** : Assure la **liaison des nucléotides**.
- **1 β'** : Assure la **liaison à la matrice d'ADN**.
- **1 σ** : Assure la **reconnaissance du promoteur**, initiation de la transcription.

Chez les eucaryotes : Il existe **3 ARN polymérase différentes**, localisées dans le noyau cellulaire. Chacun de ces ARN est spécialisé dans la synthèse de certains transcrits.

3 Enzymes	Localisation nucléaire	Transcrits	Activité cellulaire
ARN Polymérase I	Nucléole	ARNr (5.8S, 18S, 28 S)	60 à 70 %
ARN Polymérase II	Nucléoplasme	ARNm, ARNsno, LncRNA, miARN et certains ARNsn	10 à 30 %
ARN Polymérase III	Nucléoplasme	ARNt, ARNr 5S et certains ARNsn	~ 10 %

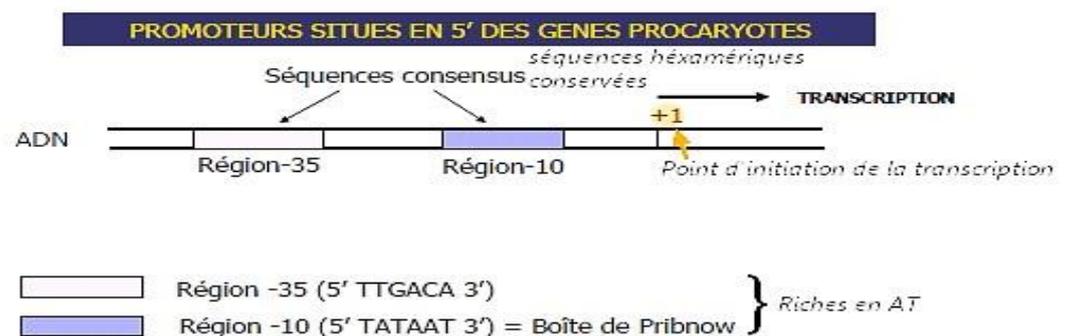
Les deux brins d'ADN peuvent servir de matrice pour la transcription le long d'un chromosome, mais pour un gène donné un seul des 2 brins est utilisé. Le choix du brin matrice se fait en fonction du sens de déplacement de l'ARN polymérase :

- Le **brin sens** est le brin **non transcrit** ou non matrice
- Le **brin anti sens** sera le brin **transcrit** et donc le brin **matrice**. Le brin anti-sens est le brin utilisé comme brin matrice par l'ARN polymérase. Il sera donc lu et transcrit.

Ex. : si l'ARN polymérase lit le brin 3' à 5', la synthèse d'ARN se fait de manière anti parallèle et complémentaire du brin anti-sens. Et donc l'ARN sera identique au brin sens (ici 5' à 3'), sauf que thymine est remplacée par de l'uracile.

Etapes de la transcription chez les procaryotes

INITIATION : L'ARN polymérase reconnaît et se fixe sur des séquences spécifiques de l'ADN, localisées avant le point d'initiation de la transcription. Ces séquences sont appelées **promoteurs** ou régions promotrices.



Pour un gène donné : **convention +1** au lieu où la transcription est initiée. Donc le **nt +1** est le premier nt transcrit sur l'ARN, et donc le ribonucléotide incorporé en 5' du futur transcrit. Avant ce site, toutes les paires de bases sont numérotées de manière négative.

Pour les gènes **procaryotes**, on retrouve dans les régions promotrices des **séquences consensus** (conservées au cours de l'évolution) localisées en position **-10** et **-35**. Ces séquences sont des séquences **hexamériques** (6nt) **riches en T et en A**, surtout celle positionnée en -10.

- **Région en -35** (3' TTGACA 5') : lieu de fixation de l'ARN polymérase.
- **Région en -10** (3' TATAAT 5') = **boite de Pribnow**. Région très importante, composée principalement de thymine et d'adénine. Il s'agit d'une séquence double brin, mais seule la séquence du brin sens est donnée par convention.

Initiation de la transcription :

Reconnaissance des séquences promotrices : L'ARN polymérase doit être sous sa forme **holoenzyme**, donc liée au facteur σ , car la forme holoenzyme est la forme active et compétente pour initier la transcription.

L'ARN sous forme holoenzyme reconnaît et se fixe sur la séquence -35 de l'ADN génomique, on obtient un complexe binaire (dit fermé) : enzyme+ ADN. Ensuite, l'ARN polymérase migre et se déplace jusqu'au promoteur -10 (boite de Pribnow).

Début de la polymérisation : L'ARN polymérase est capable d'**ouvrir la double hélice** de l'ADN génomique sur une courte région. On a donc un passage à l'état simple brin, pour permettre la lecture de l'ADN par l'ARN polymérase. C'est la boucle de transcription, qui contient **17nt** (zone dénaturée).

Sur ces 17 nt, la région hybridée (c'est-à-dire, la région où l'ADN et l'ARN sont associés par liaisons hydrogène) s'établit sur **12 bases**. L'ADN se renature après le passage de l'ARN polymérase et revient à sa forme double brin.

L'ARN polymérase se déplace le long de l'ADN et lit le brin qui doit être transcrit. La synthèse du transcrit se fait par **polymérisation**, c'est-à-dire qu'il y a incorporation de ribonucléotides et que l'ARN polymérase favorise la création des liaisons phosphodiester.

L'ATP, GTP, UTP et CTP rentrent sous forme triphosphate dans la réaction, mais dans la molécule d'ADN, ils sont sous forme monophosphate. Il y a donc élimination du pyrophosphate, ce qui apporte l'énergie nécessaire à la réaction de polymérisation.

Le premier ribonucléotide d'ARN est sous forme triphosphate à l'extrémité 5' chez les procaryotes.

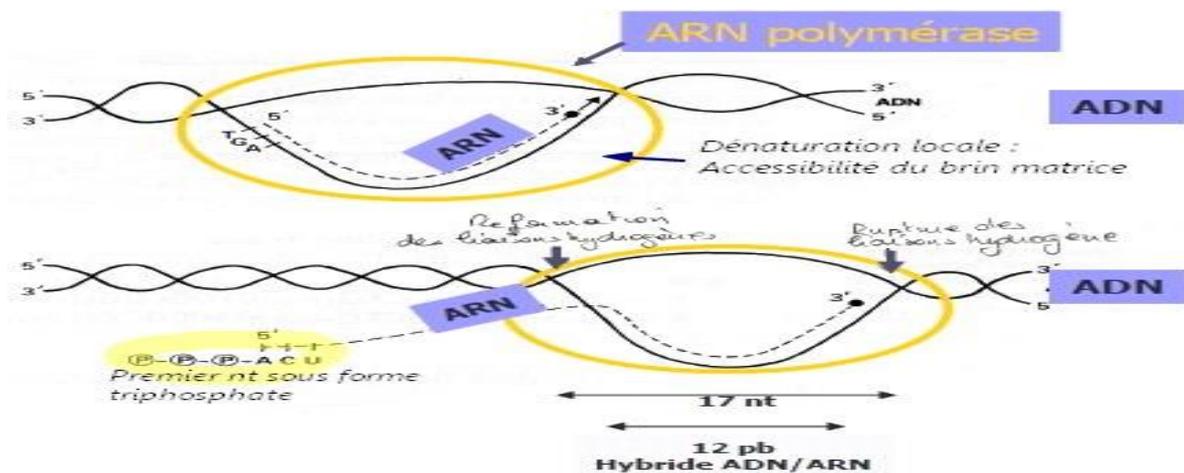
L'ARN polymérase est toujours sous forme holoenzyme. Au bout d'une dizaine de ribonucléotides incorporés, le facteur σ est libéré. L'ARN polymérase adopte donc sa conformation cœur, pour poursuivre la synthèse de l'ARN.

La synthèse se poursuit jusqu'à ce que l'ARN polymérase rencontre une séquence sur l'ADN qui correspond à un **signal de terminaison de transcription**. Lorsque l'ARN polymérase rencontre ce signal, il y a libération de l'ADN, libération de l'ARN et dissociation de l'ARN polymérase.

Une seule ARN polymérase est responsable de la synthèse des tous les ARN de la bactérie.

ELONGATION : Il y a incorporation du premier ribonucléotide sous forme triphosphate, puis création de la liaison phosphodiester au niveau du 3'OH avec un autre ribonucléotide qui va libérer son pyrophosphate, ce qui libère l'énergie nécessaire à la réaction. Donc la synthèse de l'ARN se fait bien dans le **sens 5' à 3'**.

Progression de la boucle de transcription :



TERMINAISON : L'ARN polymérase reconnaît sur l'ADN génomique en cours de lecture des séquences particulières : les **signaux de terminaison de transcription**, qui aboutissent à **l'arrêt de la transcription**, à la **libération de la molécule d'ARN** et à la **dissociation de l'enzyme** vis-à-vis de l'ADN matrice en cours de lecture.

Différences avec la transcription chez les eucaryotes

Localisation cellulaire :

- Chez la **bactérie**, elle s'effectue dans le **cytoplasme**.
- Chez les **eucaryotes**, elle s'effectue dans le **noyau** cellulaire.
- **ARN polymérase :**
- Chez les **procaryotes** : **1** seule.
- Chez les **eucaryotes** : **3**.

Chez les eucaryotes uniquement : événement de **maturation d'un précurseur d'ARN**, qui est appelé transcrit primaire et qui nécessite des modifications pour devenir mature donc fonctionnel.

Transcription et maturation des ARNm chez les eucaryotes

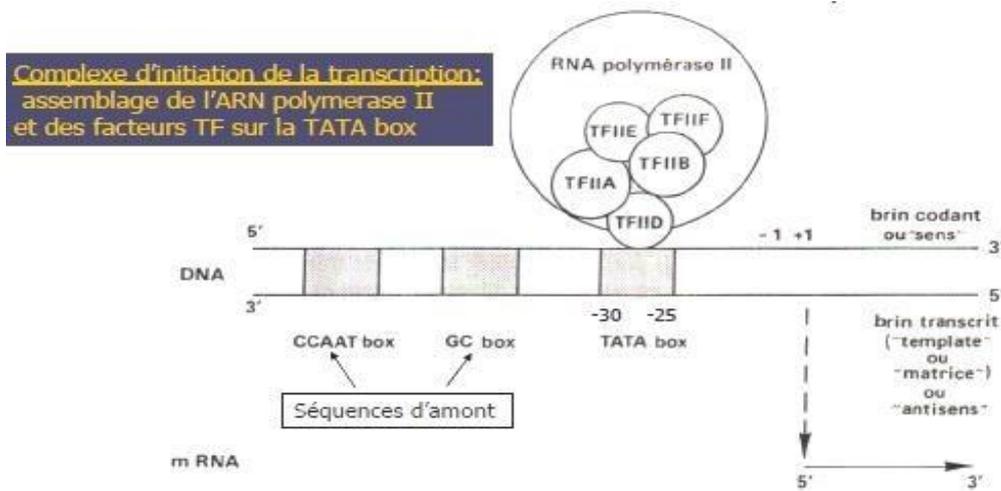
Initiation : Chez les eucaryotes, tous les ARNm sont synthétisés par l'ARN polymérase II. Comme pour les gènes procaryotes, on retrouve une **région promotrice** qui fait intervenir plusieurs régions consensus, conservée au cours de l'évolution (même nomenclature).

La plus importante est la **boîte TATA** (promoteur), qui est l'équivalente de la boîte Pribnow bactérienne. Elle se positionne **de -30 à -25** (séquence hexamérique) et est **riche en A et en T**. Cette boîte TATA représente une séquence promotrice où vient se fixer un **complexe d'initiation de la transcription** qui fait intervenir **ARN POL II** ainsi que d'autres protéines : les **facteurs généraux de la transcription**.

Cette séquence est très fréquente dans les régions promotrices des gènes transcrits en ARNm et traduit en protéines. Il existe des gènes ubiquitaires, et en particulier qui codent pour les histones, qui ont des régions promotrices **sans boîte TATA** (promoteur TATA-less).

Le complexe d'initiation de la transcription se positionne sur la TATA-box. C'est un **complexe multimérique**, qui fait intervenir notamment l'ARN POL II et différentes protéines TFII (+lettre).

Les **TF** sont les facteurs généraux de la transcription, qui s'associent dans un ordre précis pour initier la transcription du gène, contrôlée par la boîte TATA au niveau de la région promotrice.



L'ARN polymérase a besoin de s'associer à ces complexes. En effet, contrairement à l'ARN polymérase procaryote, l'**ARN POL II** seule est incapable de reconnaître le promoteur, de réaliser l'ouverture de l'ADN au niveau de la boucle de transcription et d'initier la transcription.

Chez les eucaryotes, l'ARN polymérase doit faire face aux structures chromatinienne, puisque l'ADN est associé aux nucléosomes. Elle a donc besoin de faire partie d'un complexe où elle est associée à ces facteurs généraux de la transcription. Certains de ces facteurs ont un rôle particulier :

Le facteur TFIID : sa sous-unité **TBP** reconnaît et se fixe à la **boîte TATA**, ce qui induit un **début de distorsion** de la molécule d'ADN. Ce mécanisme induit le recrutement d'autres facteurs TFII et de l'ARN polymérase II, dans une certaine conformation.

Le facteur TFIIH : il a deux fonctions :

- **Rôle d'hélicase** : favorise la **dénaturation locale** de l'ADN, en éliminant les liaisons hydrogène. Nécessite de l'ATP. Ce complexe peut commencer à initier la transcription, donc synthèse d'ARN sur une courte distance.
- **Rôle de protéine-kinase** : capable de **phosphoryler l'ARN polymérase II**, cela va induire un **changement de conformation** de l'enzyme et la dissociation des facteurs généraux (plus besoin car initiation déjà faite). L'ARN POL II réalise l'élongation, la synthèse sur la distance correspond à la séquence du gène du transcrit ARN.

ELONGATION CHEZ LES EUCARYOTES

L'élongation fait intervenir :

- Des **facteurs protéiques d'élongation**, qui permettent le maintien le plus longtemps possible de l'ARN polymérase II sur l'ADN matrice : ils augmentent la processivité.
- Des **topoisomérases**, qui permettent l'élimination des surenroulements positifs créés par l'avancée de l'ARN pol II.
- Des **facteurs protéiques de correction**, qui stimulent la fonction de correction hydrolytique de l'ARN pol II.

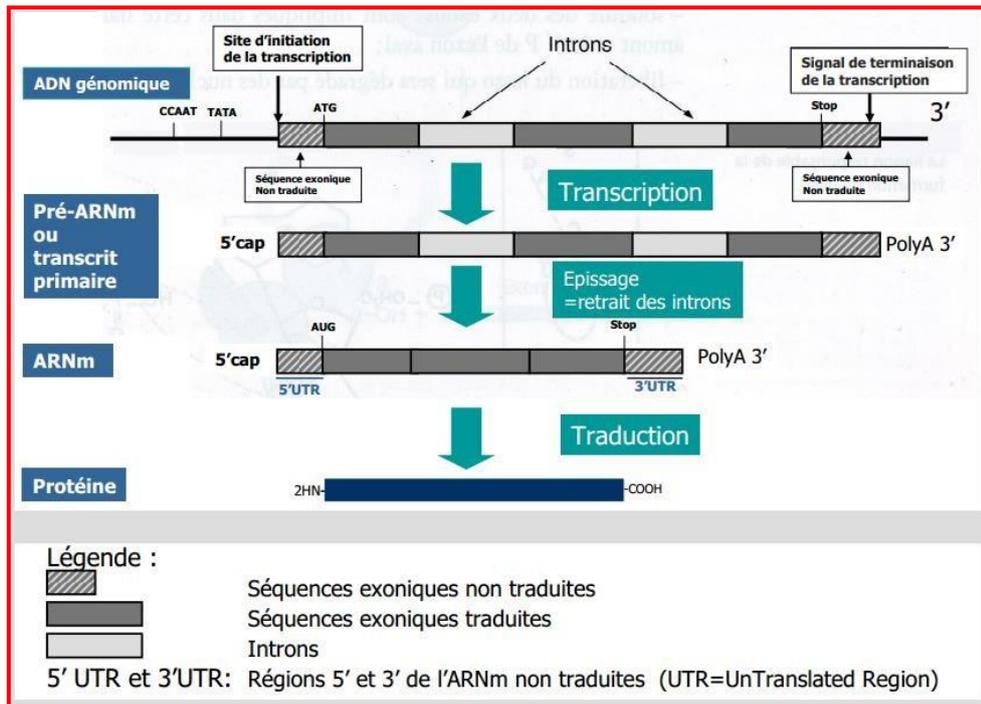
MATURATION DES TRANSCRITS

Spécificité eucaryote : l'élongation des ARNm par ARN POL II s'accompagne d'une maturation du transcrit. Certains événements de cette maturation s'effectuent en même temps que la transcription (dans le noyau cellulaire).

La maturation comprend à la fois la modification covalente des extrémités 3' et 5' de l'ARN messager et le retrait des introns grâce au mécanisme d'épissage.

La séquence transcrite d'un gène est composée de séquences **introniques** et **exoniques**. De manière générale, la plupart des **séquences exoniques** seront **exprimées**, donc **traduites** en séquence protéiques, sauf les régions 5'UTR et 3'UTR qui ne seront pas traduites. Tandis que les **introns** seront **transcrits** mais **non traduits**.

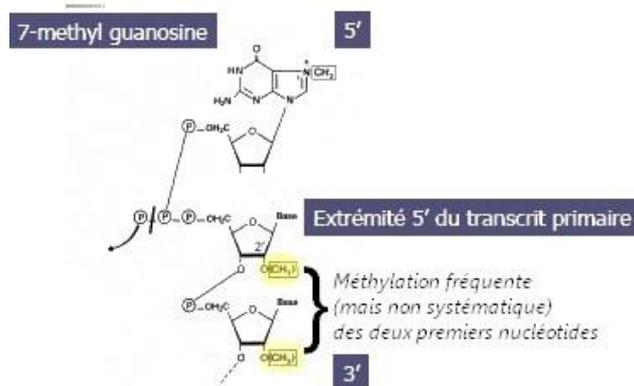
Lors de la transcription, **toute la séquence est transcrite en ARN pré messager** (ou précurseur). C'est seulement après maturation qu'on aura un **ARN messager mature**, où il y aura eu des modifications covalentes des extrémités en 3' et en 5' ainsi que le retrait des séquences introniques.



La coiffe est présente sur **tous les ARNm eucaryotes** (qui proviennent de la transcription par l'ARN POL II uniquement). Il s'agit d'une **modification covalente** qui se fait dans le noyau **au cours de l'élongation** par l'ARN pol II. Cette coiffe se met en place dès le début de la transcription, après 20 à 30 ribonucléotides incorporés dans l'ARN en cours de synthèse, en 5' du transcrit primaire.

Il y a d'abord élimination d'un phosphate en 5' du premier nucléotide triphosphate par une **phosphatase** puis soudure d'une **7-méthyl-guanine** (GMP méthylée sur l'azote en 7) par une **guanyle-transférase** et une **méthyle-transférase**.

Il y a ensuite **méthylation en 2'** sur les riboses des 2 premiers nucléotides dans la plupart des ARNm.



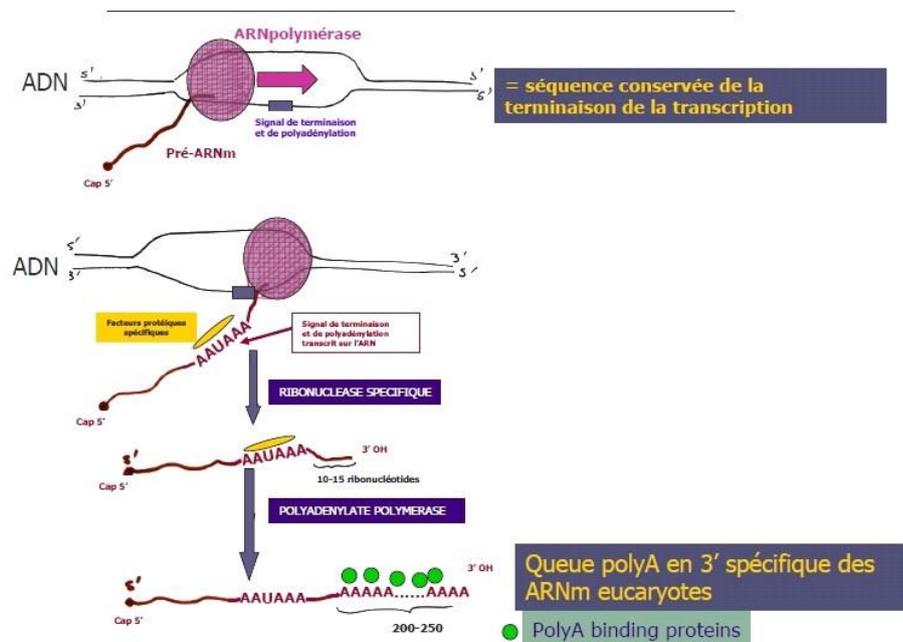
Sur cette coiffe viennent se fixer des protéines spécifiques qui jouent un rôle fonctionnel important. Les protéines qui se lient à la coiffe constituent le **cap binding complexe (CBC)**.

La coiffe a pour rôle de permettre :

- La **distinction** (au niveau de la cellule) entre les ARNm et les autres ARN
- **L'export dans le cytoplasme** de l'ARNm pour traduction
- Le **recrutement de la petite sous-unité du ribosome** à l'extrémité 5' permettant **l'initiation de la traduction**.

La queue poly(A) : coupure et polyadénylation des ARNm

Pour arrêter la transcription chez les eucaryotes, il existe une séquence signal de terminaison de transcription aussi appelée **séquence signal de polyadénylation**.



Epissage : Il s'agit d'un mécanisme qui aboutit à l'**excision des introns** et à la **ligation des exons**, pour aboutir à un ARN fonctionnel et mature.

Le code génétique

Introduction : Les protéines sont issues de l'expression de l'information génétique contenue dans l'ADN des cellules. Le fait que les protéines aussi bien que l'ADN soient constituées d'une séquence d'éléments a fait naître dans l'esprit des chercheurs que l'on pouvait comparer l'enchaînement des acides aminés, comme celui des bases azotées à l'enchaînement des lettres de l'alphabet pour former des mots. Il restait à trouver un système de correspondance permettant de traduire le langage de l'ADN avec celui des protéines : c'est **le code génétique**.

Découverte et établissement du code génétique : L'alphabet de l'ADN est constitué de quatre lettres (quatre bases azotées différentes : A-adénine, T-thymine, C-cytosine, G-guanine) tandis que celui des protéines est constitué de vingt acides aminés différents. Il est donc impossible qu'un nucléotide corresponde à un seul acide aminé.

Ainsi, il a été admis qu'un **acide aminé** de la séquence des protéines est codé par la combinaison d'un **triplet de base azotée** de la séquence de l'ADN.

L'approche expérimentale de décryptage et de l'établissement du code génétique a été réalisée par **Nirenberg et Khorana vers 1965**. Pour ce faire, ils ont mis en présence un acide nucléique de séquence connue, répétitive et constitué de l'alternance de seulement deux bases, avec des acides aminés libres dans un milieu permettant la production de protéines.

La séquence en acides aminés des protéines ainsi produites a pu être comparée à la séquence connue des acides nucléiques de départ. En réalisant l'expérience avec d'autres séquences d'acides nucléiques du même type, ils ont pu établir un tableau de correspondance entre des groupes de **trois nucléotides** (= **codons**) et les vingt acides aminés constituants des protéines.

Le code génétique donne donc la correspondance entre un codon porté par l'ARN messager et un acide aminé. Ainsi, la modification seulement d'une base azotée le long de la séquence d'acide nucléique qui constitue l'information génétique d'un individu suffit à modifier la séquence de la protéine qui en est issue.

	U		C		A		G		
U	UUU	phénylalanine	UCU	sérine	UAU	tyrosine	UGU	cystéine	U
	UUC		UCC		UAC		UGC		C
	UUA	leucine	UCA		UAA	stop	UGA	stop	A
	UUG		UCG		UAG		UGG	tryptophane	G
C	CUU	leucine	CCU	proline	CAU	histidine	CGU	arginine	U
	CUC		CCC		CAC		CGC		C
	CUA		CCA		CAA	CGA	A		
	CUG		CCG		CAG	CGG	G		
A	AUU	isoleucine	ACU	thréonine	AAU	asparagine	AGU	sérine	U
	AUC		ACC		AAC		AGC		C
	AUA	méthionine	ACA		AAA	lysine	AGA	arginine	A
	AUG		ACG		AAG		AGG		G
G	GUU	valine	GCU	alanine	GAU	acide aspartique	GGU	glycine	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		GAA	acide glutamique	GGA		A
	GUG		GCG		GAG		GGG		G

Le code génétique

Justifier pourquoi il n'est pas possible de coder une séquence de protéine en attribuant un acide aminé à une base azotée ou à une paire de bases azotées.

Le nombre de combinaisons possibles pour coder un message avec une seule base azotée est de quatre, celui pour coder un message avec une séquence de deux bases est de seize ; dans les deux cas c'est un nombre de combinaisons insuffisant pour pouvoir coder vingt objets différents, ce qui est nécessaire lorsqu'il s'agit de coder pour les vingt acides aminés qui participent à la structure des protéines.

Le code génétique, un système redondant et universel : Mathématiquement, il existe 64 possibilités (4^3) de combinaison de nucléotides différents pour former un triplet (= codons). Or, seulement 20 acides aminés entrent dans la composition des protéines. Ainsi, plusieurs triplets de nucléotides peuvent désigner le même acide aminé (ex : la sérine est codée par 6 codons différents). Le code génétique est dit **redondant** ou **dégénéré**. En fait, **un codon spécifique correspond à un seul et même acide aminé**.

N.B. : certains codons ne correspondent à aucun acide aminé, ils sont qualifiés de codons stop. Ils provoquent l'arrêt de la synthèse protéique.

L'ADN et les protéines représentent les molécules fondamentales de tous les êtres vivants, lesquels possèdent un même système de correspondance entre ces molécules. Cette **universalité** du code génétique est un argument supplémentaire en faveur d'une origine commune des êtres vivants. Cette propriété du code génétique a rendu possible les expériences de transgénèse. En effet, l'information génétique portée par une molécule d'ADN d'un organisme d'une espèce donnée peut s'exprimer dans le cytoplasme d'une cellule appartenant à n'importe quel autre organisme.

Traduction de l'information génétique

La **traduction**, qui consiste en l'assemblage des acides aminés (provenant de l'alimentation et fournis aux cellules par le sang) pour former une chaîne polypeptidique dans l'ordre imposé par l'information véhiculée par l'ARNm, a lieu dans le **cytoplasme**. Le processus de traduction fait intervenir, en plus des ARN messagers, deux autres types d'ARN :

Les **ARN de transfert** (ARNt), qui portent à leur extrémité un des 20 acides aminés et à l'autre extrémité une petite séquence, appelée **anticodon**, permettant une reconnaissance spécifique des codons.

Les **ARN ribosomiques** (ARNr) sont les constituants principaux des **ribosomes**. Ces ribosomes sont de grosses structures constituées de deux sous-unités permettant d'associer les ARNm et les ARNt de manière spécifique. Ce sont les ateliers de synthèse des protéines.

La lecture de l'ARNm, donc de l'incorporation des acides aminés, se fait toujours dans le sens 5'-3'. Le mécanisme de cette synthèse de protéines comporte trois étapes :

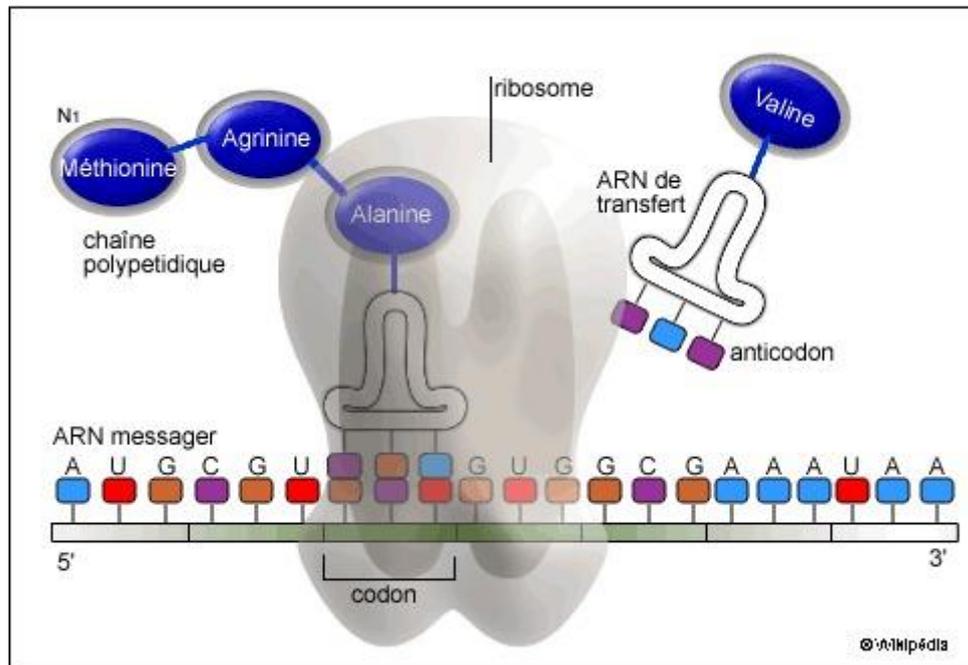
L'initiation : définie par la fixation d'un ribosome sur un triplet (trois nucléotides) de l'ARNm qui est toujours le codon AUG (codon initiateur ; méthionine).

L'élongation : caractérisée par le déplacement du ribosome sur une molécule d'ARNm après chaque incorporation d'un nouvel acide aminé (au niveau de la grande sous-unité) correspondant au codon rencontré par le ribosome. L'ordre de l'incorporation des acides aminés suit ainsi l'ordre déterminé par la séquence des nucléotides de l'ARNm.

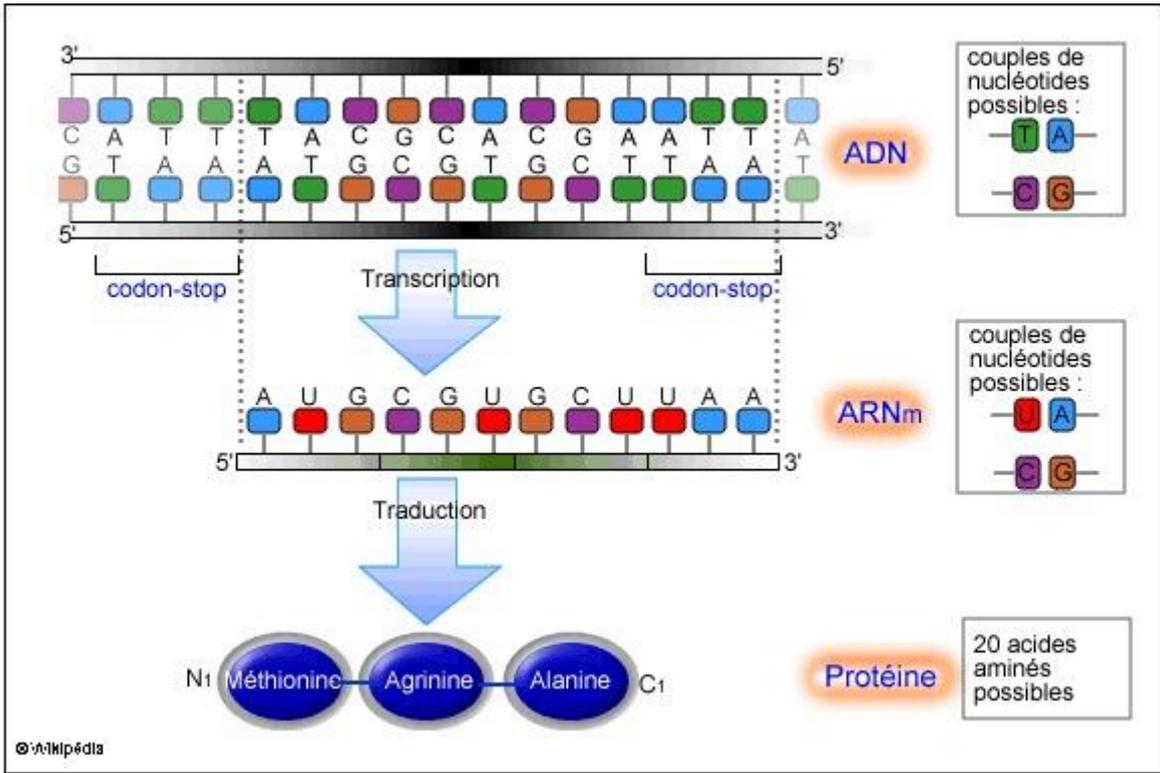
Cette élongation de la chaîne se réalise par l'établissement d'une **liaison peptidique** entre les acides aminés. Ainsi, une fois que la liaison peptidique de ce nouvel acide aminé est effectuée

avec l'acide aminé qui le précédait, le ribosome se décale d'un codon (vers 3') laissant une place libre pour un nouvel ARNt porteur d'un autre acide aminé. La longueur du polypeptide s'accroît donc au fur et à mesure que le ribosome progresse le long de l'ARNm.

La terminaison : provoquée par l'arrivée du ribosome au niveau d'un codon **stop**. La conséquence est telle que le ribosome se dissocie alors de l'ARNm et libère le polypeptide formé.



Processus de traduction de l'ARN en protéine



Différentes étapes de l'expression de l'information génétique

CHAPITRE 5 : LA REGULATION DE L'EXPRESSION GENIQUE

Le contrôle de l'expression des gènes est essentiel pour le maintien équilibré de la croissance cellulaire. Ce contrôle permet à la cellule d'ajuster ses synthèses aux conditions environnementales. La régulation des gènes se fait aux différentes étapes de l'expression :

Deux modes de régulation de l'expression d'un gène cible par une molécule régulatrice :

D'une **façon positive** : l'interaction déclenche l'expression du gène grâce à des **activateurs**

D'une **façon négative** : l'interaction empêche l'expression du gène grâce à des **répresseurs**.

En 1961 Jacob et Monod établissent le modèle de l'expression dans lequel ils distinguaient 2 sortes de séquences d'ADN : agissant en **trans** ou en **cis** :

-**Trans**: les produits sont libres de diffuser pour trouver une cible (activateur, répresseur)

-**Cis**: pour toute séquence d'ADN non transformée en une autre molécule, elle agit *in situ* et touche l'ADN auquel elle est liée (promoteur et operateur).

Les bactéries régulent l'expression de leurs gènes de manière à ne produire que ce dont la cellule a besoin

Elles peuvent faire varier la quantité des produits des gènes en faisant varier le taux d'expression

Le mieux connu parmi les mécanismes est **le contrôle de la transcription**.

Les gènes bactériens sont arrangés sous forme d'**opérons**, qui sont régulés de façon coordonnée.

Un opéron est composé d'un ensemble de gènes sous le contrôle d'un système régulateur unique.

Les gènes sont transcrits à partir d'une région régulatrice commune, sous la forme d'un ARNm **polycistronique** qui sera traduit en protéines différentes.

Cet opéron est contrôlé par une protéine de régulation : **Répresseur ou activateur**.

Opérons inductibles : codent pour des enzymes impliquées dans la voie catabolique et sont induits par le substrat de cette voie.

Exemple : **Opéron lactose (Opéron Lac)** = Synthèse des enzymes nécessaires au métabolisme du lactose.

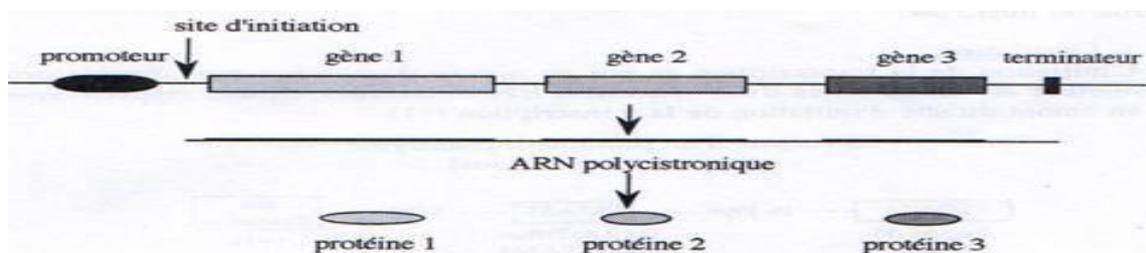
Opérons répressibles : Codent pour des enzymes impliquées dans la voie de biosynthèse et l'expression est contrôlée par le produit final de la chaîne.

Exemple : **Opéron tryptophane (Opéron Trp)** : Synthèse des enzymes nécessaires à la synthèse du tryptophane.

Par définition, un **opéron** est une **unité génétique** trouvée uniquement chez les **procaryotes** composé de **gènes** adjacents qui seront régulés et transcrit ensemble à l'aide d'un même promoteur et l'ARN messager ainsi obtenu est dit **polycistronique** (un ARN spécifique contient l'information nécessaire pour former plusieurs protéines différentes)

L'ARNm polycistronique présente plusieurs séquences codantes indépendantes dites cistrons L'opéron comprend :

- Les gènes de structure
- Un ou plusieurs gènes régulateurs codants des protéines régulatrices : répresseurs ou activateurs
- Des éléments de contrôle présents dans la séquence d'ADN ; promoteur, operateur



Opérateur : une région de l'ADN sur laquelle se fixe un répresseur pour contrôler l'expression d'un gène ou d'un groupe de gènes.

***Promoteur** : une région d'ADN en amont du site d'initiation de la transcription sur laquelle l'ARN polymérase peut se lier

***Terminateur** : fin de transcription.

Opéron lactose de la bactérie ESCHERICHIA COLI : Avec l'étude de l'opéron lactose, François Jacob, Jacques Monod et André Lwoff ont été les premiers scientifiques à décrire un système de régulation de la transcription des gènes. Ils proposent l'existence de deux classes de gènes qu'ils différencient par leur fonction : les gènes structuraux et les gènes régulateurs. C'est à partir de ces travaux qu'est né le concept de la régulation génique. (Prix Nobel de physiologie et médecine en 1965)

L'opéron lactose comprend :

Le gène **Lac I** en 5' (**gène régulateur I**) possédant son propre promoteur(PI) code pour une protéine appelée répresseur I.

Trois gènes structuraux :LacZ ,LacY,LacA

Le gène **lac Z** (code pour La **β-galactosidase**) : qui hydrolyse le lactose en glucose et galactose.

Le gène **lac Y** (code pour La **lactose perméase**) : Cette protéine membranaire permet l'entrée du lactose dans la cellule.

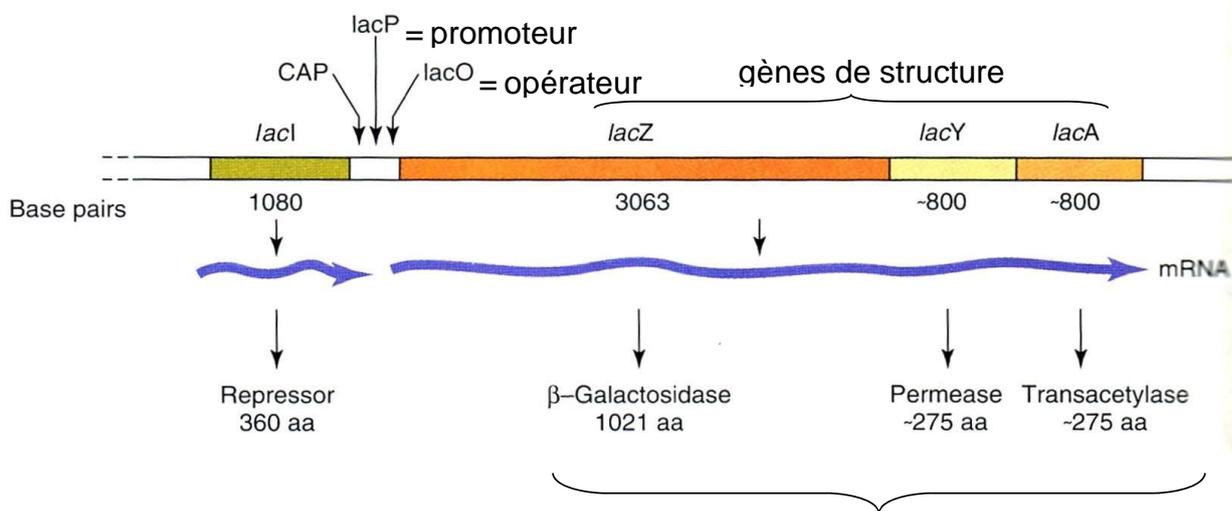
Le gène **lac A** (code pour la **thiogalactoside transacétylase**) : rôle inconnu de l'acétylation.

Une région régulatrice de ces trois gènes comprend le **promoteur(P)** et l'**opérateur(O)**.

Le promoteur de ces trois gènes possède :

En amont : un site **CAP** fixant une protéine CAP activée par l'AMPc : la fixation induit l'activation du promoteur et donc la transcription des gènes de structure.

En aval : un **opérateur**, qui est une séquence d'ADN sur laquelle se fixe des protéines régulatrices permettant la régulation transcriptionnelle de l'opéron.

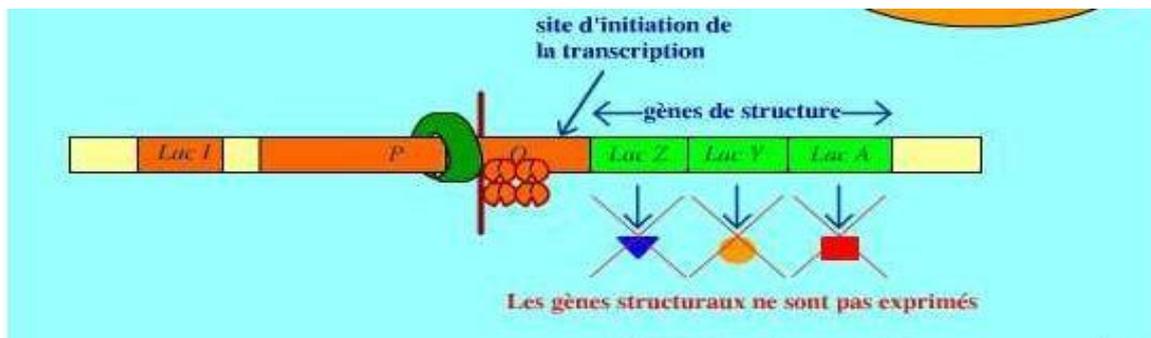


La bactérie trouve sa source de carbone dans le catabolisme des sucres. Si le glucose est la source de carbone "préférée", le **lactose** peut également être consommé par la bactérie et métabolisé en galactose et glucose.

Les enzymes nécessaires à l'utilisation du lactose ne seront synthétisées qu'en présence de ce substrat.

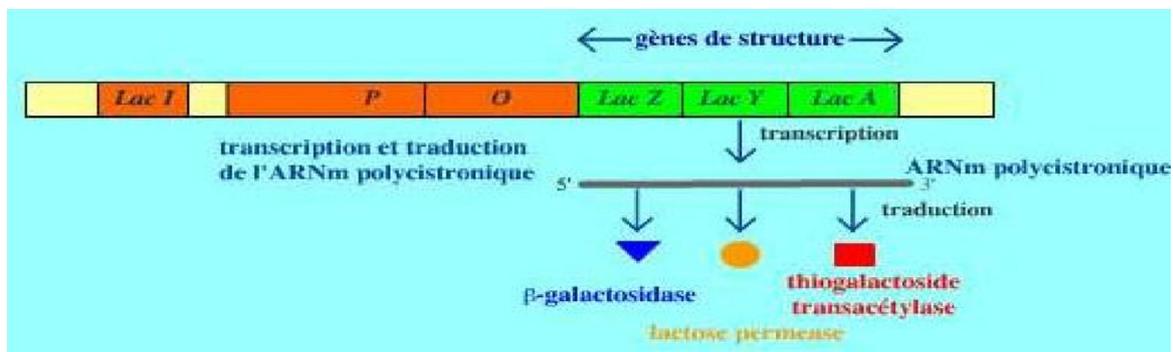
1-EN PRESENCE DE GLUCOSE ET ABSCENCE DE LACTOSE :

- Fixation du répresseur sur l'opérateur
- Blocage de la progression de l'ARN polymérase ce qui empêche la transcription des gènes



2-EN PRESENCE DE LACTOSE ET ABSCENCE DE LACTOSE :

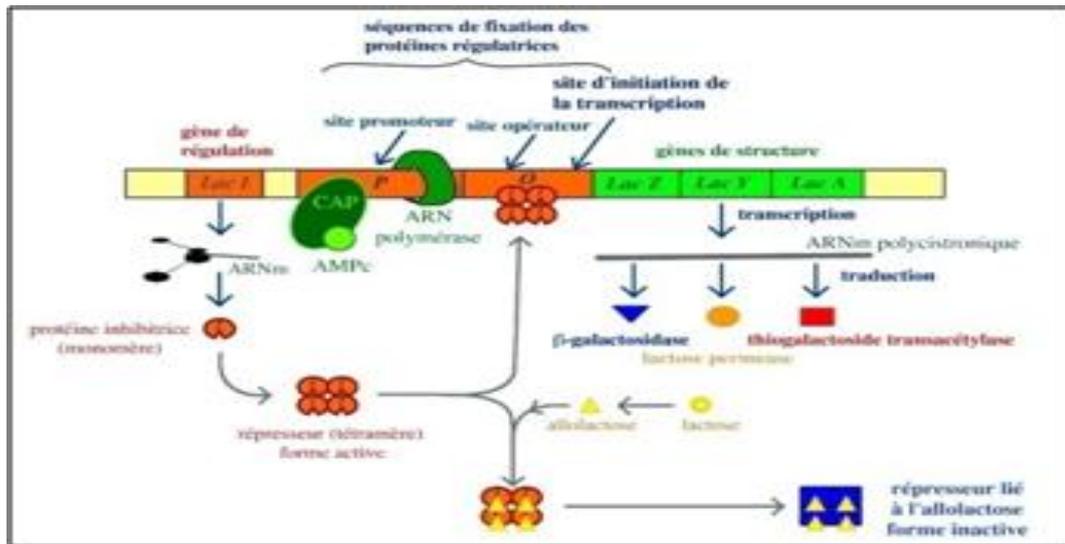
- Fixation du Lactose qui agit comme **inducteur** sur le répresseur
- Changement de la conformation du répresseur qui ne pourra plus se fixer à l'opérateur
- Ce qui conduit à la progression de l'ARN polymérase et la transcription des gènes



3- En présence de glucose et de lactose :

- Le glucose réprime l'expression de l'opéron lactose.
- Intervention de la protéine CAP «Protéine Activatrice du Catabolisme ».
- Présence d'un site activateur (Site CAP) sur le promoteur de l'opéron Lac activé par la fixation de la protéine CAP lorsqu'elle est associée à l'AMPc.
- En absence de glucose: augmentation du taux d'AMPc qui forme un complexe avec la protéine CAP et l'active.
- En présence de glucose: baisse du taux d'AMPc: empêche l'activation de CAP

Le complexe AMPc-CAP est un régulateur positif.



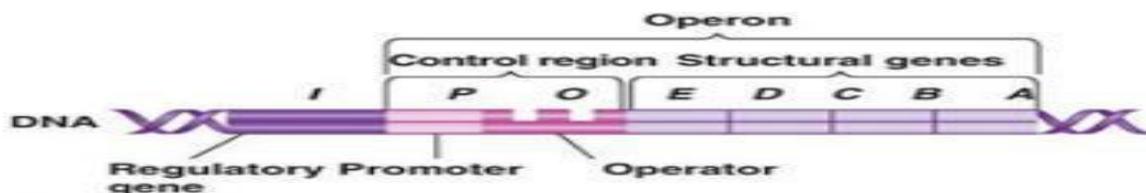
Opéron Tryptophane :

L'opéron Tryptophane est constitué des éléments suivants :

Gènes de structure : *TrpA*, *TrpB*, *TrpC*, *TrpD* et *TrpE*, qui sont des gènes qui permettent de transformer le chorismate en tryptophane.

Éléments de contrôle : représentant le site de fixation du répresseur au niveau de la région en amont des gènes de structure ou opérateur et Promoteur.

Gène régulateur: *TrpR* codant pour un apo-répresseur.



La transcription est régulée par le taux de tryptophane dans la cellule

Activation/répression de la transcription :

En absence de Trp, la transcription des 5 gènes se produit. Le répresseur seul ou apo-répresseur est inactif.

En présence de l'acide aminé, le complexe Trp-répresseur devient actif et bloque la progression de l'ARN polymérase. Il s'agit d'une régulation négative et le Trp est un co-répresseur.

